Essays in Political Economy



Adam Brzezinski

New College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Economics

Hilary Term 2023

Acknowledgements

I am grateful to many people who have directly contributed to and stimulated my work. Most is owed to my supervisor, Rick van der Ploeg, who has unwaveringly supported me throughout this journey. Frequent discussions with Professor Tim Besley have deeply shaped my views on political economy. My co-authors have been an invaluable source of help as well as an inspiration in their own regard: Roberto Bonfatti, Yao Chen, K. Kıvanç Karaman, Valentin Kecht, David van Dijcke, Felix Ward, and Austin L. Wright. A special thanks is due to Nuno Palma, a co-author of two chapters, and a great source of knowledge and encouragement. I very much thank Itzhak Rasooly and Sebastian Hager for spending a great amount of time on giving feedback on my thesis at various stages of my doctoral studies.

More people than can be listed here have helped me develop my thoughts throughout my studies. For helpful suggestions and discussions, I would like to particularly thank Robert Akerlof, Guido Ascari, Adrian Auclert, Jeremy Baskes, Christopher Bowdler, Carlos Santiago Caballero, Eric Chaney, David Chilosi, Lawrence Christiano, Olivier Coibion, Leonor Costa, Matthias Doepke, Martin Eichenbaum, Barry Eichengreen, Andrea Ferrero, Arduin Findeis, Nicola Gennaioli, Simon Gilchrist, Helios Herrera, Sanjay Jain, Willemien Kets, İlkay Kiriscioglu, Julien Labonne, Barton Lee, Jonathan Levy, Gilat Levy, Andrew Little, Nicholas J. Mayhew, Meg Meyer, Ines Moreno de Barreda, Carlos Alvarez Nogal, Pilar Nogues-Marco, Patrick O'Brien, Sevket Pamuk, Armon Rezai, Gary Richardson, Rafaelle Rossi, Nathan Sussman, Akos Valentinyi, Nikolaus Wolf, Stephane Wolton, Meng Wu, David Yang, and Noam Yuchtman. For the published papers, I would also like to thank editors and anonymous referees for their time and feedback.

Contents

Introduction **4**

1 Narratives of Blame **6**

1 Introduction . . . 7
2 Background and literature . . . 11
2.1 Model background and intuition . . . 11
2.2 Contributions to the literature . . . 19
3 The model . . . 25
4 Analysis . . . 31
4.1 Preliminary observations . . . 32
4.2 Endogenous turnout and equilibrium selection . . . 39
4.3 Populist success and the "left behind" . . . 45
4.4 Left-wing policy incentives for right-wing populists? . . . 49
4.5 Connections to the empirical literature . . . 51
5 Conclusion . . . 52
6 References . . . 54
7 Appendices . . . 59

2 Science Skepticism Reduced Compliance with COVID-19 Shelter-in-Place Policies in the United States **91**

1 Introduction . . . 92
2 Results . . . 93
3 Discussion . . . 98
4 Methods . . . 102
5 Tables . . . 106
6 Figures . . . 108
7 References . . . 113
8 Supplementary Information . . . 120

3 Monetary Capacity **133**

1 Introduction . . . 134
2 Conceptual framework . . . 140
3 Empirical analysis . . . 141
3.1 Data . . . 142
3.2 Empirical strategy . . . 144
3.3 Results . . . 147
3.4 Robustness . . . 148
4 Long-run patterns . . . 155
5 Conclusion . . . 160
6 References . . . 162
7 Appendices . . . 168

4 The Vagaries of the Sea **186**

1 Introduction . . . 187
2 Money supply shocks and econometric method . . . 189
2.1 Money and precious metal inflows in early modern Spain . . . 189
2.2 Maritime disasters . . . 191
2.3 Outcome variables . . . 194
2.4 Econometric method . . . 199
3 The causal effects of money supply shocks . . . 200
3.1 Output effects . . . 200
3.2 Monetary transmission channels . . . 203
4 Conclusion . . . 212
5 References . . . 214
6 Appendices . . . 225

Concluding remarks **281**

Introduction

Beliefs motivate actions, thereby shaping economic, political, and sociocultural outcomes. In this thesis, I examine the role of subjective beliefs across a diverse set of topics within political economy that concern both the present and the past. I discuss how subjective beliefs shape the incentives of competing parties; how they influence policy compliance; how they determine the long-term trajectory of states; and how they mediate the effect of monetary shocks. By the end of the thesis, I hope to have convinced readers of the importance of understanding how subjective beliefs are shaped, and what their effects are, for studying political economy.

Each chapter features its own introduction, discussion of relevant literature, and conclusion. Hence, I will confine myself to briefly outlining the topics covered and highlighting the common threads in this introduction. The chapters comprise both theory and empirics, and concern present occurrences as well as past developments. Subjective beliefs of agents, however, are a central consideration throughout.

The first two chapters analyse the role of shared worldviews in shaping current political outcomes. In the first chapter, **Narratives of Blame**, I formulate a new theory of political competition in narratives applicable to the study of populism. In the framework, political narratives influence beliefs about the true causes for observable economic and sociocultural outcomes. Populist narratives scapegoat outside factors for undesirable outcomes, and are preferred by lower-status voters who feel a sense of being "left behind". Whether populist narratives are adopted from (economically) left-wing or right-wing parties is determined by the relative salience of economic and non-economic issues: if the former are more salient, populist tendencies will tend to come from the left, and vice versa for the right.

In the second chapter, **Science Skepticism reduced Compliance with COVID-19 Shelter-In-Place Policies in the United States** (now published in *Nature Human Behaviour*[1]), together with co-authors I analyse a concrete implication of worldviews that inhibit effective policymaking. The chapter's title gives away the

[1]Brzezinski, Adam, Valentin Kecht, David Van Dijcke, and Austin L. Wright. 2021. Science skepticism reduced compliance with COVID-19 shelter-in-place policies in the United States. *Nature Human Behaviour*, 5(11), 1519-1527. URL: https://doi.org/10.1038/s41562-021-01227-0.

main conclusion of the paper, arrived at through an analysis of granular mobile device tracking data. The chapter clearly highlights the practical policy implications of defeatist narratives about the effectiveness of policies, such as science skepticism.

For the final two chapters, the focus shifts from the present to the early modern period, from current political outcomes to the long-run evolution of states. In the third chapter, **Monetary Capacity**, we analyse empirically and theoretically how the ability of a state to maintain high levels of monetization, its *monetary capacity*, depends on its ability to raise taxes, its *fiscal capacity*, and vice versa. The key mechanism behind the coevolution of monetary and fiscal capacity is, again, mediated by beliefs: a high fiscal capacity maintains beliefs that the state will not resort to inflationary episodes for the funding of wars or other public expenditure. This increases the willingness of agents to hold money, improving a state's monetary capacity. Conversely, a high monetization enables more effective tax collection and incentivises the state to build up fiscal capacity.

Chapter four, **The Vagaries of the Sea** (forthcoming in the *Review of Economics and Statistics*[2]), goes deeper into analysing the causal effect of monetary shocks. It focuses on the case of Spain, a first-order receiving country of precious metals in the early modern period. The shock we use for causal identification consists of the loss of monetary silver and gold, caused by maritime disasters that affected treasury fleets on the way from the Americas to Spain. The loss of precious metals impacts the economy through changing beliefs of agents about the expected money stock in the economy. We find that such shocks have real effects: prices adjust sluggishly, while real GDP responds on impact.

The work presented in this thesis is part of a wider ongoing research agenda to study the role of subjective beliefs more generally, and political narratives in particular. In the concluding remarks to this thesis, I outline my ongoing and future work that builds on the topics explored here. It is my hope that this thesis presents a building block towards a better understanding of how subjective beliefs influence, and are influenced by, politics.

[2]Brzezinski, Adam, Yao Chen, Nuno Palma, and Felix Ward. Forthcoming. The vagaries of the sea: evidence on the real effects of money from maritime disasters in the Spanish Empire. *Review of Economics and Statistics*. URL: https://doi.org/10.1162/rest_a_01223.

Narratives of Blame

Abstract

This paper formulates a new theory of political competition in narratives applicable to the study of populism. Parties act as narrators, offering voters explanations for observable outcomes that differ in the degree of blame attributed to external scapegoats. Populist *narratives of blame* appeal to voters who face unfavourable outcomes. The model sheds light on the factors that lead to left-wing versus right-wing populism. In particular, economic cleavages in society favour left-wing populism, whereas sociocultural cleavages favour right-wing populism. The relative rise of right-wing populism and an emerging sociocultural split in the electorate can jointly be interpreted as driven by an increasing salience of sociocultural issues. The theory also examines the role of "left-behind" voters in the success of populists, and explores the incentives for right-wing populists to adopt more redistributive economic policies in order to satisfy such voters.

Keywords: narratives, subjective beliefs, populism

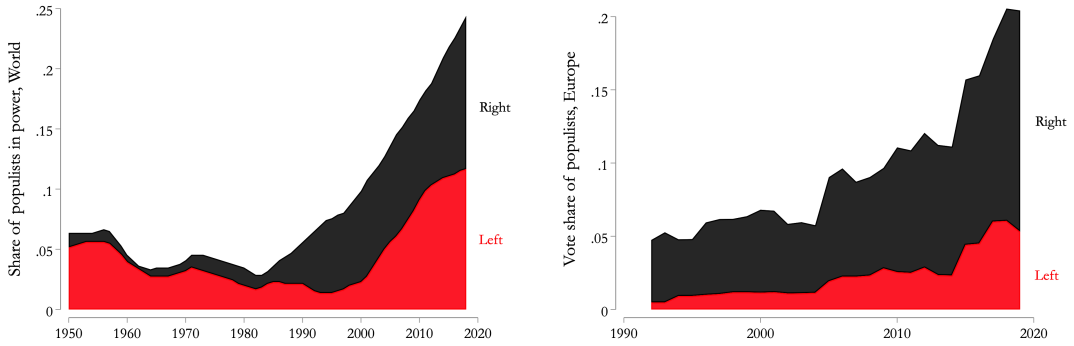
JEL Codes: D72, D84, D91

1 Introduction

An emerging sociocultural dimension has transformed political competition in Western democracies (Gethin et al., 2022). Described in the titles of recent work as a "rise in identity politics" (Besley and Persson, 2021), "cultural backlash" (Norris and Inglehart, 2019), or "white backlash" (Abrajano and Hajnal, 2015), this change is challenging to comprehensively document, not least because of the difficulty in measuring some of the concepts involved. Yet, one symptom transpires clearly: a surge of populist parties, especially of the right-wing variety, as shown in Figure 1 based on two methodologies (Funke et al., 2020; Rooduijn and Taggart, 2019). Dwarfed by left-wing populism in the 1950s-1980s (Panel *a*), right-wing populism now constitutes a, if not the, dominant populist variant in Western democracies (Panel *b*). While much debated in the literature, why populism takes an increasingly right-wing form remains an open question from a theoretical perspective (Guriev and Papaioannou, 2022).

Figure 1: The rise of populist parties

(a) Share of populists in power, World (b) Vote share of populists, Europe



Notes: *Panel (a):* based on the classification of populist parties in 60 countries in the world from Funke et al. (2020). It shows the 5-year moving average for the share of all 60 countries in the dataset which are classified as having a left-wing or right-wing populist government. Data beyond 2018 is extrapolated by using the final datapoint in order to calculate the moving average. *Panel (b):* based on the classification of populist parties in 31 European countries from Rooduijn and Taggart (2019). It shows the vote share of far-left and far-right populist parties as classified in the source.

This paper argues that the relative rise of right-wing populists can be better understood by considering the role of narratives in political competition. I follow a literature in political economy in conceptualising narratives as subjective causal

models about observable outcomes (Eliaz and Spiegler, 2020). In this view, narratives boil down to claims about whether economic downturns are shaped by government actions or inevitable crises, whether a job loss is caused by work attitude or immigration, or similar competing explanations for different outcomes. I further narrow this definition to one particular dimension in which narratives vary: the degree of blame they ascribe to external scapegoats that lie outside of the control of agents.

To make this definition more concrete, consider clashing narratives about the determinants of a strong economy offered by the candidates in the 2016 United States presidential election. Donald Trump, both as candidate and as president, blamed immigration for bad economic outcomes: "Illegal immigration hurts American workers [...] taking precious resources away from the poorest Americans who need them most. Illegal immigration costs our country billions and billions of dollars each year" (White House, 2018). He also repeatedly blamed trade relations with China, calling it "the greatest theft in the history of the world" (Stracqualursi, 2017). Hilary Clinton, instead, often highlighted the positive role of immigration for the economy: "The United States was built by immigrants, and our economy depends on immigrants," (Farrar, 2016); and when discussing the causes for economic success, she highlighted the importance of investment for economic growth, which can happen if "we support advanced manufacturing, if we are the kind of country that once again understands how important it is to build things. We are builders and we need to get back to building" (Keith et al., 2016).

The quotes above do not constitute a formal classification of the differential positions of these two candidates. Yet, they present useful illustrations of the scapegoating dimension of narrative competition I consider in this paper. One set of narratives, here voiced by Donald Trump, concentrates on *blame*: bad economic outcomes are driven by external factors, including immigration and the actions of another country. The competing set of narratives instead underscores *agency*: economic growth can be fostered by actors within the United States.

Populist tendencies as investigated in this paper refer to the employment of

narratives of blame. Scapegoating is indeed an integral trait of populist language, confirmed through the analysis of parliamentary speeches (Tzelgov and Olander, 2018; Magnusson et al., 2018). Populists cast the world as one of opposing groups, "us versus them" (Mudde and Kaltwasser, 2017; Müller, 2017), and blame "them" for a wide range of bad outcomes. "They" can refer to an economic or political elite, or an outside group such as immigrants. Who is blamed is outside the scope of this paper—instead, it is concerned with who is doing the blaming, and for what reasons.

To investigate this question, I conceive of parties as narrators who provide the set of narratives that voters choose from. To give narratives maximal bite, I assume that they are disciplined only by voter preferences and party incentives, but not by objective data. How empirical evidence may or may not discipline overly simplified narratives (Levy et al., 2022) or false narratives (Eliaz et al., 2022) has been fruitfully studied elsewhere. My conception of narratives should be viewed as a complement to such approaches, and as more applicable the more disputed facts are in public. It gives a theoretical grounding for a "post-truth" view of political competition, where parties and voters choose which "alternative fact" to put forward or believe. It does not follow, however, that "anything goes" in the model: as the framework will show, the incentives of the narrators—parties—and the audience—voters—put firm constraints on the types of narratives that can emerge in equilibrium.

Political equilibria are influenced by the narrative preference of voters, which is a function of voters' *status*. This status is determined by the economic and non-economic outcomes that voters face.[1] Based on their status, I assume that voters form beliefs in a motivated way, which shapes their narrative preference.[2] It follows that voters with low status will prefer populist narratives of blame, in order to scapegoat outside factors for their unfavourable outcomes, while those with high status will prefer narratives of agency.

[1] I leave the interpretation deliberately open of what these non-economic components may be. However, to make the difference between economic and non-economic components more concrete, I use the term "sociocultural" as a synonym to "non-economic" in this paper.

[2] The assumption of motivated reasoning follows an established approach in political economy models (Eliaz and Spiegler, 2020; Levy et al., 2022), and is also well established empirically. Section 2 aruges this in more detail.

The framework sheds light on when populist parties should be expected from the right of the economic spectrum, and when from the left. The form of populism depends on the relative importance of economic and non-economic components of voter status. In particular, economic cleavages favour left-wing populism, while sociocultural cleavages favour right-wing populism. On the one hand, if economic issues play a large role, voters' status will be highly correlated with their economic situation. Left-wing populists will then appeal to low-status voters more than right-wing populists would. Indeed, the right-wing party has an explicit incentive to abstain from narratives of blame in this case, to satisfy high-income, high-status voters. On the other, if sociocultural issues are more salient, the correlation between voters' status and their economic situation will be less strong. Once this correlation is sufficiently undermined, there is room for a right-wing populist party to emerge.

By considering the relationship between voters' economic and non-economic outcomes and giving narratives a central role, the framework can jointly interpret several observations documented in the literature. The rise of right-wing populism (Funke et al., 2020; Rodrik, 2018) can be interpreted as being caused by forces which break the correlation between a voter's economic outcomes and their overall status in society. Among such forces are movements that question the social status of an established elite, for instance rich white men (Wolton, 2023), be that on the basis of race (Black Lives Matter), gender (#MeToo), or values held (Fridays for Future). In and of itself, the rising importance of sociocultural issues documented in the literature (Gethin et al., 2022) will have a similar effect, by decreasing the role of a voter's economic outcomes in shaping their status in society. The framework thus gives an explicit theoretical foundation for the observations of Rodrik (2018), who shows that globalisation shocks that increase the salience of cultural issues will lead to a rise in right-wing populism, while those of an economic nature will lead to left-wing populism. In the framework, the same forces that bring about a right-wing populist equilibrium also lead to increasingly sociocultural cleavages in the electorate. This accords with the empirical observation that left-wing voters increasingly stem from a highly educated cultural elite, and right-wing voters from wealthier but less ed-

ucated spheres (Gethin et al., 2022). From the model's perspective, the perceived failure expressed in the media (e.g. Teixera, 2022) of left parties to concentrate on the grievances of low-income voters appears as a strategic choice to cater to a cultural elite, rather than a policy mistake. The model also sheds light on the role of the often-quoted "left-behind" voters in the success of populists (Hochschild, 2016; Norris and Inglehart, 2019), and explores the incentives for right-wing populists to adopt more redistributive economic policies in order to satisfy such voters.

The remainder of the paper is structured as follows: Section 2 gives relevant background to the model's assumptions and insights, and discusses the contributions to the literature. Section 3 then gives the formal set-up of the model. Section 4 comprises the main model analysis, and Section 5 concludes.[3]

2 Background and literature

This paper builds on an emerging theoretical literature in political economy that takes the notion of narratives seriously. In Section 2.1, I discuss how the assumptions of the model relate to the existing literature, and how they depart from it. I also give a first intuitive discussion of the model's main insights. Building on this discussion, in Section 2.2 I locate the paper in the existing literature and discuss how I contribute to it.

2.1 Model background and intuition

Narratives are the main currency of political competition I focus on in this paper. While narratives can be defined in multiple ways, for the purpose of this paper I will use a definition that fits the context of populism. Specifically, I define narratives by the following three traits: (1) they are subjective causal models of how the world works; (2) they are adopted by voters in a self-serving way from the set of political narratives offered by parties; and (3) they differ in the degree to which they ascribe the cause for outcomes to outside scapegoats, as opposed to the inherent

[3]Readers who are interested in understanding the main mechanisms while doing away with the complexities of the main model are referred to Appendix A after reading Section 3.

characteristics of an agent. Trait (1) is a common way to describe narratives in economics and political science (Eliaz and Spiegler, 2020; Levy et al., 2022); other instructive ways to define narratives exist (e.g. Bénabou et al., 2018), but are less suitable to this paper. In this formulation, a narrative is a tool that allows one to make conclusions about the unobserved causal relations behind observable outcomes.

Trait (2) is again common in the literature, reflecting the simple fact that it is more convenient to believe narratives that paint one's outcomes in a more positive light. In the context of this paper, motivated reasoning implies that voters with undesirable outcomes will prefer scapegoating narratives, while those with good outcomes will not.

The view that beliefs are frequently self-motivated has gained traction in psychology and the social sciences (Gino et al., 2016; Haidt, 2012). Motivated beliefs have also recently entered formal models of political economy, as exemplified by Levy et al. (2022) and Eliaz et al. (2022); yet, in these approaches, data disciplines belief formation. The main mechanism is then driven by the discrepancy between a true causal model and prevalent misspecified beliefs. In contrast, in my approach, motivated beliefs are disciplined only by the incentives of political parties and voters. This approach can be viewed as giving a theoretical grounding for a "post-truth" view of political competition, where parties and voters choose which "alternative fact" to believe. The purpose of this is to give narratives as large a role in political competition as possible. It should be viewed as a complement to existing approaches, rather than a substitute, which is more applicable the more disputed the facts are.[4]

Trait (3) of narratives, to focus on their scapegoating property, is a deliberate reduction in scope suitable for the study of populist parties. The use of scapegoats, be that an "elite" or an outside group such as immigrants, is integral to the definition of populism (Mudde and Kaltwasser, 2017), and confirmed in empirical work that studies scapegoating tendencies of populists based on parliamentary speeches (Tzelgov and Olander, 2018; Magnusson et al., 2018). Conceptualised in this way,

[4]The approach can be considered a special case of the approach of Little (2019), when the incentive to adopt motivated beliefs overwhelms the incentive to match observable data.

narratives lie between two extremes: *narratives of blame* which scapegoat outside factors for observed outcomes; and *narratives of agency* which instead underscore the importance of an agent's personal traits. To give a concrete example, consider the potential determinants for an individual's income level. Narratives of agency would ascribe income differentials to differences in individual characteristics, be that education, work attitude, or similar attributes. Narratives of blame would instead highlight external factors unrelated to the individual, such as immigration or globalisation shocks. Intermediate narratives allow outcomes to potentially depend on both intrinsic characteristics and external factors.

In my framework, parties aim to maximise their vote share by freely choosing their narratives. There are two parties with a fixed economic position: a left party that supports higher taxes, and a right party that opposes them. In all other ways, the parties face symmetric conditions; in particular, they do not have a way to propose policies in the non-economic dimension, nor do they have a preference for one set of narratives over another. Assuming a preference of a party for either specific narratives or non-economic policies would defeat the purpose of this exercise, as I wish to explain differences in populist tendencies solely based on the different (right or left) economic stance of parties. I also assume that parties face symmetric voter bases, at least in the baseline model—again, to not skew the results ex-ante towards one where one of the parties has an advantage. I do allow for such asymmetries in extensions to the baseline model (Section 4.3), and also examine the (local) incentives of parties to change their policies (Section 4.4), which yields additional insights that are explained towards the end of the introduction.

One apparent puzzle in the context of right-wing populism is the tendency of voters to vote against their economic self-interest (Hochschild, 2016): low-income voters backing right-wing politicians that oppose higher taxes, and high-income voters backing left-wing politicians. Such patterns cease to be real puzzles when allowing for two dimensions in voter characteristics: one economic and one non-economic. This is the approach I take in this paper, similar to other theoretical work in the area (e.g. Besley and Persson, 2021; Bonomi et al., 2021). I deliberately leave the

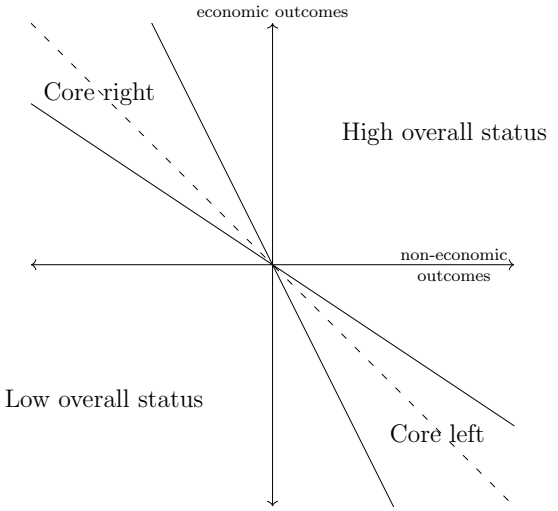
interpretation open as to what the non-economic dimension may be, but reasonable options include cultural factors, such as education (as highlighted by Gethin et al., 2022), religious beliefs, or values held. These non-economic factors may or may not correlate with economic factors; indeed, the relationship between the two will be a key factor in shaping the type of populism that arises in equilibrium. Similarly to economic factors, I assume that non-economic factors have an inherent (social) desirability that allows a ranking of better versus worse non-economic outcomes. Indeed, even if not universally agreed upon, many non-economic outcomes allow a ranking, such as educational achievements. Moreover, some cultural views (such as racism or sexism) are understood to be socially less desirable than others even by those who hold them—a factor that complicates empirical analysis of surveys as respondents may lie due to social desirability response bias (Stout et al., 2021).

In this conceptualisation, voters can be depicted as points in two-dimensional space, with "economic outcomes" on one axis and "non-economic outcomes" on the other, as shown in Figure 2. Together, the two outcomes shape a voter's *status*. Status considerations that are shaped by economic outcomes have a long tradition in models of political economy (Shayo, 2009, e.g.); the two-dimensional approach of status I explore here will be useful in analysing the forms of populism that arise.

Thus, in my framework, both economic and non-economic outcomes shape status. The weights of these two components depend on the relative *salience of cultural issues*: the higher that salience, the more do non-economic outcomes determine status. To show how economic and non-economic outcomes affect status, Figure 2 plots the "cutoff" weighted sum of economic and non-economic outcomes in a dashed line, which delineates high-status voters (above the line) from low-status voters (below the line). If the salience of cultural issues is relatively low, overall status is mainly determined by economic outcomes, and the dashed line will be flat; in contrast, it will be steep when the salience of cultural issues is high.

A voter's status determines her preference for the two instruments of political competition explored in this paper: redistributive policy and political narratives. As usual in economic analysis, preferences for the former are shaped by income,

Figure 2: Voters' status and core supporters



Notes: The figure shows the classification of voters by their non-economic and economic outcomes. The dashed line separates voters with high status (above the line) from those with low status (below the line). Within the solid lines are voters who form the core right (in the second quadrant) and core left supporters (in the fourth quadrant).

with voters that score low on the economic dimension preferring more redistribution. Narrative preference, on the other hand, is shaped by the overall status of a voter. As narratives speak to whether or not observable outcomes reflect inherent unobserved characteristics, voters with a high status will prefer narratives of agency and those with low status narratives of blame. This is how the assumption of motivated reasoning, discussed at the start of the section, enters the model.

Voters base their voting decision on both their policy and narrative preferences. They vote sincerely by backing the party whose policy-narrative pair they prefer, and are more likely to turn up for the election the more their preferences accord with what their preferred party has to offer.

Core and swing voters. The framework naturally gives rise to core and swing voters. Voters who are close to the delineation line between high and low status (dashed line in Figure 2) do not have a strong narrative preference and will base their vote only on their economic policy preference. They form the core support of the left and the right party, in the sense that they will always vote for them regardless of the political narratives chosen. In light of competition in narratives, the framework shows that the core support is formed of people whose economic and non-economic outcomes are negatively related. For the left party, the core comprises low-income

voters with desirable sociocultural outcomes; and for the right it comprises high-income voters with undesirable sociocultural outcomes.

Voters located outside of this core are swayable in their voting decision by the narratives offered by parties. They have a more clearly defined status as they are further away from the delineation line, and therefore care more about the cross-party differences in narratives. Among these swing voters are those who have correlated economic and non-economic outcomes: the "left behind" in the third quadrant and the "elite" in the first quadrant.

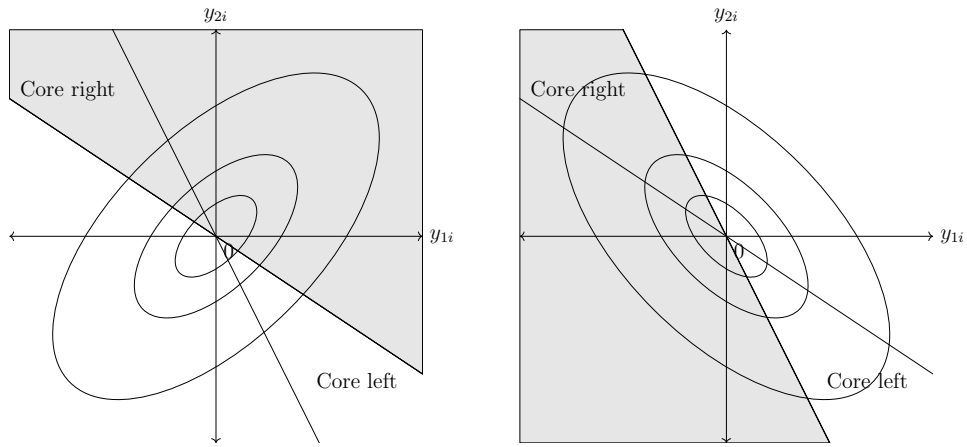
Hence, one insight of the framework is that it may not be overly surprising that left-behind voters are willing to vote against their economic interest by backing a right-wing populist. With their economic and non-economic grievances, left-behind voters place a higher weight on narratives relative to policies, and are ready to back a party that offers the right narrative even at the cost of supporting a disadvantageous economic policy. Symmetrically, the elite with its high economic and non-economic outcomes is very willing to back a left-wing party that offers narratives of agency. Core support, meanwhile, comes from those who do not have a strongly defined overall status, and thus care more about economic policies than narratives.

Left-wing and right-wing populism. The simple set-up sketched above allows one to parsimoniously capture the differences between left-wing and right-wing populism. Right-wing populism describes the situation where the (economically) right party competes with narratives of blame and the left with narratives of agency; left-wing populism, accordingly, where the left offers narratives of blame and the right narratives of agency. These two populist regimes imply a very different structure of voter support, as illustrated in Figure 3. Under left-wing populism, parties align their narratives with the logic of their economic stance: the right, whose economic policy appeals to high-income voters, offers narratives of agency, which appeals to high-status voters, while the left offers narratives of blame and appeals to low-income, low-status voters (Panel *a*, where the grey area votes for right). The only voters who go against their economic policy interest are those whose non-economic outcomes stand in opposition with their economic outcomes (the white area in the

second quadrant, and the grey in the fourth). Under right-wing populism, in contrast, parties "misalign" their economic policy stance and their narratives: the right-wing populist now appeals most to high-income, low status voters; and the left-wing opponent to low-income, high status voters. In this case, voters with a positively related economic and non-economic outcomes will vote against their economic policy preference—the elite and the left behind (quadrants 1 and 3 of Panel *b*). Thus, left- and right-wing populism induce different splits in the electorate, with the former resulting in a split across more economic lines, and the latter across more non-economic, or cultural, lines.

Figure 3: Electoral split in left-wing and right-wing populism

(a) Left-wing populism (b) Right-wing populism



Notes: The figure shows the classification of voters by their non-economic outcomes (y_{1i}) and economic outcomes (y_{2i}). The grey area shows supporters of the right party, and the white area of the left. It also shows the contours of distributions of voters. Panel *a* refers to the situation where the left party offers narratives of blame and the right narratives of agency, and vice versa for Panel *b*.

What forces decide whether populism takes a right-wing or a left-wing character? As alluded to earlier, the key factor determining the form of populism is the relationship between economic and non-economic outcomes. To illustrate this, Figure 3 shows the contours of distributions with different correlations between economic and non-economic outcomes. A positive correlation (Panel *a*) favours left-wing populism: in this case, voters' preference for more redistribution will be positively related to a preference for narratives of blame. Parties optimally respond by aligning their narratives with the logic of their economic policy: the left offers narratives of blame

and the right narratives of agency.

Right-wing populism, in contrast, arises under a negative correlation between economic and non-economic outcomes (Panel *b* of Figure 3). High-income voters then prefer scapegoating narratives and low-income voters the opposite. The right optimally caters to their high-income core by leveraging narratives of blame, and the left responds in kind with narratives of agency.

Whether populism takes a left-wing or right-wing form is also determined by the salience of cultural issues. A higher such salience serves to break a correlation between economic status and narrative preference, thereby contributing to a shift to right-wing populism. The empirical counterpart can be seen as movements that decrease the status of social groups with higher incomes. Across Western democracies, such forces would include Black Lives Matter, #MeToo, as well as anti-capitalist and pro-green movements, all of which decrease the overall status of groups with higher income, especially conservative white men (Wolton, 2023). "Wokeness", when viewed through the lens of this model, is a force that breaks the correlation between high income and high social status, shifting the incentives of right-wing parties towards narratives of blame, and left-wing parties towards narratives of agency. The framework also gives an explicit theoretical interpretation of Rodrik (2018)'s observation that cultural globalisation shocks (e.g. an influx of immigrants) tend to lead to a rise in right-wing populism, while economic globalisation shocks (e.g. financial crises) will lead to left-wing populism.

Left-behind voters and economic policy. How does the role of the often-quoted "left-behind" voters (Hochschild, 2016; Norris and Inglehart, 2019; Gest, 2016) fit into this story? The line of reasoning of this paper that the correlation between the economic and non-economic outcomes of voters determines the form of populism does not rely on an increased role of left-behind voters. Yet, an additional layer of depth is reached when allowing for such an asymmetry in voter types. In particular, when voters are symmetrically distributed across the origin, so that there is neither a majority of low-income, nor of low-status voters, left-wing and right-wing populist equilibria will generally be unique, but will lead to an equal split of

the electorate across parties. In contrast, an asymmetry that gives a larger role to left-behind voters will tend to favour one of the two parties—the populist. However, the framework highlights one caveat: while a left-wing populist always benefits from more left-behind voters, a right-wing populist's success is highest under a high, but not too high, role of such voters.

While the baseline model is silent on economic policy changes, a simple model extension allows one to consider the incentives of parties to increase or decrease their tax policy stance. These incentives crucially depend on the role of left-behind voters. Under perfect voter symmetry, populists and their opponents face exactly symmetric incentives, so that both of them want to either increase or decrease the cross-party tax difference to the same degree. However, if left-behind voters play a more prominent role, right-wing populists wish to decrease the cross-party difference relative to a non-populist opponent, while left-wing populists wish to increase it. The intuition is simple: populists gain the votes of the left behind, who prefer higher taxes. Populists—no matter whether right-wing or left-wing—therefore have a stronger incentive to increase taxes with a higher weight given to these voters. By supporting higher taxes, a right-wing populist decreases the tax gap to the left, while a left-wing populist increases it to the right. This additional prediction of the model is in line with the observation that right-wing populist parties frequently emulate an increasingly left-wing economic agenda, e.g. the Polish far-right ruling party's (PiS) flagship social policy "500 plus", which constitutes a redistribution towards lower-income voters (Graniszewski et al., 2018).

2.2 Contributions to the literature

The main contribution of this paper is theoretical, as it offers a new framework to study the conditions that favour right-wing versus left-wing populists. However, it also adds to the empirical literature by jointly making sense of some of the patterns observed in the data. Based on the discussion in the previous section which intuitively describes the model and some of its insights, I will discuss how the paper contributes to the theoretical literature, and how the model squares with empirical

evidence.

Theoretical literature. This paper offers an explanation for the recent rise of right-wing populist relative to left-wing populist parties. This puzzling fact has not been fully accounted for in the theory literature, since existing work tends to mostly focus on the political demand side— voters—rather than the supply side—parties. While existing work concentrates on the causes that drive increased demand for populism, why this demand has been increasingly satisfied by right-wing rather than left-wing populists remains less clear (Guriev and Papaioannou, 2022).

The concentration on voter preferences has yielded several valuable insights into the increased demand for populism as well as its link to cultural polarization and identity politics. Besley and Persson (2021) provide a dynamic model of two-dimensional policy competition, linking the rise of populist (anti-immigration) policies to increased polarization and higher immigration salience. They also formulate the increased nationalism of parties as a commitment device: existing parties can credibly commit to anti-immigration policies by allowing more nationalistic members to enter their party. Bonomi et al. (2021) offer a framework of endogenous belief formation where voters slant their beliefs towards that of the cultural or economic group they identify with. The authors show that lower-income voters may support lower taxation if cultural divides are sufficiently salient. Using a different framework of biased beliefs, Alesina and Stantcheva (2020) argue that misperceptions about the unwillingness of migrants to work can be another reason for support of lower tax regimes. Grossman and Helpman (2021) add an identity component to a model of trade policy preferences. In their framework, economic and cultural changes can result in a "populist revolution" with higher equilibrium trade tariffs. Their framework builds on the seminal contribution of Shayo (2009), who proposes a model of endogenous social identification based on two considerations, similarity and group status. The model characterises the existence of an equilibrium where lower income voters support lower taxes whenever they identify with the nation as a whole, since lower taxes raise average incomes and thereby the status of the nation.

What unites the theories above is that they focus on the demand side: voter

policy preferences drive political considerations, and political parties cater to the median voter (or, in the case of Grossman and Helpman, 2021, the average voter). The advantage of such models is that they can focus on the very relevant question of what drives voter demand for populist policies. The disadvantage is that they cannot account for heterogeneity in party strategies—including the difference between right-wing and left-wing populism.

Some other papers have focused on the political supply-side in the context of populism, albeit not with respect to analysing right-wing versus left-wing populist tendencies. Levy et al. (2022) show that political cycles where power alternates between non-populists and populists (who have more simplistic worldviews and advocate more extreme policies) can emerge endogenously in a model of subjective belief formation. The dynamics are driven by the fact that one group of voters has misspecified beliefs about the true causes of policy outcomes and therefore overestimates the efficacy of its own preferred policy whenever the other group is in power. In a different framework, Eliaz et al. (2022) show that false scapegoating narratives can persist and be reinforced by observable data, which will appear to buttress the case for a scapegoating narrative even if in reality the true cause for bad policy outcomes lies elsewhere. The reason for this is that a false scapegoating narrative can identify a false causal link between the scapegoated party's being in power and bad political outcomes, which will not be proved wrong as long as that party remains out of power. Crutzen et al. (2020) develop a model where a new populist party can win the backing of the (majority) common people by explicitly running an anti-elitist platform. The intuition is that existing parties may cater to the (minority) elite's interest because the elite has an informational advantage over policy efficacy. If policies have strong distributive repercussions, the common people will back a populist party precisely because it features an anti-elitist platform and thereby signals that it has the common people's interest at heart. Ash et al. (2021)'s framework highlights the complementarity between identity politics, which influences a voters' social identification, and worldview politics, which influences their beliefs about how the world works. In their model, the asymmetry between parties stems from the

fact that the incumbent has a different set of strategies compared to the political challenger.

I contribute to this literature by analysing how populism relates to the economic stance of parties. I do this by focusing on one particular aspect common to populists: the tendency to explain outcomes perceived as negative by scapegoating outside factors. Who or what is being scapegoated is not being investigated, but relevant factors include immigration, globalisation or automation (Guriev and Papaioannou, 2022). My model complements existing notions of populism explored in other papers, like the tendency to use explicitly anti-elite platforms (Crutzen et al., 2020), or to adopt simplistic worldviews and extreme policies (Levy et al., 2022).

This paper also contributes to the wider strand of literature that combines sociological, economic and political methods to understand decision-making. Recent advances have been made to incorporate identity (Akerlof and Kranton, 2000, 2010; Besley and Persson, 2021; Ash et al., 2021), values (Besley and Persson, 2019; Enke, 2020a), norms (Young, 2015), or slanted beliefs (Martin and Yurukoglu, 2017; Bonomi et al., 2021) into models of behaviour. Within this literature, the role of narratives has recently gained in importance (Shiller, 2019). While narratives have different definitions across papers, Bénabou et al. (2018, p. 1) distinguish between two main concepts: narratives as "sense-making", versus narratives as "rationales or justifications". My paper is closely related to the former definition of narratives. Yet even within this literature on narratives as subjective causal models, several differing definitions exist. My approach builds on Eliaz and Spiegler (2020), who define narratives as directed acyclical graphs (DAGs) that arrange variables into a causal structure. The authors assume that variables are arranged into such a DAG in a self-serving manner, to maximise the utility of individuals. However, given an arrangement into a directed acyclical graph, which may or may not coincide with the true model, the authors assume that individuals use data consistently with their beliefs.

My framework builds on this notion of narratives as subjective causal models, but departs from it in several ways. First, in my framework there is no uncertainty as

to what direction the causation takes. Instead, there is uncertainty about how much of a role each factor plays. This is related to the concept of subjective causal models used in Levy et al. (2022), who assume that part of the electorate have simple models that generate incorrect beliefs about the true causes of policy outcomes. My second departure from existing models is to assume that narratives are only constrained by the incentives of parties and voters to form beliefs in a self-motivated way, and not by objective data. I hereby follow evidence that voters hold beliefs inconsistent with available data because they lack processing capacity (Enke and Zimmermann, 2019; Enke, 2020b)—or because choosing to interpret it differently portrays their outcomes in a more positive light (Gino et al., 2016). This approach can be seen as a limiting case of the framework of Little (2019), when the "directional" incentive of beliefs overwhelms the need for accuracy. Finally, I explore one particular dimension of narratives: the extent to which they scapegoat outside factors relative to intrinsic outcomes. This is not dissimilar to Eliaz et al. (2022), who study narratives that can either assign an important causal role to policy decisions or to political groups. However, the dimension of scapegoating I study is not related to a particular political group within the country, but instead more generally to outside factors that are not within the control of voters.

Empirical literature. There is a large empirical literature on the causes behind the rise of populism, which can be informed by the factors considered in this framework. As in the theory of this paper, the literature generally distinguishes between economic and cultural causes of populism, but rarely disentangles right-wing from left-wing populism. Among the economic causes, Autor et al. (2020) show that more negative trade exposure is related to a vote shift from more moderate politicians towards those with more extreme views. Colantone and Stanig (2018a,b) find that rising exposure to import shocks from China led to an increase in the "Leave" vote share in the UK, and the rise of nationalism and far-right parties across Europe. Also with regards to the vote in the UK to leave the EU, Fetzer (2019) demonstrates that regions with larger public finance cuts were more likely to support leaving the EU. Funke et al. (2016) show that financial crises increase the support for the ex-

treme right for a panel of 20 advanced economies covering 140 years. In a similar vein but for Sweden, Dal Bó et al. (2017) find that regions with a higher share of those affected by financial crises or a larger difference in the disposable income between secure "insiders" and vulnerable "outsiders" are correlated with the rise of the radical-right Sweden Democrats. Guiso et al. (2021) find that the 2008 financial crisis led to an increase in populist voting in European countries, with a particularly strong increase in the supply of left-wing populist parties. The effect of the financial crisis and globalisation shocks on populism in Europe was especially pronounced in Eurozone countries (Guiso et al., 2019). On the topic of automation, Frey et al. (2018) show that job automation increased the vote for Donald Trump.

Several papers highlight the importance of cultural causes for the rise of populism. Norris and Inglehart (2019) document a generational transition in the composition of voters. They argue that because cultural views differ across generations, this generational transition leads to a "cultural backlash", which contributes to the rise of populism. Mutz (2018) argues that non-economic components of group status, rather than financial factors, explain the success of Donald Trump in the 2016 presidential election. In a similar vein, Margalit (2019) finds only limited evidence for economic causes for populism, instead highlighting the importance of cultural factors. Within such cultural factors, Margalit et al. (2022) find that empirically, the most important drivers are cultural change due to immigration, rural resentment, and status anxiety. Enke (2020a) shows that differences in the moral values of voters strongly relate to the support for Trump.

Some empirical papers consider the interaction between cultural and economic factors in shaping populism. Guiso et al. (2022) find that economic uncertainty increases the vote for populist parties both directly as well as indirectly through cultural channels, by shaping anti-migrant sentiment and trust in the government. They particularly highlight the importance of decreased turnout in the success of populist parties. Rodrik (2018) argues that negative shocks attributable to globalisation cause grievances, increasing the demand for populism. Whether that demand is satisfied by left- or right-wing populist parties then depends on the salience of

cultural versus economic factors. If globalisation shocks are cultural in nature, e.g. because they consist of increased immigration, they will lead to right-wing populism; if they are instead economic shocks, they will lead to left-wing populism. My framework accords with the prediction that an increased salience of cultural issues will lead to right-wing populism, and can be seen as complementing the observations made by Rodrik (2018), by giving them an explicit theoretical foundation.

The insights of my paper hinge on the assumption that the scapegoating tendencies of narratives play a role in shaping populism. Several papers find quantitative evidence for narratives of blame being employed by populist parties, by analysing parliamentary speeches. For instance, Magnusson et al. (2018) show that speeches made by the Swedish Democrats, a populist right-wing party, frequently address the problem of migration. Moreover, they find that after the large inflow of migrants in 2015, the discourse against migrants became more negative across parties. Tzelgov and Olander (2018) reach similar conclusions in their study of the Swedish parliament. In particular, they find that a deterioration in the economy at the district level leads to an increase in the cultural framing of migrants. Relatedly, Funke et al. (2016) show that voters are more attracted towards parties that attribute blame towards migrants after financial shocks.

3 The model

The model consists of two types of agents: a unit mass of voters, indexed by i, and two political parties, labeled by $p \in \{L, R\}$ and referred to as *Left* and *Right*, respectively. Parties differ in their economic policy, and aim to maximise their vote share by offering narratives about the unobservable causes of voters' observable outcomes. Voters support the party whose narrative-policy pair they prefer. Their preferences for narratives are shaped by their observable status, constituted by economic and non-economic observables, which is a good starting point for the model description.[5]

[5]Note that the model section is kept deliberately concise. Section 2 presents a more thorough discussion of the intuition behind the model's assumptions.

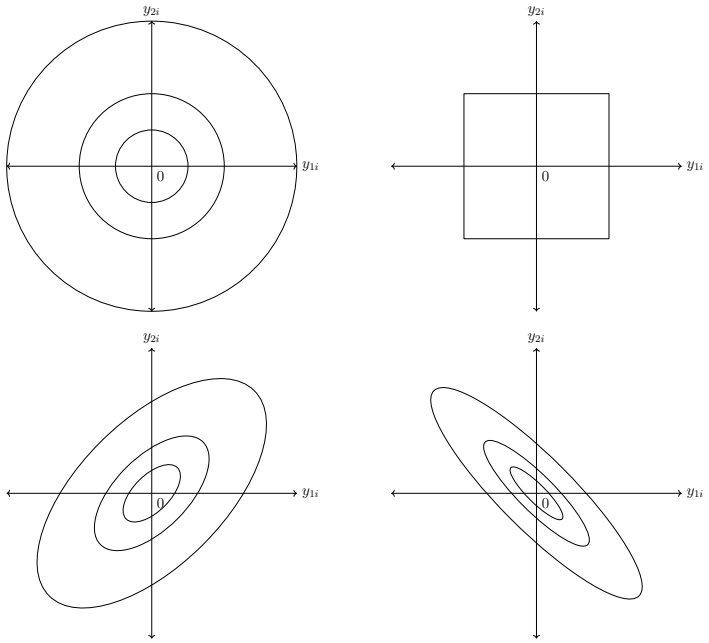
Observable status and unobservable causes. Let $y_i = (y_{1i}, y_{2i}) \in \mathbb{R}^2$ be a vector of observable outcomes, with joint density f, full support in $\mathbb{R}^2$, and $\mathbb{E}(y_{1i}) = \mathbb{E}(y_{2i}) = 0$; y_{1i} describes *non-economic outcomes* and y_{2i} *economic outcomes*. A voter i is said to have good economic (resp. non-economic) outcomes when $y_{1i} > 0$ (resp. $y_{2i} > 0$). The *overall status* of agent i is determined by the weighted sum $\chi y_{1i} + y_{2i}$, which is said to be high when it is positive and low otherwise. The weight $\chi > 0$ is referred to as the *salience of cultural issues*.

I assume that the distribution of outcomes y_i satisfies radial symmetry (Nelsen, 1993): For any pair $(y_{1i}, y_{2i}) \in \mathbb{R}^2$, $f(y_{1i}, y_{2i}) = f(-y_{1i}, -y_{2i})$. In other words, the distribution of voters is mirrored across the mean, which is located at the origin. Radial symmetry allows for a wide range of distributional patterns across y_{1i} and y_{2i}, including elliptical distributions such as the bivariate normal. Some examples are illustrated in Figure 4, which depicts contours of possible distributions. The top two subfigures illustrate examples of distributions that show no correlation between y_{1i} and y_{2i}, the left of which could come from a bivariate normal with 0 correlation, and the second from a bivariate uniform distribution. The bottom left and right subfigures constitute examples of elliptical distributions with, respectively, a positive and negative correlation.

Let $x_i = (x_{1i}, x_{2i})$ be a vector of unobservable variables, which are personal voter traits that are potential underlying causes for y_i. Similarly, let $\epsilon_i = (\epsilon_{1i}, \epsilon_{2i})$ be outside factors unrelated to i that may influence y_i. Suppose that x_i, ϵ_i are independently distributed also according to f, so the unconditional means are $\mathbb{E}(x_{1i}) = \mathbb{E}(x_{2i}) = \mathbb{E}(\epsilon_{1i}) = \mathbb{E}(\epsilon_{2i}) = 0$. Parallel to y_{1i} and y_{2i}, x_{1i} and ϵ_{1i} refer to potential causes of economic outcomes, while x_{2i} and ϵ_{2i} to causes of non-economic outcomes. While x_i and ϵ_i are unobservable, their expectations can be inferred conditional on y_i and on the narratives which define the causal relation between x_i and y_i.

Narratives and scapegoating. A narrative $n \in N$ is a function that maps unobservable latent variables x_i into observable outcomes y_i, where N is the set of all possible narratives. This definition of a narrative is very standard in the existing literature (Eliaz and Spiegler, 2020). To limit the scope of this paper, suppose that

Figure 4: Distributions of observable outcomes that satisfy radial symmetry



Notes: The figure shows the contours of possible distributions of voters by their non-economic outcomes (y_{1i}) and economic outcomes (y_{2i}), which satisfy the radial symmetry condition $f(y_{1i}, y_{2i}) = f(-y_{1i}, -y_{2i})$.

narratives take a simple "scapegoating" form: it is believed that each outcome y_{ji} with $j \in \{1, 2\}$ is either caused by an individual's unobserved characteristic x_{ji} (with some probability) or instead caused by the outside factor ϵ_{ji} (with the remaining probability). Thus, narratives differ in how much weight they assign to "scapegoats" ϵ_i, which have nothing to do with individual underlying characteristics x_i.[6] Hence, i's narrative about the data generating process for an outcome y_{ji} takes the following form:

$$y_{ji} = \begin{cases} x_{ji} & \text{w. p. } \nu_i \in [0, 1] \\ \epsilon_{ji} & \text{w. p. } 1 - \nu_i \end{cases}$$

This means that y_{ji} follows a mixture distribution, with mixture weights ν_i and $1 - \nu_i$. Since both x_i and ϵ_i follow the density f, y_i follows the same distribution no matter what the weights are.

With this formulation, a narrative n is uniquely described by the scalar ν_i, and

[6] This characterisation is similar to the notion that outcomes may depend on "luck" versus "skill", as explored in related literature (see, for instance, Little, 2019).

the space of all possible narratives is $N = [0, 1]$. Narratives lie between two extremes: *narratives of blame* at $\nu_i = 0$, which assign all the cause for outcomes to unobservable scapegoats ϵ_i; and *narratives of agency* at $\nu_i = 1$, which postulate that observable outcomes are tied to unobservable characteristics of the individual, x_i.

Beliefs. Beliefs are conditional expectations of unobservable variables, formed based on the narrative ν_i and the vector of observable variables y_i. Given that narratives specify a separate causal relation for each x_{ji} and the corresponding y_{ji}, all relevant information about x_{ji} is contained in y_{ji}, and none in $y_{\neg ji}$ for $\neg j \neq j$. In particular, $\mathbb{E}(x_{ji}|y_i, \nu_i) = \nu_i \mathbb{E}(x_{ji}|y_{ji} = x_{ji}) + (1 - \nu_i)\mathbb{E}(x_{ji}|y_{ji} = \epsilon_{ji}) = \nu_i y_{ji}$, as $\mathbb{E}(x_{ji}|y_{ji} = \epsilon_{ji}) = \mathbb{E}(x_{ji}) = 0$.

Parties. There are two parties, $p \in \{L, R\}$, which differ in their redistributive policies $\tau_p \in [0, 1]$, where $\tau_L > 0.5 > \tau_R = 1 - \tau_L$. These taxes determine the post-tax incomes of voters, $(1 - \tau_p)y_{2i}$. Therefore, τ_p constitutes a flat tax for everyone with positive income and a proportional transfer for everyone with negative income. A higher τ constitutes a higher level of redistribution, with $\tau = 1$ being perfect redistribution where everyone earns the same post-tax income of 0, while $\tau = 0$ denotes the case of no redistribution. Note that any τ_p leads to a balanced budget: because of radial symmetry, the transfer to low earners equals the total revenue from high earners, i.e. $\tau \int_{-\infty}^{\infty} \int_{-\infty}^{0} f(y_1, y_2) dy_2 dy_1 = \tau \int_{-\infty}^{\infty} \int_{0}^{\infty} f(y_1, y_2) dy_2 dy_1$.

The parties are fixed in their economic dimension and compete for votes by simultaneously choosing narratives $\nu_p \in [0, 1]$; in Section 4.4, I consider the incentives of parties to locally adjust their economic policy. Their objective is to choose the narrative that maximises their vote share.

Voter utility. Voter utility is a function of observable economic and non-economic outcomes y_i, as well as of the self-esteem gained from their beliefs about their underlying unobservable traits x_i:

$$u(y_i, \nu_i, \tau_p) = \underbrace{y_{1i}}_{\text{non-economic outcome}} + \underbrace{(1 - \tau_p)y_{2i}}_{\text{economic outcome}} + \lambda \underbrace{\mathbb{E}[\chi x_{1i} + x_{2i}|y_i, \nu_i]}_{\text{self-esteem from traits}},$$

where $\lambda > 0$ refers to the weight on self-esteem, itself a function of the expecta-

tion of the unobservable variables x_{1i} and x_{2i}; χ is as before the relative salience of non-economic issues, and ν_i is i's narrative.

Intuitively, agents prefer to hold positive beliefs about their underlying traits (high x_{1i} and x_{2i}), and more so the more important is the role of self-esteem (high λ). This feature of the model captures motivated reasoning (see Section 2 for a discussion of the literature on this topic). It accords with the view that status is an important part of individual utility by affecting the self-esteem of voters (Shayo, 2009), and that self-esteem is driven by beliefs about the underlying unobservable traits of individuals (Wolton, 2023).

Under the baseline assumption on the form of narratives, voter utility simplifies to the following:

$$u(y_i, \nu_i, \tau_p) = y_{1i} + (1 - \tau_p)y_{2i} + \nu_i\lambda\left(\chi y_{1i} + y_{2i}\right) \tag{1}$$

The effect of economic outcomes is mediated by the redistributive instrument τ_p, while the effect of self-esteem depends on the narrative of a voter ν_i. The direct effect of non-economic outcomes, on the other hand, is unmediated, as parties are assumed to not be able to compete in that dimension (or, equivalently, to offer the same fixed policy in that dimension). This simplification is a deliberate modeling choice, as the paper aims to explain the differing incentives for left-wing and right-wing parties to employ scapegoating narratives that stem only from differences in economic positions.

Timeline, voting, and equilibrium. The timeline is as follows:

1. Parties simultaneously choose narratives.

2. Voters observe narratives and independently choose which party to vote for.

This paper focuses on subgame perfect equilibria in pure strategies, where (a) each party p chooses the narrative ν_p that maximises its vote share, given the other party's choice and voter responses; and (b) each voter i votes for the party whose narrative-policy pair (ν_p, τ_p) maximises their utility $u(y_i, \nu_p, \tau_p)$. Note that strategic voting considerations pose no issue in this two-party setting.

Populism. Since this paper aims to endogenously explain whether populist tendencies come from the economic right- or left-wing, and since the only choice parties face is in their narratives, the notion of populism must be entirely related to narratives. I employ a very simple and restricted definition of populism in this paper:

Definition 1 *An equilibrium is* ***populist*** *if at least one party chooses narratives of blame.*

This definition does not do justice to populism in all its complexity, but it captures two important and salient aspects: first, that populist narratives are related to the employment of scapegoating narratives; and second, that populist narratives are "simple", in the sense that they offer simple explanations. In my framework, a populist narrative is indeed simple, as it ascribes the cause for any outcome exclusively to outside factors ϵ_i, and not to voters' intrinsic characteristics x_i.

Hence, there are three types of populism that can arise: *left-wing populism* at $\nu_L = 0, \nu_R \neq 0$; *right-wing populism* at $\nu_L \neq 0, \nu_R = 0$; and the *blame game* at $\nu_L = \nu_R = 0$. Whenever using the terms right-wing populism or left-wing populism, I refer to the case with exactly one populist party.

Cultural polarization. To investigate the connection between populism and polarization, I employ the following definition of cultural polarization:

Definition 2 *An equilibrium is (more)* ***culturally polarized*** *if a (higher) proportion of voters vote against their preferred economic policy (but in line with their preferred narrative).*

This definition, again, does not consider every conceivable form of polarization. However, it explicitly speaks to the rising importance of non-economic cleavages (Gethin et al., 2022). According to this definition, higher polarization relates to rising cleavages within an economic class, as more voters vote against their economic policy preference.

The relation between economic and non-economic outcomes. Finally, it will be helpful for future propositions to denote the relation between economic and

non-economic outcomes by $\rho \equiv \frac{\int_0^\infty \int_0^\infty f(y_1,y_2)dy_2dy_1}{\int_{-\infty}^0 \int_0^\infty f(y_1,y_2)dy_2dy_1}$, i,.e. the relative density of the first and second quadrant. By radial symmetry, this is equal to the relative density of the third and fourth quadrant, $\frac{\int_{-\infty}^0 \int_{-\infty}^0 f(y_1,y_2)dy_2dy_1}{\int_0^\infty \int_{-\infty}^0 f(y_1,y_2)dy_2dy_1}$. A positive relationship between economic and non-economic outcomes occurs when $\rho > 1$ and a negative one when $\rho < 1$. As the next section will show, the parameter ρ influences whether populism takes a right-wing or left-wing form.

4 Analysis

To isolate the effect of the model's assumptions, I will start with the most parsimonious version, and sequentially add layers of complexity. Starting with the set-up from the previous section, Section 4.1 makes some preliminary observations. It first shows that the model naturally gives rise to core and swing voters, the relative size of which depends on cross-party differences in taxes ($\Delta\tau$) and the importance of narratives (λ). It then shows that right- or left-wing populism will create the most culturally polarized political outcomes. Which of the two forms of populism will be most polarizing in this sense will depend on the salience of cultural factors, χ. It also shows that there exists a continuum of equilibria in the baseline model, where each party wins half of the votes.

Section 4.2 offers the core propositions of the paper. To select among the continuum of equilibria, it endogenises voter turnout, making it dependent on the utility that voters gain from parties' narrative-policy combinations. It finds that right-wing populism will be the unique equilibrium if economic and non-economic outcomes are negatively related (low ρ) and the salience of cultural issues is pronounced (high χ). Left-wing populism will be unique under the opposite circumstances. It also argues that equilibrium will tend to select the less culturally polarized populist equilibrium.

Section 4.3 considers what happens when "left-behind" voters, with low economic and non-economic outcomes, play a larger role in the electorate. Left-wing populists unambiguously benefit from a larger role of left-behind voters (higher β). The result is more nuanced for right-wing populists, who benefit from an intermediate, but not too high, level of such voters.

Section 4.4 considers the incentives for right- and left-wing populists to adjust their tax policy stance. It shows that right-wing populists have a stronger incentive to decrease the tax policy gap compared to a non-populist opponents. In contrast, left-wing populists have a stronger incentive to increase the gap. Section 4.5 summarises the insights of the model and links them to the empirical literature discussed in Section 2.2.

All proofs are located in Section B of the Appendix. The Appendix also includes a simpler version of the model (Section A), which gives insights into the model's main mechanism at the loss of some generality.

4.1 Preliminary observations

Voting is determined by the relative preference for the right relative to the left party. Referring back to equation 1, this is given by

$$u(y_i, \nu_R, \tau_R) - u(y_i, \nu_L, \tau_L) = \Delta\tau y_{2i} - \lambda\Delta\nu \left(\chi y_{1i} + y_{2i}\right), \tag{2}$$

where $\Delta\nu \equiv \nu_L - \nu_R \in [-1, 1]$ and $\Delta\tau \equiv \tau_L - \tau_R \in (0, 1]$. Whenever this difference is positive, i votes for *Right*, and otherwise for *Left*. If the difference is 0, the voter is indifferent, and votes for each party with equal likelihood. The model naturally gives rise to a continuum of core and swing voters, the relative size of which depends on parameter values, as the following proposition shows.

Proposition 1 *For any joint distribution f with full support in $\mathbb{R}^2$ that satisfies radial symmetry, there exists a mass of* ***core voters*** *who vote for Right and Left irrespective of political narratives. This mass is higher under a larger cross-party difference in tax policies (higher $\Delta\tau$) and a lower importance of self-esteem (lower λ).*

To understand the intuition for this proposition, note that i strictly prefers a low τ if and only if $y_{2i} > 0$, i.e. if they have good economic outcomes; and i strictly prefers high ν_i if and only if $\chi y_{1i} + y_{2i} > 0$, i.e. if they have high overall status. Referring to equation 2, one can see that voters with low absolute levels of status,

$|\chi y_{1i} + y_{2i}|$, relative to their financial preferences $\Delta\tau y_{2i}$, will always vote according to the latter. In particular, if $\chi y_{1i} + y_{2i} \in (-\frac{\Delta\tau}{\lambda} y_{2i}, \frac{\Delta\tau}{\lambda} y_{2i})$, then i will always vote for the party which gives its preferred economic policy, no matter what the narrative is. Intuitively, if the overall status of a voter is close to 0, they do not strongly prefer one narrative over the other, and cannot be swayed by narrative competition. The larger the cross-party difference in policies is or the lower the weight given to self-esteem, and hence narratives, is, the larger will this set of voters be.

In light of competition in narratives, one insight of the framework is that core voters of parties are those that face a negative relationship between their economic and non-economic outcome (as they are centered around the line $y_{2i} = -\chi y_{1i}$). Hence, high-income voters with unfavourable sociocultural outcomes form the core right support, and low-income voters with desirable sociocultural outcomes that of the left. In contrast, swing voters will consist of the "left-behind" that scores low in both dimensions, and an "elite" that scores high. This insight speaks to the puzzle as to why left-behind voters are willing to vote against their economic self interest and back a right-wing populist (Hochschild, 2016). By the logic of the model, the left behind have a more clearly defined overall status and therefore tend to care more about narratives than economic policies. To satisfy their narrative preference, they are therefore more willing to vote against their economic policy preference. The same logic applies to why an elite would back left-wing parties.

The importance of high λ for the emergence of populist narratives. The proposition hints at a factor that could explain the overall rise of populism as shown in Figure 1: a change in the importance of self-esteem (λ) in voters' preferences. For small λ, it makes little difference what narratives parties offer, as they face mostly core voters. By implication, there is a small incentive to select an extreme narrative—including a populist one. By contrast, a larger λ gives a stronger incentive to provide populist narratives, as more voters can be influenced in their voting decision. Hence, even though this is not formally explored in this paper, an increase in λ could be used to explain the overall rise in populism. This idea will be revisited in subsection 4.2 Lemma 3, where turnout is endogenous and hence there is a strict

Table 1: Definition of voter types

High-income voters	Low-income voters
$HH \equiv \{y_i \colon y_{2i} > 0 \wedge y_{1i} > 0\}$	$LL \equiv \{y_i \colon y_{2i} < 0 \wedge y_{1i} < 0\}$
$HL_1 \equiv \{y_i \colon y_{2i} \geq -\frac{\lambda\chi}{\lambda-\Delta\tau} y_{1i} \geq 0\}$	$LH_1 \equiv \{y_i \colon y_{2i} \leq -\frac{\lambda\chi}{\lambda-\Delta\tau} y_{1i} \leq 0\}$
$HL_2 \equiv \{y_i \colon -\frac{\lambda\chi}{\lambda-\Delta\tau} y_{1i} > y_{2i} \geq -\chi y_{1i} \geq 0\}$	$LH_2 \equiv \{y_i \colon -\frac{\lambda\chi}{\lambda-\Delta\tau} y_{1i} < y_{2i} \leq -\chi y_{1i} \leq 0\}$
$HL_3 \equiv \{y_i \colon -\chi y_{1i} > y_{2i} \geq -\frac{\lambda\chi}{\lambda+\Delta\tau} y_{1i} \geq 0\}$	$LH_3 \equiv \{y_i \colon -\chi y_{1i} < y_{2i} \leq -\frac{\lambda\chi}{\lambda+\Delta\tau} y_{1i} \leq 0\}$
$HL_4 \equiv \{y_i \colon -\frac{\lambda\chi}{\lambda+\Delta\tau} y_{1i} > y_{2i} \geq 0\}$	$LH_4 \equiv \{y_i \colon -\frac{\lambda\chi}{\lambda+\Delta\tau} y_{1i} < y_{2i} \leq 0\}$

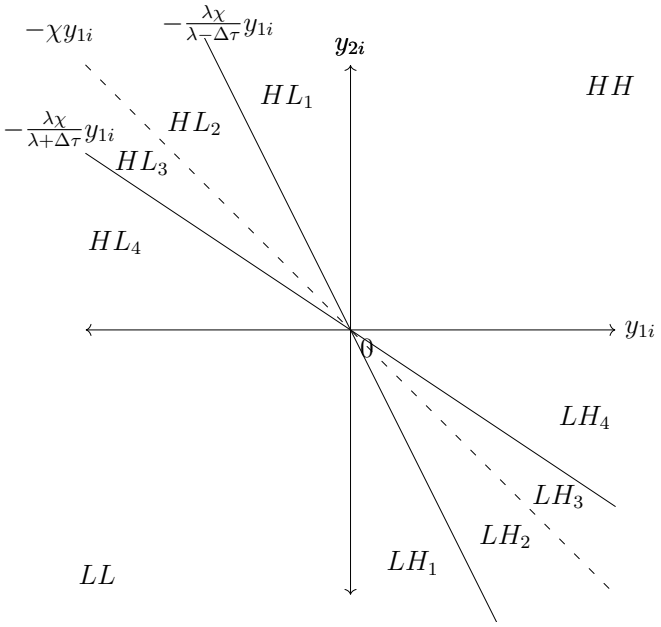
incentive to use extreme narratives.

To focus on the case where narratives play a sufficiently large role in political competition, I henceforth restrict the parameter space by assuming that $\lambda \geq 1$, which ensures a sufficiently large role of narratives for any $\Delta\tau$.

Voter types and notation. Figure 5 shows a helpful way to group voters into types based on their preferences across narratives and policies, as well as whether they are core or swing voters. The dashed line at $y_{2i} = -\chi y_{1i}$ delineates voters with high overall status from those with low overall status, and thus those who prefer narratives of agency as opposed to narratives of blame. To further distinguish between voters with different narrative-policy preference orderings, it will be helpful to group high-income and low-income voters into five groups each, as defined in Table 1 and illustrated in Figure 5. It will be useful to denote a combination of adjacent types by adding a dash in between, e.g. $LL - LH_2 = LL \cup LH_1 \cup LH_2$. Then, the density of a set of voters G is given by $\iint_G f(y_1, y_2) dA$.

Core voter types as given by proposition 1 are HL_2 and HL_3 for *Right*, and LH_2 and LH_3 for *Left*. All the remaining voters are swing voters, in the sense that their vote depends on the cross-party narrative difference $\Delta\nu$. How the remaining voters vote can be deduced from equation 2. When setting it equal to 0, we can find that the line of indifference between *Left* and *Right* is given by $y_{2i} = -\frac{\lambda\Delta\nu\chi}{\lambda\Delta\nu-\Delta\tau} y_{1i}$. Whether *Right's* supporters lie above or below that line depends on the sign and size of $\Delta\nu$. In particular, denote a voter's decision to vote for *Right* by $\mu(y_i, \Delta\nu) = 1$, and for *Left* by $\mu(y_i, \Delta\nu) = 0$, which is a function of a voter's status y_i and the cross-party narrative difference $\Delta\nu$. The vote of each i who is not indifferent between parties

Figure 5: Voter types



Notes: The figure shows the different types of voters as described in the text. Voter types are bounded by the two axes, as well as by the dashed line $y_{2i} = -\chi y_{1i}$, which delineates high-status from low-status voters, and the solid lines $y_{2i} = -\frac{\lambda\chi}{\lambda-\Delta\tau}y_{1i}$ and $y_{2i} = -\frac{\lambda\chi}{\lambda+\Delta\tau}y_{1i}$, between which are located the core voter types HL_2 and HL_3 for *Right*, and LH_2 and LH_3 for *Left*.

(i.e. $y_{2i} \neq -\frac{\lambda\Delta\nu\chi}{\lambda\Delta\nu-\Delta\tau}y_{1i}$) is then given by the following:

$$\mu(y_i, \Delta\nu) = \begin{cases} 1 \text{ if } & \Delta\nu > \frac{\Delta\tau}{\lambda} \wedge y_{2i} < -\frac{\lambda\chi\Delta\nu}{\lambda\Delta\nu-\Delta\tau}y_{1i} \\ 1 \text{ if } & \Delta\nu < \frac{\Delta\tau}{\lambda} \wedge y_{2i} > -\frac{\lambda\chi\Delta\nu}{\lambda\Delta\nu-\Delta\tau}y_{1i} \\ 1 \text{ if } & \Delta\nu = -\frac{\Delta\tau}{\lambda} \wedge y_{1i} < 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

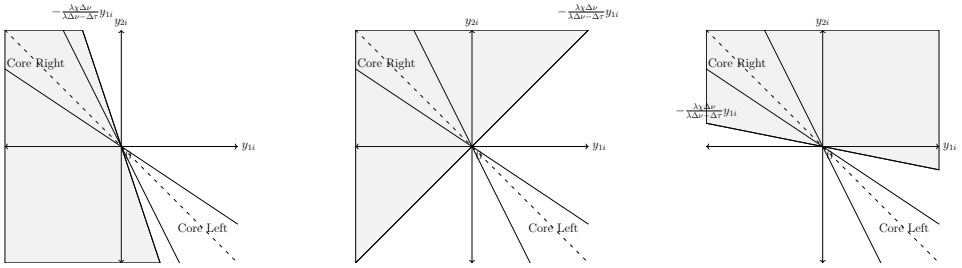
Figure 6 shows the electoral split across values of cross-party narratives $\Delta\nu$, where the grey area denotes the types who support *Right*. Moving from the left towards the right panels constitutes a decrease in $\Delta\nu$, i.e. an increase in ν_R or decrease in ν_L, which increases the proportion of types with high overall status (above the dashed line) that vote for *Right*.

It will be helpful to denote the set of left-wing voters by $L(\Delta\nu) \equiv \{y_i \colon \mu(y_i, \Delta\nu) = 0\}$, and the set of right-wing voters by $R(\Delta\nu) \equiv \{y_i \colon \mu(y_i, \Delta\nu) = 1\}$. The sets are bounded by the line $y_{2i} = -\frac{\lambda\chi\Delta\nu}{\lambda\Delta\nu-\Delta\tau}y_{1i}$, and hence depend on $\Delta\nu$. Thus, the density of *Left's* and *Right's* voters are $\iint_{L(\Delta\nu)} f(y_1, y_2)dA$ and $\iint_{R(\Delta\nu)} f(y_1, y_2)dA$, respec-

tively.

Figure 6: Voting patterns across different cross-party narrative differences $\Delta\nu$

(a) $\Delta\nu \in (\frac{\Delta\tau}{\lambda}, 1)$ (b) $\Delta\nu \in (0, \frac{\Delta\tau}{\lambda})$ (c) $\Delta\nu \in (-1, 0)$



Notes: The figure shows support for *Right* in grey versus support for *Left* in white, for different values of cross-party narrative differences $\Delta\nu$. The dashed line delineates high-status (above) from low-status (below) voters. Moving from Panel a to Panel c, $\Delta\nu$ decreases, and hence the narratives of *Left* become more scapegoating relative to to right.

The difference between right-wing and left-wing populist tendencies. Figure 6 allows a first appreciation of the differences between right-wing (high $\Delta\nu$, Panel a) and left-wing (low $\Delta\nu$, Panel c) populist tendencies. Under left-wing populism, parties align their narratives with the logic of their economic stance: *Left*, whose economic policy appeals to low-income voters, offers narratives of blame, which appeal to low-status voters. *Right*, in turn, offers narratives of agency, appealing to high-income, high-status voters. This leads to an electoral split along more economic lines (see Panel c). The only voters who go against their economic interest are those whose non-economic outcomes stands in opposition to their economic outcomes (Panel c: white area in second quadrant, and grey area in fourth quadrant).

Under right-wing populism, in contrast, parties "misalign" their narratives with their economic stance: *Right* now appeals to high-income, but low-status voters; *Left* to low-income, high-status voters. This leads to a split in the electorate along more cultural lines. Now, voters with positively related economic and non-economic outcomes vote against their economic policy preference (Panel a: first and third quadrants). As we will see next, this difference in electoral splits between left- and right-wing populism determines which of these two regimes is more culturally polarizing.

Populism and cultural polarization. If there are no cross-party differences in narratives ($\Delta\nu = 0$), the line that separates *Right's* from *Left's* voters coincides with the x-axis, and the electoral split is determined solely by economic outcomes. For any other cross-party narrative differences, the electorate will be, to some extent, split across cultural lines.

It turns out that cultural polarization (as defined by Definition 2) is intimately linked to populism (as defined by Definition 1). The next proposition shows that while any set of narratives constitutes an equilibrium, the most polarized equilibrium must be populist.

Proposition 2 *Any* (ν_L, ν_R) *constitutes an equilibrium. The most culturally polarized equilibrium is right-wing populism with* $(\nu_L, \nu_R) = (1, 0)$ *if there is a sufficiently positive relationship between status and economic outcomes* ($\iint_{HH-HL_1} f(y_1, y_2)dA \geq \iint_{HL_4} f(y_1, y_2)dA$) *and left-wing populism with* $(\nu_L, \nu_R) = (0, 1)$ *if there is a sufficiently negative relationship* ($\iint_{HH-HL_1} f(y_1, y_2)dA \leq \iint_{HL_4} f(y_1, y_2)dA$).

The fact that any combination of narratives constitutes an equilibrium can be appreciated by recalling the assumption of radial symmetry, and noting that the electorate is always split by a straight line running through the origin for any $\Delta\nu$ (see Figure 6). Hence, each party will always receive half of the vote, no matter what strategy they choose.

The second part of the proposition builds on the fact that the non-economic incentives to vote for parties are strongest when they have the largest difference in narratives. This occurs at $\Delta\nu = 1$ and $\Delta\nu = -1$, i.e. right- and left-wing populism, respectively. Therefore, cultural polarization will be highest under one of these two populist equilibria. This can also be appreciated from Figure C.1 in the Appendix: any intermediate narrative (Panel *b*) will result in lower polarisation than one of the populist regimes with maximal narrative difference (Panels *a* or *c*).

Why is right-wing populism more culturally polarized than left-wing populism if there is a sufficiently positive relation between overall status and economic outcomes? Recall from the discussion of Figure 6 that right-wing populism induces voters with positively related economic outcomes and overall status to vote against

their economic preferences. Left-wing populism, in contrast, only induces voters with negatively related economic outcomes and overall status to vote against their economic preferences. Hence, if the relationship between economic outcomes and status is sufficiently positive, a larger number of voters will vote against their economic policy preference under right-wing than under left-wing populism.

Salience of cultural issues and polarization. The relationship between the overall status and economic outcomes of voters is influenced by the salience of cultural issues, χ, as it influences the weight of non-economic factors in determining overall status. A change in the salience of cultural issues can therefore change whether right- or left-wing populism is most polarizing. This is summarised by the next corollary.

Corollary 1 *An increase in χ increases polarization under left-wing populism, and decreases polarization under right-wing populism. If economic and non-economic outcomes are positively related ($\rho > 1$), right-wing populism is always most polarized. If economic and non-economic outcomes are negatively related ($\rho < 1$), there exists a $\bar{\chi}$ such that right-wing populism is the most polarized for all $\chi \leq \bar{\chi}$, and left-wing populism under the reverse weak inequality.*

To understand this, recall that the types that vote against their economic interest, and thereby contribute to polarization, are HH- and HL_1-types under right-wing populism, and HL_4 types under left-wing populism. The corollary shows that the former set decreases and the latter increases in χ.

A higher salience of cultural issues χ makes non-economic outcomes a more important determinant of overall status. This gives non-economic outcomes a larger role in narrative preferences. It thereby expands the set of voters who are willing to go against their economic policy preference based on non-economic outcomes (LH_4 and HL_4 types). On the other hand, it reduces the set of voters who are willing to vote against their policy preference based on economic outcomes (HL_1 and LH_1 types). Hence, higher χ increases polarization under left-wing populism, and decreases it under right-wing populism.

Changes in χ, therefore, can be interpreted as changing the relationship between a voter's overall status, $\chi y_{1i} + y_{2i}$, and their economic outcomes, y_{2i}. Beyond influencing which equilibrium is most polarizing, this relationship between the different components of overall status also impacts the existence of equilibria under endogenous turnout, the topic of the next subsection.

4.2 Endogenous turnout and equilibrium selection

The assumption of radial symmetry together with full turnout leads to a continuum of equilibria, as shown in the previous section. It offered a helpful first glimpse into the intimate link between populism and cultural polarization, yet it fails to predict what, if any, form of populism we should expect to arise.

Endogenous turnout. To make progress on equilibrium selection, I introduce turnout considerations into the model. There is also an empirical foundation for doing so, as the turnout channel has been found to be an important driver of the electoral success of populist parties (Guiso et al., 2022).

To model turnout, I assume that a voter i's probability of voting for their preferred party is increasing in the utility they receive from that party's policy:

$$q(y_{1i}, y_{2i}, \nu, \tau) = \frac{u(y_i, \nu, \tau) - \underline{u}(y_i)}{\overline{u}(y_i) - \underline{u}(y_i)}, \tag{4}$$

where $\overline{u}(y_i) \equiv \max_{\nu \in [0,1], \tau \in \{\tau_L, \tau_R\}} u(y_i, \nu, \tau)$, i.e. their utility under their most preferred narrative-policy pair; and $\underline{u}(y_i) \equiv \min_{\nu \in [0,1], \tau \in \{\tau_L, \tau_R\}} u(y_i, \nu, \tau)$, the utility from their least preferred narrative-policy pair.

Hence, when considering whether to vote for their preferred party, an individual compares their utility under the narrative-policy pair offered by that party with what they would get under their most preferred and under their least preferred combination. Their turnout probability is 1 if their preferred party offers the best possible narrative-policy combination, and strictly between 0 and 1 otherwise. The logic of this turnout function is similar to the assumption that political mobilisation is increasing in individuals' intensity of preferences for political platforms, as

explored in Eliaz et al. (2022) and Levy et al. (2022).

Denote $Q_G(\tau,\nu) \equiv \iint_G q(y_1,y_2,\tau,\nu)f(y_1,y_2)dA$ as the total mass of votes cast by voters in group G under (τ,ν). $Q_{L(\Delta\nu)}(\tau_L,\nu_L)$ and $Q_{R(\Delta\nu)}(\tau_R,\nu_R)$ thus denote the total mass of votes cast for *Left* and *Right*, respectively, under narrative difference $\Delta\nu$.

The turnout function q has a helpful symmetry property: take two voters i,j that have "opposite" status, i.e. $y_i = -y_j$. Their turnout will be the same under opposite narrative-policy pairs: $q(y_{1i},y_{2i},\tau,\nu) = q(y_{1j},y_{2j},1-\tau,1-\nu)$. The implications of this property are summarised by the following lemma:

Lemma 1 *Total votes for Left and Right are equal if the parties offer opposite narrative-policy pairs:* $Q_{L(\Delta\nu)}(\nu,\tau) = Q_{R(\Delta\nu)}(1-\nu,1-\tau)$. *Moreover, for any* G, $Q_G(\nu,\tau) + Q_G(1-\nu,1-\tau) = \iint_G f(y_1,y_2)dA$.

The first statement of the lemma provides a symmetry property that is very helpful in the coming proofs, as it allows one to directly compare the voters of *Left* by recasting them in terms of voters for *Right*, and vice versa. The second statement says that for any given group, adding up the turnouts under opposite narrative-policy pairs leads to the group's full turnout, a fact that will also be used in the coming proofs.

Another lemma relating to the turnout function that will be helpful for the remaining propositions is the fact that $\frac{\partial Q_{L(\Delta\nu)}(\nu,\tau_R)}{\partial\nu}$ and $\frac{\partial Q_{R(\Delta\nu)}(\nu,\tau_L)}{\partial\nu}$ are strictly monotonic in $\Delta\nu$:

Lemma 2 $\frac{\partial Q_{L(\Delta\nu)}(\nu_L,\tau_L)}{\partial\nu_L}$ *is strictly increasing in* $\Delta\nu$ *while* $\frac{\partial Q_{R(\Delta\nu)}(\nu_R,\tau_R)}{\partial\nu_R}$ *is strictly decreasing.*

The lemma is intuitive: the higher is $\Delta\nu \equiv \nu_L - \nu_R$, the higher is ν_L relative to ν_R, and the more high-status voters will vote for *Left* relative to *Right*. Since high-status voters prefer higher narratives, the turnout of *Left*'s voters will respond more positively to an increase in narratives when $\Delta\nu$ is high. For the same reason, the turnout of *Right*'s voters will respond more negatively to an increase in narratives under high $\Delta\nu$.

λ and the emergence of populism, revisited. Before moving to the main proposition of this subsection, let us consider the role of λ for shaping the incentives to change narratives:

Lemma 3 *The higher is λ, the more responsive is voter turnout to a change in narratives.*

This lemma links back to a discussion started in Proposition 1 of Section 4.1 about how a rising overall share of populists may be related to a higher importance of self-esteem in voter utility, λ. If λ is low, parties cannot influence the turnout of individual voters via changing narratives. Their incentive to increase or decrease them is therefore limited; in particular, as $\lambda \to 0$, the influenceability of turnout via a change in narratives approaches 0. If it were costly to offer extreme narratives, for instance because this required innovation, parties would be less likely to offer such narratives—including populist ones. To keep things simple, this paper assumes that narratives are chosen at no cost, and hence this thought is not formally explored by the paper. However, this provides some intuition as to how an increase in populism may be linked to a larger importance of self-esteem considerations.

Right-wing versus left-wing populism. We are now ready for the main proposition of this section, which characterises the types of equilibria under endogenous turnout. For this purpose, define $\kappa \equiv \frac{Q_{L(0)}(0,\tau_L)}{Q_{R(0)}(0,\tau_R)}$, which captures the relationship between economic outcomes and overall status, weighted by turnout considerations: κ is increasing in the fraction of low-income voters who have a low overall status, and decreasing in the fraction of low-income voters who have high overall status. To understand why, note that from Lemma 1, it follows that we can write

$$\kappa = \frac{0.5 - Q_{LH_3 - LH_4}(1, \tau_R)}{0.5 - Q_{LL - LH_2}(0, \tau_R)}$$

This makes it explicit that κ is increasing in the weight given to voters whose status and economic outcomes are positively related ($LL - LH_2$), and decreasing in the weight of voters where it is negatively related ($LH_3 - LH_4$; see Figure 5).

This relationship between economic outcomes and overall status determines the equilibrium that will arise:

Proposition 3 *Parties only choose the extreme narratives $\nu_p \in \{0, 1\}$. In equilibrium, each party receives an equal vote share. If the relationship between economic outcomes and overall status is sufficiently positive ($\kappa > 1$), the unique equilibrium is left-wing populism at $(\nu_L, \nu_R) = (0, 1)$. If it is sufficiently negative ($\kappa < 1$), the unique equilibrium is right-wing populism at $(\nu_L, \nu_R) = (1, 0)$. If $\kappa = 1$, any combination of extreme narratives constitutes an equilibrium.*

The first part of the proposition, that parties always choose extreme narratives, is related to the incentives of individual voters: each voter (except those located at the line $y_{2i} = -\chi y_{1i}$ who are indifferent) strictly prefer one narrative to the other. It follows that each party has a strict incentive to increase or decrease narratives.

Since parties face opposite voter types, they have opposite incentives in terms of setting narratives. Hence, except for the knife-edge case at $\kappa = 1$, the unique equilibrium will be where one party specialises in narratives of blame, and the other in narratives of agency. Which of the two parties adopts populist narratives, now, depends on κ.

A high κ reflects a positive (post-turnout) relationship between voters' economic outcomes and status: voters with high income will tend to be of high status, and those of low income of low status. This incentivises parties to match their narratives with the logic of their economic policy: *Left*, whose policy appeals more to low-income voters, adopts narratives of blame, which appeal to voters of low overall status. *Right*, in turn, satisfies high-income, high overall status voters by offering narratives of agency. This explains why $\kappa > 1$ leads to the left-wing populist equilibrium.

The opposite is the case when κ is low. In that case, voters tend to have negatively related economic outcomes and status: voters with high income will tend to have low overall status, and those with low income to have high overall status. *Left* then optimally offers narratives of agency to attract the votes of a low-income, cultural "elite", while *Right* will offer narratives of blame to appeal to high-income,

culturally disgruntled voters. Hence, $\kappa < 1$ leads to right-wing populism.

The relationship between overall status and economic outcomes thus determines what form populism takes—whether it comes from the left or the right wing of the economic policy spectrum. It follows that changes in this relationship can cause a flip from one form of populism to another. One such change can be in the form of an increase or decrease in the salience of cultural issues, χ, as the following proposition summarises:

Proposition 4 *Left-wing populism is the unique equilibrium under a sufficiently low salience of cultural issues (low χ). If economic and noneconomic outcomes are negatively related ($\rho < 1$), right-wing populism is the unique equilibrium under a sufficiently high salience of cultural issues (high χ).*

The proposition highlights how the salience of cultural issues χ affects the type of populism that emerges in equilibrium. The main mechanism is that χ changes the composition of voters (see Figure 5). If χ is low, non-economic outcomes form a negligible part of status. Then, low-income voters will be mostly formed of low-status types ($LL - LH_2$), and high-income voters of high-status types ($HH - HL_2$). *Left* can then offer the most preferred policy-narrative combination to low income voters by leveraging narratives of blame. *Right* does the same to high-income voters by leveraging narratives of agency.

The opposite is the case when χ is high. In that case, a large proportion of high-income voters will be of low status (HL_3 and HL_4), and a large proportion of low-income voters of high status (LH_3 and LH_4). If there is also a negative relationship between economic and non-economic outcomes, ($\rho < 1$), right-wing populism will be the unique equilibrium. Under a high enough χ and low ρ, most high-income voters will be of low status, and most low-income voters of high status. *Right* then does best by combining their low tax stance with narratives of blame; *Left* by combining their high tax stance with narratives of agency.

Thus, the above proposition gives another interpretation of what factors could explain a rise of right-wing populism: a higher salience of cultural issues and a decrease in the relationship between economic and non-economic outcomes. Together,

such changes increase the amount of low-income voters with high overall status, and high-income voters with low overall status. This incentivises parties to strategically "misalign" their narratives with the logic of their economic stance, leading to right-wing populism as the unique outcome.

Populism and polarization, revisited. Proposition 2 in the previous section described what determines whether right-wing or left-wing populism constitutes the most polarized equilibrium in the case where turnout was not endogenous. It turns out that the same forces which make a combination of narratives the most polarized also generally prevent it from being an equilibrium once turnout is taken into account, as the next corollary describes:

Corollary 2 *The equilibrium will be the less culturally polarized populist outcome as long as the mass of core voters is not too large. In particular, this will hold iff*

a) $Q_{LL-LH_1}(0,\tau_R) - Q_{LH_4}(1,\tau_R) < \min\{0, Q_{LH_3}(1,\tau_R) - Q_{LH_2}(0,\tau_R)\}$ *(right-wing populism is unique and least polarized), or*

b) $Q_{LL-LH_1}(0,\tau_R) - Q_{LH_4}(1,\tau_R) > \max\{0, Q_{LH_3}(1,\tau_R) - Q_{LH_2}(0,\tau_R)\}$, *(left-wing populism is unique and least polarized)*

We can see this immediately by considering the implication of the uniqueness of left-wing populism: $\kappa > 1 \iff Q_{LH_3-LH_4}(1,\tau_R) < Q_{LL-LH_2}(0,\tau_R)$. Compare this with the condition that left-wing populism is less polarized: $Q_{LH_4}(1,\tau_R) < Q_{LL-LH_1}(0,\tau_R)$. Hence, a sufficient condition for equilibrium to select the less polarized outcome is that the weight of the core voters, LH_2 and LH_3 (and by symmetry HL_2 and HL_3) is low enough. Indeed, when $Q_{LH_2}(0,\tau_R) = Q_{LH_3}(1,\tau_R) = 0$, the condition that left-wing populism is less polarized is the same as it being the unique equilibrium. The same reasoning holds for the polarization of right-wing populism.

The reason why equilibrium tends to select the less polarized outcome is that higher cultural polarization by definition implies that more citizens are voting against their economic preferences, which will tend to suppress turnout of these voters. Hence, generally parties have an incentive to act in a way that limits the polarization among their voters.

For equilibrium to select the most polarized populist outcome, a very particular and somewhat odd relationship between economic and non-economic outcomes would be required: that economic and cultural outcomes are positively related among swing voters, but negatively among core voters, or vice versa. For instance, left-wing populism can be the unique equilibrium outcome if there are sufficiently many low-income and low overall status voters among its core ($Q_{LH_2}(0, \tau_R) > Q_{LH_3}(1, \tau_R)$); but it will be most polarized if there are sufficiently many low-income, high overall status swing voters ($Q_{LL-LH_1}(0, \tau_R) < Q_{LH_4}(1, \tau_R)$).

4.3 Populist success and the "left behind"

The previous section has shown the conditions under which right-wing and left-wing populism will arise. Given the importance given to disgruntled voters in the literature on populism, it is perhaps surprisingly that the shape and uniqueness of populist equilibria does not hinge on a particular role of "left-behind" voters, and instead on the relationship between the components of voters' status.

Yet, without any asymmetry in voter types, the model fails to explain populist success. Indeed, under symmetry, parties can always guarantee to win half of the votes by offering the opposite narrative to the opponent, and therefore an equal split of votes will be the equilibrium outcome.

To explain the conditions for populist success, I give a larger role to the "left-behind" voter types, LL. There is a two-fold motivation for this choice. One stems from the literature, which stresses the importance of left-behind voters with economic and cultural grievances for the success of populist parties (Hochschild, 2016; Norris and Inglehart, 2019; Gest, 2016). Another comes from the model: one defining feature of populists is that they receive backing from left-behind voters.

To that end, let the total number of votes from voter i, $vote_i$, depend on whether or not voters are the "left-behind" LL types:

$$vote_i(y_i, \nu, \tau) = \begin{cases} (1+\beta)q(y_{1i}, y_{2i}, \nu, \tau)f(y_i) & \text{if } y_i \in LL \\ q(y_{1i}, y_{2i}, \nu, \tau)f(y_i) & \text{otherwise,} \end{cases}$$

where the weight $\beta > 0$ gives a larger role to left-behind voters. There are a few ways to interpret β: for instance, if i represents a household, and status is formed at the household level, then $\beta > 0$ means that there are more household members among those with low overall status. In this interpretation, β can be seen as multiplying the household density $f(y_i)$. Another, although arguably less plausible, interpretation is that the turnout of LL types is higher than of other types.

For this section and the next, to keep the analysis simple I only consider party narrative choices of extreme narratives $\nu_p \in \{0, 1\}$, and not of intermediate narratives $\nu_p \in (0, 1)$. Recall, however, that parties generally have an incentive to offer extreme narratives, as each voter prefers one narrative over another.

The next two propositions characterise the existence criteria (Proposition 5) and form that equilibria take (Proposition 6), for different values of β. For that purpose, define $\pi_{\nu_L,\nu_R}(\beta)$ as the mass of votes that *Left* receives relative to *Right* under the narrative pair (ν_L, ν_R).

Proposition 5 *Sufficient conditions for the existence of pure-strategy equilibria are as follows:*

1. *If $\kappa > 1$ (resp. $\kappa < 1$) and no solution exists to $\pi_{0,1}(\beta) = \pi_{1,1}(\beta)$ (resp. $\pi_{0,1}(\beta) = \pi_{0,0}(\beta)$), an equilibrium exists for all $\beta \geq 0$.*

2. *If $\kappa > 1$ (resp. $\kappa < 1$) and there exists a solution to $\pi_{0,1}(\tilde{\beta}_1) = \pi_{1,1}(\tilde{\beta}_1)$ (resp. $\pi_{0,1}(\tilde{\beta}_2) = \pi_{0,0}(\tilde{\beta}_2)$), then this solution $\tilde{\beta}_1 > 0$ (resp. $\tilde{\beta}_2 > 0$) is unique. Then, an equilibrium exists $\forall \beta \in [0, \tilde{\beta}_1]$ (resp. $\forall \beta \in [0, \tilde{\beta}_2]$).*

Proposition 6 *Consider values of β for which pure strategy equilibria exist, as given by Proposition 5. For any $\beta > 0$, only populist equilibria exist, and if there is a single populist party, it wins the majority of votes. The equilibria depend on the weight of the left-behind, β, as well as on the relationship between economic outcomes and overall status, κ. Let $\underline{\beta}, \overline{\beta}$ be defined implicitly by $\pi_{1,0}(\underline{\beta}) = \pi_{0,0}(\underline{\beta})$ and $\pi_{0,0}(\overline{\beta}) = 1$, which always have unique solutions, and where $\underline{\beta} < \overline{\beta}$.*

1. *If $\kappa > 1$, the unique equilibrium is left-wing populism and Left wins a strict majority $\forall \beta > 0$.*

2. *If $\kappa < 1$, right-wing populism is the unique equilibrium $\forall \beta < \underline{\beta}$, and both parties will be populist $\forall \beta > \underline{\beta}$. Right wins a strict majority $\forall \beta < \overline{\beta}$, and Left $\forall \beta > \overline{\beta}$.*

The analysis in Proposition 6 accords with Proposition 3, while also shedding light on how left-behind voters can affect the success of populist parties. As before, whenever $\kappa > 1$, left-wing populism is the unique equilibrium for any β considered here. Since in the left-wing populist equilibrium, *Left* receives the backing of left-behind voters, $\beta > 0$ benefits the party and grants it a majority.

This is evident from considering Panel *a* of Figure 7, which plots the payoffs from each pure-strategy profile, π_{ν_L,ν_R}, against the weight of left-behind voters β.[7] The thick dashes trace the equilibrium payoffs. The dotted line at $\pi = 1$ represents the payoff where *Left's* and *Right's* vote shares are equal; *Left's* payoff is higher above, and *Right's* below that line. At $\beta = 0$, there is no asymmetry in voters, and hence $\pi_{0,1} = \pi_{1,0} = 1$; the unique equilibrium is at $(\nu_L, \nu_R) = (0, 1)$, as each party's deviation (to $(1, 1)$ and $(0, 0)$ for *Left* and *Right*, respectively) would make the deviating party worse off. With higher β, $\pi_{0,1}$, $\pi_{0,0}$ and $\pi_{1,1}$ all increase (linearly), as across these narrative profiles *Left* gains the votes of the left-behind. Equivalent reasoning explains why *Left's* payoff under right-wing populism $\pi_{1,0}$ decreases with β. For any β, however, left-wing populism remains the unique equilibrium.

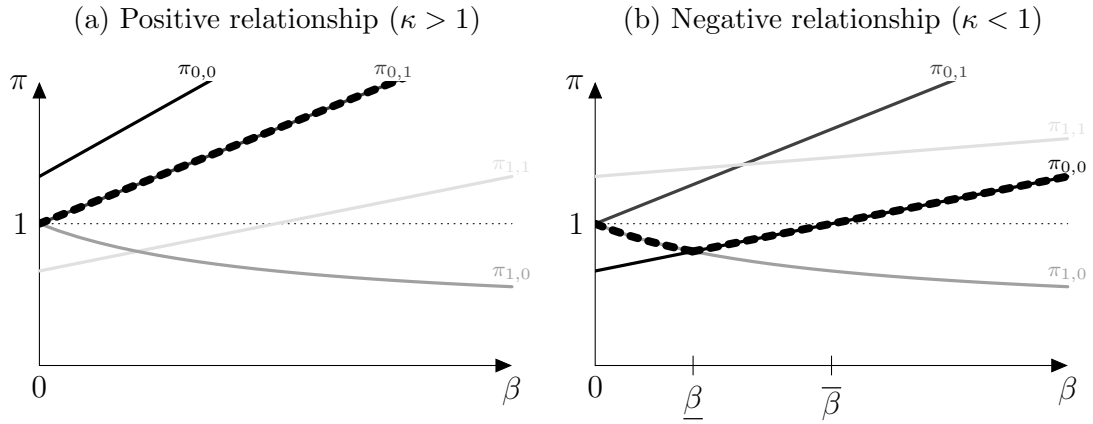
When $\kappa < 1$, the situation is more nuanced, as Panel *b* of the figure illustrates. Intermediate levels of $\beta \in [0, \underline{\beta})$ lead to right-wing populism; this is in line with Proposition 3, with the difference that the right-wing populist gains a majority of votes, for the same reasons as the left-wing populist does when $\kappa > 1$: left-behind voters back the populist, who gains from their larger role. However, once $\beta > \underline{\beta}$, *Left* is compelled to change its strategy and also offer narratives of blame in order

[7] The slopes in the example are chosen such that neither $\tilde{\beta}_1$ nor $\tilde{\beta}_2$ exists as given by Proposition 5, and hence a pure-strategy equilibrium exists for any β. This requires assumptions about the relative slopes of *Left's* relative votes across different narrative profiles. In particular, note that $\kappa > 1$ implies a flat slope for $\pi_{1,1}$; for more details, see the proof of Propositions 5-6.

to gain the votes of the left-behind. In general, that is not the optimal strategy for *Left* here: $\kappa < 1$ implies a negative post-turnout relationship between overall status and economic outcomes, so without an asymmetry in voter types, *Left* would wish to offer narratives of agency. Yet, the cost of doing so is too high when $\beta > \underline{\beta}$, and hence *Left* switches to offering narratives of blame. This strategy wins the mass of left-behind voters for *Left*, but at the cost of not offering the preferred narrative to voters with negatively related economic and overall status.

Hence, even in the case when $\kappa < 1$, once left-behind voters make up a large enough part of the electorate ($\beta > \underline{\beta}$), *Left* changes its strategy and also employs narratives of blame, so that both parties are populist. In the range $\beta \in (\underline{\beta}, \overline{\beta})$, *Left* now wins the votes of the left-behind, but loses the election. However, in this equilibirum any increase in β now benefits *Left*. Hence, the right only benefits from intermediate, but not too high, levels of β. The next corollary formalises this result.

Figure 7: *Left's* relative vote share (π) across different status relationships (κ) and weight of left-behind (β)



Notes: The figure plots *Left's* relative votes π_{ν_L,ν_R} under different narrative pairs, across the weight of left-behind voters β. The dotted line at $\pi = 1$ indicates an equal split of votes. The thick dashes trace equilibria payoffs. Panel a considers the case when $\kappa > 1$ and hence the unique equilibrium is left-wing populism. Panel b considers the case when $\kappa < 1$, where the equilibrium is right-wing populism for $\beta \in [0, \underline{\beta}]$, and where both parties are populist for $\beta \geq \underline{\beta}$; *Right* wins a strict majority if $\beta < \overline{\beta}$, and vice versa for *Left*.

Corollary 3 *Consider values of β for which pure strategy equilibria exist, as given by Proposition 5. A right-wing populist achieves the maximum vote share at an intermediate weight of left-behind voters, $\beta = \underline{\beta} > 0$. In contrast, a left-wing populist always benefits from the highest possible weight of left-behind voters.*

The corollary can be appreciated by again consulting Figure 7: under left-wing populism (Panel a), *Left's* equilibrium payoff (in thick dashes) is increasing in β. In the case of a right-wing populist party (Panel b), on the other hand, *Left's* payoff is lowest (and hence *Right's* highest) at $\beta = \underline{\beta}$, at which point *Left* is indifferent between choosing narratives of blame and of agency.

Left-behind voters, thus, are a double-edged sword for a right-wing populist. From its perspective, it is ideal if there are just enough left-behind voters to make *Left* indifferent between choosing either narrative, i.e. when $\beta = \underline{\beta}$. Beyond that point, however, more left-behind voters act to the detriment of a right-wing populist. This qualifies the argument stressed by the literature about the importance of left-behind voters for the rise of right-wing populist parties: some, but not too many, left-behind voters help a right-wing populist towards success.

4.4 Left-wing policy incentives for right-wing populists?

Another factor that may contribute to right-wing populist success is their adoption of policies usually associated with the left. A case in point is Poland's PiS party. While it is a right-wing party, its flagship social policy constitutes a redistribution towards larger, lower income families (Graniszewski et al., 2018). Similar arguments have been made with regard to Viktor Orbán's expansionist policies in Hungary (Ádám and Csaba, 2022). Boris Johnson's repeated pledge that the "NHS is not for sale", among his most frequent messages in his 2019 run for office (Lacatus and Meibauer, 2022), can be interpreted in a similiar light.

To study the incentives of parties to change their economic policies, I consider the local incentives to increase or decrease tax stances, conditional on being in an equilibrium as given by Proposition 6. For this purpose, I need to slightly adjust the turnout function, to account for the fact that the best and worst policy stance are now in principle the extreme policies $\tau \in \{0, 1\}$, rather than $\tau \in \{\tau_L, \tau_R\}$. The adjusted turnout function is then given by

$$\tilde{q}(y_{1i}, y_{2i}, \nu, \tau) = \frac{u(y_i, \nu, \tau) - \underline{\tilde{u}}_i}{\overline{\tilde{u}}_i - \underline{\tilde{u}}_i}, \tag{5}$$

where the difference relative to the previous turnout function given in equation 4 is how the best and worst utilities are defined, since they are defined relative to the set of all possible tax rates $\tau \in [0,1]$: $\bar{\tilde{u}}_i \equiv \max_{\nu\in[0,1],\tau\in[0,1]} u(y_i,\nu,\tau)$ and $\underline{\tilde{u}}_i \equiv \min_{\nu\in[0,1],\tau\in[0,1]} u(y_i,\nu,\tau)$. The two functions given by equations 4 and 5 are equivalent if $\tau_L = 1 - \tau_R = 1$.

The next proposition considers the incentives of populists who face a non-populist opponent to locally (in a derivative sense) change their tax stance.

Proposition 7 *Consider the equilibria with a single right-wing or left-wing populist party given by Proposition 6, when $\beta > 0$. A right-wing populist has a stronger incentive to decrease the cross-party tax difference $\Delta\tau$ than its non-populist opponent, while a left-wing populist has a stronger incentive to increase $\Delta\tau$ than its non-populist opponent.*

As a starting point for understanding this proposition, it is helpful to first consider incentives under no voter asymmetry, i.e. when $\beta = 0$. In that case, populist parties and their non-populist opponents have exactly symmetric incentives, as they face the opposite voter types with equal densities. It follows in this case that either both parties wish to increase the cross-party tax policy difference, or to decrease it. In other words, it cannot be that a populist party has different incentives in terms of changing $\Delta\tau$ to a non-populist when $\beta = 0$.

This changes when $\beta > 0$. In this case, the populist party, who sources votes from the left-behind LL-type voters, faces different incentives to a non-populist opponent who instead sources from HH-type voters. So the populist has a higher weight on the group that prefers higher taxes than its opponent has on the group that prefers lower taxes. It follows that populists have a higher incentive to increase taxes than its opponent has to decrease them.

The stronger incentive to increase taxes has different implications for different types of populists. For right-wing populists, stronger incentives to increase taxes mean incentives to decrease the policy gap; for a left-wing populist, it means increasing the policy gap. The higher weight of left-behind voters therefore shapes populists' incentives to change their policy stance. In particular, it sheds light on

why a right-wing populist may have an incentive to close the gap between its policy and that offered by its non-populist opponent.

4.5 Connections to the empirical literature

The theory developed in this paper can help piece together some of the regularities discovered in the burgeoning empirical literature on populism and changing cleavages discussed in Section 2. The rise of populist parties generally documented in the literature (Funke et al., 2020; Rooduijn and Taggart, 2019; Rodrik, 2018) can be seen as resulting from an increased importance of narratives—an increased λ in the parlance of the model (see Proposition 1 and Lemma 3). This increases the returns to specialising in extreme narratives, including populist ones. The model links populism to increased cultural polarization (Proposition 2), also an empirically well-documented pattern over past decades (Enke, 2020a).

The fact that in past decades populist parties increasingly stem from the right can be interpreted as being driven by an increasing importance of sociocultural issues, as found by Gethin et al. (2022) (lower ρ and higher χ; see Propositions 3 and 4). The same paper empirically documents an increasingly non-economic split in the electorate, with a "Brahmin" highly educated elite voting for the left party and a "Merchant" voter base for the right. The framework gives an intuition for why such a voter split should be expected precisely under the same conditions that lead to a surge of right-wing populist parties (again Propositions 3 and 4).

The prediction that right-wing populism should be expected when cultural issues are salient, while left-wing populism should be expected when economic issues are salient (see Propositions 3 and 4), squares with several papers in the literature. It is in line with Rodrik (2018)'s observation that globalisation shocks of a cultural nature tend to lead to right-wing populism, while those of an economic nature lead to left-wing populism. In so far as the financial crisis in 2008 increased the relative salience of economic cleavages, it also accords with the finding of Guiso et al. (2021) of a particularly strong rise in left-wing populism after the crisis.

The model also adds to the literature on the role of "left-behind" voters in the

success of populist parties (Hochschild, 2016; Norris and Inglehart, 2019; Gest, 2016, see Propositions 5-6). Indeed, the framework adds some nuance to the discussion by showing that a high, but not too high, role of left-behind voters is optimal for right-wing populists (intermediate $\beta = \underline{\beta}$; see Corollary 3). Finally, the framework can make sense of the observation that some right-wing populists adopt policies typically associated with the left (Graniszewski et al., 2018, see Proposition 7).

5 Conclusion

The framework presented in this paper offers a way to jointly interpret some recent changes in the nature of political competition across Western democracies. Through the lens of the model, the rise of right-wing relative to left-wing populist parties (Funke et al., 2020; Rooduijn and Taggart, 2019) is intimately linked to an increased non-economic dimension of political competition. As argued by Gethin et al. (2022) and paralleled in my framework, this gives weight to a new "cultural" elite among lower-income voters. My framework suggests that, by breaking the correlation between voters' overall status in society and their economic outcomes, the emerging cultural dimension favoured the relative rise of right-wing populists. The framework also weighs in on the debate about the importance of "left-behind" voters for the success of populists (Hochschild, 2016). Right-wing populists benefit from a higher mobilisation of left-behind voters only up to a point; if it is too high, the left party will adopt populist tendencies and capture a higher total share of votes. The framework can also explain why some right-wing populist parties adopt more left-leaning economic policies (Graniszewski et al., 2018), since that mobilises their left-behind supporters.

The model developed in this paper can be extended in several ways. For example, the key factors that drive the change in the shape of populism—salience of cultural issues and the correlation between voters' economic outcomes and overall status in society—are taken as exogenous in this paper. While explaining the dynamic evolution of such factors lies outside the scope of this paper, it would be worthwhile to explore how a feedback loop from political outcomes to the correlational structure

in voters' status, or the salience of cultural issues, could endogenously explain a dynamic transition towards right-wing populism. One potential avenue would be to endogenise the χ-parameter, which captures the relative salience of cultural issues. The effect of political outcomes on the salience of cultural issues and voters' status is plausibly mediated by media outlets, which again is disregarded in this paper but could be explored by future work.

The framework makes several predictions that are in principle testable. Some of the predictions seem to match empirical observations, as discussed at various points in the paper. However, a formal test of the model's predictions would be a valuable avenue for future work. An empirical decomposition of voter status into an economic and non-economic part would directly allow one to analyse whether their correlation relates to the type of populism that emerges, both across countries and over time. For the political supply side, parliamentary speeches offer a promising avenue to capture political narratives across parties and over time. Narratives of blame and of agency could be classified for several topics based on supervised machine learning tools, like the naive Bayes classifier, starting with attitudes towards migration and globalisation. Some of this is being done in current work-in-progress (Brzezinski and Findeis, 2021). Putting together new empirical measures of the demand side—voter's components of status—and supply side—parties' propensity to employ scapegoating narratives—would yield not only ways to check the model's predictions, but would constitute interesting empirical observations in themselves.

References

Abrajano, Marisa, and Zoltan L Hajnal. 2015. *White Backlash*. Princeton University Press.

Akerlof, George A, and Rachel E Kranton. 2000. Economics and identity. *The Quarterly Journal of Economics* 115(3): 715–753.

Akerlof, George A, and Rachel E Kranton. 2010. *Identity economics*. Princeton University Press.

Alesina, Alberto, and Stefanie Stantcheva. 2020. Diversity, Immigration, and Redistribution. *AEA Papers and Proceedings* 110: 329–34.

Ash, Elliott, Sharun Mukand, and Dani Rodrik. 2021. Economic Interests, Worldviews, and Identities: Theory and Evidence on Ideational Politics. *National Bureau of Economic Research Working Paper*.

Autor, David, David Dorn, Gordon Hanson, and Kaveh Majlesi. 2020. Importing political polarization? The electoral consequences of rising trade exposure. *American Economic Review* 110(10): 3139–3183.

Besley, Timothy, and Torsten Persson. 2019. Democratic Values and Institutions. *American Economic Review: Insights* 1(1): 59–76.

Besley, Timothy, and Torsten Persson. 2021. The Rise of Identity Politics: Policy, Political Organization, and Nationalist Dynamics. *LSE Working Paper*.

Bonomi, Giampaolo, Nicola Gennaioli, and Guido Tabellini. 2021. Identity, beliefs, and political conflict. *The Quarterly Journal of Economics* 136(4): 2371–2411.

Brzezinski, Adam, and Arduin Findeis. 2021. Patterns of Blame. *Unpublished Manuscript*.

Bénabou, Roland, Armin Falk, and Jean Tirole. 2018. Narratives, Imperatives, and Moral Reasoning. *IZA Discussion Paper 11665*.

Colantone, Italo, and Piero Stanig. 2018a. Global Competition and Brexit. *American Political Science Review* 112(2): 201–218.

Colantone, Italo, and Piero Stanig. 2018b. The Trade Origins of Economic Nationalism: Import Competition and Voting Behavior in Western Europe. *American*

Journal of Political Science 62(4): 936–953.

Crutzen, Benoit, Dana Sisak, and Otto H. Swank. 2020. Left Behind Voters, Anti-Elitism and Popular Will. Technical Report. *Tinbergen Institute Discussion Paper*.

Dal Bó, Ernesto, Frederico Finan, Olle Folke, Torsten Persson, and Johanna Rickne. 2017. Economic Losers and Political Winners: Sweden's Radical Right. *Working paper*.

Eliaz, Kfir, Simone Galperti, and Ran Spiegler. 2022. False Narratives and Political Mobilization. *arXiv preprint arXiv:2206.12621* .

Eliaz, Kfir, and Ran Spiegler. 2020. A Model of Competing Narratives. *American Economic Review* 110(12): 3786–3816.

Enke, Benjamin. 2020a. Moral values and voting. *Journal of Political Economy* 128(10): 3679–3729.

Enke, Benjamin. 2020b. What you see is all there is. *The Quarterly Journal of Economics* 135(3): 1363–1398.

Enke, Benjamin, and Florian Zimmermann. 2019. Correlation neglect in belief formation. *The Review of Economic Studies* 86(1): 313–332.

Farrar, Cynthia. 2016. The Video That Helps Explain Clinton's Loss. *Politico Magazine*, October 26. https://www.politico.com/magazine/story/2016/10/hillary-clinton-immigration-story-video-214392/.

Fetzer, Thiemo. 2019. Did austerity cause Brexit? *American Economic Review* 109(11): 3849–3886.

Frey, Carl Benedikt, Thor Berger, and Chinchih Chen. 2018. Political machinery: did robots swing the 2016 US presidential election? *Oxford Review of Economic Policy* 34(3): 418–442.

Funke, Manuel, Moritz Schularick, and Christoph Trebesch. 2016. Going to extremes: Politics after financial crises, 1870–2014. *European Economic Review* 88: 227 – 260.

Funke, Manuel, Moritz Schularick, and Christoph Trebesch. 2020. Populist leaders and the economy. *CEPR Discussion Paper* 15405.

Gest, Justin. 2016. *The new minority: White working class politics in an age of immigration and inequality*. Oxford University Press.

Gethin, Amory, Clara Martínez-Toledano, and Thomas Piketty. 2022. Brahmin left versus merchant right: Changing political cleavages in 21 Western Democracies, 1948–2020. *The Quarterly Journal of Economics* 137(1): 1–48.

Gino, Francesca, Michael I. Norton, and Roberto A. Weber. 2016. Motivated Bayesians: Feeling Moral While Acting Egoistically. *Journal of Economic Perspectives* 30(3): 189–212.

Graniszewski, Leszek, et al.. 2018. Janus-face of the family 500 Plus Programme. *Prace naukowe Uniwersytetu Ekonomicznego we Wrocławiu* (510): 67–74.

Grossman, Gene M, and Elhanan Helpman. 2021. Identity politics and trade policy. *The Review of Economic Studies* 88(3): 1101–1126.

Guiso, Luigi, Helios Herrera, Massimo Morelli, and Tommaso Sonno. 2019. Global crises and populism: the role of Eurozone institutions. *Economic Policy* 34(97): 95–139.

Guiso, Luigi, Helios Herrera, Massimo Morelli, Tommaso Sonno, et al.. 2022. Economic insecurity and the demand of populism in Europe. Working paper available at https://heliosherrera.com/populism.pdf.

Guiso, Luigi, Massimo Morelli, Tommaso Sonno, and Helios Herrera. 2021. The financial drivers of populism in europe. *BAFFI CAREFIN Centre Research Paper* (2021-166).

Guriev, Sergei, and Elias Papaioannou. 2022. The political economy of populism. *Journal of Economic Literature* 60(3): 753–832.

Haidt, Jonathan. 2012. *The righteous mind: Why good people are divided by politics and religion*. Vintage.

Hochschild, Arlie Russell. 2016. *Strangers in Their Own Land: Anger and Mourning on the American Right*. The New Press.

Keith, Tamara, Danielle Kurtzleben, Sarah McCammon, Meg Anderson, Brett Neely, and Jonquilyn Hill. 2016. FACT CHECK: Clinton Lays Out Economic Plan In Contrast To Trump. *NPR*, August 11.

https://www.npr.org/2016/08/11/489563362/clinton-to-lay-out-economic-plan-in-contrast-to-trump.

Lacatus, Corina, and Gustav Meibauer. 2022. 'Saying it like it is': Right-wing populism, international politics, and the performance of authenticity. *The British Journal of Politics and International Relations* 24(3): 437–457.

Levy, Gilat, Ronny Razin, and Alwyn Young. 2022. Misspecified politics and the recurrence of populism. *American Economic Review* 112(3): 928–62.

Little, Andrew T. 2019. The Distortion of Related Beliefs. *American Journal of Political Science* 63(3): 675–689.

Magnusson, Måns, Richard Öhrvall, Katarina Barrling, and David Mimno. 2018. Voices from the far right: a text analysis of Swedish parliamentary debates. *SocarXiv preprint*.

Margalit, Yotam. 2019. Economic insecurity and the causes of populism, reconsidered. *Journal of Economic Perspectives* 33(4): 152–170.

Margalit, Yotam, Shir Raviv, and Omer Solodoch. 2022. The Cultural Origins of Populism. *SSRN Working Paper 4001543*.

Martin, Gregory J., and Ali Yurukoglu. 2017. Bias in Cable News: Persuasion and Polarization. *American Economic Review* 107(9): 2565–99.

Mudde, Cas, and Cristóbal Rovira Kaltwasser. 2017. *Populism: A very short introduction*. Oxford University Press.

Müller, Jan-Werner. 2017. *What is populism?* Penguin UK.

Mutz, Diana C. 2018. Status threat, not economic hardship, explains the 2016 presidential vote. *Proceedings of the National Academy of Sciences* 115(19): E4330–E4339.

Nelsen, Roger B. 1993. Some concepts of bivariate symmetry. *Journal of Nonparametric Statistics* 3(1): 95–101.

Norris, Pippa, and Ronald Inglehart. 2019. *Cultural backlash: Trump, Brexit, and the rise of authoritarian populism.* Cambridge University Press.

Rodrik, Dani. 2018. Populism and the Economics of Globalization. *Journal of International Business Policy* 1: 12–33.

Rooduijn, Van Kessel S. Froio C. Pirro A. De Lange S. Halikiopoulou D. Lewis P. Mudde C., M., and P. Taggart. 2019. The PopuList: An Overview of Populist, Far Right, Far Left and Eurosceptic Parties in Europe. *www.popu-list.org*.

Shayo, Moses. 2009. A model of social identity with an application to political economy: Nation, class, and redistribution. *American Political Science Review* 147–174.

Shiller, Robert J. 2019. *Narrative Economics*. Princeton University Press.

Stout, Christopher T, Keith Baker, and Madelyn Baker. 2021. How social desirability response bias may lead to an overestimation of obama-trump voters. *Public Opinion Quarterly* 85(2): 694–705.

Stracqualursi, Veronica. 2017. 10 times Trump attacked China and its trade relations with the US. *ABC News*, November 9. https://abcnews.go.com/Politics/10-times-trump-attacked-china-trade-relations-us/story?id=46572567.

Teixera, Ruy. 2022. Democrats' Long Goodbye to the Working Class. *The Atlantic*, November 6. https://www.theatlantic.com/ideas/archive/2022/11/democrats-long-goodbye-to-the-working-class/672016/.

Tzelgov, Eitan, and Petrus Olander. 2018. Economic Decline and Extreme-Right Electoral Threat: How District-Level Factors Shape the Legislative Debate on Immigration. *Legislative Studies Quarterly* 43(4): 649–679.

White House. 2018. Remarks by President Trump on the Illegal Immigration Crisis and Border Security. *White House Archive*, November 1. https://trumpwhitehouse.archives.gov/briefings-statements/remarks-president-trump-illegal-immigration-crisis-border-security/.

Wolton, Stephane. 2023. White, Male, and Angry: A Reputation-based Rationale. *SSRN Working Paper 4206475*.

Young, H. Peyton. 2015. The Evolution of Social Norms. *Annual Review of Economics* 7(1): 359–387.

Ádám, Zoltán, and Iván Csaba. 2022. Populism unrestrained: Policy responses of the Orbán regime to the pandemic in 2020–2021. *European Policy Analysis* 8(3): 277–296.

A An example based on a simplified framework

This section presents a special case of the framework outlined in Section 3 of the paper. It hereby gives a simpler illustration of some of the main results. For brevity, a discussion of the intuition for the results is omitted; please refer to Sections 2.1 and 4 for thorough discussions of the intuition.

Following the same notation as outlined in Section 3, suppose that there are only four potential outcome combinations, $y_{1i} \in \{-1, 1\}$ and $y_{2i} \in \{-1, 1\}$. Denote the four voter sets of voters by HH if $y_{2i} = y_{1i} = 1$, LL if $y_{2i} = y_{1i} = -1$, HL if $y_{2i} = 1, y_{1i} = -1$, and LH if $y_{2i} = -1, y_{1i} = 1$, with densities $f(1,1) = w_{HH}$, $f(-1,-1) = w_{LL}$, $f(-1,1) = w_{HL}$, and $f(1,-1) = w_{LH}$, respectively. To satisfy radial symmetry, assume that $w_{HH} = w_{LL}$ and $w_{HL} = w_{LH}$.

Further, suppose that $\chi > 1$ and $\Delta\tau > \lambda(\chi - 1)$, which means that HL and LH will be core voters of *Right* and *Left*, respectively: they will vote for the same party regardless of the narratives offered.

Preliminaries. The utility difference from voting for *Right* versus *Left* is given as in the main text equation 2 by

$$u(y_i, \nu_R, \tau_R) - u(y_i, \nu_L, \tau_L) = \Delta\tau y_{2i} - \lambda\Delta\nu\left(\chi y_{1i} + y_{2i}\right), \tag{A.1}$$

where $\Delta\nu \equiv \nu_L - \nu_R \in [-1, 1]$ and $\Delta\tau \equiv \tau_L - \tau_R \in (0, 1]$. Whenever this difference is positive, i votes for *Right*, and otherwise for *Left*. If the difference is 0, the voter is indifferent, and votes for each party with equal likelihood.

With the simplification of this section, we can find this difference in utility $\Delta u_j(\Delta\nu, \Delta\tau)$ for each type j:

$$\Delta u_{HH}(\Delta\nu, \Delta\tau) = \Delta\tau - \lambda\Delta\nu\left(\chi + 1\right) \tag{A.2}$$

$$\Delta u_{LL}(\Delta\nu, \Delta\tau) = -\Delta\tau + \lambda\Delta\nu\left(\chi + 1\right) \tag{A.3}$$

$$\Delta u_{LH}(\Delta\nu, \Delta\tau) = -\Delta\tau - \lambda\Delta\nu\left(\chi - 1\right) \tag{A.4}$$

$$\Delta u_{HL}(\Delta\nu, \Delta\tau) = \Delta\tau + \lambda\Delta\nu\left(\chi - 1\right) \tag{A.5}$$

We can now recreate a simpler version of Proposition 1 from the main text:

Proposition A.1 *Whenever $\Delta\tau > \lambda(\chi-1)$, HL and LH are **core voters** of Right and Left, irrespective of the narrative difference $\Delta\nu$,.*

Proof. Since $\Delta\tau > \lambda(\chi - 1)$, Δu_{HL} is positive for any $\Delta\nu \in [-1, 1]$, and Δu_{LH} always negative. ■

The proposition shows the role of a high $\Delta\tau$ and low λ in generating core voters, as in the proposition of the main text.

We can now move to a simplified version of Proposition 2, employing the same definition of polarization as introduced in Definition 2 in the main text.

Proposition A.2 *Any (ν_L, ν_R) constitutes an equilibrium. The most polarized equilibrium is right-wing populism with $(\nu_L, \nu_R) = (1, 0)$.*

Proof. By equations A.2-A.3, HH votes for *Right* and LL for *Left* whenever $\frac{\Delta\tau}{\lambda(\chi+1)} > \Delta\nu$, and vice versa under the reverse inequality. Hence, for any $\Delta\nu \neq \frac{\Delta\tau}{\lambda(\chi+1)}$, each party receives either LL or HH, which have equal densities ($w_{LL} = w_{HH}$), in addition to their core voters, which also have equal densities ($w_{LH} = w_{HL}$). They therefore receive the same vote share. If $\Delta\nu = \frac{\Delta\tau}{\lambda(\chi+1)}$, then LL and HH are indifferent, and again the parties receive an equal vote share. Hence, parties always receive half of the total votes, regardless of $\Delta\nu$.

To see why right-wing populism will be most polarized, note that under right-wing populism ($\Delta\nu > \frac{\Delta\tau}{\lambda(\chi+1)}$), HH and LL vote against their economic self-interest and in line with their cultural preference. Under $\Delta\nu < \frac{\Delta\tau}{\lambda(\chi+1)}$, no type votes against their self interest, and therefore there is no polarization as defined in Definition 2. ■

The proposition is very similar to Proposition 2, with the exception that right-wing populism is always the most polarized outcome here. This is due to the restriction on χ imposed in this example, which implies that there cannot be any cultural polarization in a left-wing populist equilibrium.

Endogenous turnout. Let us now turn to the analysis of section 4.2, by considering the same turnout function as given in the main text:

$$q(y_{1i}, y_{2i}, \nu, \tau) = \frac{u(y_i, \nu, \tau) - \underline{u}(y_i)}{\overline{u}(y_i) - \underline{u}(y_i)}, \tag{A.6}$$

where as in the main text, $\overline{u}(y_i) \equiv \max_{\nu \in [0,1], \tau \in \{\tau_L, \tau_R\}} u(y_i, \nu, \tau)$, i.e. the utility of i under their most preferred narrative-policy pair; and $\underline{u}(y_i) \equiv \min_{\nu \in [0,1], \tau \in \{\tau_L, \tau_R\}} u(y_i, \nu, \tau)$. The main section goes into the intuition of this turnout function, which, for brevity, will not be repeated here.

With only four types, we can easily find the turnout function $q_j(\nu, \tau)$ for each type j:

$$q_{HH}(\nu, \tau) = \frac{1 - \tau - (1 - \tau_L) + \nu\lambda(\chi + 1)}{\Delta\tau + \lambda(\chi + 1)} \tag{A.7}$$

$$q_{LL}(\nu, \tau) = \frac{-(1 - \tau) + (1 - \tau_R) + (1 - \nu)\lambda(\chi + 1)}{\Delta\tau + \lambda(\chi + 1)} \tag{A.8}$$

$$q_{HL}(\nu, \tau) = \frac{(1 - \tau) - (1 - \tau_L) + (1 - \nu)\lambda(\chi - 1)}{\Delta\tau + \lambda(\chi - 1)} \tag{A.9}$$

$$q_{LH}(\nu, \tau) = \frac{-(1 - \tau) + (1 - \tau_R) + \nu\lambda(\chi - 1)}{\Delta\tau + \lambda(\chi - 1)} \tag{A.10}$$

It can easily be verified that the symmetry condition from Lemma 1 holds here too, so for opposing types with $y_i = -y_j$, we have that $q_i(\nu, \tau) = q_j(1 - \nu, 1 - \tau)$.

We can now recreate the main Proposition 3. For this purpose, define $\kappa \equiv \frac{w_{LL} q_{LL}(0,\tau_L) + w_{LH}(0,\tau_L)}{w_{HH} q_H(0,\tau_R) + w_{HL}(0,\tau_R)}$, analagously to the main text. Note that $\kappa = \frac{0.5 - w_{LH} q_{LH}(1,\tau_R)}{0.5 - w_{LL} q_{LL}(0,\tau_R)}$, so κ is higher under a more positive relationship between economic and overall status (higher $\frac{w_{LL}}{w_{LH}}$).

Proposition A.3 *Whenever $\kappa \neq 1$, parties only choose the extreme narratives $\nu_p \in \{0, 1\}$. If the correlation between economic and overall status is sufficiently positive ($\kappa > 1$), the unique equilibrium is left-wing populism at $(\nu_L, \nu_R) = (0, 1)$. If it is sufficiently negative ($\kappa < 1$), the unique equilibrium is right-wing populism at $(\nu_L, \nu_R) = (1, 0)$. If $\kappa = 1$, any combination of extreme narratives constitutes an equilibrium.*

Proof. Paralleling the proof for Proposition 3 in the main text, we will con-

struct this proof in multiple steps. First, we argue that in any equilibrium each party must gain the same vote share. Second, we show that there cannot be an equilibrium with intermediate narratives at $\kappa \neq 1$. Third, we characterise the equilibria at $\kappa \neq 1$. Finally, we show that any extreme narrative constitutes an equilibrium when $\kappa = 1$.

1. Equal split in equilibrium. To see that each party can guarantee half the vote share, consider the symmetry property of the turnout function, which leads to the fact that $w_{HH}q_{HH}(\nu,\tau) = w_{LL}q_{LL}(1-\nu,1-\tau)$ and $w_{HL}q_{HL}(\nu,\tau) = w_{HL}q_{HL}(1-\nu,1-\tau)$. Since each party wins either HH or LL as well as its set of core voters, and since $\tau_L = 1-\tau_R$, each party can guarantee itself half the votes by offering the opposing narrative of the other party. Hence, any equilibrium must give an equal split to each party.

2. No intermediate narratives when $\kappa \neq 1$. To see that there cannot be an equilibrium with intermediate narratives, suppose that $\nu_R \in (0,1)$. We can then show that either $\nu_L = 0$ or $\nu_L = 1$ must be winning more than half the vote, and hence ν_R cannot be an equilibrium strategy.

Denote the relative vote share of *Left* by π_{ν_L,ν_R} (and hence that of *Right* by $\pi^{-1}_{\nu_L,\nu_R}$). We then have the following payoffs:

$$\pi_{1,\nu_R} = \begin{cases} \frac{w_{LL}\frac{\Delta\tau}{\Delta\tau+\lambda(\chi+1)}+w_{LH}}{w_{HH}\frac{\Delta\tau+\nu_R\lambda(\chi+1)}{\Delta\tau+\lambda(\chi+1)}+w_{HL}\frac{\Delta\tau+(1-\nu_R)\lambda(\chi-1)}{\Delta\tau+\lambda(\chi-1)}} & \text{if } \Delta\nu \leq \frac{\Delta\tau}{\lambda(\chi+1)} \\ \frac{w_{HH}\frac{\lambda(\chi+1)}{\Delta\tau+\lambda(\chi+1)}+w_{LH}}{w_{LL}\frac{(1-\nu_R)\lambda(\chi+1)}{\Delta\tau+\lambda(\chi+1)}+w_{HL}\frac{\Delta\tau+(1-\nu_R)\lambda(\chi-1)}{\Delta\tau+\lambda(\chi-1)}} & \text{if } \Delta\nu > \frac{\Delta\tau}{\lambda(\chi+1)} \end{cases}$$

$$\pi_{0,\nu_R} = \frac{w_{LL} + w_{LH}\frac{\Delta\tau}{\Delta\tau+\lambda(\chi-1)}}{w_{LL}\frac{\Delta\tau+\nu_R\lambda(\chi+1)}{\Delta\tau+\lambda(\chi+1)} + w_{LH}\frac{\Delta\tau+(1-\nu_R)\lambda(\chi-1)}{\Delta\tau+\lambda(\chi-1)}}$$

By part 1 of the proof, we know that ν_R cannot be an equilibrium if *Left's* relative vote share is more than 1. We can already see that the π_{1,ν_R} is strictly greater than 1 if $\Delta\nu > \frac{\Delta\tau}{\lambda(\chi+1)}$, as the payoff of each type in the numerator is larger than of the corresponding type in the denominator, while the densities are equal ($w_{LL} = w_{HH}$ and $w_{LH} = w_{HL}$). For the remaining deviations, we

have the following:

$$\begin{aligned}
\pi_{1,\nu_R} > 1 \wedge \Delta\nu \leq \frac{\Delta\tau}{\lambda(\chi+1)} \\
\iff w_{LH}\frac{\nu_R\lambda(\chi-1)}{\Delta\tau+\lambda(\chi-1)} > w_{LL}\frac{\nu_R\lambda(\chi+1)}{\Delta\tau+\lambda(\chi+1)} \\
\iff \frac{w_{LH}}{w_{LL}} > \frac{(\chi+1)\Delta\tau+\lambda(\chi+1)(\chi-1)}{(\chi-1)\Delta\tau+\lambda(\chi+1)(\chi-1)} \iff \kappa < 1 \\
\pi_{0,\nu_R} > 1 \iff w_{LL}\frac{(1-\nu_R)\lambda(\chi+1)}{\Delta\tau+\lambda(\chi+1)} > w_{LH}\frac{(1-\nu_R)\lambda(\chi-1)}{\Delta\tau+\lambda(\chi-1)} \\
\iff \frac{w_{LL}}{w_{LH}} > \frac{(\chi-1)\Delta\tau+\lambda(\chi+1)(\chi-1)}{(\chi+1)\Delta\tau+\lambda(\chi+1)(\chi-1)} \iff \kappa > 1
\end{aligned}$$

Since $\kappa \neq 1$, one of the two deviations must be profitable. By symmetry and the same arguments, there is a winning strategy for *Right* when $\nu_L \in (0,1)$. Hence, there cannot be an equilibrium where one party offers an intermediate narrative.

3. Equilibria characterisation when $\kappa \neq 1$. The payoffs from different combinations of extreme narratives are given in Table 2. This illustrates that the set of possible equilibria in extreme narratives depends on κ.

Table 2: *Left's* relative vote share under extreme narratives

	$\nu_R = 0$	$\nu_R = 1$
$\nu_L = 0$	κ	1
$\nu_L = 1$	1	$\frac{1}{\kappa}$

Consider first the case where $\kappa > 1 \iff \frac{w_{LL}}{w_{LH}} > \frac{(\chi-1)\Delta\tau+\lambda(\chi+1)(\chi-1)}{(\chi+1)\Delta\tau+\lambda(\chi+1)(\chi-1)}$. In that case, the only equilibrium in extreme narrative that is possible is left-wing populism at $(\nu_L, \nu_R) = (0,1)$, which gives $\Delta\nu = -1$. It is immediate that there is no profitable deviation to one of the extreme narratives. Let us now prove that there is no profitable deviation to any intermediate narrative. Note first that under any deviation, $\Delta\nu \leq 0$, and hence each party retains the same voter types: LL and LH vote for *Left*, HH and HL for *Right*. Hence, the derivative with respect to narratives is constant. Note first that $\frac{\partial\pi_{0,1}}{\partial\nu_L} = \frac{w_{LL}\frac{\lambda(\chi+1)}{\Delta\tau+\lambda(\chi+1)} - w_{LH}\frac{\lambda(\chi-1)}{\Delta\tau+\lambda(\chi-1)}}{w_{LL}\frac{\Delta\tau}{\Delta\tau+\lambda(\chi+1)} + w_{LH}}$. Now, $\kappa > 1$ implies that the numerator is negative (while the denominator is always positive), and hence $\frac{\partial\pi_{0,1}}{\partial\nu_L} < 0$. *Left*

would wish to further lower it's narratives, but it already offers the lowest possible narratives, so it cannot profitably deviate. The same reasoning shows that $\kappa > 1 \iff \frac{\partial \pi_{0,1}^{-1}}{\partial \nu_R} > 0$, so *Right* also does not have a profitable deviation.

Now consider the case where $\kappa < 1$. The only potential equilibrium is now at $(\nu_L, \nu_R) = (1, 0)$, so that $\Delta\nu = 1$. In that case, LL and HL vote for *Right* and HH and LH for *Left*. We should separately consider deviations to $\Delta\nu > \frac{\Delta\tau}{\lambda(\chi+1)}$, and those to $\Delta\nu \leq \frac{\Delta\tau}{\lambda(\chi+1)}$. Deviations of the former type lead to the same vote split, with LL and HL voting for Right and HH and LH for Left. Such a deviation implies that either *Right* increased or *Left* decreased their narrative. But neither of these can be profitable, as all *Right's* voters strictly prefer lower narratives, and *Left's* higher narratives.

Now consider the second type of deviation, to $\Delta\nu \leq \frac{\Delta\tau}{\lambda}$. This induces a change in voter behaviour where now HH and HL vote for *Right* and LL and LH vote for *Left*. Consider the change in payoffs for Left from choosing ν_L with the effect that $\Delta\nu \leq \frac{\Delta\tau}{\lambda(\chi+1)}$. The necessary condition for *Left* such that this is a profitable deviation is that this leads to a payoff greater than 1. Hence, for the deviation to be profitable it must be that

$$\begin{aligned} & \frac{w_{LL}\frac{\Delta\tau+(1-\nu_L)\lambda(\chi+1)}{\Delta+\lambda(\chi+1)} + w_{LH}\frac{\Delta\tau+\nu_L\lambda(\chi-1)}{\Delta\tau+\lambda(\chi-1)}}{w_{LL}\frac{\Delta\tau}{\Delta\tau+\lambda(\chi+1)} + w_{LH}} > 1 \\ \iff \; & w_{LL}\frac{(1-\nu_L)\lambda(\chi+1)}{\Delta\tau+\lambda(\chi+1)} > w_{LH}\frac{(1-\nu_L)\lambda(\chi-1)}{\Delta\tau+\lambda(\chi-1)} \\ & \iff \kappa > 1, \end{aligned}$$

which contradicts the assumption that $\kappa < 1$. Hence, there cannot be a profitable deviation for *Left*. Equivalent reasoning shows that the existence of a profitable deviation for *Right* requires $\kappa > 1$, leading to the same contradiction.

In summary, we have proved that right-wing populism at $(\nu_L, \nu_R) = (1, 0)$ is the unique equilibrium under $\kappa < 1$, left-wing populism $(0, 1)$ under $\kappa > 1$. We will now consider what happens when $\kappa = 1$.

4. Extreme narrative equilibria when $\kappa = 1$. When $\kappa = 1$, $\pi_{0,0} = \pi_{1,1} = \pi_{1,0} = \pi_{1,1}$. Furthermore, the deviations considered in part 2 of the proof also

give a payoff of 1. Hence $\pi_{0,0}$ and $\pi_{1,1}$ are now also additional equilibria in extreme narratives.[8] ■

The proposition above follows the logic of the main Proposition 3 closely, with the exception that since types are coarse, we need to take more explicit account of the discontinuity at $\kappa = 1$. We can now move to restating Proposition 4 for the case of the simple model considered here.

Proposition A.4 *Left-wing populism is the unique equilibrium under a sufficiently low salience of cultural issues (low χ). If economic and noneconomic outcomes are negatively related ($\frac{w_{LL}}{w_{LH}} < 1$), right-wing populism is the unique equilibrium under a sufficiently high salience of cultural issues (high χ).*

Proof. Referring to the proof of the previous proposition, we know that right-wing populism is unique when $\kappa < 1$, and left-wing populism under the reverse inequality. We know that $\kappa < 1 \iff \frac{w_{LL}}{w_{LH}} < \frac{(\chi-1)\Delta\tau+\lambda(\chi+1)(\chi-1)}{(\chi+1)\Delta\tau+\lambda(\chi+1)(\chi-1)}$. The right-hand side of the inequality is strictly increasing in χ and reaches 1 in the limit as $\chi \to \infty$, while the left-hand side is increasing in $\frac{w_{LL}}{w_{LH}}$. Hence, if$\frac{w_{LL}}{w_{LH}} \geq$, Left-wing populism is unique for any χ. If $\frac{w_{LL}}{w_{LH}} < 1$, Hence, right-wing populism will be the unique equilibrium high enough χ, and left-wing populism under low enough χ. ■

Voter asymmetry. Let us now consider what happens when the turnout of the left-behind voters is adjusted to $(1+\beta)w_{LL}q_{LL}(\nu,\tau)$, as considered in Section 4.3 of the main text. The next two propositions recreate Propositions 5 and 6 from the main text; the proof considers both of these propositions jointly.

Proposition A.5 *For any $\beta \geq 0$ there always exists an equilibrium in pure strategies.*

Proposition A.6 *For any $\beta > 0$, only populist equilibria exist, and if there is a single populist party, it wins the majority of votes. The equilibria depend on the weight of the left-behind, β, as well as on the relationship between economic*

[8] There will also be additional equilibria at some but not all intermediate narratives when $\kappa = 1$, which are not considered here for brevity.

outcomes and overall status, κ. Let $\underline{\beta}, \overline{\beta}$ be defined implicitly by $\pi_{1,0}(\underline{\beta}) = \pi_{0,0}(\underline{\beta})$ and $\pi_{0,0}(\overline{\beta}) = 1$, which always have unique solutions, and where $\underline{\beta} < \overline{\beta}$.

1. *If $\kappa > 1$, the unique equilibrium is left-wing populism and Left wins a strict majority $\forall \beta > 0$.*

2. *If $\kappa < 1$, right-wing populism is the unique equilibrium $\forall \beta < \underline{\beta}$, and both parties will be populist $\forall \beta > \underline{\beta}$. Right wins a strict majority $\forall \beta < \overline{\beta}$, and Left $\forall \beta > \overline{\beta}$.*

Proof. Payoffs π_{ν_L,ν_R} under extreme-narrative combinations change with β in the following way:

$$\frac{\partial \pi_{0,0}}{\partial \beta} = \frac{w_{LL}}{w_{HH}\frac{\Delta\tau}{\Delta\tau+\lambda(\chi+1)} + w_{HL}}$$

$$\frac{\partial \pi_{0,1}}{\partial \beta} = \frac{w_{LL}}{w_{HH} + w_{HL}\frac{\Delta\tau}{\Delta\tau+\lambda(\chi-1)}}$$

$$\frac{\partial \pi_{1,0}}{\partial \beta} < 0$$

$$\frac{\partial \pi_{1,1}}{\partial \beta} = \frac{w_{LL}\frac{\Delta\tau}{\Delta\tau+\lambda(\chi+1)}}{w_{HH} + w_{HL}\frac{\Delta\tau}{\Delta\tau+\lambda(\chi-1)}}$$

Note that $\frac{\partial \pi_{0,1}}{\partial \beta} > \frac{\partial \pi_{1,1}}{\partial \beta}$, and $\frac{\partial \pi_{0,0}}{\partial \beta} > \frac{\partial \pi_{0,1}}{\partial \beta} \iff \kappa > 1$. We should separately consider the case when $\kappa > 1$ and $\kappa < 1$.

Suppose that $\kappa > 1$. This implies that $\frac{\partial \pi_{0,0}}{\partial \beta} > \frac{\partial \pi_{0,1}}{\partial \beta} > \frac{\partial \pi_{1,1}}{\partial \beta}$. Note also that at $\beta = 0$, $\pi_{0,0} > \pi_{0,1} = \pi_{1,0} = 1 > \pi_{1,1}$. Hence, $\forall \beta$, *Left's* relative payoffs are $\pi_{0,0} > \pi_{0,1} > \pi_{1,1}$, while $\pi_{1,0} \leq 1$. It follows that $(\nu_L, \nu_R) = (0, 1)$ is the unique equilibrium.

Suppose that $\kappa < 1$. This implies that $\frac{\partial \pi_{0,1}}{\partial \beta} > \max[\frac{\partial \pi_{1,1}}{\partial \beta}, \frac{\partial \pi_{0,0}}{\partial \beta}]$. Note also that at $\beta = 0$, $\pi_{1,1} > \pi_{0,1} = \pi_{1,0} = 1 > \pi_{0,0}$. Hence, $\forall \beta$, *Left's* payoffs satisfy $\pi_{0,1} > \pi_{0,0}$. Moreover, since $\frac{\partial \pi_{0,0}}{\partial \beta} > 0$ and $\frac{\partial \pi_{1,0}}{\partial \beta} < 0$, while at $\beta = 0$ $\pi_{1,0} > \pi_{0,0}$, there exist $\underline{\beta}, \overline{\beta}$ with the property that $\pi_{1,0}(\underline{\beta}) = \pi_{0,0}(\underline{\beta})$ and $\pi_{0,0}(\overline{\beta}) = 1$. Note also that since $\pi_{1,0}(\underline{\beta}) < 1$, it follows that $\underline{\beta} < \overline{\beta}$.

For $\beta \in (0, \underline{\beta})$, $\pi_{0,1} > 1 > \pi_{1,0} > \pi_{0,0}$, so *Right* can guarantee itself more votes than *Left* by choosing $\nu_R = 0$. The unique equilibrium in this case is $(\nu_L, \nu_R) = (1, 0)$, where *Right* always wins a majority. For $\beta = \underline{\beta}$, *Left* is

indifferent between $\nu_L = 0$ and $\nu_L = 1$, while *Right* chooses $\nu_R = 0$; again *Right* wins a majority.

For $\beta > \underline{\beta}$, $\pi_{0,1} > \pi_{0,0} > \pi_{1,0}$, so $(0,0)$ is the unique equilibrium. *Right* wins a strict majority when $\beta < \overline{\beta}$, and *Left* under the reverse inequality.

With regards to Proposition A.5, note that the proof considered all cases of $\beta \geq 0$, and has shown that a pure-strategy equilibrium always exists. For Proposition 6, the characterisation for $\kappa > 1$ and $\kappa < 1$ has been done below the relevant subheading. This completes the proof. ■

Incentives to change redistributive policies. Finally, consider the incentives of parties to adjust their economic policies, as discussed in Section 4.4. As in that section, take the same turnout equation 5. We can now recreate the final proposition of the main section:

Proposition A.7 *Consider the equilibria with a single right-wing or left-wing populist party given by Proposition A.6, when $\beta > 0$. A right-wing populist has a stronger incentive to decrease the cross-party tax difference $\Delta\tau$ than a non-populist opponent, while a left-wing populist has a stronger incentive to increase $\Delta\tau$ than a non-populist opponent.*

Proof. Consider the right-wing populist equilibrium at $(\nu_L, \nu_R) = (1, 0)$. As before, denote *Left's* relative votes by $\pi_{1,0}$, giving *Right's* relative votes as $\pi_{1,0}^{-1}$. The responses of *Left's* and *Right's* relative vote shares are then given by

$$\frac{\partial \pi_{1,0}}{\partial \tau_L} = \frac{-w_{HH}\frac{1}{1+\lambda(\chi+1)} + w_{LH}\frac{1}{1+\lambda(\chi-1)}}{(1+\beta)w_{LL}\frac{\Delta\tau}{\Delta\tau+\lambda(\chi+1)} + w_{HL}}$$

$$\frac{\partial \pi_{1,0}^{-1}}{\partial \tau_R} = \frac{(1-\beta)w_{LL}\frac{1}{1+\lambda(\chi+1)} - w_{HL}\frac{1}{1+\lambda(\chi-1)}}{w_{HH}\frac{\Delta\tau}{\Delta\tau+\lambda(\chi+1)} + w_{LH}}$$

For $\beta = 0$, $\frac{\partial \pi_{1,0}^{-1}}{\partial \tau_R} = -\frac{\partial \pi_{1,0}}{\partial \tau_L}$ since $w_{HH} = w_{LL}$ and $w_{HL} = w_{LH}$. However, for $\beta > 0$, $\frac{\partial \pi_{1,0}^{-1}}{\partial \tau_R} > -\frac{\partial \pi_{1,0}}{\partial \tau_L}$. Hence, in that case, *Right* has a stronger incentive to increase taxes than *Left* has to decrease them. Since increasing τ_R and decreasing τ_L both lead to a lower $\Delta\tau$, we can conclude that under right-wing populism, *Right* has a stronger incentive to decrease $\Delta\tau$ than *Left*.

We can next consider the left-wing populist equilibrium at $(\nu_L, \nu_R) = (0, 1)$. Following the same steps, we find that for $\beta > 0$, $\frac{\partial \pi_{0,1}}{\partial \tau_L} > -\frac{\partial \pi_{0,1}^{-1}}{\partial \tau_R}$. Hence, the left-wing populist has a stronger incentive to decrease $\Delta\tau$ (by increasing τ_L) than its right-wing opponent (by decreasing τ_R). ■

B Proofs

B.1 Proposition 1

Proof. We characterise the set of core voters by considering equation 2: i is a core right voter *iff* $\forall \Delta\nu, u(y_i, \nu_R, \tau_R) > u(y_i, \nu_L, \tau_L)$, and a core left voter iff $\forall \Delta\nu, u(y_i, \nu_R, \tau_R) < u(y_i, \nu_L, \tau_L)$. The set of core right voters is thus given by $\{y_i \colon y_{2i} > \frac{\lambda}{\Delta\tau}|(\chi y_{1i} + y_{2i})|\}$, which is a non-empty set that always contains the points along the line $y_{2i} = -\chi y_{1i} > 0$. Similarly, the set of core left voters is given by $\{y_i \colon y_{2i} < -\frac{\lambda}{\Delta\tau}|(\chi y_{1i} + y_{2i})|\}$, which always has $y_{2i} = -\chi y_{1i} < 0$ among its elements. It is evident that the size of these sets decreases in λ and increases in $\Delta\tau$. The mass of core right and core left voters is always strictly positive: if $\lambda > \Delta\tau$, the mass of core right and core left voters is given by $\int_{-\infty}^{0}\int_{-\frac{\lambda\chi}{\lambda+\Delta\tau}y_1}^{-\frac{\lambda\chi}{\lambda-\Delta\tau}y_1} f(y_1, y_2)dy_2dy_1$ and $\int_{0}^{\infty}\int_{-\frac{\lambda\chi}{\lambda-\Delta\tau}y_1}^{-\frac{\lambda\chi}{\lambda+\Delta\tau}y_1} f(y_1, y_2)dy_2dy_1$, respectively. If $\lambda = \Delta\tau$, the masses are given by $\int_{-\infty}^{0}\int_{-\frac{\lambda\chi}{\lambda+\Delta\tau}y_1}^{\infty} f(y_1, y_2)dy_2dy_1$ and $\int_{0}^{\infty}\int_{-\infty}^{-\frac{\lambda\chi}{\lambda+\Delta\tau}y_1} f(y_1, y_2)dy_2dy_1$, respectively. Finally, if $\lambda \in (0, \Delta\tau)$, they are given by $\int_{-\infty}^{0}\int_{-\frac{\lambda\chi}{\lambda+\Delta\tau}y_1}^{\infty} f(y_1, y_2)dy_2dy_1 + \int_{0}^{\infty}\int_{\frac{\lambda\chi}{\Delta\tau-\lambda}y_1}^{\infty} f(y_1, y_2)dy_2dy_1$ for the core right and $\int_{-\infty}^{0}\int_{-\infty}^{\frac{\lambda\chi}{\Delta\tau-\lambda}y_1} f(y_1, y_2)dy_2dy_1 + \int_{0}^{\infty}\int_{-\infty}^{-\frac{\lambda\chi}{\lambda+\Delta\tau}y_1} f(y_1, y_2)dy_2dy_1$ for the core left. ■

B.2 Proposition 2

Proof. First, to see that any combination of narratives constitutes an equilibrium, recall that $f(y_{1i}, y_{2i}) = f(-y_{1i}, -y_{2i})\forall y_i$ by radial symmetry. Moreover, note that the line $y_{2i} = -\frac{\lambda\chi\Delta\nu}{\lambda\Delta\nu-\Delta\tau}y_{1i}$ splits the voter space into supporters of *Right* and *Left* and goes through the origin (where voters located on the line vote for each party with equal likelihood). Hence, for any voter that *Right* wins with density $f(y_{1i}, y_{2i})$ on one side of the line, *Left* wins a type with equal density $f(-y_{1i}, -y_{2i})$ on the other side. Recalling that $L(\Delta\nu)$ denotes *Left's* set of supporters and $R(\Delta\nu)$ *Right's*, it follows that $\iint_{L(\Delta\nu)} f(y_{1i}, y_{2i})dA = \iint_{R(\Delta\nu)} f(y_{1i}, y_{2i})dA$, i.e. that each party wins an equal

vote share.

Second, to consider which equilibrium is most culturally polarized, note first that the least culturally polarized case is at $\Delta\nu = 0$, where everyone votes according to their economic preference: all high-income voters back *Right*, and low-income voters back *Left*. Now, we will show that polarization is strictly decreasing in $\Delta\nu$ whenever $\Delta\nu < 0$, and strictly increasing in $\Delta\nu$ whenever $\Delta\nu > 0$, and hence that the most polarized equilibria must be at $\Delta\nu = \{-1, 1\}$.

Consider first the case of $\Delta\nu < 0$. The lower is $\Delta\nu$, the more negative is the slope of $-\frac{\lambda\chi\Delta\nu}{\lambda\Delta\nu-\Delta\tau}y_{1i}$ which separates the voters of *Left* and *Right*, and the more voters of the HL_4 type will back *Left* (and analogously LH_4 for *Right*). Similarly, when $\Delta\nu > 0$, the higher is $\Delta\nu \in (0, \frac{\Delta\tau}{\lambda})$, the more positive is the slope that separates voters of *Left* and *Right*, and the more HH types will back *Left*. The more $\Delta\nu$ rises above $\frac{\Delta\tau}{\lambda}$, the more HL_1 types will back the left party. Hence, the fraction of voters who vote against their economic policy preference is higher under $\Delta\nu = 1$ than under any $\Delta\nu \in [0, 1)$, and higher under $\Delta\nu = -1$ than under any $\Delta\nu \in (-1, 0]$.

Finally, whether the polarization is highest under right-wing populism at $\Delta\nu = 1$ or left-wing populism at $\Delta\nu = -1$ depends on which of these two generate a higher mass of voters who vote against their economic policy preference. It therefore depends on whether the density of $HH - HL_1$ is greater than of HL_4 (and thus, by radial symmetry, $LL - LH_1$ is greater than LH_4). The two densities are given by $\iint_{HH-HL_1} f(y_1, y_2)dA$ and $\iint_{HL_4} f(y_1, y_2)dA$, respectively. This is visualised in Figure C.1, which for different values of $\Delta\nu$ shows the area of voters who vote against their economic interest with dashed lines. Right-wing populism will be more polarized than left-wing populism if the density of the area in Panel *a* is larger than that in Panel *c*. ■

B.3 Corollary 1

Proof. The proof for this corollary follows from the proof above for Proposition 2. For the first part, referring back to the plot of voter types in Figure 5,

note that the area of HL_1 is bounded from below by the line $y_{2i} = -\frac{\lambda\chi}{\lambda-\Delta\tau}y_{1i}$, and of HL_4 bounded from above by $y_{2i} = -\frac{\lambda\chi}{\lambda+\Delta\tau}y_{1i}$. Since the absolute value of the slope of each of these lines is increasing in χ, a higher χ decreases the set of HL_1 voter types and increases the set of HL_4 voter types. It thus increases polarization of left-wing populism and decreases that of right-wing populism. This can be appreciated by comparing Panels *a* and *b* of Figure C.2, which for different levels of χ indicates the voters who vote against their economic policy preference under right-wing populism in dashed lines.

For the second statement, note that under radial symmetry, a positive relation between y_{1i} and y_{2i} ($\rho > 1$) means that the total density of the first quadrant is higher than that of the second. Hence, $\iint_{HH} f(y_1, y_2)dA > \iint_{HL_4} f(y_1, y_2)dA$, and right-wing populism must be the most polarized.

For the third statement, note that a negative relationship between y_{1i} and y_{2i} ($\rho < 1$) means that the total density of the second quadrant is strictly higher than of the first. Now, since $\lim_{\chi\to\infty} \iint_{HL_4} f(y_1, y_2)dA = \int_{-\infty}^{0}\int_{0}^{\infty} f(y_1, y_2)dy_2dy_1$ and $\lim_{\chi\to\infty} \iint_{HL_1} f(y_1, y_2)dA = 0$, we have that $\lim_{\chi\to\infty} \iint_{HL_4} f(y_1, y_2)dA$ is greater than $\lim_{\chi\to\infty} \iint_{HH-HL_1} f(y_1, y_2)dA$. Hence, in the limit as $\chi \to \infty$, left-wing populism is most polarized. On the other hand, $\lim_{\chi\to 0} \iint_{HL_4} f(y_1, y_2)dA = 0$ and $\lim_{\chi\to 0} \iint_{HL_1} f(y_1, y_2)dA = \int_{-\infty}^{0}\int_{0}^{\infty} f(y_1, y_2)dy_2dy_1$, and hence right-wing populism is most polarized when $\chi = 0$. Since $\iint_{HL_4} f(y_1, y_2)dA$ is strictly increasing and $\iint_{HL_1} f(y_1, y_2)dA$ strictly decreasing in χ, and since each is continuous in χ, there must in this case be a cutoff $\bar{\chi}$ where polarization is exactly equal at right-wing and left-wing populism. It follows that polarization is highest for right-wing populism under $\chi \leq \bar{\chi}$, and highest for left-wing populism under $\chi \geq \bar{\chi}$. ■

B.4 Lemma 1

Proof. We start by showing that $q(y_{1i}, y_{2i}, \nu, \tau) = q(y_{1j}, y_{2j}, 1-\nu, 1-\tau)$ for $y_i = -y_j$: the turnout of opposite types is the same for taking the "opposing" platform. To see this, note first that $u(y_i, \nu, \tau) = -u(y_j, \nu, \tau)$. From

that if follows that $\overline{u}(y_i) = -\underline{u}_j$ and $\underline{u}(y_i) = -\overline{u}_j$. Hence, $\frac{u(y_j,\nu_j,\tau_j)-\underline{u}_j}{\overline{u}_j-\underline{u}_j} = \frac{\overline{u}(y_i)-u(y_i,\nu_j,\tau_j)}{\overline{u}(y_i)-\underline{u}(y_i)}$. Hence, turnouts of i and j are equal *iff* $u(y_i,\nu_i,\tau_i) - \underline{u}(y_i) = \overline{u}(y_i) - u(y_i,\nu_j,\tau_j) \iff u(y_i,\nu_i,\tau_i) + u(y_i,\nu_j,\tau_j) = \underline{u}(y_i) + \overline{u}(y_i) = (2-\tau_L-\tau_R)y_{1i}+\lambda(\chi y_{1i}+y_{2i})$. It follows that at equal turnout of i and j, $\nu_i = 1-\nu_j$ and $\tau_i = 1-\tau_j$ (since $\tau_R = 1-\tau_L$). Hence, $q(y_{1i},y_{2i},\nu,\tau) = q(y_{1j},y_{2j},1-\nu,1-\tau)$.

For the first statement of the lemma, we show that the above implies that $Q_{L(\Delta\nu)}(\nu,\tau) = Q_{R(\Delta\nu)}(1-\nu,1-\tau)$. Note that by radial symmetry, $f(y_i) = f(-y_j)$ for $y_i = -y_j$. Hence, $q(y_{1i},y_{2i},\nu,\tau)f(y_i) = q(-y_{1i},-y_{2i},1-\nu,1-\tau)f(-y_i)$, which implies that for any $a<b$ and $c<d$,

$$\int_a^b\int_c^d q(y_1,y_2,\nu,\tau)f(y_1,y_2)dy_2dy_1 = \int_{-b}^{-a}\int_{-d}^{-c} q(y_1,y_2,1-\nu,1-\tau)f(y_1,y_2)dy_2dy_1$$

Therefore, $Q_G(\nu,\tau) = Q_{\neg G}(1-\nu,1-\tau)$, for any subset of types G and the subset of opposing types $\neg G$, implying that $Q_{L(\Delta\nu)}(\nu,\tau) = Q_{R(\Delta\nu)}(1-\nu,1-\tau)$.

For the second statement of the lemma, note that $q(y_{1i},y_{2i},\nu,\tau)+q(y_{1i},y_{2i},1-\nu,1-\tau) = 1$, as $u(y_i,\tau,\nu) + u(y_i,1-\tau,1-\nu) = \overline{u}(y_i) + \underline{u}(y_i)$. Hence, $q(y_{1i},y_{2i},\nu,\tau)f(y_i)+q(y_{1i},y_{2i},1-\nu,1-\tau)f(y_i) = f(y_i)$, and it then follows that $\int_a^b\int_c^d q(y_1,y_2,\nu,\tau)f(y_1,y_2)dy_2dy_1 + \int_a^b\int_c^d q(y_1,y_2,1-\nu,1-\tau)f(y_1,y_2)dy_2dy_1 = \int_a^b\int_c^d f(y_1,y_2)dy_2dy_1$. In our notation, this can be written as $Q_G(\nu,\tau)+Q_G(1-\nu,1-\tau)dA = \iint_G f(y_1,y_2)dA$. ■

B.5 Lemma 2

Proof. Note first that since the turnout function is linear in ν and is additively separable in ν and τ. Thus, for any G, $\frac{\partial^2 Q_G(\nu,\tau)}{\partial\nu^2} = 0$ and $\frac{\partial Q_G(\nu,\tau_L)}{\partial\nu} = \frac{\partial Q_G(\nu,\tau_R)}{\partial\nu}$, i.e. the partial derivative of the turnout with respect to ν does not depend on the level of ν or on τ. To see that $\frac{\partial Q_{L(\Delta\nu)}(\nu_L,\tau_L)}{\partial\nu_L}$ is strictly increasing in $\Delta\nu$, consider first the total votes for *Left* at some $\Delta\nu$, $Q_{L(\Delta\nu)}(\nu_L,\tau_L)$. Now compare this with what happens when we decrease $\Delta\nu$ to $\Delta\nu-\zeta$, with $-1 \leq \Delta\nu-\zeta < \Delta\nu$ — i.e. when *Left's* narratives become more scapegoating relative

to *Right's*. The new total votes for *Left* can be written as

$$Q_{L(\Delta\nu-\zeta)} = Q_{L(\Delta\nu)}(\nu, \tau_L) + Q_{L_1(\Delta\nu)}(\nu_L, \tau_L) - Q_{R_1(\Delta\nu)}(\nu, \tau_L),$$

where R_1 denotes the types who no longer vote for *Left* and L_1 those who now switched to *Left* (see Figure C.3).

Since a decrease in $\Delta\nu$ means a lower ν_L relative to ν_R, those who switched from *Left* to *Right*, i.e. R_1-types, must do so because they prefer higher narratives, while L_1 must prefer lower narratives. Since R_1 and L_1 denote opposite types, by Lemma 1 $\frac{\partial Q_{L_1(\Delta\nu)}(\nu,\tau_L)}{\partial\nu} = -\frac{\partial Q_{R_1(\Delta\nu)}(\nu,\tau_L)}{\partial\nu} < 0$. It follows that

$$\begin{aligned}\frac{\partial Q_{L(\Delta\nu-\zeta)}(\nu,\tau_L)}{\partial\nu} &= \frac{\partial\left(Q_{L(\Delta\nu)}(\nu,\tau_L) + Q_{L_1(\Delta\nu)}(\nu,\tau_L) - Q_{R_1(\Delta\nu)}(\nu,\tau_L)\right)}{\partial\nu} \\ &= \frac{\partial Q_{L(\Delta\nu)}(\nu,\tau_L)}{\partial\nu} + 2\frac{\partial Q_{L_1(\Delta\nu)}(\nu,\tau_L)}{\partial\nu} < \frac{\partial Q_{L(\Delta\nu)}(\nu,\tau_L)}{\partial\nu}\end{aligned}$$

Equivalent reasoning shows that $\frac{\partial Q_{R(\Delta\nu-\zeta)}(\nu,\tau_R)}{\partial\nu} > \frac{\partial Q_{R(\Delta\nu)}(\nu,\tau_L)}{\partial\nu}$, which can also be appreciated from the fact that $\frac{\partial Q_{L(\Delta\nu)}(\nu,\tau_L)}{\partial\nu} = -\frac{\partial Q_{R(\Delta\nu)}(\nu,\tau_R)}{\partial\nu} \forall \Delta\nu$. ■

B.6 Lemma 3

Proof. For any i, the absolute responsiveness is $|\frac{\partial q(y_{1i},y_{2i},\nu,\tau)}{\partial\nu}| = \frac{\lambda(|\chi y_{1i}+y_{2i}|)}{|\Delta\tau y_{1i}|+\lambda(|\chi y_{1i}+y_{2i}|)}$, which is increasing in λ. Hence, for any voter with y_i: $y_{1i} + \chi y_{2i} \neq 0$ responsiveness of each i to a change in ν is increasing in λ. For the remaining voters, responsiveness is unchanging in λ. ■

B.7 Proposition 3

Proof. The proof for this proposition proceeds in four steps:

1. Each party can guarantee itself half the vote through $\nu_L = 1 - \nu_R$. Hence, any equilibrium outcome must give the same share of votes to each party.
2. There is at most one level of $\Delta\nu \equiv \bar{\Delta\nu}$ with equal share of votes for each party when $\nu_L \neq 1 - \nu_R$.
3. There exist no interior equilibria where either party offers $\nu_p \in (0, 1)$.

Hence, the only possible equilibria are at the extreme narrative combinations $(1,0),(0,1)$ if $\bar{\Delta\nu} \neq 0$; and additionally $(0,0),(1,1)$ if $\bar{\Delta\nu} = 0$.

4. It is sufficient to check deviations to extreme narratives from equilibria $\nu_p \in \{0,1\}$, and hence pure strategy equilibria exist if no deviation to another extreme narrative is profitable.

The first step follows directly from the first statement of Lemma 1: If a party offers the opposite narrative to its opponent, it will receive an equal vote share. Hence, it follows that in any equilibrium it must be the case that each party receives an equal vote share.

For the second step we wish to show that there is a unique $\bar{\Delta\nu}$ that satisfies $\frac{Q_{L(\bar{\Delta\nu})}(\nu_L,\tau_L)}{Q_{R(\bar{\Delta\nu})}(\nu_R,\tau_R)} = 1$ for $\nu_L \neq 1-\nu_R$. From Lemma 1, we know that $Q_{R(\bar{\Delta\nu})}(\nu_R,\tau_R) = Q_{L(\bar{\Delta\nu})}(1-\nu_R,\tau_L)$, and hence $\bar{\Delta\nu}$ satisfies $\frac{Q_{L(\bar{\Delta\nu})}(\nu_L,\tau_L)}{Q_{L(\bar{\Delta\nu})}(1-\nu_R,\tau_L)} = 1$. Now, since $\nu_L \neq 1-\nu_R$, the only way in which this can be satisfied is if turnout at $\bar{\Delta\nu}$ is the same for these two different narratives. Because turnout is linear in ν, this implies that $\frac{\partial Q_{L(\bar{\Delta\nu})}(\nu,\tau_L)}{\partial\nu} = 0$. By the strict monotonicity of $\frac{\partial Q_{L(\Delta\nu)}(\nu,\tau_L)}{\partial\nu}$ in $\Delta\nu$ shown in Lemma 2, there can be at most one $\bar{\Delta\nu}$ that satisfies this. Moreover, $\forall\Delta\nu > \bar{\Delta\nu}$, $\frac{\partial Q_{L(\Delta\nu)}(\nu,\tau_L)}{\partial\nu} > 0$ and $\forall\Delta\nu < \bar{\Delta\nu}$, $\frac{\partial Q_{L(\Delta\nu)}(\nu,\tau_L)}{\partial\nu} < 0$.

For the third step, we can prove that if one player offers an intermediate narrative, then the other player has a winning strategy. For this purpose, suppose that $\nu_R \in (0,1)$. We now show that it must be that either $\nu_L = 0$ or $\nu_L = 1$ is winning. Denote $\Delta\nu_0 \equiv 0-\nu_R$ and $\Delta\nu_1 \equiv 1-\nu_R$, with $\Delta\nu_1 > \Delta\nu_0$. Then the relative vote share of *Left* is $\frac{Q_{L(\Delta\nu_0)}(0,\tau_L)}{Q_{L(\Delta\nu_0)}(1-\nu_R,\tau_L)}$ from $\nu_L = 0$ and $\frac{Q_{L(\Delta\nu_1)}(1,\tau_L)}{Q_{L(\Delta\nu_1)}(1-\nu_R,\tau_L)}$ from $\nu_L = 1$. From step 2 of the proof we know that whether $\Delta\nu$ is higher or lower than $\bar{\Delta\nu}$ determines the sign of $\frac{\partial Q_{L(\Delta\nu)}(\nu,\tau_L)}{\partial\nu}$. Using this, we can prove that at least one strategy of *Left* must be winning for any intermediate strategy $\nu_R \in (0,1)$ of *Right*:

a) If $\bar{\Delta\nu} \leq \Delta\nu_0$, then $\forall\Delta\nu \in (\Delta\nu_0,\Delta\nu_1]\ \frac{\partial Q_{L(\Delta\nu)}(\nu,\tau_L)}{\partial\nu} > 0$ and therefore $\nu_L = 1$ wins more than half the votes ($\frac{Q_{L(\Delta\nu_1)}(1,\tau_L)}{Q_{L(\Delta\nu_1)}(1-\nu_R,\tau_L)} > 1$.

b) If $\bar{\Delta\nu} \geq \Delta\nu_1$, then $\forall\Delta\nu \in [\Delta\nu_0,\Delta\nu_1)\ \frac{\partial Q_{L(\Delta\nu)}(\nu,\tau_L)}{\partial\nu} < 0$ and therefore $\nu_L = 1$ wins more than half the votes ($\frac{Q_{L(\Delta\nu_0)}(0,\tau_L)}{Q_{L(\Delta\nu_0)}(1-\nu_R,\tau_L)} > 1$).

c) If $\Delta\nu_0 < \bar{\Delta\nu} < \Delta\nu_1$, then both $\nu_L = 0$ and $\nu_L = 1$ win more than half the votes ($\frac{Q_{L(\Delta\nu_0)}(0,\tau_L)}{Q_{L(\Delta\nu_0)}(1-\nu_R,\tau_L)} > 1$ and $\frac{Q_{L(\Delta\nu_1)}(1,\tau_L)}{Q_{L(\Delta\nu_1)}(1-\nu_R,\tau_L)} > 1$).

Following steps 1-3, we know that there can only be equilibria in extreme narratives. It also follows that equilibria where both parties offer the same narrative can only exist if $\bar{\Delta\nu} = 0$, since otherwise the payoffs of parties would not be equal, which by step 1 cannot be an equilibrium. In the last step, we prove that we only need to consider deviations to other extreme narratives, since if a deviation to an intermediate narrative is profitable, a deviation to an extreme narrative must be profitable as well. This allows us to characterise the conditions for different equilibria.

We start by considering deviation from $\nu_L = \nu_R \in \{0, 1\}$. As stated before, the precondition for this to be an equilibrium is that $\bar{\Delta\nu} = 0$. Here, we can show that any deviation to an intermediate level $\nu_L \in (0, 1)$ will lose. Suppose first that we start at $\nu_L = \nu_R = 0$. If *Left* deviates to $\nu_L \in (0, 1)$, then $\Delta\nu > \bar{\Delta\nu}$, which implies that $\frac{\partial Q_{L(\Delta\nu)}(\nu,\tau_L)}{\partial\nu} > 0$ by Lemma 2, so the payoff is $\frac{Q_{L(\nu_L-0)}(\nu_L,\tau_L)}{Q_{L(\nu_L-0)}(1,\tau_L)} < 1$. The relative payoff from deviating to $\nu_L = 1 = 1 - \nu_R$, meanwhile, is 1. Parallel logic applies to the case where $\nu_L = \nu_R = 1$: if *Left* deviates $\nu_L \in (0, 1)$, then $\Delta\nu < \bar{\Delta\nu}$, which implies that $\frac{\partial Q_{L(\Delta\nu)}(\nu,\tau_L)}{\partial\nu} < 0$ by Lemma 2, so the relative payoff is $\frac{Q_{L(\nu_L-1)}(\nu_L,\tau_L)}{Q_{L(\nu_L-1)}(0,\tau_L)} < 1$. A deviation to the extreme narrative $\nu_L = 0$ would again give a relative payoff of 1.

Let us now consider deviations from $\nu_L = 1-\nu_R \in \{0, 1\}$. We can show that if an intermediate deviation is profitable, then an extreme deviation is more profitable. To see this, suppose that we start at $\nu_L = 1$ and $\nu_R = 0$. Consider three strategies of *Left*: *Initial* at $\nu_L = 1$, *Deviation 1* to $\nu_L = \tilde{\nu}_L \in (0, 1)$, and *Deviation 2* to $\nu_L = 0$. Denote the following six voter types (see also Table 3): L_c vote for *Left* in all three cases; L_1 votes for *Left iff* $\nu_L = 1$; L_2 votes for *Left iff* $\nu_L \geq \tilde{\nu}_L$; L_3 votes for *Left iff* $\nu_L \leq \tilde{\nu}_L$; L_4 votes for *Left iff* $\nu_L = 0$; and R_c never votes for *Left*. Note that L_c and R_c, L_1 and L_3, as well as L_2 and L_4, respectively, define opposite types (mirrored across the origin), and hence by Lemma 1 $Q_{L_c}(\nu, \tau_L) = Q_{R_c}(1 - \nu, \tau_R)$, and so on for other opposing types.

Table 3: Voter type behaviour following deviations

	Initial ($\nu_L = 1$)	Deviation 1 ($\nu_L = \tilde{\nu}_L \in (0,1)$)	Deviation 2 ($\nu_L = 0$)
L_c	Left	Left	Left
L_1	Left	Right	Right
L_2	Left	Left	Right
L_3	Right	Left	Left
L_4	Right	Right	Left
R_c	Right	Right	Right

The narrative preference of L_c and R_c is uncertain, but we for the other types that $\frac{\partial Q_{L_1}(\nu,\tau_L)}{\partial \nu} > 0$ and $\frac{\partial Q_{L_2}(\nu,\tau_L)}{\partial \nu} > 0$ (and the opposite for the opposing types L_3 and L_4). Because *Initial* gives equal vote shares, *Deviation 1* is profitable *iff* $\frac{Q_{L_c}(\tilde{\nu}_L,\tau_L)+Q_{L_3}(\tilde{\nu}_L,\tau_L)+Q_{L_2}(\tilde{\nu}_L,\tau_L)}{Q_{L_c}(1,\tau_L)+Q_{L_3}(1,\tau_L)+Q_{L_2}(1,\tau_L)} > 1$. Because for each set of voters G $\frac{\partial Q_G(\nu,\tau_L)}{\partial \nu}$ is constant in ν, we can write the turnout for any ν as $Q_G(\nu,\tau_L) = Q_G(1,\tau_L) - \frac{\partial Q_G(\nu,\tau_L)}{\partial \nu}(1-\nu)$.[9] Hence, Deviation 1 is profitable if and only if

$$\begin{aligned} & Q_{L_c}(1,\tau_L) + Q_{L_3}(1,\tau_L) + Q_{L_2}(1,\tau_L) \\ & - \left(\frac{\partial Q_{L_c}(\nu,\tau_L)}{\partial \nu} + \frac{\partial Q_{L_3}(\nu,\tau_L)}{\partial \nu} + \frac{\partial Q_{L_2}(\nu,\tau_L)}{\partial \nu} \right)(1-\tilde{\nu}_L) \\ & > Q_{L_c}(1,\tau_L) + Q_{L_3}(1,\tau_L) + Q_{L_2}(1,\tau_L) \\ \iff & \frac{\partial Q_{L_c}(\nu,\tau_L)}{\partial \nu} + \frac{\partial Q_{L_3}(\nu,\tau_L)}{\partial \nu} + \frac{\partial Q_{L_2}(\nu,\tau_L)}{\partial \nu} < 0 \end{aligned}$$

By similar logic, *Deviation 2* is profitable if and only if $\frac{\partial Q_{L_c}(\nu,\tau_L)}{\partial \nu} + \frac{\partial Q_{L_3}(\nu,\tau_L)}{\partial \nu} + \frac{\partial Q_{L_4}(\nu,\tau_L)}{\partial \nu} < 0$. But note that *Deviation 1* being profitable implies that *Deviation 2* is profitable, since $\frac{\partial Q_{L_2}(\nu,\tau_L)}{\partial \nu} > 0$ and $\frac{\partial Q_{L_4}(\nu,\tau_L)}{\partial \nu} = -\frac{\partial Q_{L_2}(\nu,\tau_L)}{\partial \nu} < 0$. Hence, if any deviation to an intermediate narrative is profitable, a deviation to an extreme narrative is profitable as well. Parallel logic proves this for *Right*, as well as for the starting situation with $\nu_L = 1 - \nu_R = 0$.

Table 4: *Left's* relative payoffs of extreme narratives

	$\nu_R = 0$	$\nu_R = 1$
$\nu_L = 0$	κ	1
$\nu_L = 1$	1	$\frac{1}{\kappa}$

[9] Turnout under (ν,τ) is $Q_G(\nu,\tau) = \iint_G \frac{\lambda\nu(\chi y_1 + y_2) + (1-\tau)y_2 - \underline{u}(y)}{\overline{u}(y) - \underline{u}(y)} f(y_1,y_2)dA$. The derivative w.r.t ν is thus $\frac{\partial Q_G(\nu,\tau)}{\partial \nu} = \iint_G \frac{\lambda(\chi y_1 + y_2)(y)}{\overline{u}(y) - \underline{u}(y)} f(y_1,y_2)dA$. Hence, $Q_G(\nu,\tau) = Q_G(1,\tau) - (1-\nu)\frac{Q_G(\nu,\tau)}{\partial \nu}$.

Hence, equilibria can only be in extreme narratives, and we only need to consider deviations to other extreme narratives. We can therefore depict the game in the 2×2 matrix given by Table 4, which gives *Left's* vote share relative to *Right's*. Left chooses to maximise its payoff, and *Right* to minimise Left's payoff. if $\kappa > 1$, there is a unique equilibrium at $(\nu_L, \nu_R) = (0, 1)$ (left-wing populism); if $\kappa < 1$, there is a unique equilibrium at $(\nu_L, \nu_R) = (1, 0)$ (right-wing populism). If $\kappa = 1$, each of these two is an equilibrium, as well as $(0, 0)$ and $(1, 1)$. ■

B.8 Proposition 4

Proof.

For this proof, we consider how $\kappa(\chi)$ changes in χ. Using Proposition 3, we can then show that $\kappa(\chi) > 1$ for low values of χ, and $\kappa(\chi) < 1$ for high values of χ. This leads to left-wing and right-wing populism being the unique equilibrium, respectively.

For this purpose, recall that

$$\kappa = \frac{Q_{LL}(0, \tau_L) + Q_{LH_1-LH_2}(0, \tau_L) + Q_{LH_3-LH_4}(0, \tau_L)}{Q_{LL}(1, \tau_L) + Q_{LH_1-LH_2}(1, \tau_L) + Q_{LH_3-LH_4}(1, \tau_L)},$$

and define the following:

$$\underline{\kappa}(\chi) \equiv \frac{Q_{LL}(0, \tau_L) + Q_{LH_1-LH_2}(0, \tau_L) + Q_{LH_3-LH_4}(0, \tau_L)}{Q_{LL}(0, \tau_L) + Q_{LH_1-LH_2}(1, \tau_L) + Q_{LH_3-LH_4}(1, \tau_L)}$$
$$\overline{\kappa}(\chi) \equiv \frac{Q_{LL}(0, \tau_L) + Q_{LH_1-LH_2}(0, \tau_L) + Q_{LH_3-LH_4}(0, \tau_L)}{Q_{LH_1-LH_2}(1, \tau_L) + Q_{LH_3-LH_4}(1, \tau_L)},$$

where the denominator of $\underline{\kappa}(\chi)$ is strictly larger than of κ (since $Q_{LL}(0, \tau_L) > Q_{LL}(1, \tau_L)$, and the denominator of $\overline{\kappa}(\chi)$ is strictly smaller (as $0 < Q_{LL}(1, \tau_L)$). The numerators of both are the same as that of κ. Hence, for any finite $\chi > 0$ we can bound $\kappa(\chi)$ between $\underline{\kappa}(\chi) < \kappa(\chi) < \overline{\kappa}(\chi)$. We can then show that $\frac{\partial \underline{\kappa}(\chi)}{\partial \chi} < 0$ and $\frac{\partial \overline{\kappa}(\chi)}{\partial \chi} < 0$. This allows us to find $\underline{\chi}$ for which $\underline{\kappa}(\underline{\chi}) = 1$, implying that $\forall \chi \leq \underline{\chi}$, $\kappa(\chi) > 1$. This means that Left-wing populism will always be the unique equilibrium under low enough χ.

If $\rho < 1$, we can also find $\overline{\kappa}(\overline{\chi}) = 1$. This implies that $\forall \chi \geq \overline{\chi}, \kappa(\chi) < 1$, and hence that right-wing populism will be the unique equilibrium under high enough χ.

For the purpose of taking derivatives of the two boundaries, it will be helpful to write $\underline{\kappa}(\chi) = \frac{Z}{N}$ and $\overline{\kappa}(\chi) = \frac{Z}{\tilde{N}}$ where Z, N and $\tilde{N}$ are given as follows:

$$
\begin{aligned}
Z \equiv & \int_{-\infty}^{0}\int_{-\infty}^{0} f(y_1,y_2)dy_2dy_1 + \int_{0}^{\infty}\int_{-\infty}^{-\chi y_1} f(y_1,y_2)dy_2dy_1 \\
& + \int_{0}^{\infty}\int_{-\chi y_1}^{0} \frac{-\Delta\tau y_2}{-\Delta\tau y_2 + \lambda(\chi y_1 + y_2)} f(y_1,y_2)dy_2dy_1 \\
N \equiv & \int_{-\infty}^{0}\int_{-\infty}^{0} f(y_1,y_2)dy_2dy_1 \\
& + \int_{0}^{\infty}\int_{-\infty}^{-\chi y_1} \frac{-\Delta\tau y_2}{-\Delta\tau y_2 - \lambda(\chi y_1 + y_2)} f(y_1,y_2)dy_2dy_1 + \int_{0}^{\infty}\int_{-\chi y_1}^{0} f(y_1,y_2)dy_2dy_1 \\
\tilde{N} \equiv & N - \int_{-\infty}^{0}\int_{-\infty}^{0} f(y_1,y_2)dy_2dy_1,
\end{aligned}
$$

Note that $Z, N, \tilde{N}$ are strictly positive. Consider first the following derivatives:

$$
\begin{aligned}
\frac{\partial Z}{\partial \chi} &= -\int_{0}^{\infty} f(y_1,-\chi y_1)y_1dy_1 + \int_{0}^{\infty} f(y_1,-\chi y_1)y_1dy_1 \\
&\quad - \int_{0}^{\infty}\int_{-\chi y_1}^{0} \lambda y_1 \frac{-\Delta\tau y_2}{(-\Delta\tau y_2 + \lambda(\chi y_1 + y_2))^2} f(y_1,y_2)dy_2dy_1 \\
&= -\int_{0}^{\infty}\int_{-\chi y_1}^{0} \lambda y_1 \frac{-\Delta\tau y_2}{(-\Delta\tau y_2 + \lambda(\chi y_1 + y_2))^2} f(y_1,y_2)dy_2dy_1 \\
\frac{\partial N}{\partial \chi} = \frac{\partial \tilde{N}}{\partial \chi} &= -\int_{0}^{\infty} f(y_1,-\chi y_1)y_1dy_1 + \int_{0}^{\infty} f(y_1,-\chi y_1)y_1dy_1 \\
&\quad \int_{0}^{\infty}\int_{-\chi y_1}^{0} \lambda y_1 \frac{-\Delta\tau y_2}{(-\Delta\tau y_2 + \lambda(\chi y_1 + y_2))^2} f(y_1,y_2)dy_2dy_1 \\
&= \int_{0}^{\infty}\int_{-\chi y_1}^{0} \lambda y_1 \frac{-\Delta\tau y_2}{(-\Delta\tau y_2 + \lambda(\chi y_1 + y_2))^2} f(y_1,y_2)dy_2dy_1
\end{aligned}
$$

This allows us to calculate the derivatives of $\underline{\kappa}$ and $\overline{\kappa}$:

$$\begin{aligned}
\frac{\partial \underline{\kappa}}{\partial \chi} N^2 =& - N \left(\int_0^{\infty} \int_{-\chi y_1}^{0} \lambda y_1 \frac{-\Delta\tau y_2}{(-\Delta\tau y_2 + \lambda(\chi y_1 + y_2))^2} f(y_1, y_2) dy_2 dy_1 \right) \\
& - Z \left(\int_0^{\infty} \int_{-\infty}^{-\chi y_1} \lambda y_1 \frac{-\Delta\tau y_2}{(-\Delta\tau y_2 - \lambda(\chi y_1 + y_2))^2} f(y_1, y_2) dy_2 dy_1 \right) \\
\frac{\partial \overline{\kappa}}{\partial \chi} \tilde{N}^2 =& - \tilde{N} \left(\int_0^{\infty} \int_{-\chi y_1}^{0} \lambda y_1 \frac{-\Delta\tau y_2}{(-\Delta\tau y_2 + \lambda(\chi y_1 + y_2))^2} f(y_1, y_2) dy_2 dy_1 \right) \\
& - Z \left(\int_0^{\infty} \int_{-\infty}^{-\chi y_1} \lambda y_1 \frac{-\Delta\tau y_2}{(-\Delta\tau y_2 - \lambda(\chi y_1 + y_2))^2} f(y_1, y_2) dy_2 dy_1 \right)
\end{aligned}$$

It follows that $\frac{\partial \underline{\kappa}}{\partial \chi}$ and $\frac{\partial \overline{\kappa}}{\partial \chi}$ are negative: $\left(\int_0^{\infty} \int_{-\chi y_1}^{0} \lambda y_1 \frac{-\Delta\tau y_2}{(-\Delta\tau y_2 + \lambda(\chi y_1 + y_2))^2} f(y_1, y_2) dy_2 dy_1 \right)$ is positive, $\left(\int_0^{\infty} \int_{-\infty}^{-\chi y_1} \lambda y_1 \frac{-\Delta\tau y_2}{(-\Delta\tau y_2 - \lambda(\chi y_1 + y_2))^2} f(y_1, y_2) dy_2 dy_1 \right)$ is positive, and $\tilde{N}, N, Z$ are positive.

Hence, the boundaries $\underline{\kappa}$ and $\overline{\kappa}$ are strictly decreasing in χ. We can now compute the limits to find the conditions under which there exist values of χ for which $\underline{\kappa}(\chi) = 1$ and $\overline{\kappa}(\chi) = 1$:

$$\begin{aligned}
\lim_{\chi \to \infty} \underline{\kappa} &= \frac{\int_{-\infty}^{0} \int_{-\infty}^{0} f(y_1, y_2) dy_2 dy_1}{\int_{-\infty}^{0} \int_{-\infty}^{0} f(y_1, y_2) dy_2 dy_1 + \int_0^{\infty} \int_{-\infty}^{0} f(y_1, y_2) dy_2 dy_1} < 1 \\
\lim_{\chi \to \infty} \overline{\kappa} &= \frac{\int_{-\infty}^{0} \int_{-\infty}^{0} f(y_1, y_2) dy_2 dy_1}{\int_0^{\infty} \int_{-\infty}^{0} f(y_1, y_2) dy_2 dy_1} = \rho \\
\lim_{\chi \to 0} \underline{\kappa} &= \frac{\int_{-\infty}^{0} \int_{-\infty}^{0} f(y_1, y_2) dy_2 dy_1 + \int_0^{\infty} \int_{-\infty}^{0} f(y_1, y_2) dy_2 dy_1}{\int_{-\infty}^{0} \int_{-\infty}^{0} f(y_1, y_2) dy_2 dy_1 + \int_0^{\infty} \int_{-\infty}^{0} \frac{-\Delta\tau y_2}{-\Delta\tau y_2 - \lambda y_2} f(y_1, y_2) dy_2 dy_1} > 1 \\
\lim_{\chi \to 0} \overline{\kappa} &= \frac{\int_{-\infty}^{0} \int_{-\infty}^{0} f(y_1, y_2) dy_2 dy_1 + \int_0^{\infty} \int_{-\infty}^{0} f(y_1, y_2) dy_2 dy_1}{\int_0^{\infty} \int_{-\infty}^{0} \frac{-\Delta\tau y_2}{-\Delta\tau y_2 - \lambda y_2} f(y_1, y_2) dy_2 dy_1} > 1
\end{aligned}$$

Because $\lim_{\chi \to 0} \underline{\kappa} > 1$ and $\lim_{\chi \to \infty} \underline{\kappa} < 1$, and the cutoff is continuously decreasing in χ, we can always define implicitly $\underline{\chi}$ via $\underline{\kappa}(\underline{\chi}) = 1$. We thus have that $\forall \chi \leq \underline{\chi}$, $\underline{\kappa} \geq 1$, and hence $\kappa > 1$, giving left-wing populism as the unique equilibrium.

We also have that $\lim_{\chi \to 0} \overline{\kappa} > 1$. If $\rho < 1$, we have that $\lim_{\chi \to \infty} \overline{\kappa} < 1$. Hence, if $\rho < 1$, we can implicitly define $\overline{\chi}$ by $\overline{\kappa}(\overline{\chi}) = 1$. In that case, $\forall \chi \geq \overline{\chi}$, $\overline{\kappa} \leq 1$, and hence $\kappa < 1$, which gives right-wing populism as the unique equilibrium. ■

B.9 Corollary 2

Proof. The proof for this corollary follows directly from combining the equilibrium condition from Proposition 3 with the condition that an equilibrium is the most polarized. In particular, the first condition given, that $Q_{LL-LH_1}(0,\tau_R) - Q_{LH_4}(1,\tau_R) < \min\{0, Q_{LH_3}(1,\tau_R) - Q_{LH_2}(0,\tau_R)\}$, is simply the condition that $\kappa < 1$ (right-wing populism is unique) combined with the condition that $Q_{LL-LH_1}(0,\tau_R) < Q_{LH_4}(1,\tau_R)$ (right-wing populism is less polarized). Similarily, the second condition given, $Q_{LL-LH_1}(0,\tau_R) - Q_{LH_4}(1,\tau_R) > \max\{0, Q_{LH_3}(1,\tau_R) - Q_{LH_2}(0,\tau_R)\}$, is the combination of $\kappa > 1$ and $Q_{LL-LH_1}(0,\tau_R) > Q_{LH_4}(1,\tau_R)$. As the mass of core voters approaches zero, $Q_{LH_3}(1,\tau_R) - Q_{LH_2}(0,\tau_R)$ becomes negligible, and the unique equilibrium will be least polarized. This completes the proof. ■

B.10 Proposition 5-6

Proof. This proof relates to both propositions 5 and 6. It will characterise the existence and set of equilibria for different levels of β. The relative size of the following derivatives will matter in shaping the existence of narratives, where π_{ν_L,ν_R} again denotes the relative votes of *Left*:

$$\frac{\partial \pi_{0,0}}{\partial \beta} = \frac{Q_{LL}(0,\tau_L)}{Q_{HH-HL_4}(0,\tau_R)} < \kappa$$
$$\frac{\partial \pi_{0,1}}{\partial \beta} = \frac{Q_{LL}(0,\tau_L)}{Q_{HH}(1,\tau_R) + Q_{HL_{1-3}}(1,\tau_R) + Q_{LH_4}(1,\tau_R)} < \frac{Q_{LL}(0,\tau_L)}{Q_{HH-HL_4}(1,\tau_R)} < 1$$
$$\frac{\partial \pi_{1,0}}{\partial \beta} < 0$$
$$\frac{\partial \pi_{1,1}}{\partial \beta} = \frac{Q_{LL}(1,\tau_L)}{Q_{HH-HL_4}(1,\tau_R)} < \frac{1}{\kappa}$$

We will consider the existence and characterisation of equilibria separately for the cases of $\kappa > 1$ and $\kappa < 1$.

Suppose first that $\kappa > 1$. Note that $\frac{\partial \pi_{0,0}}{\partial \beta} > \frac{\partial \pi_{1,1}}{\partial \beta}$ *iff* $\kappa > \frac{Q_{LL}(1,\tau_L)}{Q_{LL}(0,\tau_L)}$, where the last term is strictly less than 1. Note further that $\frac{\partial \pi_{0,0}}{\partial \beta} > \frac{\partial \pi_{0,1}}{\partial \beta}$ *iff* $\frac{Q_{HH}(1,\tau_R)+Q_{HL_{1-3}}(1,\tau_R)+Q_{LH_4}(1,\tau_R)}{Q_{HH-HL_4}(0,\tau_R)} > 1$, where the left-hand side is strictly

greater than κ. Hence $\kappa > 1$ implies that $\frac{\partial \pi_{0,0}}{\partial \beta} > \max[\frac{\partial \pi_{1,1}}{\partial \beta}, \frac{\partial \pi_{0,1}}{\partial \beta}]$. We therefore have two cases to consider: (a) $\frac{\partial \pi_{1,1}}{\partial \beta} \leq \frac{\partial \pi_{0,1}}{\partial \beta}$, and (b) $\frac{\partial \pi_{1,1}}{\partial \beta} > \frac{\partial \pi_{0,1}}{\partial \beta}$. The two cases are depicted in Figure C.4, which plot the payoff of *Left* for different combinations of narratives.

Case (a) is shown in Panel *a* of this Figure, which coincides with Panel *a* shown in Figure 7 of the main text. In that case, for any β, $\pi_{0,0} > \pi_{0,1} > \pi_{1,1}$, while $\pi_{1,0} \leq 1$. Hence, *Right* minimises *Left's* maximal payoff by choosing $\nu_R = 1$, and *Left* does so by choosing $\nu_L = 0$. There does not exist a profitable deviation nor any other equilibria for any $\beta \geq 0$. Note also that *Left* wins a strict majority for $\beta \geq 0$, and *Left's* equilibrium payoff is increasing in β.

Case (b) is shown in Panel *b* of the Figure, which keeps the slope of $\pi_{0,0}$ constant, but changes the relative slopes of $\pi_{1,1}$ and $\pi_{0,1}$. Now, $\exists! \tilde{\beta}_1 : \pi_{1,1}(\tilde{\beta}_1) = \pi_{0,1}(\tilde{\beta}_1)$. For any $\beta \leq \tilde{\beta}_1$, $\pi_{0,0} > \pi_{0,1} \geq \pi_{1,1}$, the analysis is exactly the same as for case (a). However, for any $\beta > \tilde{\beta}_1$, $\pi_{0,0} > \pi_{1,1} > \pi_{0,1}$. Hence, left-wing populism at $(\nu_L, \nu_R) = (0, 1)$ is not an equilibrium, since *Left* can make itself better off by switching narratives to $\nu_L = 1$. Indeed, no pure strategy is an equilibrium in this case, since one party can make itself better off for any combination of pure strategies in extreme narratives.

Suppose next that $\kappa < 1$. Here there are now three relevant cases to consider: (a) $\frac{\partial \pi_{0,1}}{\partial \beta} \geq \frac{\partial \pi_{0,0}}{\partial \beta}$; (b) $\frac{\partial \pi_{0,1}}{\partial \beta} < \frac{\partial \pi_{0,0}}{\partial \beta}$ and $\frac{\partial \pi_{1,1}}{\partial \beta} \geq \frac{\partial \pi_{0,1}}{\partial \beta}$; and (c) $\frac{\partial \pi_{1,1}}{\partial \beta} < \frac{\partial \pi_{0,1}}{\partial \beta} < \frac{\partial \pi_{0,0}}{\partial \beta}$. The corresponding cases are shown in Figure C.5, where again each panel corresponds with the matching case.

Panel (a) of the figure corresponds to Panel *b* of Figure 7 from the main text. Here, the equilibrium in pure strategies is $(\nu_L, \nu_R) = (1, 0)$ if $\beta \leq \underline{\beta}$,and $(0, 0)$ if $\beta \geq \underline{\beta}$. Note that $\pi_{1,1}$ may cut across $\pi_{0,0}$ for high enough β, depending on the slopes, but it does not make a difference for the equilibria for this case; as long as $\frac{\partial \pi_{0,1}}{\partial \beta} \geq \frac{\partial \pi_{0,0}}{\partial \beta}$, $\pi_{0,1} > \pi_{0,0}$ for any β. Note that *Right* wins a strict majority of votes for $0 < \beta < \overline{\beta}$, and *Left* for $\beta > \overline{\beta}$.

Panel (b) now considers what happens when $\frac{\partial \pi_{0,1}}{\partial \beta} < \frac{\partial \pi_{0,0}}{\partial \beta}$ and $\frac{\partial \pi_{1,1}}{\partial \beta} \geq \frac{\partial \pi_{0,1}}{\partial \beta}$. This implies that $\pi_{1,1} > \pi_{0,1} \forall \beta$, and that $\exists! \tilde{\beta}_2 : \pi_{0,0}(\tilde{\beta}_2) = \pi_{0,1}(\tilde{\beta}_2)$. The

analysis is exactly the same as for case (a) whenever $\beta \leq \tilde{\beta}_2$. However, once $\beta > \tilde{\beta}_2$, we have that $\pi_{0,0} > \pi_{0,1}$ and $\pi_{1,1} > \pi_{0,1}$. Thus, there is no equilibrium in pure strategies: $\pi_{0,0} > \pi_{0,1}$, so *Right* deviates from $(\nu_L, \nu_R) = (0, 0)$; $\pi_{1,1} > \pi_{0,1}$ so *Left* deviates from $(0, 1)$; $\pi_{1,1} > \pi_{1,0}$ so *Right* deviates from $(1, 1)$; and $\pi_{0,0} > \pi_{1,0}$, so *Left* deviates from $(1, 0)$.

Panel (c) now considers the case when $\frac{\partial \pi_{1,1}}{\partial \beta} < \frac{\partial \pi_{0,1}}{\partial \beta} < \frac{\partial \pi_{0,0}}{\partial \beta}$. Now, $\exists! \tilde{\beta}_3 : \pi_{1,1}(\tilde{\beta}_3) = \pi_{0,1}(\tilde{\beta}_3)$. The analysis is the same as for case (b) whenever $\beta < \max[\tilde{\beta}_2, \tilde{\beta}_3]$. However, whenever $\beta \geq \max[\tilde{\beta}_2, \tilde{\beta}_3]$, there exists an equilibrium at $(\nu_L, \nu_R) = (0, 1)$: since $\beta > \tilde{\beta}_3$, $\pi_{1,1} < \pi_{0,1}$, and hence *Left* does not wish to deviate from $(0, 1)$; moreover, because $\beta > \tilde{\beta}_2$, $\pi_{0,0} > \pi_{0,1}$, and hence *Right* also does not wish to deviate.

To summarise, the existence and characterisiation of equilibria is as follows:

1. If $\kappa > 1$
    a) $\frac{\partial \pi_{1,1}}{\partial \beta} \leq \frac{\partial \pi_{0,1}}{\partial \beta}$: An equilibrium exists for any β. $(\nu_L, \nu_R) = (0, 1)$ is the unique equilibrium. *Left* wins a strict majority of votes $\forall \beta > 0$.
    b) $\frac{\partial \pi_{1,1}}{\partial \beta} > \frac{\partial \pi_{0,1}}{\partial \beta}$: There exists $\tilde{\beta}_1 > 0$ such that a unique equilibrium in pure strategies exists at $(\nu_L, \nu_R) = (0, 1)$ $\forall \beta \leq \tilde{\beta}_1$ where *Left* wins a strict majority of votes $\forall \beta > 0$; and no pure strategy equilibrium exists $\forall \beta > \tilde{\beta}_1$.
2. If $\kappa < 1$:
    a) $\frac{\partial \pi_{0,1}}{\partial \beta} \geq \frac{\partial \pi_{0,0}}{\partial \beta}$: An equilibrium exists for any β. There exists a $\underline{\beta} > 0$ such that $(\nu_L, \nu_R) = (1, 0)$ is an equilibrium *iff* $\beta \leq \underline{\beta}$, and $(0, 0)$ is an equilibrium *iff* $\beta \geq \underline{\beta}$. There exists a $\overline{\beta} > \underline{\beta}$ such that *Right* receives a strict majority for $\beta < \overline{\beta}$, and *Left* for $\beta > \overline{\beta}$.
    b) $\frac{\partial \pi_{0,1}}{\partial \beta} < \frac{\partial \pi_{0,0}}{\partial \beta}$ and $\frac{\partial \pi_{1,1}}{\partial \beta} \geq \frac{\partial \pi_{0,1}}{\partial \beta}$: There exists a $\tilde{\beta}_2 > \overline{\beta}$ such that a unique equilibrium in pure strategies exists as given by point (a) for $\beta \leq \tilde{\beta}_2$; and no pure strategy equilibrium exists $\forall \beta > \tilde{\beta}_2$.
    c) $\frac{\partial \pi_{1,1}}{\partial \beta} < \frac{\partial \pi_{0,1}}{\partial \beta} < \frac{\partial \pi_{0,0}}{\partial \beta}$: There exists $\tilde{\beta}_3$ such that the analysis is the same as in point b) for all $\beta < \max[\tilde{\beta}_2, \tilde{\beta}_3]$. If $\beta \geq \max[\tilde{\beta}_2, \tilde{\beta}_3]$, then there exists an equilibrium at $(\nu_L, \nu_R) = (0, 1)$, where *Left* wins a

majority of votes.

For Proposition 5, note that a sufficient condition for equilibrium existence is that either $\kappa > 1$ (resp. $\kappa < 1$) and no solution exists to $\pi_{0,1}(\beta) = \pi_{1,1}(\beta)$ (resp. $\pi_{0,1}(\beta) = \pi_{0,0}(\beta)$); or that $\beta \leq \tilde{\beta}_1$ (resp. $\beta \leq \tilde{\beta}_2$) when the solution does exist. Proposition 6 follows the analysis given above whenever β is below the relevant cutoff such that existence is guaranteed. This completes the proof. ■

B.11 Corollary 3

Proof. The proof for this corollary follows from the proof for Propositions 5-6. In particular, note that whenever $\kappa > 1$, the equilibrium payoff (if it exists) is $\pi_{0,1}$, which is strictly increasing in β (thick dashed lines of Figure C.4). Hence, for $\kappa > 1$, *Left's* pure-strategy equilibrium payoff is always higher with higher β. Whenever $\kappa < 1$, *Left's* equilibrium payoff is $\pi_{1,0} \forall \beta \leq \underline{\beta}$, which is strictly decreasing in β; it is either $\pi_{0,0}$ or $\pi_{0,1}$ for $\beta > \underline{\beta}$, whenever the payoff exists, each of which is strictly increasing in β (thick dashed lines of Figure C.5). Hence, for $\kappa < 1$, *Left's* pure-strategy equilibrium payoff is lowest, and Right's highest, at $\beta = \underline{\beta}$. ■

B.12 Proposition 7

Proof. Consider first right-wing populism, $\Delta\nu = 1$. The total number of votes for *Right* and *Left* (after transposing *Left's* voters into *Right's* by using the symmetry condition of Lemma 1, which also holds for the new turnout function $\tilde{q}$) are given by the following:

$$
\begin{aligned}
Q_{R(1)}(0,\tau_R) &= (1+\beta)Q_{LL}(0,\tau_R) + Q_{LH_1}(0,\tau_R) \\
&\quad + Q_{HL_3-HL_4}(0,\tau_R) + Q_{HL_2}(0,\tau_R) \\
&= (1+\beta)\int_{-\infty}^{0}\int_{-\infty}^{0} \frac{-\tau_R y_2 - \lambda(\chi y_1 + y_2)}{-y_2 - \lambda(\chi y_1 + y_2)} f(y_1,y_2)dy_2dy_1 \\
&\quad + \int_{0}^{\infty}\int_{-\infty}^{g_1(\Delta\tau,y_1)} \frac{-\tau_R y_2 - \lambda(\chi y_1 + y_2)}{-y_2 - \lambda(\chi y_1 + y_2)} f(y_1,y_2)dy_2dy_1 \\
&\quad + \int_{-\infty}^{0}\int_{0}^{-\chi y_1} \frac{(1-\tau_R) y_2 - \lambda(\chi y_1 + y_2)}{y_2 - \lambda(\chi y_1 + y_2)} f(y_1,y_2)dy_2dy_1 \\
&\quad + \int_{-\infty}^{0}\int_{-\chi y_1}^{g_1(\Delta\tau,y_1)} \frac{(1-\tau_R) y_2}{y_2 + \lambda(\chi y_1 + y_2)} f(y_1,y_2)dy_2dy_1
\end{aligned}
$$

$$
\begin{aligned}
Q_{L(1)}(1,\tau_L) &= Q_{LL}(0,1-\tau_L) + Q_{LH_1}(0,1-\tau_L) + Q_{HL_3-HL_4}(0,1-\tau_L) \\
&\quad + Q_{HL_2}(0,1-\tau_L) \\
&= \int_{-\infty}^{0}\int_{-\infty}^{0} \frac{-(1-\tau_L) y_2 - \lambda(\chi y_1 + y_2)}{-y_2 - \lambda(\chi y_1 + y_2)} f(y_1,y_2)dy_2dy_1 \\
&\quad + \int_{0}^{\infty}\int_{-\infty}^{g_1(\Delta\tau,y_1)} \frac{-(1-\tau_L) y_2 - \lambda(\chi y_1 + y_2)}{-y_2 - \lambda(\chi y_1 + y_2)} f(y_1,y_2)dy_2dy_1 \\
&\quad + \int_{-\infty}^{0}\int_{0}^{-\chi y_1} \frac{\tau_L y_2 - \lambda(\chi y_1 + y_2)}{y_2 - \lambda(\chi y_1 + y_2)} f(y_1,y_2)dy_2dy_1 \\
&\quad + \int_{-\infty}^{0}\int_{-\chi y_1}^{g_1(\Delta\tau,y_1)} \frac{\tau_L y_2}{y_2 + \lambda(\chi y_1 + y_2)} f(y_1,y_2)dy_2dy_1,
\end{aligned}
$$

where $g_1(\Delta\tau, y_{1i}) \equiv -\frac{\lambda\chi}{\lambda-\Delta\tau} y_{1i}$ separates the supporters of *Left* and *Right* at right-wing populism.

We can compute the derivatives with respect to τ_R (noting that $\Delta\tau' = -1$):

$$
\begin{aligned}
&\frac{\partial Q_{L(1)}(1,\tau_L)}{\partial \tau_R} = \\
&\int_{0}^{\infty} -g_1'(\Delta\tau,y_1) \frac{-(1-\tau_L) g_1(\Delta\tau,y_1) - \lambda(\chi y_1 + g_1(\Delta\tau,y_1))}{-g_1(\Delta\tau,y_1) - \lambda(\chi y_1 + g_1(\Delta\tau,y_1))} f(y_1, g_1(\Delta\tau,y_1))dy_1 \\
&\quad + \int_{-\infty}^{0} -g_1'(\Delta\tau,y_1) \frac{\tau_L g_1(\Delta\tau,y_1)}{g_1(\Delta\tau,y_1) + \lambda(\chi y_1 + g_1(\Delta\tau,y_1))} f(y_1, g_1(\Delta\tau,y_1))dy_1 = 0
\end{aligned}
$$

where the equality to 0 holds because opposite types cancel each other out. To see this, note first that $g_1(\Delta\tau, y_{1i}) = -g_1(\Delta\tau, -y_{1i})$ and hence $g_1'(\Delta\tau, y_{1i}) = -g_1'(\Delta\tau, -y_{1i})$. Moreover, along the line $y_{2i} = g_1'(\Delta\tau, y_{1i})$ both of the following

two hold:

$$q(y_{1i}, g_1(\Delta\tau, y_{1i}), \tau, \nu) = q(-y_{1i}, g_1(\Delta\tau, -y_{1i}), 1-\tau, 1-\nu)$$
$$f(y_{1i}, g_1(\Delta\tau, y_{1i})) = f(-y_{1i}, g_1(\Delta\tau, -y_{1i}))$$

Hence, the following follow:

$$g_1'(\Delta\tau, y_{1i}) q(y_{1i}, g_1(\Delta\tau, y_i), \tau, \nu) f(y_{1i}, g_1(\Delta\tau, y_i) =$$
$$-g_1'(\Delta\tau, -y_{1i}) q(-y_{1i}, g_1(\Delta\tau, -y_{1i}), 1-\tau, 1-\nu) f(-y_{1i}, g_1(\Delta\tau, -y_1)$$
$$\to \int_a^b g_1'(\Delta\tau, y_1) q(y_1, g_1(\Delta\tau, y_1), \tau, \nu) f(y_1, g_1(\Delta\tau, y_1) dy_1$$
$$= -\int_{-b}^{-a} g_1'(\Delta\tau, y_1) q(y_1, g_1(\Delta\tau, y_1), 1-\tau, 1-\nu) f(y_1, g_1(\Delta\tau, y_1) dy_1$$
$$\to \int_0^\infty g_1'(\Delta\tau, y_1) q(y_1, g_1(\Delta\tau, y_1), \tau, \nu) f(y_1, g_1(\Delta\tau, y_1) dy_1$$
$$= -\int_{-\infty}^0 g_1'(\Delta\tau, y_1) q(y_1, g_1(\Delta\tau, y_1), 1-\tau, 1-\nu) f(y_1, g_1(\Delta\tau, y_1) dy_1$$

The derivative of *Right's* payoff with respect to τ_R is then given by the following (where again, two terms cancel out just like with the derivative for *Left*):

$$\frac{\partial Q_{R(1)}(0, \tau_R)}{\partial \tau_R} = (1+\beta) \int_{-\infty}^0 \int_{-\infty}^0 \frac{-y_2}{-y_2 - \lambda(\chi y_1 + y_2)} f(y_1, y_2) dy_2 dy_1$$
$$+ \int_0^\infty \int_{-\infty}^{g(\Delta\tau, y_1)} \frac{-y_2}{-y_2 - \lambda(\chi y_1 + y_2)} f(y_1, y_2) dy_2 dy_1$$
$$+ \int_{-\infty}^0 \int_0^{-\chi y_1} \frac{-y_2}{y_2 - \lambda(\chi y_1 + y_2)} f(y_1, y_2) dy_2 dy_1$$
$$+ \int_{-\infty}^0 \int_{-\chi y_1}^{g(\Delta\tau, y_1)} \frac{-y_2}{y_2 + \lambda(\chi y_1 + y_2)} f(y_1, y_2) dy_2 dy_1$$

Note that the first two terms are positive and the last two terms are negative.

Following the same steps, we find the derivative with respect to *Left's* policy

change:

$$\frac{\partial Q_{R(1)}(0,\tau_L)}{\partial \tau_L} = 0$$

$$\begin{aligned}\frac{\partial Q_{L(1)}(1,\tau_L)}{\partial \tau_L} &= \int_{-\infty}^{0}\int_{-\infty}^{0} \frac{y_2}{-y_2-\lambda(\chi y_1+y_2)} f(y_1,y_2)dy_2dy_1 \\ &+ \int_{0}^{\infty}\int_{-\infty}^{g(\Delta\tau,y_1)} \frac{y_2}{-y_2-\lambda(\chi y_1+y_2)} f(y_1,y_2)dy_2dy_1 \\ &+ \int_{-\infty}^{0}\int_{0}^{-\chi y_1} \frac{y_2}{y_2-\lambda(\chi y_1+y_2)} f(y_1,y_2)dy_2dy_1 \\ &+ \int_{-\infty}^{0}\int_{-\chi y_1}^{g(\Delta\tau,y_1)} \frac{y_2}{y_2+\lambda(\chi y_1+y_2)} f(y_1,y_2)dy_2dy_1,\end{aligned}$$

where the first two terms of the latter expression are negative and the last two terms are positive.

Putting the above together, the change in total payoff from changing τ_L and τ_R is, for *Left* and *Right* respectively, $\frac{\partial Q_{L(1)}(1,\tau_L)}{\partial \tau_L}\left(Q_{R(1)}(0,\tau_R)\right)^{-1}$ and $\frac{\partial Q_{R(1)}(0,\tau_R)}{\partial \tau_R}\left(Q_{L(1)}(1,\tau_L)\right)^{-1}$.

We can now see that when $\beta = 0$, $Q_{L(1)}(1,\tau_L) = Q_{R(1)}(0,\tau_R)$ and also $\frac{\partial Q_{L(1)}(1,\tau_L)}{\partial \tau_L}\left(Q_{R(1)}(0,\tau_R)\right)^{-1} = -\frac{\partial Q_{R(1)}(0,\tau_R)}{\partial \tau_R}\left(Q_{L(1)}(1,\tau_L)\right)^{-1}$. Hence, in that case the two parties have opposite incentives in terms of increasing their tax rate, which means that they have exactly the same incentives to increase or decrease $\Delta\tau$: either both wish to decrease $\Delta\tau$ or to increase it. However, for $\beta > 0$, $Q_{R(1)}(0,\tau_R) > Q_{L(1)}(1,\tau_L)$, and also $\frac{\partial Q_{R(1)}(0,\tau_R)}{\partial \tau_R} > -\frac{\partial Q_{L(1)}(1,\tau_L)}{\partial \tau_L}$. Hence, in that case we have that $\frac{\partial Q_{R(1)}(0,\tau_R)}{\partial \tau_R}\left(Q_{L(1)}(1,\tau_L)\right)^{-1} > -\frac{\partial Q_{L(1)}(1,\tau_L)}{\partial \tau_L}\left(Q_{R(1)}(0,\tau_R)\right)^{-1}$.

Right has more of an incentive to decrease $\Delta\tau$ (by increasing τ_R) than *Left* (by decreasing τ_L). In particular, this means that it is possible for the parties to have opposing incentives with $\frac{\partial Q_{R(1)}(0,\tau_R)}{\partial \tau_R}\left(Q_{L(1)}(1,\tau_L)\right)^{-1} > 0 > -\frac{\partial Q_{L(1)}(1,\tau_L)}{\partial \tau_L}\left(Q_{R(1)}(0,\tau_R)\right)^{-1}$: In this case both *Right* and *Left* want to increase their tax rate.

Now let us consider left-wing populism, which yields payoffs for *Left* and *Right* of $Q_{L(-1)}(0,\tau_L)$ and $Q_{R(-1)}(1,\tau_R)$, respectively. Denoting $g_2(\Delta\tau, y_{1i}) \equiv -\frac{\lambda\chi}{\lambda+\Delta\tau}y_{1i}$ and following the same steps as before, we can find the derivative

with respect to τ_L:

$$\frac{\partial Q_{R(-1)}(1,\tau_R)}{\partial \tau_L} = 0$$

$$\begin{aligned}\frac{\partial Q_{L(-1)}(0,\tau_L)}{\partial \tau_L} = &\int_0^{\infty}\int_{g_2(\Delta\tau,y_1)}^{0} \frac{y_2}{-y_2+\lambda(\chi y_1+y_2)} f(y_1,y_2)dy_2dy_1 \\ &+(1+\beta)\int_0^{\infty}\int_0^{\infty} \frac{y_2}{y_2+\lambda(\chi y_1+y_2)} f(y_1,y_2)dy_2dy_1 \\ &+\int_{-\infty}^{0}\int_{-\chi y_1}^{\infty} \frac{y_2}{y_2+\lambda(\chi y_1+y_2)} f(y_1,y_2)dy_2dy_1 \\ &+\int_{-\infty}^{0}\int_{g_2(\Delta\tau,y_1)}^{-\chi y_1} \frac{y_2}{y_2-\lambda(\chi y_1+y_2)} f(y_1,y_2)dy_2dy_1,\end{aligned}$$

where the first term of $\frac{\partial Q_{L(-1)}(0,\tau_L)}{\partial \tau_R}$ is negative, and all others are positive. We also have that

$$\frac{\partial Q_{L(-1)}(0,\tau_L)}{\partial \tau_R} = 0$$

$$\begin{aligned}\frac{\partial Q_{R(-1)}(1,\tau_R)}{\partial \tau_R} = &\int_0^{\infty}\int_{g_2(\Delta\tau,y_1)}^{0} \frac{-y_2}{-y_2+\lambda(\chi y_1+y_2)} f(y_1,y_2)dy_2dy_1 \\ &+\int_0^{\infty}\int_0^{\infty} \frac{-y_2}{y_2+\lambda(\chi y_1+y_2)} f(y_1,y_2)dy_2dy_1 \\ &+\int_{-\infty}^{0}\int_{-\chi y_1}^{\infty} \frac{-y_2}{y_2+\lambda(\chi y_1+y_2)} f(y_1,y_2)dy_2dy_1 \\ &+\int_{-\infty}^{0}\int_{g_2(\Delta\tau,y_1)}^{-\chi y_1} \frac{-y_2}{y_2-\lambda(\chi y_1+y_2)} f(y_1,y_2)dy_2dy_1\end{aligned}$$

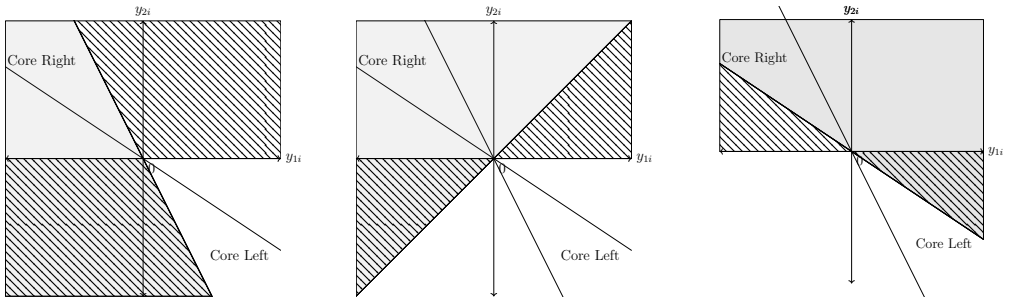
where of the latter expression, the first term is positive, and all others are negative.

By the same logic as discussed for right-wing populism, for any $\beta > 0$, we have that $\frac{\partial Q_{L(-1)}(0,\tau_L)}{\partial \tau_L}\left(Q_{R(-1)}(1,\tau_R)\right)^{-1} > -\frac{\partial Q_{R(-1)}(1,\tau_R)}{\partial \tau_R}\left(Q_{L(-1)}(0,\tau_L)\right)^{-1}$. Hence, the left-wing populist has stronger incentives to increase $\Delta\tau$ (by increasing τ_L) than the right-wing opponent has to increase $\Delta\tau$ (by decreasing τ_R). ■

C Additional Figures

Figure C.1: Polarization across different cross-party narratives

(a) $\Delta\nu = 1$ (b) $\Delta\nu \in (0, 1)$ (c) $\Delta\nu = -1$

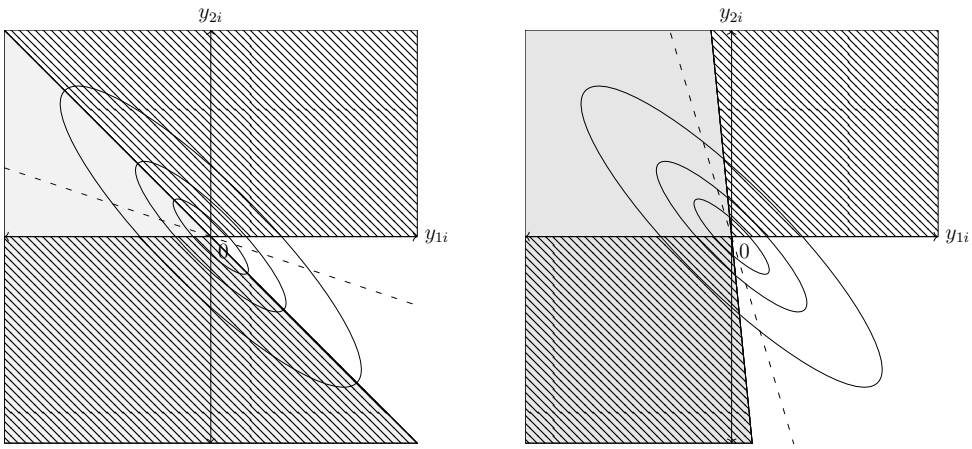


Notes: The figure shows support for *Right* in grey versus support for *Left* in white, for different values of cross-party narrative differences $\Delta\nu$. The dashed areas denote "culturally polarized" voters, who vote against their economic policy preferences. Moving from Panel *a* to Panel *c*

, the narratives of *Left* become more scapegoating relative to to *Right*.

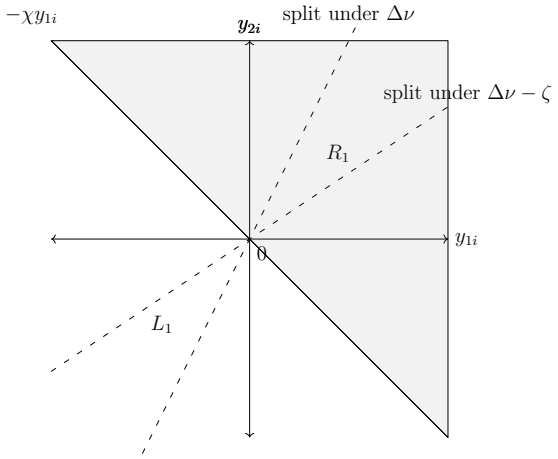
Figure C.2: Right-wing populism polarization under different χ

(a) $\chi = 1$ (b) $\chi > 1$



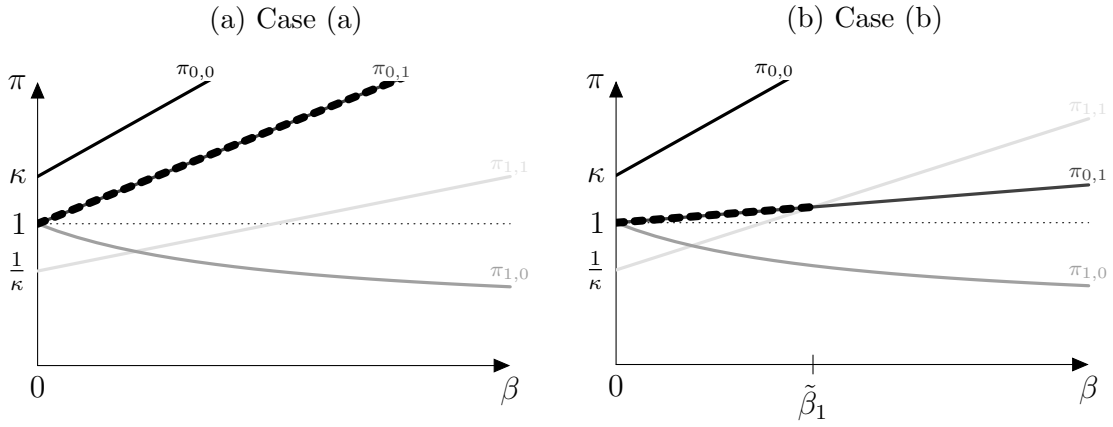
Notes: The figure shows support for *Right* in grey versus support for *Left* in white, for different values of salience of cultural issues χ. The dashed areas denote "culturally polarized" voters, who vote against their economic policy preferences.

Figure C.3: Monotonicity of $\frac{\partial Q_{L(\Delta\nu)}(\nu,\tau_L)}{\partial \nu}$ (Lemma 2)



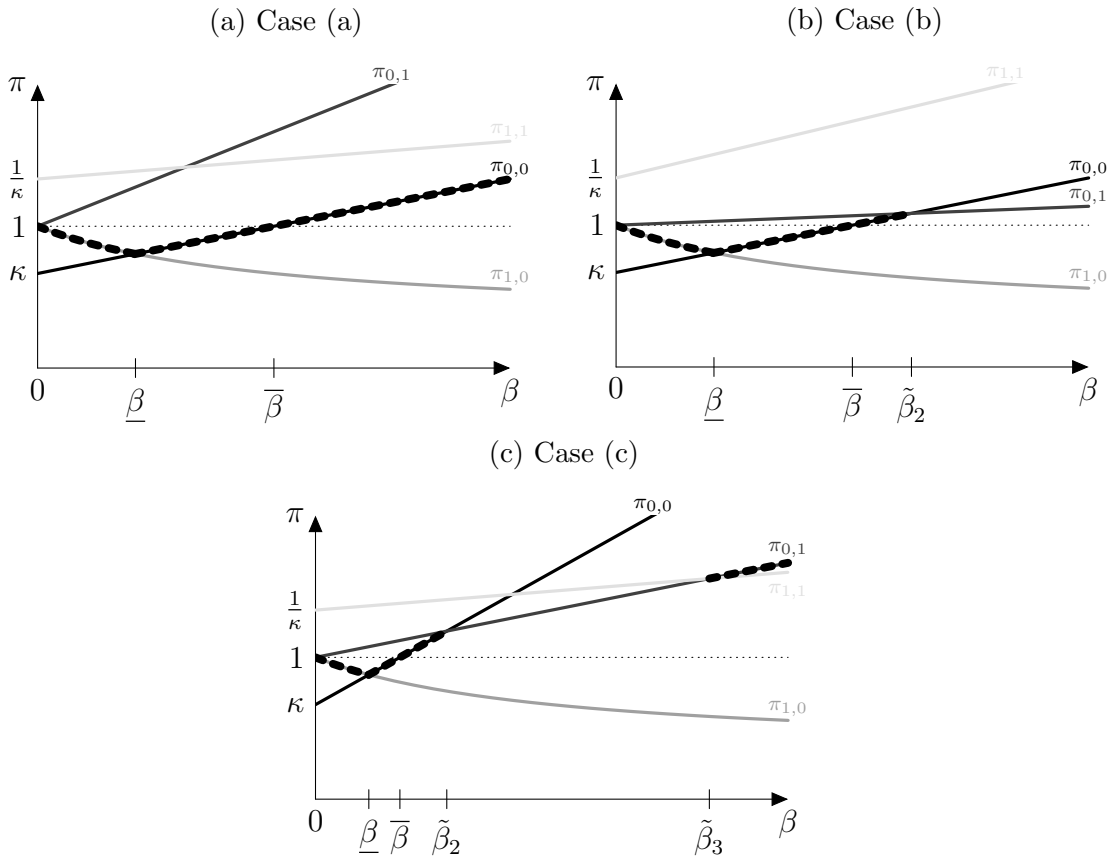
Notes: The figure shows voters with high status (who prefer narratives of agency) in grey, and those with low status (who prefer narratives of blame) in white. The two dashed lines electoral splits across different cross-party narrative differences, $\Delta\nu$ and $\Delta\nu - \zeta$. Voters above the dashed line vote for *Right* and those below for *Left*. A smaller proportion of *Left's* voters prefer narratives of agency under a lower $\Delta\nu$, and hence the turnout responds less positively to an increase in ν in that case.

Figure C.4: *Left's* payoff when $\kappa > 1$



Notes: The figure shows *Left's* votes relative to *Right's*, π_{ν_L,ν_R}, for different choices of narratives for the case when $\kappa > 1$. The dotted line at $\pi = 1$ indicates an equal split of votes. The thick dashes trace equilibria payoffs, whenever a pure strategy equilibrium exists. The panels refer to the different cases discussed in the Proof to Proposition 3.

Figure C.5: *Left's* payoff when $\kappa < 1$



Notes: The figure shows *Left's* votes relative to *Right's*, π_{ν_L,ν_R}, for different choices of narratives for the case when $\kappa < 1$. The dotted line at $\pi = 1$ indicates an equal split of votes. The thick dashes trace equilibria payoffs, whenever a pure strategy equilibrium exists. The panels refer to the different cases discussed in the Proof to Proposition 3.

Science Skepticism Reduced Compliance with COVID-19 Shelter-in-Place Policies in the United States*

Published in: *Nature Human Behaviour*†

Abstract

Physical distancing reduces transmission risks and slows the spread of COVID-19. Yet compliance with shelter-in-place policies issued by local and regional governments in the United States was uneven and may have been influenced by science skepticism and attitudes towards topics of scientific consensus. Using county-day measures of physical distancing derived from cellphone location data, we demonstrate that the proportion of people who stayed at home after shelter-in-place policies went into effect in March and April 2020 in the United States was significantly lower in counties with a high concentration of science skeptics. These results are robust to controlling for other potential drivers of differential physical distancing, such as political partisanship, income, education and COVID severity. Our findings suggest public health interventions that take local attitudes toward science into account in their messaging may be more effective.

*This chapter is co-authored with Valentin Kecht (Department of Economics, Bocconi University), David Van Dijcke (Department of Economics, University of Michigan; Institute for New Economic Thinking at the Oxford Martin School, University of Oxford), and Austin L. Wright (Harris School of Public Policy, University of Chicago).

†Brzezinski, Adam, Valentin Kecht, David Van Dijcke, and Austin L. Wright. 2021. Science skepticism reduced compliance with COVID-19 shelter-in-place policies in the United States. *Nature Human Behaviour*, 5(11), 1519-1527. URL: https://doi.org/10.1038/s41562-021-01227-0.

1 Introduction

Physical distancing reduces interpersonal transmission risks related to the COVID-19 virus [1, 2, 3]. Government policies that mandate physical distancing slow the spread of COVID-19 [4]. Local non-compliance with these shelter-in-place orders creates public health risks and may cause regional spread [5, 6]. Understanding which local factors impact compliance is a first-order public policy concern and informs evidence-based policy interventions to heighten their efficacy and mitigate the effects of the pandemic.

Recent research highlights several factors that influence compliance: partisanship [7, 8, 9, 10, 3, 11, 12], political polarization [13], poverty and economic dislocation [14], and differences in risk perception [15, 7, 16, 17]. These factors also influence physical distancing in the absence of government mandates [18]. Our central contribution is to highlight the role of science skepticism and attitudes regarding topics of scientific consensus in shaping patterns of physical distancing.

We leverage granular, representative data on science skepticism in the United States—beliefs about the anthropogenic (human-made) causes of global warming [19]—to study how physical distancing patterns vary with skepticism towards science. We combine this county-level science skepticism measure with location trace data on the movement of around 40 million mobile devices as well as data on state-level shelter-in-place policies (sample period: March 1 to April 19 2020). Our findings suggest that science skepticism is an important determinant of local compliance with government shelter-in-place policies, even after accounting for the role of partisanship, population density, education, and income, among other factors. We also benchmark our measure of science skepticism against other measures of belief in science available at the state-level, illustrating that our measure captures a more general notion of skepticism towards topics of scientific consensus.

Adapting science-based policy communication to account for the target audience's bias [20] can mitigate the risk that the message is rejected and the associated policy undermined. This can be achieved by correcting [21] or pre-emptively debunk-

ing [22] falsely held beliefs about science, thereby countering the dissemination of disinformation through modern media [23]. Taken together, our results underline the importance of tailoring public health interventions and associated messaging campaigns to account for local attitudes towards science.

2 Results

Main Results We analyze how physical distancing changes with the imposition of a state-wide shelter-in-place policy. First, we provide descriptive evidence for the role of science skepticism in influencing physical distancing behavior. Second, we use an event study design with a sample split implemented for counties with high and low degree of science skepticism. We then complement this approach with a staggered difference-in-differences (DiD) design, where we additionally control for a series of potentially confounding factors.

Figure 1 provides a descriptive view of the data. The left panel depicts the percentage of devices that stayed home each day of the sample period (March 1 to April 19 2020), with a polynomial trend fitted for the subsample of counties above (dark) and below (light) the median belief in anthropogenic global warming. We document a substantial uptick in physical distancing in most counties over the sample period (panel a)) that appears more pronounced in areas with lower levels of science skepticism (dark line). These trends depict a clear difference in physical distancing outcomes between counties with contrasting attitudes towards climate change—a topic of scientific consensus. Panel b) documents the substantial variation in the geographic distribution of such attitudes.

Panel a) in Figure 2 shows the results from the event study approach alongside 95% confidence intervals (sample period: March 1 to April 19 2020). The coefficients, t-statistics and p-values are listed in Table 1. Before the introduction of a shelter-in-place mandate, day-to-day changes in physical distancing remain consistently close to zero. Crucially, any slight upward or downward shifts in movement prior to the onset of a policy are mirrored across high-belief and low-belief counties. 2

out of the 18 estimated pre-trend dummies are close to but statistically significantly different from 0 (Table 1). However, there is no systematic pattern in the pre-trends and it is not unexpected to have some false positives when estimating a large number of pre-implementation effects. The parallel evolution of the pre-trends indicates that the difference in effects between the two event studies (high- and low-skepticism) is consistently estimated. After the first full day of a shelter-in-place order, non-skeptic counties see an increase in physical distancing of 3.9 percentage points, which is statistically significantly different from zero in a two-tailed test (t-statistic$(42324) = 4.88$, $p < 0.001$, $b_{\text{non_skeptic}} = 0.039$, 95 % Confidence Interval (CI) $[0.229, 0.551]$; see Table 1). Skeptic counties see an increase of 1.8 percentage points (t-statistic$(42284) = 5.09$, $p < 0.001$, $b_{\text{skeptic}} = 0.018$, 95 % CI $[0.011, 0.025]$; see Table 1).

We now turn to the staggered DiD approach, which allows us to control for a wide range of possible factors that could confound the relationship between belief in science and physical distancing during the sample period. The results are shown in Panel b) of Figure 2. The figure shows, for several different specifications, the point estimates alongside 95% confidence intervals for the marginal effect of science skepticism, i.e. the additional physical distancing response in high-belief relative to low-belief counties. As a benchmark, we start by estimating the same specification as in the event study approach, pooling the post-treatment lags and leaving out further controls except for the day and county fixed effects. This corresponds to the first point estimate on panel b). The marginal effect of being a non-skeptic county is 2.1 p.p., which is statistically different from zero according to a two-tailed test (t-statistic$(84692) = 5.469$, $p < 0.001$, $b_{\text{did_benchmark}} = 0.021$, 95 % CI $[0.013, 0.028]$; see Table 2), and similar in magnitude to the difference in responses between skeptic and non-skeptic counties in Panel a). To put this in context, the average share of devices staying home on a given day in February 2020 was around 25% (mean (M) $= 0.246$, standard deviation (SD) $= 0.057$, minimum (min) $= 0.02$, maximum (max) $= 0.667$), so this constitutes an additional increase for non-skeptic counties of about 9% ($= 0.021/0.246$). Considering that a device is only counted as staying home if it

is not seen anywhere outside its $153m \times 153m$ designated nighttime location on a given day, this is a sizeable difference.

Panel b) of Figure 2 also shows the estimates under different sets of control variables, interacted with the post-treatment period, to confirm that the estimated relationship is not driven by confounding factors. One potential concern is that differences in belief in science just serve as a proxy for partisan divides. Supplementary Figure 1 shows that this is a reasonable concern, as counties that are majority democrat have a higher fraction of non-skeptics. This partisan gap across parties has also become more polarized over the last decade, for a variety of proxies of science skepticism (Supplementary Figure 2). To take into account this potential source of bias, we allow the physical distancing response to a shelter-in-place policy to depend on whether the 2016 Republican vote share exceeded 50% of the total two-party ballot count (model 2 in Figure 2 panel b). The marginal effect estimate remains similar at 1.5 p.p. (t-statistic$(84691) = 4.663$, $p < 0.001$, $b_{\text{did_voting}} = 0.015$, 95 % CI $[0.009, 0.022]$) and statistically indistinguishable from the benchmark (Wald test that coefficient is equal to benchmark estimate: $F(1, 45) = 2.78$, $p = 0.102$). Thus, even after partialling out the differential effects due to partisanship, there is a significantly higher response to the shelter-in-place policy in non-skeptic counties.

We consider several other potential sources of bias by sequentially adding further (interacted) control variables, and show the coefficient sizes, t-statistics as well as p-values in Table 2. Across these specifications we account for how rural (model 3), educated (model 4), wealthy (model 5), religious (model 6), and institutionally robust (model 7) a county is by interacting these variables with the shelter-in-place policy dummy. This approach allows the estimated effect of shelter-in-place policies to vary with all these additional factors. We also take into consideration whether related business or school closure policies have been introduced (model 8), the local severity of COVID-19 cases and deaths (model 9), and state-wide cases and deaths (model 10). In model 11, we additionally allow the effect of local and state-wide COVID-19 cases and deaths on physical distancing to vary with the introduction of the shelter-in-place policy. Finally, in model 12, we also control for the proportion of

essential workers in a county, again interacted with the shelter-in-place policy. This helps to mitigate the potential confounding factor that some people may leave their homes because their work demands it, rather than because they do not comply with the shelter-in-place policy.

Even though several of these measures, especially religiosity, income, and education, may over-control for variation in the outcome of interest (i.e., these covariates are colliders), the estimate remains robust and statistically significant, even in the fully saturated model that includes the complete set of interactive control variables (last model in Table 2). The robustness of the point estimates supports the idea that the patterns of non-compliance will likely also hold for alternative measures of science skepticism, as discussed in the theory section.

Robustness In Supplementary Information, we investigate several potential concerns with our approach. We begin by demonstrating the robustness of the main results to a number of alternative sampling and model specifications. We introduce these results in Extended Data Figure 1. In panels a) and b), we demonstrate that the main results hold if we leverage a longer pre-sample window of 10 days. Use of the 10-day window means that the number of days used as a baseline period balances the number of leads and lags. In panel c), we estimate the main event study without an omitted treatment lead. In panel d), we estimate the main specification with a saturated set of covariates-by-date fixed effects. In panels e) and f), we use a one-day pre-sample window. This means the number of omitted treatment leads (one) matches the number of omitted pre-window dates. Panels g) and h) replicate the main event study and difference-in-differences models without weighting units by population. In each of these additional variants, we observe a similar pattern of results: parallel pretrends in physical distancing across high-belief and low-belief counties; large increases in physical distancing among high-belief counties after the introduction of a shelter-in-place; robust and statistically significant marginal effects of belief in science after accounting for the battery of potential confounders discussed earlier (partisanship, rurality, education, income, religiosity, institutional health, related government policies, local and state-wide COVID-19 severity).

Construct Validity We next consider the construct validity of our measure of belief in science. It is possible that attitudes towards the anthropogenic causes of climate change are bundled together with a range of other ideological considerations, especially partisanship [24]. We assess how our measure of science skepticism relates to other measures of attitudes towards science, scientific inquiry, and scientific breakthroughs.

To confirm that our measure of skepticism about the human causes of global warming proxies more general attitudes towards topics of scientific consensus, we test to what extend it correlates with other measures of belief in science. As an initial benchmark we use the Pew AVS survey, pooled over the years 2002-2009. The sample period has the advantage that it precedes the pronounced increase from 2008 onwards in both climate change skepticism [31, 32], driven by partisan polarization, as well as polarization in attitudes towards science, which we document in Supplementary Figure 2. We supplement this datasource with information from the World Values Survey (WVS). The data are pooled over two waves (2011 and 2017) and report average levels of trust in science at the state level. The latter is determined as the first polychoric principal component of 6 questions concerning beliefs on science and technology. We also leverage data on the average state-level number of MMR vaccinations received by children at the age of 36 months. The use of MMR vaccinations in this age group has been found to reflect broader vaccine hesitancy, [33, 34].

We visualize the relationship between these alternative measures of attitudes towards science in Figure 3. Panels a) and b) present the principal component from the WVS survey (trust in science) and healthier life factor loading respectively. Notice the robust, positive relation. In panel c), we illustrate the association between the Pew AVS measure and the [19] measure. In panel d), we introduce the scatterplot of MMR vaccination rates and beliefs on global warming. We find similarly consistent positive associations. In Supplementary Table 1, we introduce regression-based estimates. The estimated associations are large, ranging from 0.331 ($R^2 = 0.11$, WVS factor loading for 'science is important for daily life') to 0.635 ($R^2 = 0.4$,

WVS PCA). These results suggest our primary measure of belief in science is a representative measure of the more general public attitude towards science.

Science Skepticism, mask use and vaccine hesitancy Although our analysis is primarily focused on physical distancing, the role of science skepticism in public health policy is much broader, potentially influencing other mitigating behaviors such as mask use and vaccine hesitancy. Science skepticism may undermine public mask use and reinforce hesitancy, thwarting attempts to establish herd immunity. To explore these additional policy-relevant issues, we gather data on county-level mask use and vaccine hesitancy across the United States. In Figure 4, we show that the correlation between science skepticism and mask use at the height of the pandemic in the United States (July 2020) is negative and robust to accounting for local partisanship, the primary confounding factor of concern in our main analysis. In Figure 5, we replicate this analysis for vaccine hesitancy in April 2021. We find a similarly striking pattern with science skepticism and local hesitancy being positively correlated (strong skepticism increases hesitancy), which is robust to conditioning out the role of partisanship, an especially poignant concern in light of recent evidence that hesitancy is disproportionately high among polled Republicans. Although data limitations restrict an investigation of the role of mask use and hesitancy that is comparable to our main estimates, these findings suggest a persistent effect of belief in science on public health measures beyond physical distancing.

3 Discussion

Our main result suggests that counties with lower levels of skepticism about topics of scientific consensus (non-skeptics) have complied with shelter-in-place mandates at a higher rate than counties with higher levels of science skepticism (skeptics). We argue that this is because skeptics are less likely to accept the motivations underlying science-based public policies.

When a policy has imperfect enforcement, compliance will be incomplete among people who believe that the policy has no beneficial or harm-reducing effects. In

the case of the COVID-19 pandemic, estimates of such effects hinged crucially on scientific assessments and medical research that indicated the virus spread through interpersonal contact and could be mitigated by physical distancing (see [25] for an early review of this literature). Moreover, in the United States, policy enforcement was indeed imperfect [26, 27]. As a result, the success of shelter-in-place policies in combating the spread of the virus necessarily relied for a large part on voluntary compliance [28, 29, 30]. In areas where scientific skepticism is widespread, we hypothesize that fewer people would believe in the benefits of social distancing, reducing policy effectiveness.

Science skepticism can influence potentially costly social behaviors through several plausible mechanisms. Skepticism may reduce the willingness of individuals to engage with scientific information that contradicts their prior beliefs [32]. Skeptics may also engage in less information-seeking behavior, preferring to rely on anecdotal or "common sense" assessments of the risks they or their community face. Further, skeptics may specifically seek out slanted coverage that reinforces their existing beliefs or biases. In any of these cases, subsequent risk perceptions will remain unchanged by new information or conform to existing assessments of potential threats [35]. Research on science communication during the COVID-19 pandemic provides ample evidence of these mechanisms. Other research finds that experimental subjects with lower cognitive reflection and scientific knowledge were more willing to share fake or false stories about the pandemic [36]. In another paper, survey respondents were more likely to believe COVID-19 risks were exaggerated and that the virus was purposefully manufactured and spread if they had a psychological predisposition to reject expert assessments and engage in conspiratorial thinking [37]. [38] provides evidence that anti-intellectualism—which is closely related to what we describe as science skepticism—is linked to diminished concern regarding COVID-19, perceptions of risk factors, and willingness to engage with and seek out information from experts. Finally, [39] introduce a theoretical model whereby information seeking is slanted, leading skeptics to seek out media coverage that minimizes the reported threat

posed by non-compliance with shelter-in-place mandates. These studies provide an empirical and theoretical foundation for our core hypothesis.

There remain important unresolved questions about the psychological, political, and social origins of science skepticism. It is possible that skepticism emerges due to a real or perceived challenge to an existing set of beliefs or principles ("identity-protective cognition") [40]. These beliefs may be informed by misconceptions about causal relationships [41]; or motivated by political affiliations [42], exposure to slanted sources of information [43, 44], conspiracy theories [32], or related cultural mythologies [45]. An individual's principles may be ideological heuristics that are challenged by scientific inquiry, triggering psychological reactance and, as a consequence, a rejection of scientific expertise [46, 47]. Political and social institutions may also emerge to reinforce these dynamics. [48] provides a review of these dynamics, noting that skepticism may lead to either the rejection of a narrow set of scientific facts or of scientific expertise as an entire epistemic system. Indeed, the drivers of science skepticism have been shown to be more heterogeneous than previously thought, with political partisanship being an important predictor of climate skepticism [40, 49, 50], while religiosity better predicts vaccine and evolution skepticism [51, 52].

Supplementary Figure 1 illustrates that political partisanship is indeed related to science skepticism using our county-level measure of climate change skepticism from [19], while Supplementary Figure 2 documents an across-the-board increase in political polarization around science attitudes over the last decade in the US. [24] provides a psychological explanation for self-reported climate skepticism as being rooted in individuals' cultural or political ties, rather than their cognitive capacity. Taken together, it is clear that a robust exploration of science skepticism through the lens of climate dynamics (our primary measure) requires a careful consideration of the cultural and political systems within which this skepticism is embedded. These systems include partisanship, religiosity, rurality and urbanicity, and educational attainment. We take these factors into account empirically so as to address important concerns about the construct validity of our measure—whether the patterns of non-compliance we observe are likely to hold for alternative measures of science skepticism.

Our findings complement recent research that finds COVID-19 skepticism to have similar antecedents as vaccine and climate skepticism [53].

By documenting how science skepticism can undermine public health interventions, our research complements earlier experimental and survey-based results on the drivers of such skepticism, and affirms their importance [36, 37, 38]. Our findings suggest that it would benefit the effective implementation of public policies to tailor public messaging so as to correct or preempt falsely held beliefs about the scientific evidence these policies are based on [23, 20]. The science communication literature offers many potential strategies to achieve these goals [54]. Similarly, science institutions can attempt to pre-empt skeptical reception of and non-compliance with science-based policy by "pre-bunking" [22] false beliefs about science and promoting education and scientific literacy training [55]. Such "psychological inoculation" [56] has been applied with some success in the context of vaccination [57] and climate change [58]. We contribute to this body of research by documenting that substantial behavioral harm can arise when science skepticism undermines compliance with public policy.

Our study has several limitations. Although we present credible evidence that movement patterns followed parallel trends prior to shelter-in-place policies, these interventions are not strictly exogenous. These orders also entered into effect early during the pandemic and our analysis focuses on compliance during this initial phase of the crisis in the United States. Social distancing over the course of the pandemic and the potential role of so-called "pandemic fatigue" may be of interest to future work [59] but are not examined in this study. Furthermore, while we introduce measures of mask use and vaccine hesitancy observed later in the pandemic, current mask use and vaccine utilization may be driven by other factors that we do not fully explore here. We also anticipate future research could meaningfully contribute to our understanding of scientific skepticism by gathering less polarized measures of understanding of and beliefs in scientific concepts and facts than climate skepticism.

4 Methods

Data We analyze differential changes in county-level movement patterns and physical distancing after the implementation of shelter-in-place policies in the United States (see Supplementary Information for more detail). Daily panel data are aggregated to the county-level from GPS pings of more than 40 million mobile devices, obtained from SafeGraph, a data company that aggregates anonymized location data from numerous applications in order to provide insights about physical places, via the PlaceKey Community. These data allow us to track the percentage of devices that stayed home all day. The latter is defined as the ratio of the number of devices that were not seen outside of their home all day in a given county over the total number of devices observed in the county that day. A device's home is determined as the common night-time location (delineated by a 153m×153m geo-fence) over a 6-week period. The underlying anonymized data were subjected to exhaustive data processing with the aim to guarantee reliability, granularity, anonymity and accuracy. The panel of devices in the sample is designed to be geographically and demographically representative, with a 97% correlation between the panel's population density and the American Census's population density at the county level.

To mitigate the spread of COVID-19, county and state governments implemented shelter-in-place policies during the second half of March 2020. We collect implementation dates of state- and county-level school and business closures, state-of-emergency declarations and shelter-in-place policies from various sources. When a policy goes into effect after 12pm on a given day, we assign it an implementation date one day later. We measure science skepticism using data assembled by [19] at the county level. Our main measure is the estimated percentage of people who agree with the statement that global warming is caused by humans. Despite scientific consensus that humans are the primary cause of present and projected climate change dynamics, there is substantial variability in local beliefs about this topic in the United States.

We leverage this variation to study how patterns of physical distancing differ across counties with lower levels of science skepticism (higher belief in science).

Research Design We adopt two related empirical approaches to analyze how physical distancing behavior depends on whether a county exhibits above- or below-median levels of science skepticism (see Supplementary Information for more detail). First, we implement a county-day level event study design where we focus on the impact of shelter-in-place policies on physical distancing, in line with related studies [8, 18, 14]. We estimate the response separately for counties with above-median and below-median levels of science skepticism. The event study design serves to check for the presence of pre-treatment (anticipation) effects. The main effects will be consistently estimated even if there is some evidence of anticipation effects as long as those anticipation effects are parallel across the two event studies. An additional benefit of the event study design is that it allows to estimate and visualize how the policy response changes dynamically after the introduction of shelter-in-place mandates.

Second, we use a staggered Difference-in-Differences (DiD) approach to estimate the differential response to shelter-in-place policies for counties with above-median versus below-median science skepticism. The split-sample event study is a powerful approach when stratifying over one dimension with a well-balanced binary classification. This is the case with the event study design we use for our measure of science skepticism. However, the introduction of additional dimensions requires stratification over all of them simultaneously. With a large number of potential confounding factors, the split-sample event study design is statistically underpowered. Instead, to address the fact that there likely is a range of potential sources of bias, as noted in the Discussion section, we adopt a staggered DiD design. This approach is closely aligned with the event study approach but includes a single dummy that is equal to one at or after the implementation date, which is interacted with a dummy for whether the county is marked by above- or below-median belief in science. Moreover, we interact a vector of potential confounders with the treatment indicator, enabling us to control for the effect of additional covariates on the physical distancing response

to a shelter-in-place policy. One primary control variable is local partisanship, which is correlated with science skepticism (see Supplementary Figure 1). Political polarization with respect to climate change and attitudes towards science generally in the United States have been amplified over the past decade (see Supplementary Figure 2). We attempt to address this concern by partialling out any marginal effects associated with partisanship, using voting records from the 2016 Presidential election. The effect of science skepticism may also be confounded by how rural, educated, religious, and wealthy a given community is. Institutional health, the share of essential workers and local and regional exposure to COVID-19 may similarly influence our primary estimates. We incorporate these additional controls as well.

For each of the two approaches, we weigh each county by its population within the regression framework in order to provide representative results. In Extended Data Figure 1 we show that the main results are robust to estimating each design without population weights and varying baseline periods.

Data availability The mobility data provided by SafeGraph are proprietary but available to academic researchers via https://www.safegraph.com. All other data that support the findings of this study are available at https://www.openicpsr.org/openicpsr/project/144861.

Code availability The code for all analyses of this study can be found at https://github.com/Davidvandijcke/science_skepticism_nature_hb.

Acknowledgements We thank SafeGraph for providing their data and support. We also thank Rick Van der Ploeg, Cyrus Samii and Kaan Cem Ketenci for support and discussions, and Bakdaulet Baitan for excellent research assistance. The authors received no specific funding for this work.

Author contributions All authors conceived the idea. V.K., D.V.D. and A.L.W. provided and prepared the data. All authors developed the baseline model. A.B., D.V.D. and A.L.W. completed the robustness checks. V.K. and A.L.W. constructed the descriptive figures and performed the analyses on construct validity. A.B. and

A.L.W. drafted the manuscript. D.V.D. and V.K. provided critical revisions. All authors approved the final version of the manuscript for submission.

Competing interests The authors declare no competing interests.

5 Tables

Table 1: Baseline Event-Study Approach, Full Results

	Low Science Skepticism				High Science Skepticism			
	(1) Coeff.	(2) t-Stat.	(3) p-val.	(4) 95 % CI.	(5) Coeff.	(6) t-Stat.	(7) p-val.	(8) 95 % CI.
Time -10	-0.003	-0.403	0.689	[-0.017,0.011]	-0.005	-1.503	0.140	[-0.012,0.002]
Time -9	-0.003	-0.476	0.636	[-0.016,0.010]	-0.006	-1.248	0.218	[-0.016,0.004]
Time -8	-0.002	-0.344	0.732	[-0.012,0.009]	-0.004	-1.143	0.259	[-0.011,0.003]
Time -7	-0.000	-0.078	0.938	[-0.011,0.010]	-0.003	-0.734	0.467	[-0.010,0.005]
Time -6	0.001	0.354	0.725	[-0.006,0.008]	-0.006	-1.808	0.077	[-0.012,0.001]
Time -5	0.008	2.657	0.011	[0.002,0.015]	-0.000	-0.061	0.952	[-0.010,0.010]
Time -4	0.003	0.705	0.485	[-0.005,0.010]	-0.003	-0.948	0.348	[-0.009,0.003]
Time -2	-0.002	-0.546	0.588	[-0.010,0.006]	-0.010	-2.711	0.009	[-0.017,-0.003]
Time -1	-0.001	-0.270	0.789	[-0.010,0.008]	-0.006	-1.589	0.119	[-0.014,0.002]
Time 0	0.022	3.590	0.001	[0.010,0.035]	0.016	2.978	0.005	[0.005,0.026]
Time 1	0.039	4.880	<0.001	[0.023,0.055]	0.018	5.090	<0.001	[0.011,0.025]
Time 2	0.031	5.741	<0.001	[0.020,0.042]	0.019	3.589	0.001	[0.008,0.029]
Time 3	0.026	4.389	<0.001	[0.014,0.039]	0.014	2.697	0.010	[0.004,0.025]
Time 4	0.028	5.182	<0.001	[0.017,0.039]	0.012	2.475	0.017	[0.002,0.022]
Time 5	0.028	5.065	<0.001	[0.017,0.038]	0.010	1.820	0.075	[-0.001,0.020]
Time 6	0.030	3.887	<0.001	[0.014,0.046]	0.008	1.664	0.103	[-0.002,0.018]
Time 7	0.034	5.093	<0.001	[0.020,0.047]	0.013	2.763	0.008	[0.003,0.022]
Time 8	0.032	4.793	<0.001	[0.019,0.046]	0.007	1.304	0.199	[-0.004,0.018]
Time 9	0.028	3.952	<0.001	[0.014,0.042]	0.006	1.029	0.309	[-0.006,0.017]
Time 10	0.028	3.851	<0.001	[0.013,0.043]	0.002	0.215	0.831	[-0.013,0.016]
R^2	0.928				0.897			
DoF	42324				42284			
Obs.	43912				43828			

This Table corresponds to the results shown in panel a) of Figure 2. Dependent variable: percent of devices fully at home. Difference-in-difference estimates with pre-policy event study set-up. Columns (1) - (4) display the results for the sample restricted to counties with above-median belief in anthropogenic global warming (skeptics); Columns (5) - (8): below-median belief in the same (non-skeptics). See Supplementary Information for a more detailed discussion of the event study approach. Sample goes from 15 days before to 10 days after the policy for each county; effect sizes shown are changes relative to the period 15-11 days, as well as 3 days, before the policy. County and date fixed effects are included. 95% confidence intervals are shown, based on standard errors triple-clustered by county, date and state-week. The regressions employ population weights. Sample period: March 1 to April 19 2020.

Table 2: DiD Approach, Full Results

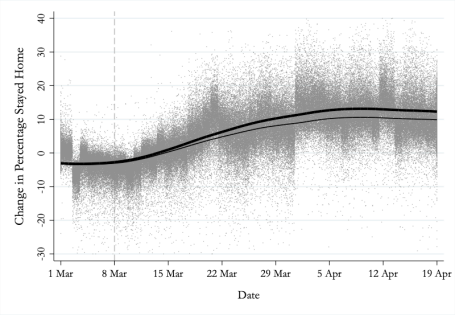
	Coeff.	t-Stat.	p-val.	95 % CI		R^2	DoF	Obs.
				Lower Bound	Upper Bound			
Benchmark	0.021	5.469	0.000	0.013	0.028	0.922	84692	87740
+Voting	0.015	4.663	0.000	0.009	0.022	0.923	84691	87740
+Rural	0.013	4.264	0.000	0.007	0.019	0.923	84690	87740
+Education	0.009	3.123	0.003	0.003	0.014	0.924	84689	87740
+Income	0.008	2.898	0.006	0.002	0.013	0.924	84688	87740
+Religiosity	0.008	3.013	0.004	0.003	0.013	0.924	84687	87740
+InstHealth	0.008	2.988	0.005	0.003	0.013	0.924	83971	87002
+GovtPolicies	0.008	3.031	0.004	0.003	0.013	0.925	83968	87002
+LocalCOVID	0.007	2.848	0.007	0.002	0.012	0.929	83966	87002
+StateCOVID	0.006	2.696	0.010	0.002	0.011	0.930	83964	87002
+SiPCOVID	0.007	3.078	0.004	0.002	0.011	0.931	83960	87002
+EssentialWorkers	0.007	2.999	0.004	0.002	0.011	0.931	81404	84372

This Table corresponds to the results shown in panel b) of Figure 2. Dependent variable: percent of devices fully at home. Difference-in-differences estimates for the differential effect of a shelter-in-place policy for counties with above-median (non-skeptics) relative to below-median belief in science (skeptics) across different specifications. The coefficients refer to the interaction $post_{s,t} \times BiS_c$, the post-treatment period in state s and a dummy that takes value 1 if a county c exhibits above-median belief in science. "Benchmark model" controls for date and county fixed effects. Further models are specified as follows, progressively expanding the benchmark model: "+Voting" additionally controls for $post_{s,t} \times vote_c$, where $vote_c = 1$ if c was majority Republican in the 2016 presidential election; "+Rural" additionally controls for $post_{s,t} \times rural_c$, where $rural_c = 1$ if c has below-median population density; "+Education" additionally controls for $post_{s,t} \times educ_c$, where $educ_c = 1$ if c has above-median college education levels; "+Income" additionally controls for $post_{s,t} \times income_c$, where $income_c = 1$ if c has above-median income; "+Religiosity" additionally controls for $post_{s,t} \times relig_c$, where $relig_c = 1$ if c has above-median levels of religiosity; "+Inst. Health" additionally controls for $post_{s,t} \times health_c$, where $health_c = 1$ if c has above-median levels of institution health;"+Govt. Policies" additionally controls for whether the c had a school closure, business closure, or state of emergency declaration; "+Local COVID" additionally controls for the number of confirmed COVID cases and deaths in c; "+State COVID" additionally controls for the number of confirmed COVID cases and deaths in the state of county c; and "SiP by COVID" additionally controls for the interaction of COVID cases and deaths with $post_{s,t}$; "Essential Workers" further controls for the interaction of the share of essential workers with $post_{s,t}$. Standard errors are triple-clustered by county, date and state-week. The regressions employ population weights. Sample period: March 1 to April 19 2020.

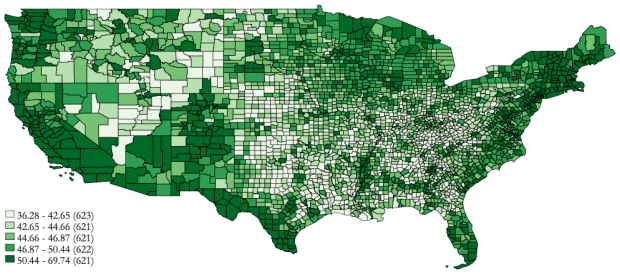
6 Figures

Figure 1: Science Skepticism and Physical Distancing: Descriptive Evidence

(a) Data: Full Sample



(b) Belief In Anthropogenic Global Warming

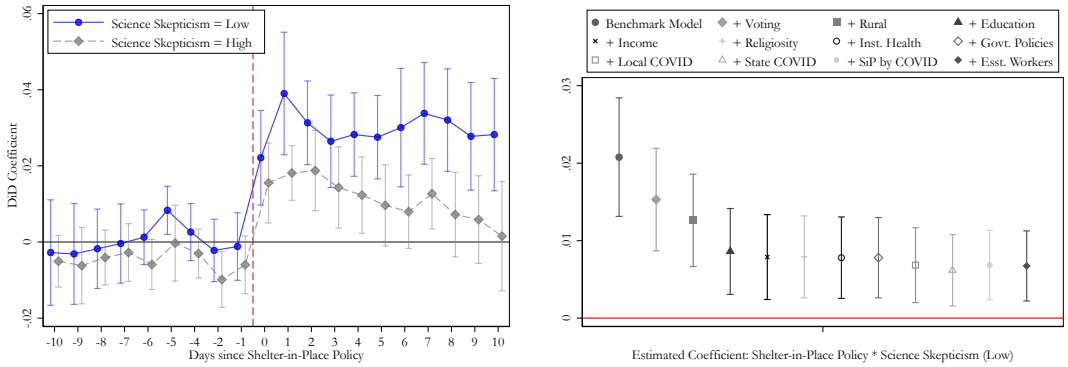


Panel a): *Light*: below-median belief in anthropogenic global warming (skeptic); *Dark*: above-median belief in the same (non-skeptic). Plots show the percentage difference in devices that stayed home during the sample period (March 1 to April 19) from the day-of-week-specific average during February. Solid lines are from local polynomial smoothing with bandwidth 5. N = 161,363. **Panel b)**: percent of people in county who believe that global warming is anthropogenic [19]. The source of the shapefile is provided in Supplementary Information. N = 3,006. Sample period: March 1 to April 19 2020.

Figure 2: Compliance With Shelter-in-Place Policy Depends on Science Skepticism

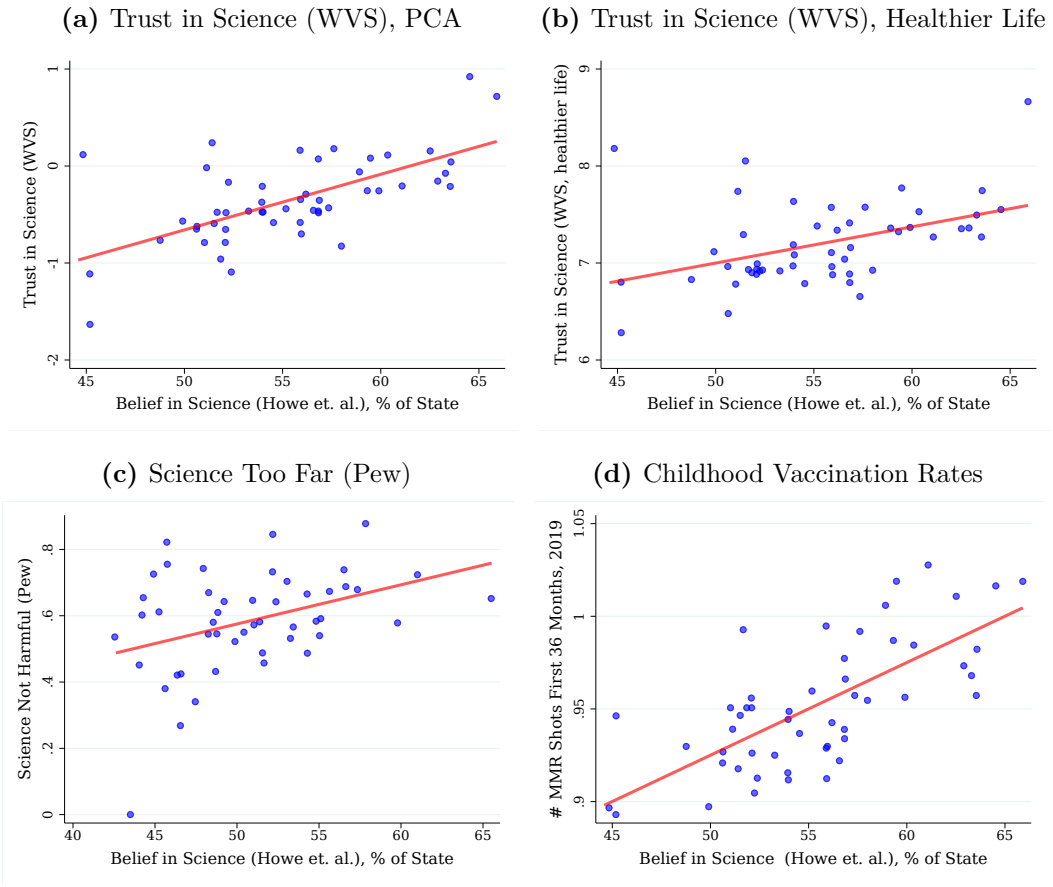
(a) Event Study

(b) Difference-In-Difference



Dependent variable: percent of devices fully at home (see Supplemental Information for details). 95% confidence intervals are shown, based on standard errors triple-clustered by county, date and state-week. The regressions employ population weights; for unweighted results, see panels g) and h) of Extended Data Figure 1. **Panel a)**: Difference-in-difference estimates with pre-policy event study approach *Light grey*: below median belief in anthropogenic global warming (skeptics); *Dark blue*: above-median belief in the same (non-skeptics). Sample goes from 15 days before to 10 days after the policy for each county; effect sizes shown are changes relative to the period 15-11 days and 3 days before the policy. County and date fixed effects are included. **Panel b)**: Difference-in-differences estimates for the differential effect of a shelter-in-place policy for counties with above-median (non-skeptics) relative to below-median belief in science (skeptics) across different specifications. Point estimates and confidence interval come from interactions $post_{s,t} \times BiS_c$, the post-treatment period in state s and a dummy equal to 1 for counties c with above-median belief in science. "Benchmark model" controls for date and county fixed effects. The following models progressively expand the benchmark model by introducing the interaction of $post_{s,t}$ with the following variables: "+Voting": $vote_c$ (= 1 if c was majority Republican in the 2016 presidential election); "+Rural": $rural_c$ (= 1 if c has below-median population density); "+Education": $educ_c$ (= 1 if c has above-median college education levels); "+Income": $income_c$ (= 1 if c has above-median income); "+Religiosity": $relig_c$ (= 1 if c has above-median levels of religiosity); "+Inst. Health": $health_c$ (= 1 if c has above-median levels of institution health). "+Govt. Policies" additionally controls for whether c had a school closure, business closure, or state of emergency declaration; "+Local COVID" additionally controls for the number of confirmed COVID cases and deaths in c; "+State COVID" additionally controls for the number of confirmed COVID cases and deaths in the state of county c; "SiP by COVID" adds the interaction of COVID cases and deaths with $post_{s,t}$; "Essential Workers" adds the interaction of the share of essential workers with $post_{s,t}$. N = 87,040. Sample period: March 1 to April 19 2020.

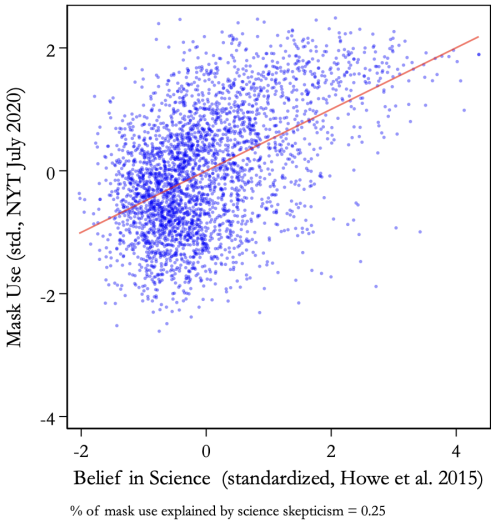
Figure 3: Benchmarking Science Skepticism (Main Specification) with Alternative Measures of Trust in Science



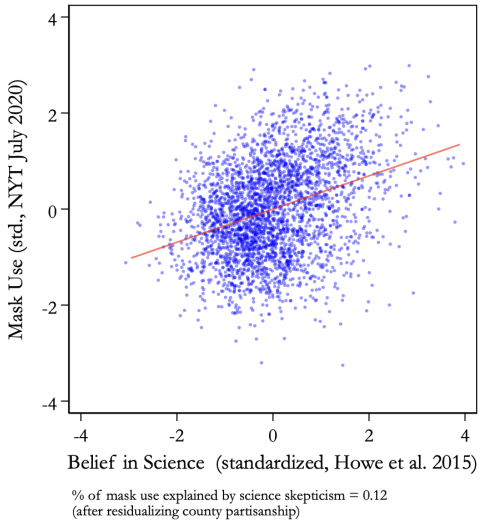
Panel a), b), c): scatterplots of baseline proxy for belief in science from [19], aggregated to state-level using population weights, against various alternative measures. Higher values of belief in science correspond to weaker skepticism. *Trust in Science (WVS)*: first polychoric principal component of six variables measuring attitudes toward science from the World Values Survey (2011 and 2017 waves pooled). All variables were standardized, survey-weighted and averaged at the state level. *Trust in Science (WVS), Healthier Life*: "Science and technology are making our lives healthier, easier, and more comfortable." Higher values refer to higher levels of trust in science. See Table 1 for more details. *Science Too Far (Pew)*: percent of state sample population in Pew Research Center's American Values Survey (2002-2009) reporting disagreement with the statement 'I am worried that science is going too far and is hurting society rather than helping it". **Panel d)**: using state-level vaccination rates to benchmark our baseline proxy for Science Skepticism. *Child Vaccination*: average number of MMR vaccination shots received by children during the first 36 months of their lives at the state level.

Figure 4: Correlation with Mask Use (County Level)

(a) Raw Data



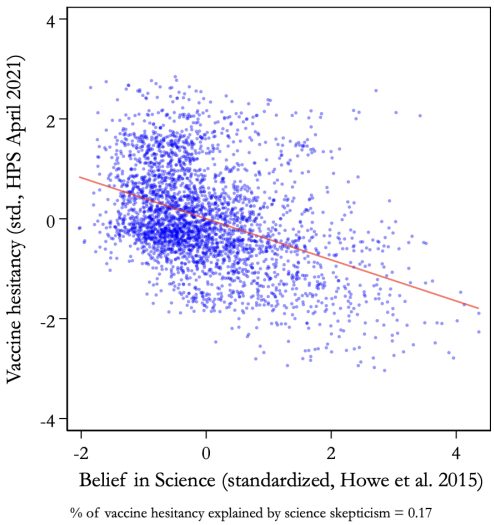
(b) After Controlling for Partisanship



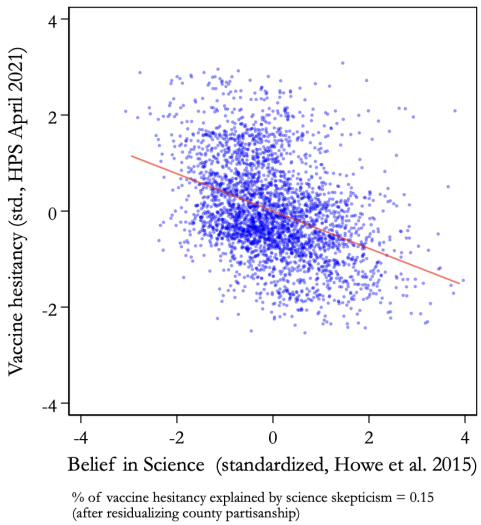
Panel a): Figure plots belief in science from [19] against county-level mask use in July 2020. Higher values of belief in science correspond to weaker skepticism. Mask use is defined as the share of people that stated to always wear a mask, which is then standardized to have mean of zero and standard deviation of one. These data have been obtained from a large-scale survey commissioned by the New York Times, which was conducted online between July 2 and July 14 and includes 250,000 survey responses. For details on the data, see [?]. **Panel b)**: Figure equals the one from Panel a), but shows residuals from regressions on an indicator of a county's partisanship. Each county is classified according to which party had a higher vote share than the other party in the 2016 presidential election.

Figure 5: Correlation with Vaccine Hesitancy (County Level)

(a) Raw Data



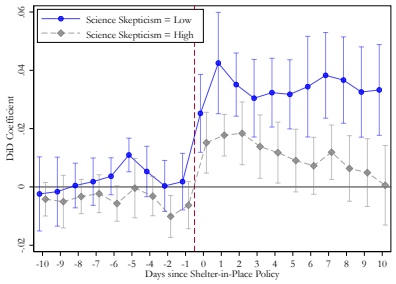
(b) After Controlling for Partisanship



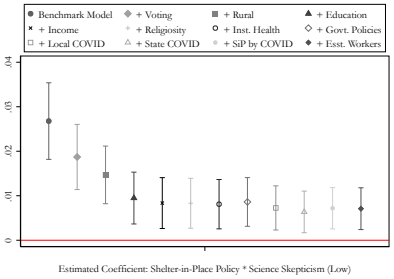
Panel a): Figure plots belief in science from [19] against county-level vaccine hesitancy in April 2021. Higher values of belief in science correspond to weaker skepticism. Hesitancy is measured as the share of county residents with little or no interest in receiving a COVID-19 vaccine, which is then standardized to have mean of zero and standard deviation of one. These data have been obtained from the county-level estimates provided by U.S. Census Bureau, leveraging their Household Pulse Survey. For additional information, see https://bit.ly/3osf1M5. **Panel b)**: Figure equals the one from Panel a), but shows residuals from regressions on an indicator of a county's partisanship. Each county is classified according to which party had a higher vote share than the other party in the 2016 presidential election.

Extended Data Figure 1: Compliance With Shelter-in-Place Policy (Additional Results)

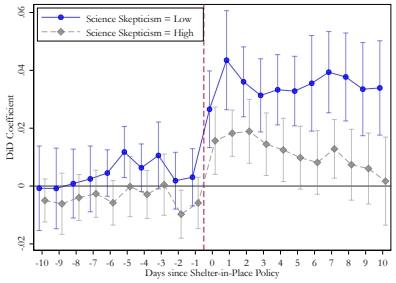
(a) Event Study (10 Day Baseline)



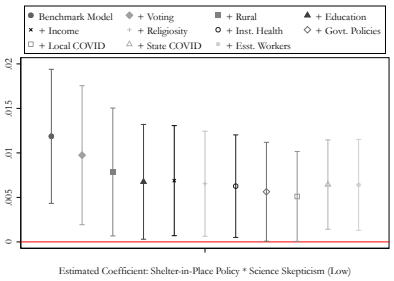
(b) Diff.-In-Diff. (10 Day Baseline)



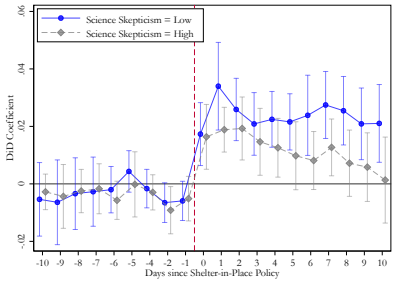
(c) Event Study (All Leads)



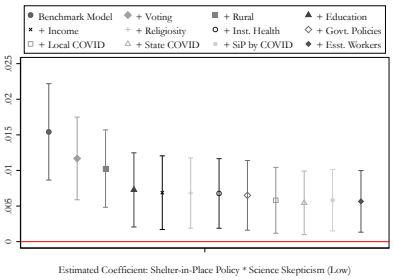
(d) Diff.-In-Diff. (Saturated Fixed Effects)



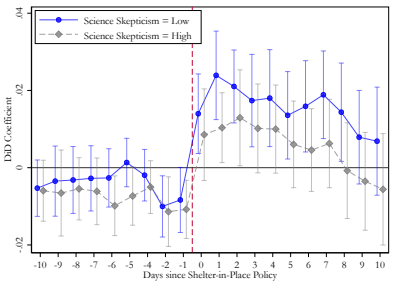
(e) Event Study (1 Day Baseline)



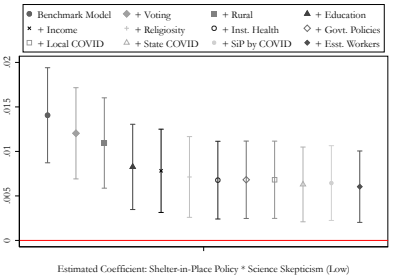
(f) Diff.-In-Diff. (1 Day Baseline)



(g) Event Study (Unweighted)



(h) Diff.-In-Diff. (Unweighted)



Panel a)-h): The specifications are the same as in the corresponding panels in Figure 2, except: panels a) and b) use a 10 day baseline period (rather than five); panel c) includes all leads in the event study specification; panel d) includes saturated covariates × date fixed effects; panels e) and f) use an alternative one day baseline period (rather than five); panel g) and h) are not population weighted. Sample period: March 1 to April 19 2020. See Figure 2 for additional details.

References

[1] Viner, R. M. *et al.* School closure and management practices during coronavirus outbreaks including covid-19: a rapid systematic review. *The Lancet Child & Adolescent Health* **5**, 397–404 (2020). URL http://www.sciencedirect.com/science/article/pii/S235246422030095X.

[2] Anderson, R. M., Heesterbeek, H., Klinkenberg, D. & Hollingsworth, T. D. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet* **395**, 931–934 (2020).

[3] Bai, Y. *et al.* Presumed Asymptomatic Carrier Transmission of COVID-19. *JAMA Research Letter, Published Online February 21, 2020* (2020).

[4] Hsiang, S. *et al.* The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature* **584**, 262–267 (2020).

[5] Lewnard, J. A. & Lo, N. C. Scientific and ethical basis for social-distancing interventions against covid-19. *The Lancet. Infectious Diseases* **20**, 631–633 (2020).

[6] Chen, S., Yang, J., Yang, W., Wang, C. & Bärnighausen, T. Covid-19 control in china during mass population movements at new year. *The Lancet* **395**, 764–766 (2020).

[7] Allcott, H. *et al.* Polarization and Public Health: Partisan Differences in Social Distancing during the Coronavirus Pandemic. *NBER Working Paper* (2020).

[8] Painter, M. & Qiu, T. Political beliefs affect compliance with government mandates. *Journal of Economic Behavior & Organization* **185**, 688–701 (2021).

[9] Grossman, G., Kim, S., Rexer, J. M. & Thirumurthy, H. Political partisanship influences behavioral responses to governors' recommendations for covid-19 prevention in the united states. *Proceedings of the National Academy of Sciences* **117**, 24144–24153 (2020).

[10] Gadarian, S. K., Goodman, S. W. & Pepinsky, T. B. Partisanship, health behavior, and policy attitudes in the early stages of the covid-19 pandemic. *PLOS ONE* **16**, e0249596 (2021).

[11] Gollwitzer, A. *et al.* Partisan differences in physical distancing are linked to health outcomes during the covid-19 pandemic. *Nature Human Behaviour* **4**, 1186–1197 (2020).

[12] Clinton, J., Cohen, J., Lapinski, J. & Trussler, M. Partisan pandemic: How partisanship and public health concerns affect individuals' social mobility during covid-19. *Science advances* **7**, eabd7204 (2021).

[13] Cornelson, K. & Miloucheva, B. Political polarization, social fragmentation, and cooperation during a pandemic. *University of Toronto Department of Economics Working Paper 663* (2020).

[14] Wright, A. L., Sonin, K., Driscoll, J. & Wilson, J. Poverty and economic dislocation reduce compliance with covid-19 shelter-in-place protocols. *Journal of Economic Behavior & Organization* **180**, 544–554 (2020).

[15] Akesson, J., Ashworth-Hayes, S., Hahn, R., Metcalfe, R. D. & Rasooly, I. Fatalism, beliefs, and behaviors during the COVID-19 pandemic. *NBER Working Paper* (2020).

[16] Barrios, J. M. & Hochberg, Y. V. Risk perception through the lens of politics in the time of the COVID-19 pandemic. *NBER Working Paper* (2020). URL https://doi.org/10.3386/w27008.

[17] Engle, S., Stromme, J. & Zhou, A. Staying at home: Mobility effects of covid-19. *Covid Economics: Vetted and Real-Time Papers* **4**, 86–102 (2020).

[18] Brzezinski, A., Deiana, G., Kecht, V. & Van Dijcke, D. The COVID-19 Pandemic: Government vs. Community Action Across the United States. *Covid Economics: Vetted and Real-Time Papers* **7**, 115–156 (2020). URL https://cepr.org/file/9873/download?token=KuyXlx6b.

[19] Howe, P., Mildenberger, M., Marlon, J. & Leiserowitz, A. Geographic variation in opinions on climate change at state and local scales in the USA. *Nature Climate Change* **5**, 596–603 (2015).

[20] Jamieson, K. H. & Hardy, B. W. Leveraging scientific credibility about arctic sea ice trends in a polarized political environment. *Proceedings of the National Academy of Sciences* **111**, 13598–13605 (2014).

[21] Carey, J. M., Chi, V., Flynn, D., Nyhan, B. & Zeitzoff, T. The effects of corrective information about disease epidemics and outbreaks: Evidence from Zika and yellow fever in Brazil. *Science advances* **6**, eaaw7449 (2020).

[22] Roozenbeek, J., Van Der Linden, S. & Nygren, T. Prebunking interventions based on "inoculation" theory can reduce susceptibility to misinformation across cultures. *Harvard Kennedy School Misinformation Review* **1** (2020).

[23] Iyengar, S. & Massey, D. S. Scientific communication in a post-truth society. *Proceedings of the National Academy of Sciences* **116**, 7656–7661 (2019).

[24] Kahan, D. M. Climate-science communication and the measurement problem. *Political Psychology* **36**, 1–43 (2015).

[25] Chu, D. K. *et al.* Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *The Lancet* **395**, 1973–1987 (2020).

[26] Gostin, L. & Wetter, S. Why there's no national lockdown. *The Atlantic* (2020). URL https://www.theatlantic.com/ideas/archive/2020/03/why-theres-no-national-lockdown/609127/.

[27] Rusch, K. & Smith, C. How do you enforce a law that tramples the land of the free? *The New York Times* (2020). URL https://www.nytimes.com/2020/05/11/us/coronavirus-california-lockdowns.html.

[28] Brzezinski, A., Kecht, V. & Van Dijcke, D. The cost of staying open: voluntary social distancing and lockdowns in the US. *Oxford Department of Economics Working Papers* (2020).

[29] Gupta, S., Simon, K. & Wing, C. Mandated and voluntary social distancing during the covid-19 epidemic. *Brookings Papers on Economic Activity* **2020**, 269–326 (2020).

[30] Barrios, J. M., Benmelech, E., Hochberg, Y. V., Sapienza, P. & Zingales, L. Civic capital and social distancing during the COVID-19 pandemic. *Journal of Public Economics* **193** (2021).

[31] Merkley, E. & Stecula, D. A. Party cues in the news: Democratic elites, republican backlash, and the dynamics of climate skepticism. *British Journal of Political Science* 1–18 (2020).

[32] Lewandowsky, S., Gignac, G. E. & Oberauer, K. The role of conspiracist ideation and worldviews in predicting rejection of science. *PLOS ONE* **8**, e75637 (2013).

[33] The Lancet Child Adolescent Health. Vaccine hesitancy: a generation at risk. *The Lancet: Child & Adolescent health* **3**, 281 (2019).

[34] Torracinta, L., Tanner, R. & Vanderslott, S. MMR Vaccine Attitude and Uptake Research in the United Kingdom: A Critical Review. *Vaccines* **9**, 402 (2021).

[35] Jolley, D. & Douglas, K. M. The effects of anti-vaccine conspiracy theories on vaccination intentions. *PLOS ONE* **9**, e89177 (2014).

[36] Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G. & Rand, D. G. Fighting covid-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science* **31**, 770–780 (2020).

[37] Uscinski, J. E. *et al.* Why do people believe COVID-19 conspiracy theories? *Harvard Kennedy School Misinformation Review* **1** (2020).

[38] Merkley, E. & Loewen, P. J. Anti-intellectualism and the mass public's response to the covid-19 pandemic. *Nature Human Behaviour* **5**, 706–715 (2021). URL https://doi.org/10.1038/s41562-021-01112-w.

[39] Gitmez, A., Sonin, K. & Wright, A. L. Political economy of crisis response. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2020).

[40] McPhetres, J. & Pennycook, G. Science beliefs, political ideology, and cognitive sophistication. *mimeo* (2019).

[41] Whitson, J. A. & Galinsky, A. D. Lacking control increases illusory pattern perception. *Science* **322**, 115–117 (2008).

[42] Blank, J. M. & Shaw, D. Does partisanship shape attitudes toward science and public policy? the case for ideology and religion. *The ANNALS of the American Academy of Political and Social Science* **658**, 18–35 (2015).

[43] Scheufele, D. A. & Krause, N. M. Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences* **116**, 7662–7669 (2019).

[44] Nisbet, M. C. *et al.* Knowledge, reservations, or promise? a media effects model for public perceptions of science and technology. *Communication Research* **29**, 584–608 (2002).

[45] Oliver, J. E. & Wood, T. J. Conspiracy theories and the paranoid style(s) of mass opinion. *American Journal of Political Science* **58**, 952–966 (2014).

[46] Nisbet, E. C., Cooper, K. E. & Garrett, R. K. The partisan brain: How dissonant science messages lead conservatives and liberals to (dis) trust science. *The ANNALS of the American Academy of Political and Social Science* **658**, 36–66 (2015).

[47] Ma, Y., Dixon, G. & Hmielowski, J. D. Psychological reactance from reading basic facts on climate change: The role of prior views and political identification. *Environmental Communication* **13**, 71–86 (2019).

[48] Rekker, R. The nature and origins of political polarization over science. *Public Understanding of Science* **30**, 352–368 (2021).

[49] Hornsey, M. J., Harris, E. A. & Fielding, K. S. Relationships among conspiratorial beliefs, conservatism and climate scepticism across nations. *Nature Climate Change* **8**, 614–620 (2018).

[50] Rutjens, B. T., Heine, S. J., Sutton, R. M. & van Harreveld, F. Attitudes towards science. In *Advances in Experimental Social Psychology*, vol. 57, 125–165 (Elsevier, 2018).

[51] Rutjens, B. T. & van der Lee, R. Spiritual skepticism? Heterogeneous science skepticism in the Netherlands. *Public Understanding of Science* **29**, 335–352 (2020).

[52] Rutjens, B. T. *et al.* Science skepticism across 24 countries. *Social Psychological and Personality Science* 19485506211001329 (2021).

[53] Rutjens, B. T., van der Linden, S. & van der Lee, R. Science skepticism in times of covid-19. *Group Processes & Intergroup Relations* **24**, 276–283 (2021).

[54] Fischhoff, B. The sciences of science communication. *Proceedings of the National Academy of Sciences* **110**, 14033–14039 (2013).

[55] McPhetres, J., Rutjens, B. T., Weinstein, N. & Brisson, J. A. Modifying attitudes about modified foods: Increased knowledge leads to more positive attitudes. *Journal of Environmental Psychology* **64**, 21–29 (2019).

[56] McGuire, W. J. Inducing resistance to persuasion. some contemporary approaches. In *Self and Society. An Anthology of Readings*, 192–230 (Ginn Custom Publishing, 1964).

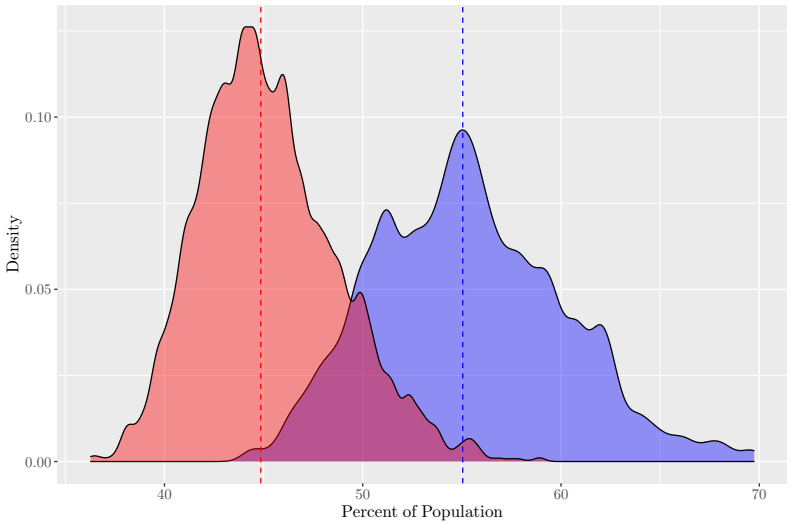
[57] Jolley, D. & Douglas, K. M. Prevention is better than cure: Addressing anti-vaccine conspiracy theories. *Journal of Applied Social Psychology* **47**, 459–469 (2017).

[58] Van der Linden, S., Leiserowitz, A., Rosenthal, S. & Maibach, E. Inoculating the public against misinformation about climate change. *Global Challenges* **1**, 1600008 (2017).

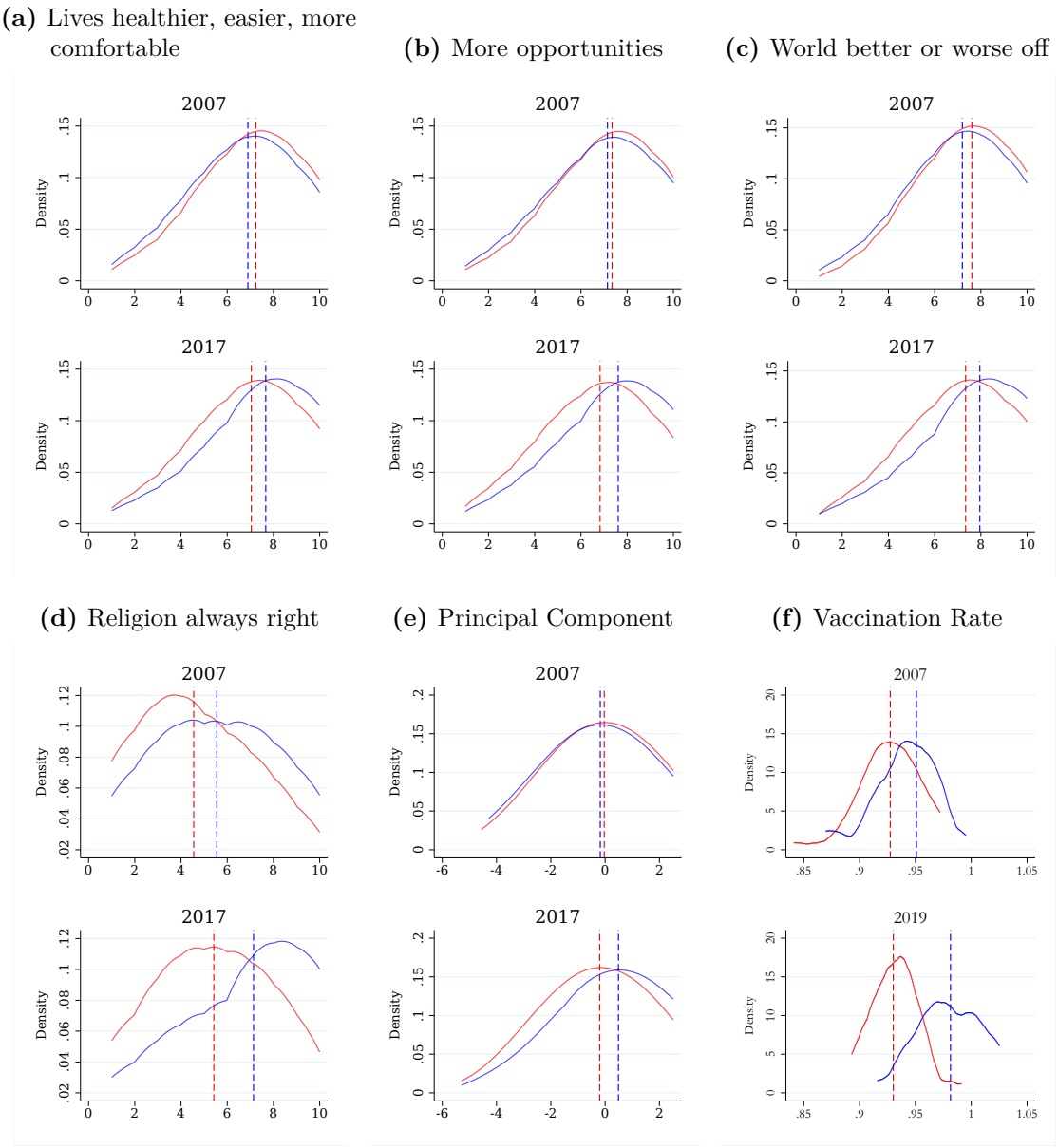
[59] Michie, S., West, R. & Harvey, N. The concept of "fatigue" in tackling covid-19. *bmj* **371** (2020).

Supplementary Figures

Science Skepticism and Partisanship



Supplementary Figure 1 – Distribution of Belief in Anthropogenic Global Warming, by Party: The figure shows smoothed Gaussian kernel densities using Silverman bandwith. Data comes from [1] and measures percent of county population that believes in anthropogenic global warming from surveys conducted over 2008-2013. The left (red) distribution is for all Republican counties, the right (blue) distribution is for all Democratic counties. Dotted vertical lines are party-specific medians. A county is classified according to which party had a higher vote share than the other party in the 2016 presidential election.



Supplementary Figure 2 – Distribution of Other Measures of Science Skepticism, by Party: **Panels a) - d)**: plots the density of measures for trust in science obtained from the World Value Survey (WVS), smoothed using the Epanechnikov kernel function. The red distributions refer to all individuals that identify as Republicans, the blue distributions refer to Democrats. Dashed vertical lines are party-specific means. See Table 1 for more details on the variables. **Panel e)**: shows the first polychoric principal component of the other four variables from the WVS. **Panel f)**: shows the density of the number of MMR vaccination shots received by children during the first 36 months, calculated at the state level. A state is classified according to which party had a higher two-party vote share in the 2016 presidential election. Sample period: March 1 to April 19 2020.

Supplementary Tables

Supplementary Table 1: Benchmarking Science Skepticism (main specification) with Alternative Measures of Trust in Science (Regression-Based Estimates)

	Coeff.	t-Stat.	p-val.	95 % CI		R^2	DoF	Obs.
				Lower Bound	Upper Bound			
Pew Measure	0.384	2.825	0.007	0.110	0.658	0.148	46	48
Principal Component WVS	0.635	5.760	<0.001	0.414	0.857	0.404	49	51
Life Better	0.432	3.354	0.002	0.173	0.691	0.187	49	51
Opportunities	0.476	3.785	<0.001	0.223	0.728	0.226	49	51
Better Off	0.423	3.263	0.002	0.162	0.683	0.179	49	51
Faith	0.584	5.036	<0.001	0.351	0.817	0.341	49	51
Religion always right	0.613	5.430	0.000	0.386	0.840	0.376	49	51
Science important	0.331	2.458	0.018	0.061	0.602	0.110	49	51

Shows coefficients from state-level regressions of science skepticism from [1], aggregated to state-level using population weights, on different measures for trust in science derived from Pew surveys and the World Value Survey (WVS). Higher values of belief in science correspond to weaker skepticism. All variables have been standardized to have mean zero and standard deviation one. The WVS data have been pooled over survey waves 6 and 7 (conducted in 2011 and 2017, respectively) and denote survey-weighted averages at the state level. The WVS variables are: (2) *Principal Component WVS*: "The first polychoric principal component of the other variables"; (3) *Life Better*: "Science and technology are making our lives healthier, easier, and more comfortable"; (4) *Opportunities*: "Because of science and technology, there will be more opportunities for the next generation"; (5) *Better Off*: "The world is better off, or worse off, because of science and technology; (6) *Faith*: "We depend too much on science and not enough on faith"; (7) *Religion always right*: "Whenever science and religion are in conflict, religion is always right"; (8) *Science important*: "It is not important for me to know about science in my daily life." If necessary, the measures have been inverted such that higher values refer to higher levels of trust in science. Standard errors in parentheses.

Supplementary Methods

Data

Physical distancing. We leverage data from SafeGraph, linking GPS pings from up to 40 million mobile devices across the US. This data has previously been used in [2], [3] and [4] to study behavioral responses to the outbreak of COVID-19. We rely on SafeGraph's physical distancing Metrics dataset to assess the differential responses to the shelter-in-place policies at the county level. The dataset is based on an underlying panel of mobile devices in nearly all 200,000+ Census Block Groups (CBG) in the United States, which we aggregate to the county-day level. Only CBGs with less than 5 devices are excluded to preserve privacy. A device's home is defined as its common nighttime location narrowed down to Geohash-7 (153m $\times$ 153m) precision. As our main outcome variable, we determine the *percentage of devices that stayed home all day* by taking the ratio of all such devices for each day over the total number observed in a county during a given day. As discussed, geographic bias of the data is small. The absolute difference between the panel's density and the true population density according to the US census remains below 1% for all counties in the sample; the correlation between both measures is 0.97. The data is similarly representative at the county level in terms of race, demographic and income groups [5].

Science skepticism. To proxy science skepticism (i.e., belief in science), we leverage data on county-level attitudes towards global warming as described in [1]. In particular, we use the item *human*, the estimated percentage of people who think that global warming is caused mostly by human activities (see panel b) in Figure 1). The data builds on 12 nationally representative climate change opinion surveys conducted between 2008 and 2013 that have been combined to an overall dataset (n=12,061). Based on geographic and demographic covariates, [1] apply multilevel regression and post-stratification to predict county-level opinion data. Their estimates from cross-validation and external validation exercises yield county-level error margins

of ±8 percentage points at the 95% confidence level. [6] confirm the validity of this approach.

Alternative belief in science proxies. We collect several alternative measures of belief in science to test the validity of the belief in global warming proxy. For this purpose we conducted a comprehensive search of all datasets listed by the Inter-university Consortium for Political and Social Research and the Roper Center for Public Opinion Research, two large-scale social science database archives. We identified two databases with questions about attitudes towards science as well as geographical information. These are the General Social Survey (GSS) at the NORC (https://bit.ly/3yFvZew) and the American Values Survey (AVS) by Pew Research Center (https://pewrsr.ch/3oRdHlZ). Geographical identifiers in the GSS are only available under restricted access. Moreover, neither survey has a sufficiently large or representative sample to replicate our main county-level results. We use these additional measures to confirm that our main measure of skepticism about the human causes of global warming proxies more general attitudes towards topics of scientific consensus.

- American Values Survey: as an alternative proxy for belief in science, we calculate the percent of the population in each state for which we have respondents that disagree with the statement, "I am worried that science is going too far and is hurting society rather than helping it". The sample consist of 6,292 unique responses pooled from several large national surveys carried out over the years 2002-2009. The surveys were administered over landlines and cell phones. We weight each response by its respective survey weight, as calculated by Pew based on demographic characteristics and survey design.*

- World Values Survey: we also proxy belief in science using a number of variables from the WVS. The questions read: (1) "Science and technology are making our lives healthier, easier, and more comfortable"; (2) "Because of science and technology, there will be more opportunities for the next generation"; (3) "The

*For more information, see https://bit.ly/3vgSl3P.

world is better off, or worse off, because of science and technology"; (4) "We depend too much on science and not enough on faith"; (5) "Whenever science and religion are in conflict, religion is always right"; (6) "It is not important for me to know about science in my daily life". If necessary, the measures have been inverted such that higher values refer to higher levels of belief in science. State identifiers are only contained in the latest two waves of the WVS (2011 and 2017). We pool the data from these waves together in order to increase the sample size and extract the first polychoric principal component from all variables (N=4,582). Two of the questions ((4) and (6)) have not been asked in earlier waves of the survey. For this reason, Supplementary Figure 2 is only based on the remaining variables and their corresponding first polychoric principal component.

- Attitudes towards vaccination: We rely on vaccine hesitancy rates estimated at the county-level using the Census Bureau's Household Pulse Survey (HPS) data. These data are provided by the Assistant Secretary for Planning and Evaluation under the Department of Health and Human Services. ASPE uses the Census Bureau's 2019 American Community Survey (ACS) 1-year Public Use Microdata Sample (PUMS) and a PUMS-to-county crosswalk from the Missouri Census Data Center to estimate county-level variation. The survey collection period was from March 3, 2021 to March 15, 2021. The HPS question used was: "Once a vaccine to prevent COVID-19 is available to you, would you...get a vaccine?" Hesitancy was calculated as the rate of respondents responding probably or definitely not.

- The New York Times commissioned Dynata, an online market research firm, to collect a large scale survey in the United States about mask use. The survey was conducted online between July 2 and July 14, 2020 and includes approximately 250,000 survey responses. Each participant was asked "How often do you wear a mask in public when you expect to be within six feet of another person?" with answer options of "Never," "Rarely," "Sometimes," "Frequently," and "Always." These responses are then normalized to create our

primary outcome of interest described in The New York Times's introduction of the data: the probability (chance) that, if one has five random encounters, all people encountered are wearing masks. County-level data was retrieved from the article's online data repository. For additional information, see: 'A Detailed Map of Who Is Wearing Masks in the U.S.' By Josh Katz, Margot Sanger-Katz and Kevin Quealy. July 17, 2020. *The New York Times*.

- Child vaccination data: we draw data on the number of MMR-containing shots received by children at age 36 months from the National Immunization Survey for the years 2007 and 2019.

POLICIES. We compiled a comprehensive dataset on county-specific and state-wide policies from the National Governors' Association (NGA)[†]; the National Association of Counties (NACO)[‡]; and Education Week[§], an independent news organization that compiled school closure data from government websites, staff reporting and the National Center for Education Statistics. This data incorporates information about the onset of school closures, closures of non-essential businesses and the introduction of shelter-in-place policies.

COVID-19 CASES AND DEATHS. County-day and state-day level information on COVID-related deaths and cases are collected by the Johns Hopkins Coronavirus Research Center.[¶]

ELECTION RESULTS. Data on county-level vote shares by party in the 2016 presidential election was drawn from the MIT Election Lab.[‖] We classify a county as Democratic if the Democratic vote share exceeds the Republican one (and vice versa).

[†]See https://bit.ly/3wyM5ER
[‡]See https://bit.ly/3wy8JNI
[§]See https://bit.ly/3hUkN7B
[¶]See https://bit.ly/3oXDaKE
[‖]See https://bit.ly/3fkYy9l

County Shapefile. The map shown in Figure 1 is based on the 2016 TIGER/Line county-level shapefile provided by the US Census Bureau (Department of Commerce).**

Essential Workers. Data on the share of essential workers have been taken from [7] based on the Quarterly Census on Employment and Wages for the third quarter of 2019.††

**See https://bit.ly/2JU4ZQe
††See https://bit.ly/3yMiAAD

Event study design

The following presents the event study methodology employed for the analysis of policy-related physical distancing. Our specification aims at isolating the differential impact of shelter-in-place policies on physical distancing. We use an event study methodology to compare counties within states with shelter-in-place policies to those without any measure implemented. Our empirical strategy follows a difference-in-differences approach with variation in treatment timing as proposed in [8] and previously applied to the study of behavioral responses to COVID-19 in [3, 4, 9].

To disentangle differential movement patterns according to our covariates of interest, we estimate the following equation separately for the counties with above-median (non-skeptics) and below-median (skeptics) values of the variable *human*:

$$pd_{c,t} = \gamma + \alpha_c + \delta_t + \sum_{j=-10}^{-4} \rho_j D_{s,t_0+j} + \sum_{j=-2}^{10} \rho_j D_{s,t_0+j} + u_{c,t}, \tag{1}$$

where $pd_{c,t}$ is the percentage of devices that stay home all day in county c at time t. α_c and δ_t refer to county and date fixed effects, respectively. The dummies D_{s,t_0+j} are centered around the state policy's implementation date t_0, such that D_{s,t_0+j} equals 1 at time t if the state policy was enacted j days ago. As such, we construct the dummies $D_{s,t+j}$ for the 10 days preceding as well as the 10 days following the implementation date. Panel a in Figure 2 plots our coefficients of interest, the sequence $\{\rho_j\}_{j=-10}^{10}$ for both subgroups. As a baseline period, we use the 15 to 11th day before our state-specific policy was enacted, as well as the third day before the implementation of the policy. This choice of baseline periods follows the recommendation of [10] to omit two (sets of) treatment leads that are as far and as close as possible to the first treatment period, as this leads to the most conservative (most precisely estimated) placebo effects. We omit the third treatment lead since policy implementation in some states occurred in the afternoon of the previous date (we use the first full date of each shelter-in-place policy as the zero date for centering) and members of the public may have been able to anticipate the implementation of the policy due to announcements by government officials in the days immediately before the policy entered effect. As recommended by [4], we drop those counties from

the sample that moved ahead of their state in implementing shelter-in-place policies. This is primarily done to counter concerns that the timing of local policy adoption is endogenous to local virus severity and physical distancing patterns (counties moving ahead of states in implementing a local shelter-in-place).

The different roll-out dates of policies across counties and states allow to construct control groups composed of areas that have not yet experienced the event. To account for the staggered treatment, we incorporate county- and day fixed effects in our analysis. On one side, this eliminates latent, time-invariant heterogeneity across counties that may be correlated with movement patterns, such as differences in income, political partisanship or geographic features. On the other side, the day fixed effects rule out the possibility that changes in physical distancing may be driven by nation-wide common shocks in terms of information or protocols.

Difference-in-differences design

The Difference-in-Differences (DiD) research design follows closely the logic of the event study, with three main differences. First, instead of allowing the policy impact to be different in each of the pre- and post-implementation periods (leads and lags), we estimate a single $post_{s,t}$ effect, whereby $post_{s,t} = 1$ at or after the policy implementation in state s. Second, we do not estimate the equation separately for counties with above- and below-median belief in science, but instead we estimate the interactive effect $post_{s,t} \times BiS_c$, whereby $BiS_c = 1$ if county c has an above-median measure of *human* (non-skeptics). Finally, we also include a vector of control variables $\mathbf{x_{c,t}}$, which includes interactions with the $post_{s,t}$-dummy, in order to control for potential confounding factors.[‡‡] This leads to the following specification for our DiD:

$$pd_{c,t} = \tilde{\gamma} + \tilde{\alpha}_c + \tilde{\delta}_t + \rho post_{s,t} + \psi post_{s,t} \times BiS_c + \mathbf{\Phi x_{c,t}} + \tilde{u}_{c,t} \tag{2}$$

[‡‡]Note that the vector $\mathbf{x_{c,t}}$ contains some time-varying as well as some constant controls. The control variables which are constant over time will be perfectly collinear with the county dummies α_c, and hence for these we only include their interactions with $post_{s,t}$. For the time-varying controls, we include the levels as well as their interactions (model 11).

The key parameter of interest is ψ, which estimates the additional physical distancing response to shelter-in-place policies of counties with above-median belief in science (non-skeptics) relative to counties with below-median belief in science (skeptics). This is the parameter that is plotted on panels b) of Figure 2. The vector $\mathbf{x_{c,t}}$ includes different sets of control variables and their interactions with $post_{s,t}$ depending on the specification, as is described in the Notes to Figure 2.

Scatterplots

The scatterplots we introduce in Figures 4 and 5 rely on standardized residuals. That is, we isolate residual variation in mask use, vaccine hesitancy, and belief in science with respect to a constant (panel a) and county vote share in 2016 (panel b). The residuals of each measure are standardized with mean zero and standard deviation one for consistency across figures. A line of best fit is added and the R^2 is reported for the underlying residualized regressions (graphically depicted as a red line).

References

[1] Howe, P., Mildenberger, M., Marlon, J. & Leiserowitz, A. Geographic variation in opinions on climate change at state and local scales in the USA. *Nature Climate Change* **5**, 596–603 (2015). URL https://doi.org/10.1038/nclimate2583.

[2] Allcott, H. *et al.* Polarization and Public Health: Partisan Differences in Social Distancing during the Coronavirus Pandemic. *NBER Working Paper* (2020). URL https://doi.org/10.3386/w26946.

[3] Brzezinski, A., Deiana, G., Kecht, V. & Van Dijcke, D. The COVID-19 Pandemic: Government vs. Community Action Across the United States. *Covid Economics: Vetted and Real-Time Papers* **7**, 115–156 (2020). URL https://doi.org/10.13140/RG.2.2.10115.84006.

[4] Painter, M. & Qiu, T. Political beliefs affect compliance with government mandates. *Journal of Economic Behavior & Organization* **185**, 688–701 (2021). URL https://doi.org/10.1016/j.jebo.2021.03.019.

[5] Coston, A. *et al.* Leveraging administrative data for bias audits: Assessing disparate coverage with mobility data for covid-19 policy. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* 173–184 (2021). URL https://doi.org/10.1145/3442188.3445881.

[6] Hamilton, L. C., Hartter, J. & Safford, T. G. Validity of county-level estimates of climate change beliefs. *Nature Climate Change* **5**, 704–704 (2015). URL https://doi.org/10.1038/nclimate2720.

[7] Garnier, R., Benetka, J. R., Kraemer, J. & Bansal, S. Socioeconomic disparities in social distancing during the covid-19 pandemic in the united states: observational study. *Journal of medical Internet research* **23**, e24591 (2021). URL https://doi.org/10.2196/24591.

[8] Goodman-Bacon, A. Difference-in-differences with variation in treatment timing. *NBER Working Paper* (2018). URL https://doi.org/10.3386/w25018.

[9] Wright, A. L., Sonin, K., Driscoll, J. & Wilson, J. Poverty and economic dislocation reduce compliance with covid-19 shelter-in-place protocols. *Journal of Economic Behavior & Organization* **180**, 544–554 (2020). URL https://doi.org/10.1016/j.jebo.2020.10.008.

[10] Borusyak, K. & Jaravel, X. Revisiting event study designs. *Available at SSRN 2826228* (2017). URL http://dx.doi.org/10.2139/ssrn.2826228.

Monetary Capacity*

Abstract

Monetary capacity refers to the maximum level of monetization attainable by a state, which depends on demand- and supply-side factors. We argue that monetary and fiscal capacity, and by extension, markets and states have a symbiotic relationship. Long-run European evidence corroborates this mutual dependence, with money and taxes moving in close synch. History also offers a natural experiment to estimate the causal effect of monetary capacity on fiscal capacity. The discovery and extraction of precious metals in the Americas increased money stocks and in turn tax revenues in early modern Europe, a finding that is robust to controlling for economic growth.

JEL Codes: E50, E60, H21, N10, O11

Keywords: monetary capacity, fiscal capacity, monetization, inflation, taxation, quantity theory of money, monetary non-neutrality.

*This chapter was co-authored with Roberto Bonfatti (University of Padua; University of Nottingham), K. Kıvanç Karaman (Bogazici University), Nuno Palma (University of Manchester; ICS, Universidade de Lisboa; CEPR).

1 Introduction

Even a cursory review of long-run evidence hints at a close relationship between monetization and taxation. Breakdowns of monetary systems tend to bring about falling tax revenues and the collapse of government authority, while fiscally troubled states fail to issue and circulate their own money. Understanding this two-way relationship is central to both the study of economics and politics, since markets cannot function without money, and states without taxes. Beyond the anecdotal evidence, however, there is a lack of studies that formally model how monetization and taxation interact, empirically identify the causal effects, and put this relationship in historical perspective.

In this study, we address each of these three points in turn. We begin by introducing the concept of monetary capacity, which we define as the maximum level of monetization that a state can sustain. We then study the interaction of monetary capacity with fiscal capacity, defined by the literature as a state's administrative and political capacity to raise taxes (Besley and Persson, 2009, 2011). Building on our concept of monetary capacity, we provide three contributions. First, we lay out a framework to explain why monetary and fiscal capacity should be seen as complements. This implies that, empirically, one should observe that monetization and taxation move together in the long run. Second, relying on a natural experiment associated with the massive inflow of silver and gold from the Americas—an exogenous, supply-side shock to the European states' monetary capacity—we estimate the causal effect of monetization on taxation for early modern England, France and Spain. Finally, we put the empirical results in historical perspective, relying on monetary and fiscal data for Europe since antiquity. The results indicate that economic and political development are interdependent, and money is not neutral in the long run.

We start by laying out a framework that highlights the dynamically reinforcing relationship between monetary and fiscal capacity. This relationship is mediated by agents' beliefs about the government's ability to sustain a stable currency. On the one hand, greater fiscal capacity influences public beliefs about the sustainability of a

stable currency, because the government is seen as less likely to resort to inflationary episodes to fund its expenditure on defense and other public goods. This increases the monetary capacity of the state. On the other, greater monetary capacity induces a government to invest in building fiscal capacity, because it is more convenient to tax a monetized economy. In the long run, monetary and fiscal capacity depend on each other. Furthermore, any exogenous increase or disruption in monetary capacity spills over to fiscal capacity, and vice versa.

On the empirical front, we estimate the causal impact of monetary capacity on fiscal capacity. We focus on the impact of monetary on fiscal capacity, because it has received less attention in the empirical literature than the converse impact of fiscal on monetary capacity.[1] As is standard in the literature, we proxy for fiscal capacity (which is not observable) using per capita real tax revenues. Similarly, we proxy for monetary capacity using per capita real money stocks. For identification, we rely on the unique historical experiment offered by the discovery and exploitation of precious metals in the Americas across the early modern period. These inflows, which were overwhelmingly privately-owned, led to exogenous variation in European money stocks (Palma, 2022). As historians have repeatedly pointed out, for silver-starved European economies, these inflows meant an increased capacity of governments to issue money, and a dramatic increase in the share of transactions conducted using money.[2] In terms of our framework, they were a positive supply-side shock to monetary capacity, which resulted in greater monetization, resulting in higher fiscal capacity in the long run. For the sample period, we restrict attention to the period after 1550, when silver from the Americas began to arrive in large quantities, and before 1790, when paper monies increasingly replaced commodity money and weakened the link between precious metals and money stocks. For the sample countries, we restrict attention to England, France and Spain, as these countries were heavily impacted by the precious metals shock, and have detailed historical annual data for both monetization and taxation. For estimation,

[1] For the latter impact, see, among others, Catao and Terrones (2005), Karaman et al. (2020) and Sargent (1982).

[2] "[E]ndless books have been written about the dangers of government printing too much money. But for centuries the opposite problem was just as common: governments often couldn't mint enough coins... to meet their subjects' needs" (Pomeranz and Topik, 1999, p.14)

we rely on a Local Projection Instrumental Variable (LP-IV) approach, which yields consistent estimates even at long horizons.

The empirical results identify a significant and substantial causal impact of monetary capacity on fiscal capacity. In the first stage, an increase in the production of precious metals in the Americas increased monetary capacity, as proxied by the per capita real money stock. In the second stage, a 1% increase in monetary capacity increased fiscal capacity (as proxied by per capita real tax revenues) by 0.5%- 1% within a decade, and this impact did not diminish over the course of subsequent decades. These findings are robust after controlling for income, wars and political regime. Likewise, the findings remain robust after accounting for potential violations of the exclusion restriction, relying on the adjusted impulse responses methodology of Jordà et al. (2020a).

We put these empirical results into broader historical context by collecting the fragmentary data available for earlier periods and other parts of Europe. The resulting patterns support the notion of a close relationship between precious metals, monetary capacity and fiscal capacity (again proxied by per capita real money stocks and tax revenues). Before 1500, we find large swings in monetary and fiscal capacity, with peaks at the 5th century BC Athens, the first two centuries AD Rome, and early 14th century, and with troughs in between. We also find that each of these three peaks was associated with surges in silver production, respectively at Laurion, Iberian and central European mines.

For the period between 1500 and 1800, we collect the fragmentary monetary and fiscal estimates available for the Dutch Republic, Portugal, the Ottomans, Poland-Lithuania and Russia, and put them in a comparative perspective with the English, French and Spanish data analyzed in the empirical section. To give a first insight, Figure 1 summarizes the resulting patterns. The top panel plots the relationship between money stock per capita and tax revenue per capita, in grams of silver, for eight states over three centuries. The middle panel plots the money stock per capita and tax revenue per capita, after dividing them both by the cost of a standard consumption basket, in order to account for the changes in price levels. Perhaps most

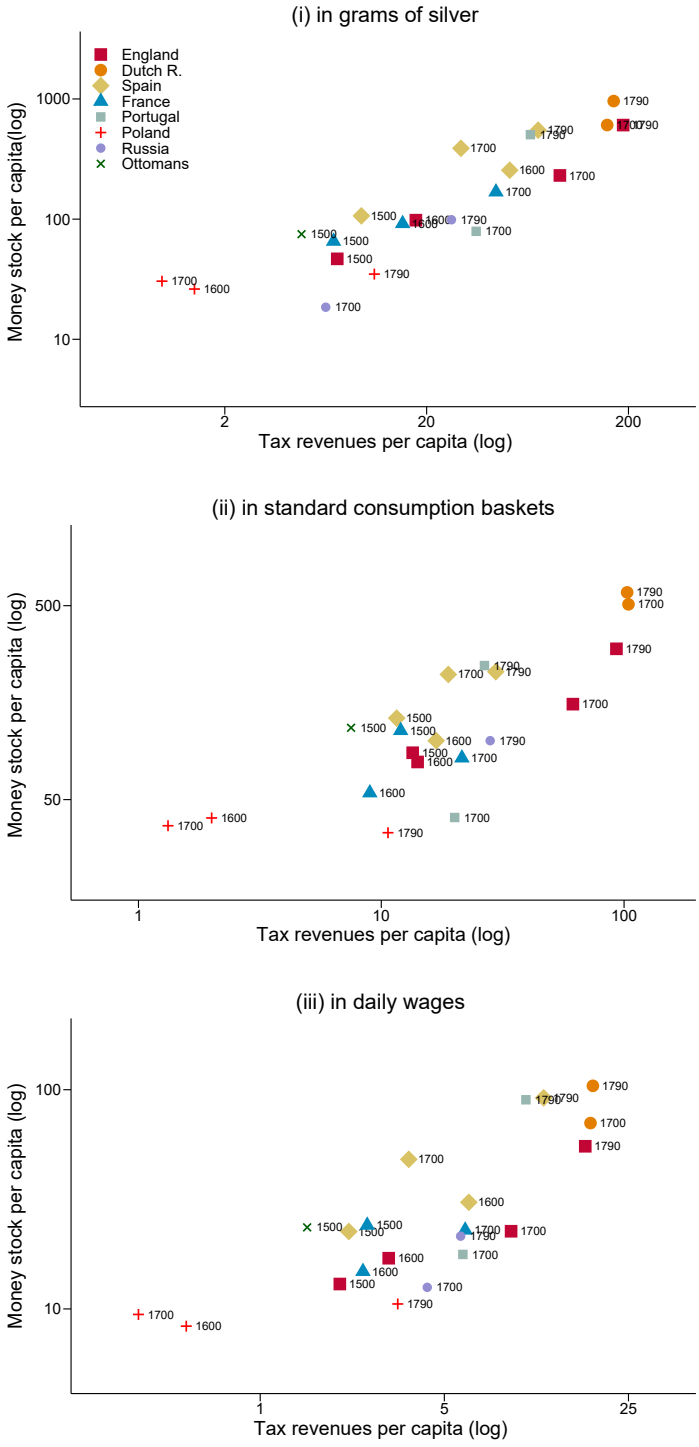
importantly, the bottom panel plots the two variables after dividing them by the nominal wages, which is the commonly used proxy for per capita GDP in historical studies. This normalization reassures against the concern that both variables might be driven by economic growth.

The review of the comparative evidence identifies several patterns. For one, the evidence corroborates the notion that, starting from the latter half of the 16th century, the inflow of precious metals from the Americas resulted in an unprecedented increase in monetization levels, breaking the cycles of monetization and demonetization that had characterized European history until then. Secondly, the impact of the shock was real, in the sense that money stocks and tax revenues increased more than prices and incomes. Third, the impact of the shock was asymmetric, in the sense that it primarily affected Western Europe, and diffused to the East only in limited amounts. Taken together, the continent-wide evidence corroborates the close relationship between monetary and fiscal capacity.

The main contribution of the paper is to establish theoretically and empirically that monetary and fiscal capacity, and in more general terms, markets and states are interdependent. Markets and states are often conceptualized as alternative methods of allocating resources and governing social interactions. This characterization is incomplete, however, as they also depend on each other to function. Markets, and by extension, a monetized and commercialized economy facilitates state-building, while a strong state fosters monetization. We also document that this interdependence is not merely a theoretical possibility. Rather, the co-movement of monetary and fiscal capacity has been a defining and robust characteristic of the European political economy since antiquity.

The co-evolution of monetary and fiscal capacity matters because they shape economic performance. Fiscal capacity allows states to integrate domestic markets, provide public goods, and solve externality problems. It also promotes growth by financing legal capacity and the protection of property rights (Besley et al., 2013; Besley and Persson, 2009, 2011). Monetary capacity, on the other hand, is itself a public good that pervades all sectors of the economy. It lowers transaction costs and

Figure 1: Per Capita Money Stocks and Tax Revenues



Note: The figure plots the relationship between money stock per capita and tax revenue per capita in nominal terms (Panel i), real terms (Panel ii), and normalized by a proxy for per capita GDP (Panel iii). The figure is not intended to establish causation, but rather to review the long-run patterns. See Section 3 for the identification strategy.

facilitates trade and investment (Irigoin, 2009a,b, 2013; Palma, 2018a, 2019). Hence, understanding the evolution of monetary and fiscal capacity helps understand the patterns of long-run economic growth.

The findings also offer new insights on one of the central ideas in macroeconomics, the long-run neutrality of money. The common wisdom in the literature is that money is neutral in the long run: an increase in money supply has no real effects and only leads to a proportional increase in the price level.[3] In contrast, we find that it did have a real long-run effect, in the form of an increase in fiscal capacity. The reason for this positive effect is that historically large segments of economies were not sufficiently monetized. Following a positive monetary shock, money penetrated into these under-monetized segments. Prices increased, but less than proportionally, because now a greater share of the economic activity relied on money. The growing monetization in turn facilitated market transactions and increased tax revenues. Hence, the effects of the increase in money supply were real and had economic as well as political implications.

In the same vein, our findings relate to the literature on the trade-off between monetary expansion and taxation (Mankiw, 1987; Sargent and Wallace, 1981; Sims, 1994). The common wisdom in the literature is that states can raise revenue by either increasing taxes or increasing money supply, the latter of which in turn increases the price level. Hence, in the short run, increasing taxes and increasing money supply are substitutes. We find in this study that, over the long run, the relationship is more complex. In under-monetized economies, if the money supply expands in a way that provides liquidity to new regions or sectors, it also improves the tax collection, triggering a virtuous cycle of monetary and fiscal capacity building. Hence, in the long run, monetization and taxation are in fact complements.

The remainder of the paper is structured as follows. Section 2 sets out our conceptual framework, summarising the insights of our model from Appendix A. Section 3 discusses the new dataset and empirical results. Section 4 puts the evolution of monetary and fiscal capacity in historical context. Section 5 concludes.

[3]See, however, Brzezinski et al. (Forthcoming); Jordà et al. (2020b); Palma (2018a, 2022).

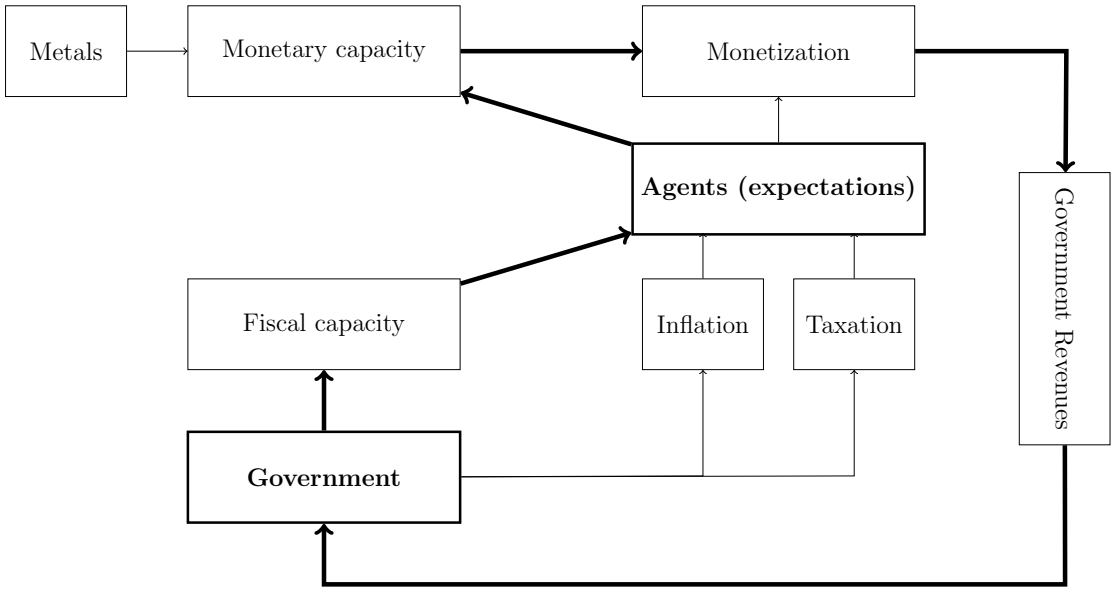
2 Conceptual framework

We define monetary capacity as the maximum level of monetization that a state can sustain. Monetization refers to the amount of transactions that are conducted in money as opposed to barter. What maximum level of monetization is attainable depends on the state's fiscal capacity as well as the availability of precious metals.

Our theoretical framework builds on two insights. The first insight is that the fiscal system is a more efficient way to tax the economy than the inflation tax, given that agents can relatively easily choose not to hold money. The second is that a higher monetization level eases the collection and administration of taxes. It eases the collection, because it expands the role of markets, and markets are easier to monitor and tax than nonmarket activities. Monetization also eases the administration of taxes, because once collected, money taxes can be conveniently transported and governed in a centralized way, whereas taxes in kind cannot.

In a two-period model presented in Appendix section A, we show that these insights together imply a dynamically reinforcing relationship between monetary and fiscal capacity. An overview of the mechanism is shown in Figure 2. In the model, the government (bottom of the figure) can invest in fiscal capacity, and decides on taxation rates as well as inflation in the economy through currency debasements. The government's actions influence agents by changing their tax burden and inflationary expectations, thereby shaping their willingness to hold money. This leads to a reinforcing mutual dependency of monetary capacity and fiscal capacity, highlighted by bold arrows in the figure. On the one hand, greater fiscal capacity increases monetary capacity, because a greater capacity to tax the economy through the fiscal system reduces economic agents' expected inflation, and hence increases the public's willingness to accept state-issued money. The latter is, in turn, a key demand-side determinant of monetary capacity. On the other, greater monetary capacity (and hence greater monetization) induces a government to invest in building fiscal capacity, because the government's revenues from any given tax rate are higher in a more monetised economy. In the long run, monetary and fiscal capacity depend on each other.

Figure 2: Framework overview



Note: The figure represents our framework from Appendix A. The bold arrows represent the mutually enforcing relationship between monetary capacity and fiscal capacity.

Furthermore, any exogenous increase or disruption in monetary capacity spills over to fiscal capacity, and vice versa. Consider, for instance, a change in the availability of precious metals in the economy (top left corner of Figure 2). This directly increases the monetary capacity of the economy, allowing a higher level of equilibrium monetization to be attained. Higher monetization leads to increased government revenues from any given tax rate, which incentivizes government investment into fiscal capacity. The higher fiscal capacity, in turn, lowers inflationary expectations, leading back to higher monetary capacity.

3 Empirical analysis

Our framework argues that an exogenous positive shock to monetary capacity increases fiscal capacity, and a positive shock to fiscal capacity increases monetary capacity. This section tests the first of these predictions, the causal impact of monetary capacity on fiscal capacity. The reason we focus on the impact of monetary capacity is that it is less well-understood and has received less attention in the empirical literature than the impact of fiscal capacity.[4] Moreover, the natural ex-

[4]For the early modern period, empirical support for the argument that fiscal capacity had a positive impact on monetary capacity is provided by Karaman et al. (2020). For 11 European states between 1500 and 1900, the paper finds that a higher fiscal capacity stabilized monetary units,

periment associated with the discovery of and exploitation of massive amounts of silver in the Americas allows for a unique opportunity to identify this impact.

The first subsection that follows discusses the construction of the data series for monetary and fiscal capacity for early modern European states. The second subsection discusses the empirical methodology. The causal effect of monetary capacity on fiscal capacity is estimated through a Local Projection Instrumental Variable (LP-IV) approach, where monetary capacity is instrumented by the discovery and extraction of precious metals from the Americas. The third subsection discusses the results. For Spain, England and France, monetary capacity has a positive, significant and persistent effect on fiscal capacity. The last subsection discusses the robustness of the results.

3.1 Data

We test the impact of monetary capacity on fiscal capacity, instrumenting the former with the silver and gold output in the Americas. These privately-extracted and owned[5] outputs resulted in exogenous variation in European money stocks. Because we rely on a natural experiment for identification, we focus on the period in which this experiment was most relevant, between 1550 and 1790.[6] We also restrict the sample to England, France and Spain, the only countries for which detailed monetary and fiscal data is available. Furthermore, these three countries were among the most affected by the inflow of metals from the Americas.

The proxy for monetary capacity is the yearly per capita real money stock.[7] To calculate this proxy, we first collect the data on the money stocks of different coun-

as states no longer depreciated their currencies for seigniorage. This impact breaks down only at very low levels of fiscal capacity, because weak states lacked the capacity to circulate currency, and could not generate seigniorage revenues in the first place. As for the modern period, among others, Sargent (1982) and Catao and Terrones (2005) provide empirical support for the impact of fiscal capacity on monetary capacity.

[5]The industry was taxed by the Spanish Crown but also received subsidies. Hence, only for Spain was there a direct effect of metal outputs on government revenues. See Section 3.2 for the treatment of potential consequences of this effect.

[6]1550 marks the beginning of the arrival of American precious metals in Europe in large quantities, whereas 1790 marks the growing importance of paper money and a weakening relationship between precious metals and money stocks.

[7]We focus on state-issued money because private forms of money rarely circulated at the national level, did not last long without government backing, and complemented, rather than displaced, state-issued money.

tries.[8] We then convert the money stock from local currency into silver, the universal measure of value for the early modern period.[9] We next divide the money stock by the population[10] to get the per capita money stock in grams of silver. Finally, we divide the per capita money stock by the daily cost of a standard consumption basket in grams of silver[11] to arrive at the per capita real money stock.[12]

Note that monetary capacity is defined as the capacity to monetize the economy, but we proxy for it with the actual monetization level. In order for this proxy to be valid, states should choose to monetize their economies at the highest level they can. This is indeed the case in the equilibrium of the model of Appendix A. It is also empirically plausible for our time period, because early modern economies were chronically starved of money, and states sought to remedy undermonetization by all policy means available.

The proxy for fiscal capacity is yearly per capita real tax revenues. It is calculated by dividing the central government revenues[13] by the population and price level of each country. Consequently, fiscal capacity is measured in terms of standard consumption baskets, the same unit as monetary capacity.

Figure 3 shows the per capita real money stock and real tax revenue series for the three countries in our sample. The two series appear to be closely related, and

[8]Annual money stock estimates for England are based on Palma (2018b) and for Spain based on Brzezinski et al. (Forthcoming). For France, annual estimates for 1493-1680 are based on Glassman and Redish (1985). Between 1680-1788, the series are interpolated based on the six benchmark estimates by Riley and McCusker (1983). The money stock estimates include silver and gold commodity money, but not other types of fiat and paper monies, which were insignificant until the 19th century.

[9]The value of monetary units in terms of silver are based on Karaman et al. (2020) and the sources cited therein.

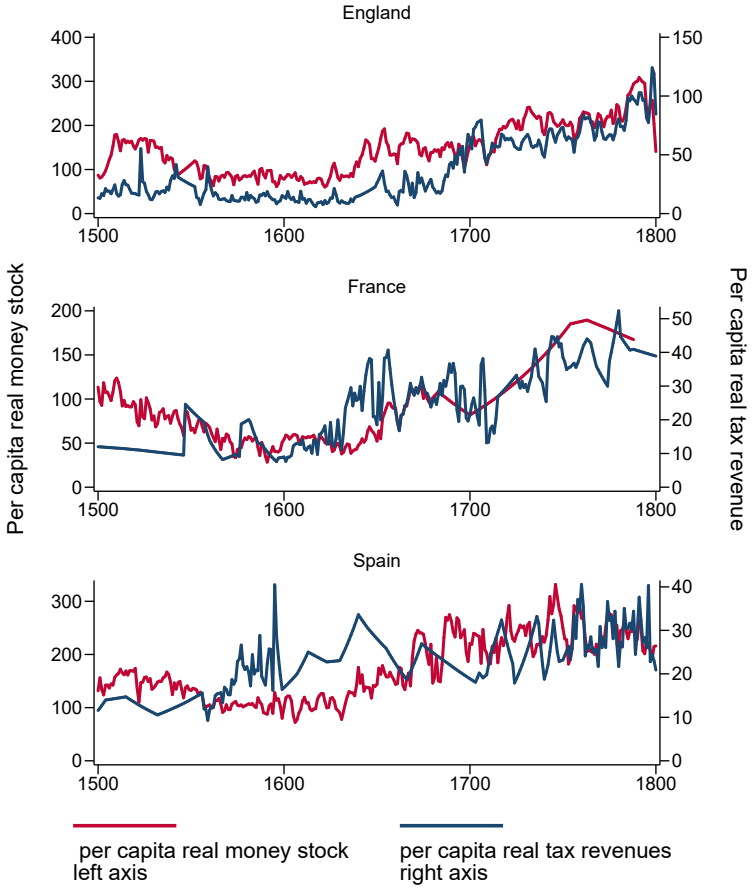
[10]Population data is based on Broadberry et al. (2015), who in turn rely from 1541 on Wrigley et al. (1997), for England, McEvedy and Jones (1978) for France, and Santiago Caballero et al. (2020) for Spain.

[11]Price level data is based on Allen (2001b). The price series tracks the daily cost of a standard consumption basket that includes food items totaling 1941 calories per day, fuel and clothing, and reflects the consumption patterns of an adult male for the early modern period.

[12]Money stocks (which for premodern Europe were predominantly composed of coin stocks) have been reconstructed using a combination of assumptions. These rely on the annual observation of mint output flows combined with the occasionally observed stock (usually at the moment of total or partial recoinages) and other supporting information, which varies from study to study. For more details on the reconstruction method of money stocks, see Palma (2018b).

[13]For Portugal, the revenue series are interpolated based on Costa et al. (2020). The other three states are based on Karaman and Pamuk (2010) and the sources therein. English revenue data is available annually with few gaps. There are gaps for French revenue data in the 16th and Spanish revenue data in the 17th century, which are interpolated. These revenues are converted to silver, once again based on the rates in Karaman et al. (2020).

Figure 3: Money Stock and Tax Revenues Per Capita (in standard consumption baskets)



both increased significantly through the early modern period. This close relationship, however, does not by itself imply a causal relationship between the two, since monetary and fiscal capacity may both be determined by other variables.

In the empirical approach, the production of precious metals in the Americas serves as the instrument for monetary capacity. The precious metal production is the sum of silver and gold production in metric tons, with the latter translated to silver units. The production series is based on Palma (2022), which in turn largely relies on TePaske (2010).

3.2 Empirical strategy

The empirical specification follows the LP-IV methodology. This methodology uses an excluded instrument to estimate the causal effects for different time horizons.

The methodology is well-suited for our aim of estimating long-run effects, as it yields consistent estimates even at long horizons.[14]

Following the LP-IV methodology, we proceed in two stages. In the first stage, we use the production of precious metals in the Americas as an instrument for monetary capacity. In the second stage, we estimate the causal effect of monetary capacity on fiscal capacity, using only the exogenous variation in the former variable that was generated by the production of precious metals.

The first-stage equation, estimated separately for each country i, is of the following form:

$$ln(monetary_{i,t}) = \alpha_i + \gamma_i ln(metals_{t-1}) + \boldsymbol{\phi_i} \mathbf{x_{i,t}} + e_{i,t}, \tag{1}$$

where $metals_t$ is the production of precious metals in American mines, expressed in grams of silver.[15]

The second-stage relationship is estimated separately for each country i for each time horizon h:

$$\ln(fiscal_{i,t+h}) - \ln(fiscal_{i,t-1}) = \alpha_{i,h} + \beta_{i,h} ln(\widehat{monetary}_{i,t}) + \boldsymbol{\psi_{i,h}} \mathbf{x_{i,t}} + u_{i,t+h}. \tag{2}$$

The outcome variable on the left-hand side is the cumulative growth of real per capita fiscal capacity between period $t+h$ and $t-1$ in country i. The main explanatory variable on the right-hand side, $ln(\widehat{monetary}_{i,t})$, is the log of real per capita monetary capacity, instrumented by the production of precious metals in the first stage. $\mathbf{x_{i,t}}$ is a vector of control variables whose components are described below for each local projection result. $u_{i,t+h}$ is a horizon-specific error term. To address the potential serial correlation in the error term (Jordà, 2005), we use HAC-robust standard errors based on Newey and West (1987) throughout.[16]

[14] See Jordà et al. (2020b) for a general discussion of the methodology and the consistency of the estimates.

[15] Note that the excluded instrument has no i-subscript, since it captures the production of precious metals in the Americas and hence is the same across the European countries in the analysis.

[16] Montiel Olea and Plagborg-Møller (2021) note that the inclusion of lagged variables in the

The main coefficient of interest is $\beta_{i,h}$, estimated for each time horizon h in the second stage. This coefficient captures the impact of exogenous changes in monetary capacity on the cumulative growth of fiscal capacity up to period h in country i. For example, $\beta_{i,0}$ measures the impact of monetary capacity on fiscal capacity between periods t and $t-1$; $\beta_{i,1}$ measures the impact between periods $t+1$ and $t-1$; and so on.

There is a close correspondence between the theoretical model of Appendix A and our empirical approach. In the theoretical model, an exogenous positive shock to the money supply (ϕ_μ) increases monetary capacity (μ_2) and the equilibrium level of monetization (m_2^*). The first stage of the empirical model tests this impact, with American precious metal output as the exogenous shock, and money stock per capita as the monetization level. The theoretical model also establishes that the higher monetization level in turn increases fiscal capacity ($\tau_2 - \tau_1$) and the equilibrium tax level (t_2^*). This impact is tested in the second stage, which estimates the impact of real money stock per capita on real tax revenue per capita.

One concern with the empirical strategy is whether the proposed instrument, precious metal production in the Americas, satisfies the two requirements of the LP-IV methodology.[17] The first requirement is relevance: the instrument, precious metal production, should be sufficiently correlated with the European money stocks, after controlling for other variables. The existence of a strong correlation between the two is confirmed by both historical and statistical evidence. In the early modern period, the bulk of the European money stock was coins made of silver and gold, and America was by far the most important producer of these two metals (Barrett, 1990). As for the statistical evidence, the first-stage estimation results presented in the next section corroborate the close relationship between American precious metal production and European money stocks.

The second requirement is the exclusion restriction: after controlling for other covariates, American precious metal production should affect fiscal capacity only through monetary capacity. There are several ways that this condition can be vio-

estimation can eliminate the need to use HAC-robust standard errors. In our specification, we include lagged variables and additionally allow for autocorrelation in the error term.

[17] For a formal discussion, see Stock and Watson (2018).

lated. One potential violation would occur if the conditions in Europe, such as the state of the economy, affected both the precious metal production and fiscal capacity simultaneously. The historical evidence suggests, however, that the variations in American precious metal production were independent of European conditions. In particular, Palma (2022) shows that both the discovery of the mines and the mining intensity were driven by local conditions and accidents, with a high degree of randomness. A second potential violation is that American metal production might have affected fiscal capacity through channels other than monetary capacity. To address this concern, we include the main political and economic characteristics of European states as covariates in the regressions: GDP, warfare and political regime. A final issue, only relevant for Spain, is that the precious metal production was itself a source of tax revenue and thereby increased fiscal capacity directly. To address this specific issue, and as a general robustness exercise, we estimate a modified version of the LP-IV method based on the methodology outlined by Jordà et al. (2020a). The modified method assumes that the exclusion restriction does get violated, and adjusts the estimates for this violation. The results of these exercises, reported in the following sections, support the robustness of the empirical findings.

3.3 Results

Table 1 reports the first-stage results. The F-statistic is above the rule-of-thumb value of ten for all three countries in the sample. Likewise, the elasticity of the money stock with respect to metal production is positive and significant for all three, with a higher point estimate for Spain. In particular, a one percent increase in the production of precious metals in the Americas is estimated to have increased Spanish real per capita money supply by 0.4 percent after one year. This greater impact on Spanish money supply is expected, since the precious metals from the Americas first arrived in Spain, and diffused to England and France only indirectly through capture or trade.[18]

[18]Note that, in principle, the higher elasticity could also be explained by differing initial money stock levels: if England and France had higher monetary stocks per capita, then an equal increase in precious metals would mechanically lead to a less strong response in these two countries. However, for our period of consideration, the opposite was the case, as Spain had a higher per capita monetary

Table 1: First-stage results

Dependent variable: real money p. c.	England (1550-1790)	France (1550-1788)	Spain (1550-1790)
lagged metals production	0.189***	0.261***	0.431***
	(0.058)	(0.054)	(0.053)
Observations	241	239	241
First-stage F	10.39	23.36	65.89

Note: The table shows the first-stage results, where the endogenous variable is the log real per capita money stock. Regressions control for two lags of log real tax per capita, as well as contemporaneous and lagged real GDP. The LP-IV approach uses Newey-West standard errors throughout. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 4 presents the second-stage results. For each horizon h and each country i, the vertical axis reports the point estimate of the coefficient $\beta_{i,h}$, together with the confidence intervals. For all three countries in our sample, an exogenous increase in monetary capacity increased fiscal capacity. For the timing of impact, it was instantaneous in Spain, but worked with a lag for France and England. For its magnitude, it was the strongest for England, where a one percent increase in real money stock per capita increased real tax revenues per capita by 0.8 percent in five years and more than one percent in twelve years. As for its persisence, in all three countries, the impact did not diminish over time.

3.4 Robustness

We next evaluate the robustness of the results. For this purpose, we estimate the impulse responses with different sets of control variables and different time horizons. We also estimate adjusted impulse response functions that take into account potential violations of the exclusion restriction, following the methodology of Jordà et al. (2020a).

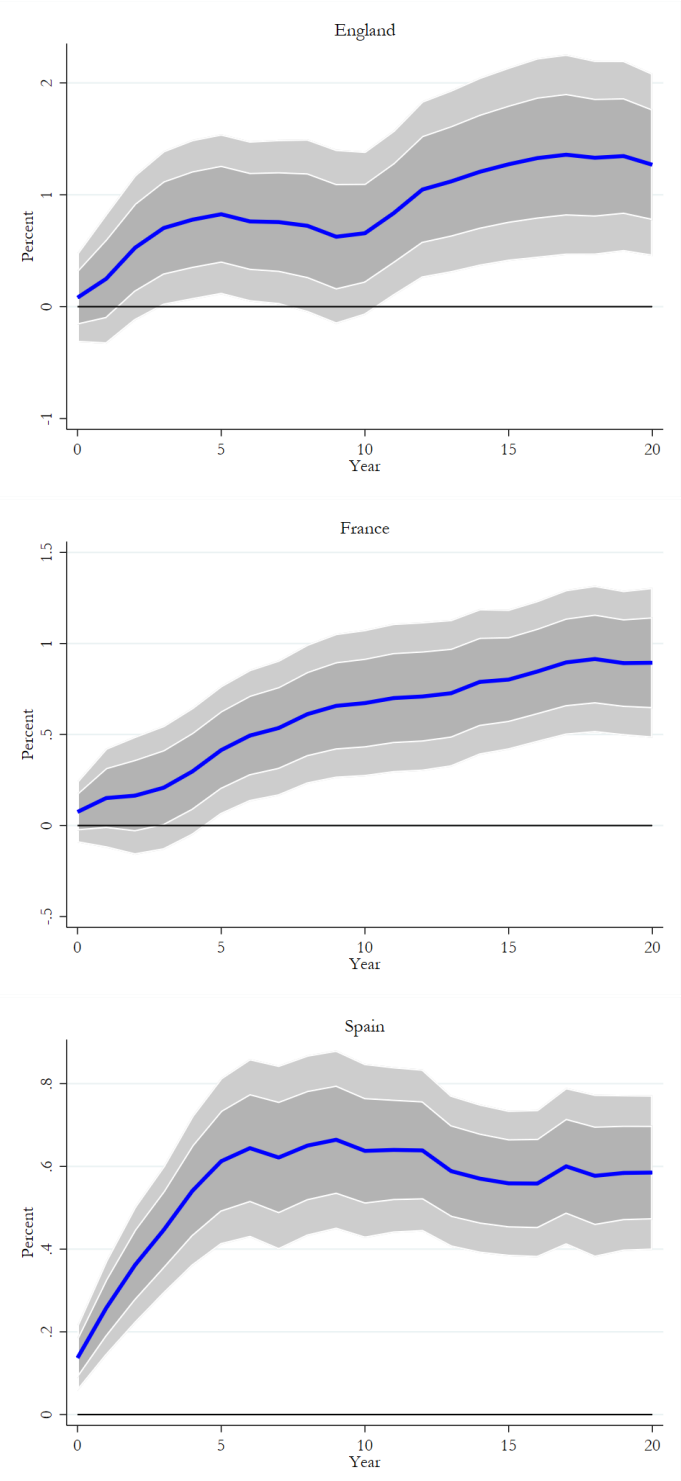
Robustness to control variables

Table 2 presents the estimation results for different sets of control variables. For each specification, the size of the impact and standard errors are reported for horizons $h = 10$ and $h = 20$. The full set of impulse responses are reported in Appendix B.

Column 1 is the baseline specification, presented earlier in Figure 4. Columns

stock (see Figure 1).

Figure 4: % change in per capita real tax revenues due to a 1% exogenous increase in per capita real monetary stock



Note: LP-IV impulse responses showing the cumulative response of real tax per capita to a 1% increase in real money stock per capita, instrumented by the production of precious metals. 90% (light gray) and 1 standard deviation (dark gray) Confidence Intervals shown, based on Newey-West standard errors. The regressions control for two lags of log of real tax revenues per capita, as well as contemporaneous and lagged real GDP. Source: see text.

2 and 3 report the results when the money stock is instrumented with the contemporaneous or the second lag of metals production instead of the first lag. Across both specifications, the estimated coefficients are very similar in size and statistical significance to the baseline results.

Column 4 modifies the baseline specification by dropping GDP from the set of control variables. The estimated coefficients increase across all time horizons, except for Spain, where they remain similar. Note that in the baseline specification, the estimated coefficient for monetary capacity captures only its direct positive impact on fiscal capacity. When GDP is omitted from the set of control variables, the estimated coefficient also captures any indirect positive impact that may be transmitted through a higher GDP, which is consistent with a higher estimate.

Column 5 controls for the impact of warfare. Bellicist theories of state formation argue that wars in early modern Europe set in motion innovations in military and fiscal administration that resulted in ever higher tax revenues and the birth of the modern state (Hoffman and Norberg, 2002; Karaman and Pamuk, 2013; Levi, 1988; North, 1981; O'Brien and Palma, 2020). To account for any bias that the impact of warfare might introduce to the estimation results, we include a proxy for it in the regression equations. To construct the war proxy, we first calculate the total number of wars each state participated in each year, based on Brecke (1999). We then calculate a 3-year moving average of this variable, since the impact of war built up over time. When the proxy for warfare is included in the regressions, the estimates remain close to those in the baseline specification.

Column 6 controls for the level of parliamentary activity. The contract theory of the state argues that parliaments increased tax revenues, by making it easier for the executive and economic elites to settle on a tax-for-public-services deal, solving associated collective action problems, and lending legitimacy to taxation (Hoffman and Norberg, 2002; Levi, 1988; North, 1981). To account for the potential impact of the parliaments, we include in the regressions an indicator variable for parliamentary meetings for each year and state, based on Henriques and Palma (2020). For England and France, the estimation results remain similar to the baseline, while for Spain the

estimates double. This result is consistent with the argument that American silver made Spanish kings less dependent on local elites and undermined the representative government (Abad and Palma, 2021; Drelichman and Voth, 2008). If precious metals had two separate effects on tax revenues, a positive effect through monetization and a negative effect through the collapse of parliaments, controlling for the latter would only leave the former to be estimated and result in a higher coefficient. Column 7 controls for the impacts of warfare and parliamentary activity simultaneously. The estimates remain very similar.

Table 2: Robustness checks

	Baseline	No Lag Instrument	Second Lag Instrument	No GDP Control	War	Parliaments	War and Parliaments
England							
$h = 10$	0.656	0.636	0.739*	1.325***	0.557	0.708	0.612
	(0.444)	(0.448)	(0.435)	(0.478)	(0.511)	(0.463)	(0.541)
$h = 20$	1.268**	1.181**	1.210***	1.842***	1.336**	1.345***	1.427**
	(0.496)	(0.515)	(0.458)	(0.500)	(0.606)	(0.507)	(0.627)
France							
$h = 10$	0.673***	0.637***	0.744***	1.160***	0.718**	0.686***	0.753**
	(0.245)	(0.224)	(0.274)	(0.271)	(0.326)	(0.249)	(0.341)
$h = 20$	0.894***	0.834***	0.983***	1.299***	1.095***	0.898***	1.111***
	(0.251)	(0.225)	(0.285)	(0.259)	(0.354)	(0.255)	(0.369)
Spain							
$h = 10$	0.638***	0.630***	0.652***	0.627***	0.736***	1.370***	1.491***
	(0.128)	(0.122)	(0.130)	(0.129)	(0.169)	(0.353)	(0.431)
$h = 20$	0.585***	0.571***	0.615***	0.607***	0.572***	1.133***	1.083***
	(0.114)	(0.111)	(0.119)	(0.121)	(0.140)	(0.310)	(0.342)

This table shows second-stage results for horizons $h = 10$ and $h = 20$ for alternative specifications, with Newey-West standard errors in parentheses. The first column shows the baseline specification. The second and third column show alternative specifications where the contemporaneous value or the second lag of the metals production variable are used in the first stage, respectively. The fourth column shows the baseline specification without any GDP controls. The fifth, sixth and seventh column show results when adding war controls, parliament controls, and both of these controls to the baseline, respectively.

Finally, we consider the validity of our results for longer time horizons. Note that, the further we extend our horizon, the smaller the sample becomes, and thus the less reliable and precisely estimated the results are. Nevertheless, Appendix Figure A7 clearly shows that the estimated effects remain strong and significant for the subsequent decade, at horizons of 20-30 years.

Robustness to violations of the exclusion restriction

One concern with the LP-IV approach is a potential violations of the exclusion restriction. In our case, such a violation would occur if the production of precious

metals increased fiscal capacity through a channel other than monetary capacity. This concern is particularly relevant for Spain, where American precious metal production also generated tax revenues for the treasury. In this section, we investigate the robustness of our findings under such a violation.

To construct impulse responses that are robust to violations of the exclusion restriction, we adopt the procedure proposed by Jordà et al. (2020a). To see how this approach works, suppose that the excluded instrument has a direct effect on fiscal capacity that does not operate through monetary capacity and other control variables. In that case, the true two-stage relationship for each country i changes to the following system of equations:

$$y_{i,t+h} = \tilde{\alpha}_{i,h} + \tilde{\beta}_{i,h} ln(\widehat{monetary}_{i,t}) + \tilde{\delta}_{i,h} ln(metals_{t-1}) + \tilde{\boldsymbol{\psi}}_{\boldsymbol{i,h}} \mathbf{x_{i,t}} + \tilde{u}_{i,t+h} \quad (3)$$

$$ln(monetary_{i,t}) = \tilde{\alpha}_i + \tilde{\gamma}_i ln(metals_{t-1}) + \tilde{\boldsymbol{\phi}}_{\boldsymbol{i}} \mathbf{x_{i,t}} + \tilde{e}_{i,t}, \quad (4)$$

where $y_{i,t+h} \equiv \ln(fiscal_{i,t+h}) - \ln(fiscal_{i,t-1})$. The only difference to equations 1-2 is that the term $ln(metals_{t-1})$ now enters the second-stage equation 3 directly, violating the exclusion restriction.

Jordà et al. (2020a) offer a potential way to test the robustness of the results to such violations of the exclusion restriction. In particular, suppose that $\tilde{\beta}_{i,h} = \lambda_{i,h} * \tilde{\delta}_{i,h}$. In our setup, $\tilde{\beta}_{i,h}$ is the impact of monetary capacity on fiscal capacity, whereas $\tilde{\delta}_{i,h}$ is the impact of American precious metal production on fiscal capacity that does not go through an increase in monetary capacity. Hence, a higher $\lambda_{i,h}$ corresponds to a greater relative impact for monetary capacity.

With $\lambda_{i,h}$ in hand, one can adjust the left-hand-side variable such that the exclusion restriction is satisfied (Jordà et al., 2020a, Appendix B). In particular, define $y^{adj}_{i,t+h}$ as:

$$y^{adj}_{i,t+h} = y_{i,t+h} - ln(metals_{t-1}) * \frac{\widehat{\delta}_{i,h} + \widehat{\gamma} \times \widehat{\beta}_{i,h}}{1 + \lambda_{i,h} \times \widehat{\gamma}} \tag{5}$$

where the "$\widehat{\ }$"-symbol stands for OLS estimates from equations 3-4. After replacing the left-hand-side variable with the adjusted variable from equation 5, the LP-IV procedure will yield valid results.

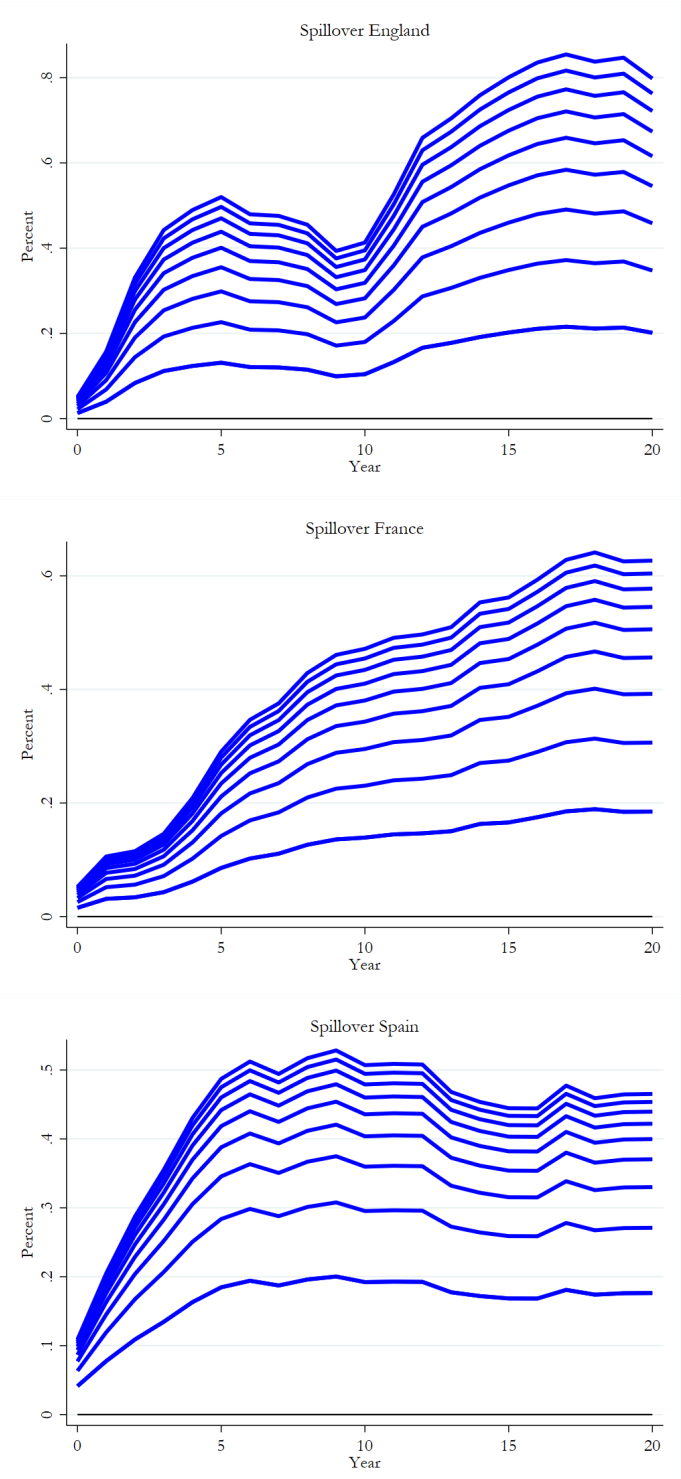
The difficulty in implementing this procedure is that the true value of $\lambda_{i,h}$ is not known and cannot be determined from the data. Hence, Jordà et al. (2020a) advocate performing the adjustments for various reasonable levels of $\lambda_{i,h}$. These levels can be viewed as the bounds on the true impulse responses that would have arisen if the exclusion restriction did not hold.

In our case, we estimate the model for $\lambda_{i,h}$ values between 1 and 10. A $\lambda_{i,h}$ value of less than 1 is unlikely, as it would imply an impact of monetary on fiscal capacity that is smaller than the effect of precious metals on tax revenues which goes through channels other than changing monetization. For Spain, the main beneficiary of American silver, the revenues from taxation of the domestic economy greatly outweighed the revenues associated with the colonies and silver (Coman and Yun-Casalilla, 2012). As for England and France, if American silver did have any impact on fiscal capacity other than by affecting monetization, it must have been much less important than for Spain. A $\lambda_{i,h}$ value of greater than 10 is plausible, but as $\lambda_{i,h}$ increases, the estimation results converge to the baseline estimates, and so having very high values does not add additional insights to the robustness exercise.

Figure 5 presents the estimation results for $\lambda_{i,h} = 1, 2, ...10$. For each country in our analysis, the spillover-adjusted responses are very similar in shape to the baseline results. As before, the effect does not fade away at longer horizons. For the more realistic larger values of $\lambda_{i,h}$, the effect sizes are within the confidence bounds of the baseline results presented in Figure 4.

Taken together, the results in this section confirm that the increase in monetary capacity caused by the massive inflow of precious metals from the Americas had a positive impact on fiscal capacity. This empirical support for this impact is robust

Figure 5: Spillover-adjusted Impulse Responses



Note: Spillover-adjusted impulse responses (Jordà et al., 2020a). Spillover terms are $\lambda_{i,h} = 1, 2, ...10$. Lower lines correspond to lower values of $\lambda_{i,h}$. Specifications follow the baseline LP-IV approach. Sources: see text.

to taking into account the effects of other relevant variables, and potential violations of the exclusion restriction.

4 Long-run patterns

This section puts the theoretical model and empirical analysis in the preceding sections into a broader historical context by reviewing the European evidence since antiquity. The long-run evidence supports the notion that precious metal production, monetary capacity and fiscal capacity had a mutually reinforcing relationship and moved together through several cycles of ups and downs from antiquity to the Middle Ages. It also indicates that the American silver shock ended these cycles and had a real, substantial, asymmetric and transformative impact on the European economy and politics.

The first well-documented historical peak in monetization and taxation was in the mid-5th century BC Athenian Empire. In this period, the estimated money stock per capita reached a level of around 200 grams of silver, and tax revenues per capita a level of around 10-15 grams. These levels were made possible, in part, by the Laurion silver mines located nearby the city of Athens, which had an estimated annual peak output of about 80 tons (Ober, 2015; Patterson, 1972).

After the decline of Athens, the second peak was reached during the Pax Romana. In the first two centuries AD, money stocks per capita rose to 50-100 grams of silver, and tax revenues per capita to 10-15 grams. Roman monetization was supported by the Iberian mines, which had an estimated annual output of 200 tons of silver (Duncan-Jones, 1998; Goldsmith, 1987; Hopkins et al., 2009; Patterson, 1972).

The Roman peak was followed by a collapse in the third century AD. The exact line of causation is difficult to unravel. In terms of precious metal availability, the exhaustion of the Iberian mines and the resulting decline in annual silver output to about 30 tons arguably played a role (Patterson, 1972). Fiscally, the growing spending on war triggered depreciation of silver coinage and inflation, which in turn induced demonetization of the economy (Bransbourg, 2015; Harris, 2008; Kohn,

2005).[19]

The early Middle Ages marked the nadir of both monetary and fiscal capacity. On the monetary front, annual silver output in Europe is estimated to have decreased to a few tens of tons, and money stocks per capita to about 15 grams of silver (Patterson, 1972). On the fiscal front, centralized taxation collapsed, leaving little historical record of treasury revenues. These two processes were inherently related. Demonetization necessitated breaking up the armies into smaller units that could be supported by the local produce, and central governments ceded de facto control over the countryside (Kohn, 2005; Spufford, 1988). The collapse of central governments in turn undermined monetization, as mining and minting required large-scale investments and centralized coordination.

Another cycle of expansion and contraction of silver output occurred from the 11th to the 15th centuries. With the discovery of new mines in northern and central Europe, annual silver output gradually increased to about 50 tons by around 1300. However, as soon as what could be extracted from these mines with the technology of the day was exhausted, the output began to decline, and collapsed back to around 15 tons by 1450s (Velde and Weber, 2000).

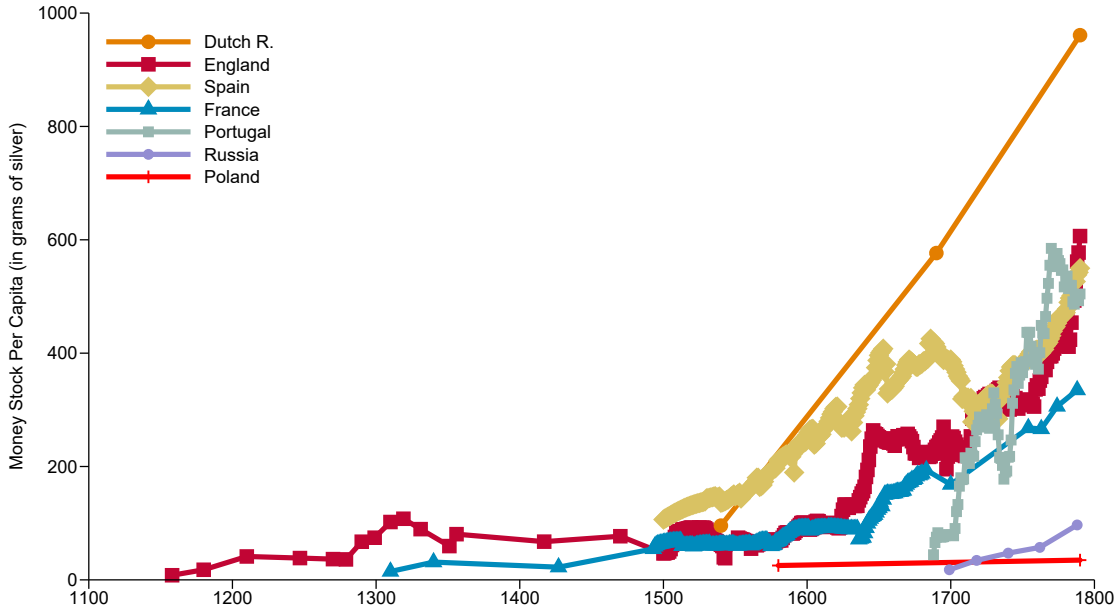
Estimates for English money stock are available for this period, allowing us to track how its trajectory mimicked that of the medieval silver production cycle. Figure 6 summarizes the available per capita money stock data. The English money stock per capita increased from 10 grams to 150 grams from 1150 to 1300, but fell back to around 70 grams by 1450, in line with the silver production estimates.[20]

The 16th century marked a dramatic increase in the availability of precious metal. This increase was driven by the silver and gold production in the Americas. The total silver production in the Americas between the 16th and 18th centuries is estimated to be at around 100,000 tons, or, to compare with annual output estimates presented earlier, around 3,300 tons a year. The silver arrived in Europe in two

[19]During the Late Roman Empire, gold increasingly replaced silver, and perhaps constituted more than half of the money stock in value, but whether it was a perfect substitute for silver is debatable. Gold had a much slower velocity, and was only convenient for very large transactions (Goldsmith, 1987). Gold production itself declined after the fourth century and largely disappeared by the 8th century (Spufford, 1988, p. 18-21).

[20]The English money stock series for 1158-1247 is based on Allen (2001a) and for 1270-1470 on Mayhew (2013). The pre-1500 French money stock series is based on Spufford (1988).

Figure 6: Money Stock Per Capita (in grams of silver)



waves, with Potosí silver peaking in the early 17th and Mexican silver in the late 18th century. The total gold production, mainly in Brazil, was about 2,500 tons, and peaked around the mid-18th century (Barrett, 1990).

This drastic increase in precious metals availability marked the end of the cycles of monetization and demonetization and the beginning of a dramatic upsurge in money stocks in Western Europe. Figure 6 plots the fragmentary evidence on the per capita money stocks of Portugal, the Dutch Republic, Russia, Poland-Lithuania, in addition to the more detailed estimates for England, Spain and France discussed earlier.[21]

Three patterns stand out. First, the money stocks in Western Europe surged following the inflow of precious metals from the Americas. From 1500 to 1800, monetary stocks per capita of Western European states increased from 50-100 grams to 300-1000 grams of silver.

Second, the trajectory of money stocks once again mimicked that of precious metal production. For Spain, England and France, there were two local peaks, in

[21]The money stock data is based on De Vries and Van der Woude (1997) and Weber (2000) for Dutch Republic, Wojtowicz et al. (2005) for Poland-Lithuania, Blanchard (1989) and Kahan and Hellie (1985) for Russia, (Sousa, 2006) for Portugal.

the mid-17th and late 18th centuries, coinciding with the peaks in silver output. For Portugal and England, the main beneficiaries of the Brazilian gold, per capita money stocks increased rapidly over the 18th century.

Finally, Figure 6 also makes it clear that American precious metals had an asymmetric impact across Western and Eastern Europe. The precious metals first arrived in Western Europe, diffused to the East through trade, but only to a limited extent, leading to a divergence in monetization levels. As the figure makes clear, monetization levels in Poland and Russia remained low.[22]

One limitation of the per capita monetary stock series presented in Figure 6 is that they are in nominal terms and only prove a nominal impact. Establishing a real impact requires showing that the money stocks increased more than the prices and incomes, so that the real money stocks increased.

To investigate this question, the top panel in Figure 7 plots the ratio of per capita money stocks to the nominal wages of unskilled workers. The results are consistent with a real impact. For Western European states, money stocks per capita increased from 15 to 20 days worth of daily unskilled workers' wages in 1500 to around 45 to 100 days by 1800, while in Eastern Europe they remained stagnant.

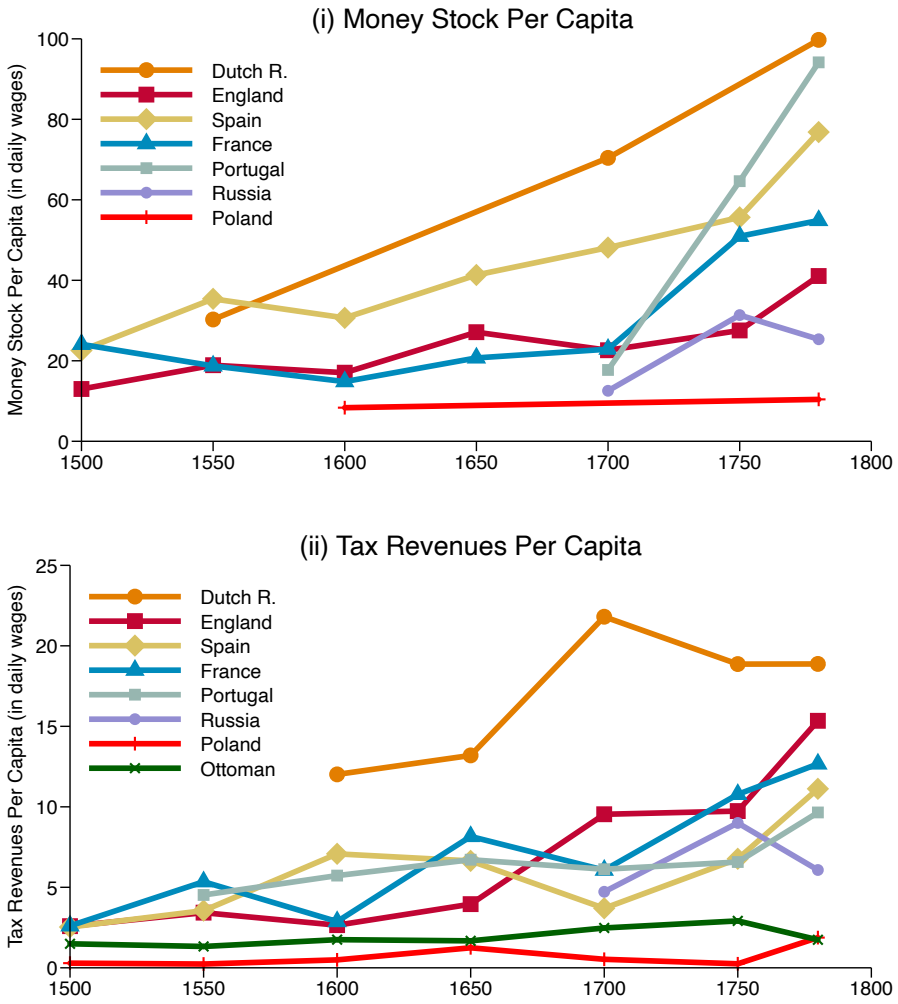
Finally, to explore the fiscal counterpart to this increase in monetary capacity, the bottom panel in Figure 7 plots the ratio of per capita tax revenues to nominal wages. The figure confirms a corresponding real increase in fiscal capacity. In Western Europe, per capita tax revenues increased from about 3 daily wages by 1500 to about 10-15 by 1800, while they lingered at lower levels in the East.[23]

The virtuous cycle of monetary and fiscal capacity building in early modern Western Europe ultimately altered the nature of money, but not the close relationship between the two. The money stock series presented in this section keep track of the levels of commodity silver and gold money over time. Increasingly, however, public banks began to experiment with paper money convertible to silver and gold,

[22]Russia received limited American silver, and continued to rely on barter and furs in addition to foreign coins (Bonney, 1995; LeDonne, 1991; Seljak, Seljak; Spasskii, 1970). Monetization levels only picked up in the 18th century, after opening of local silver mines (Danila, 2006; Kahan and Hellie, 1985; Kotilaine, 2005).

[23]Revenue data is based on Costa et al. (2020), Karaman and Pamuk (2010), Karaman and Pamuk (2013) and the sources cited therein.

Figure 7: Money Stock and Tax Revenues Per Capita (in daily wages)



arguably made possible by the growing fiscal credibility of states. These experiments culminated in the successful paper note issue by the Bank of England in 1694, a model that spread across the continent. The growing adoption of convertible and later fiat paper money in turn weakened the historical dependence of the money supply on precious metals production. The value of paper money, however, still depended on the fiscal capacity of the state, perhaps even more so than commodity money. Hence, while the relationship between precious metals and monetary capacity weakened, the relationship between monetary and fiscal capacity persisted.

All in all, the long-run historical evidence identifies large contemporaneous swings in monetary and fiscal capacity, which continued until the massive liquidity shock of American silver and gold. This shock transformed the European economy by

increasing monetization levels in Western Europe, and ultimately undermining the dependence on precious metals for the money supply itself. There is also overwhelming evidence that the shock triggered a virtuous cycle of monetary and fiscal capacity building, consistent with the theoretical and empirical analysis presented in earlier sections.

5 Conclusion

In this paper we have argued that monetary and fiscal capacity are jointly determined. This fact has implications for both macroeconomics and development.[24] In macroeconomics, higher taxes and monetary expansions are generally modeled as substitutes (Mankiw, 1987; Sargent and Wallace, 1981; Sims, 1994). In contrast, we argue that, in the long run, they are complements: monetization is a precondition for the building of fiscal capacity, and, in turn, countries with high fiscal capacity are capable of building more efficient monetary systems. This co-dependence also suggests that, in the context of undermonetized economies, money is not neutral, even in the long run.

The descriptive and theoretical arguments of our paper are supported by causal estimates of the effects of monetary capacity on fiscal capacity. For this purpose, we rely on what was arguably the most significant exogenous shock to monetary capacity in history: the inflow of silver and gold from the Americas. Our rich data for England, France and Spain allows us to pursue an instrumental variable approach, where we instrument money supply through the production of silver and gold in American mines. We find that a one percent increase in the real per capita money stock led to a 0.5-1% increase in fiscal capacity over a decade. Importantly, this effect did not diminish over the course of subsequent decades. Thus, the level of monetization significantly affected the real economy in the long run. Our historical discussion suggests that the cases of England, France and Spain are far from exceptional: complementarity between monetary and fiscal capacity has been the rule

[24] It was previously known that more advanced economies tend to have more advanced monetary and financial systems, but the direction of causality has been a matter of debate (Levine, 2005).

across the ancient, medieval and early modern world.

For economic development, our results relate to the growing emphasis on the role of state capacity, both as a force which historically allowed Western European countries to surge ahead, and as a constraint in developing countries today.[25] We highlight one neglected cost imposed on societies where the state is too weak: the inadequate provision of a liquid means of exchange, which hinders the collection of taxes, increases transaction costs, and undermines specialization and market participation. For the empirical evidence, we rely on the premodern period, as it allows identifying long-run trends, and provides a natural experiment for establishing the causal effects. The argument, however, is more general, and relevant for undermonetized economies both historical and modern.

[25]For a sample of the literature which considers the links between state-building and economic growth, see Besley and Persson (2011), Brewer (1988), Dincecco (2009), O'Brien (2011), and Rosenthal (1990).

References

Abad, Leticia Arroyo, and Nuno Palma. 2021. The Fruits of El Dorado: The Global Impact of American Precious Metals. In *The Fruits of the Early Globalization: An Iberian Perspective*, edited by Doblado, R., and A. Garcia-Hiernaux. Palgrave.

Allen, Martin. 2001a. The Volume of the English Currency, 1158-1470. *The Economic History Review* 54(4): 595–611.

Allen, Robert C. 2001b. The Great Divergence in European Wages and Prices from the Middle Ages to the First World War. *Explorations in Economic History* 38(4): 411–447.

Barrett, W. 1990. World bullion flows, 1450-1800. In *The Rise of Merchant Empires*. Cambridge University Press.

Besley, Timothy, Ethan Ilzetzki, and Torsten Persson. 2013. Weak states and steady states: The dynamics of fiscal capacity. *American Economic Journal: Macroeconomics* 5(4): 205–35.

Besley, Timothy, and Torsten Persson. 2009. The origins of state capacity: Property rights, taxation, and politics. *American Economic Review* 99(4): 1218–44.

Besley, Timothy, and Torsten Persson. 2011. *Pillars of Prosperity*. Princeton University Press.

Blanchard, I. 1989. *Russia's Age of Silver: Precious-Metal Production and Economic Growth in the Eighteenth Century*. Routledge.

Bonney, Richard, editor. 1995. *Economic systems and state finance*. Clarendon Press Oxford.

Bransbourg, Gilles. 2015. *Fiscal Regimes and the Political Economy of Premodern States*, chapter The later Roman Empire. Cambridge.

Brecke, Peter. 1999. Violent conflicts 1400 AD to the present in different regions of the world. In *1999 Meeting of the Peace Science Society*.

Brewer, J. 1988. *The Sinews of Power: War, Money, and the English State, 1688-1783*. Routledge.

Broadberry, Stephen, Bruce MS Campbell, Alexander Klein, Mark Overton, and Bas van Leeuwen. 2015. *British economic growth, 1270–1870*. Cambridge University

Press.

Brzezinski, Adam, Felix Ward, Yao Chen, and Nuno Palma. Forthcoming. The vagaries of the sea: Evidence on the real effects of money from maritime disasters in the Spanish Empire. *Review of Economics and Statistics*.

Catao, Luis A.V., and Marco E. Terrones. 2005. Fiscal Deficits and Inflation. *Journal of Monetary Economics* 52(3): 529–554.

Coman, Francisco, and Bartolom Yun-Casalilla. 2012. Spain: From Composite Monarchy to Nation-State, 1492-1914. An Exceptional Case? In *The Rise of Fiscal States: A Global History, 1500-1914*, edited by O'Brien, Patrick K., and Bartolom Yun-Casalilla, 233–266. Cambridge University Press.

Costa, L. C., A. Henriques, and N. Palma. 2020. Portugal's early modern state capacity: a comparative approach.

Danila, R. 2006. Credibility and Gradualism in a Monetary Reform: Token Copper Coins in the 17th-Early 18th Century Russia. In *Meeting papers, Ronald Coase Institute*.

De Vries, J, and A Van der Woude. 1997. *The First Modern Economy: Success, Failure and Perseverance of the Dutch Economy, 1500-1815*. Cambridge University Press.

Depeyrot, Georges. 2020. Monetary System of the Ancient Régime (Third to Eighteenth Centuries). *Handbook of the History of Money and Currency* 107–158.

Dincecco, Mark. 2009. Fiscal Centralization, Limited Government, and Public Revenues in Europe, 1650-1913. *The Journal of Economic History* 69(1): 48–103.

Drelichman, Mauricio, and Hans-Joachim Voth. 2008. Institutions and the resource curse in early modern Spain. In *Institutions and Economic Performance*, edited by Helpman, Elhanan, 120–47. Cambridge University Press.

Duncan-Jones, Richard. 1998. *Money and government in the Roman Empire*. Cambridge University Press.

Glassman, D, and A Redish. 1985. New Estimates of the Money Stock in France, 1493-1680. *The Journal of Economic History* 45(1): 31–46.

Goldsmith, Raymond W. 1987. *Premodern Financial Systems: A Historical Com-*

parative Study. Cambridge University Press.

Harris, W., editor. 2008. *The Monetary Systems of the Greeks and Romans*. Oxford University Press.

Henriques, Antonio, and Nuno Palma. 2020. Comparative European institutions and the little divergence, 1385-1800 CEPR Discussion Paper No. DP14124 (v.3).

Hoffman, Philip T, and Kathryn Norberg. 2002. *Fiscal crises, liberty, and representative government, 1450-1789*. Stanford University Press.

Hopkins, Keith, et al.. 2009. The political economy of the Roman empire. In *The dynamics of ancient empires: State power from Assyria to Byzantium*, edited by Morris, Ian, and Walter Scheidel, 178–204. Oxford University Press.

Irigoin, M. A. 2009a. The End of A Silver Era: The Consequences of the Breakdown of the Spanish Peso Standard in China and the United States, 1780s-1850s. *Journal of World History* 20(2): 207–244.

Irigoin, M. A. 2009b. Gresham on Horseback: The Monetary Roots of Spanish American Political Fragmentation in the Nineteenth Century. *The Economic History Review* 62(3): 551–575.

Irigoin, M. A. 2013. A trojan horse in Daoguang China? Explaining the flows of silver in and out of China. *LSE Economic History Working Paper*.

Jordà, Òscar. 2005. Estimation and inference of impulse responses by local projections. *The American Economic Review* 95(1): 161–182.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2020a. The effects of quasi-random monetary experiments. *Journal of Monetary Economics* 112: 22–40.

Jordà, Òscar, Sanjay R Singh, and Alan M Taylor. 2020b. The long-run effects of monetary policy. *NBER Working Paper*.

Kahan, A, and R Hellie. 1985. *The Plow, the Hammer, and the Knout: An Economic History of Eighteenth-Century Russia*. University of Chicago Press.

Karaman, K. K, and S Pamuk. 2010. Ottoman State Finances in European Perspective, 1500-1914. *The Journal of Economic History* 70(3): 593–629.

Karaman, K Kivanc, and Şevket Pamuk. 2013. Different paths to the modern state in Europe: the interaction between warfare, economic structure, and political

regime. *American Political Science Review* 107(3): 603–626.

Karaman, K. Kıvanç, Şevket Pamuk, and Seçil Yıldırım-Karaman. 2020. Money and monetary stability in Europe, 1300-1914. *Journal of Monetary Economics* 115: 279 – 300.

Kohn, Meir G. 2005. The Organization of Government in Pre-Industrial Europe.

Kotilaine, J. T. 2005. *Russia's Foreign Trade and Economic Expansion in the Seventeenth Century: Windows on the World*. BRILL.

LeDonne, John P. 1991. *Absolutism and Ruling Class: The Formation of the Russian Political Order 1700-1825*. Oxford University Press.

Levi, Margaret. 1988. *Of rule and revenue*. Univ of California Press.

Levine, R. 2005. Finance and Growth: Theory and Evidence. In *The Handbook of Economic Growth*, edited by Aghion, P, and S Durlauf. North-Holland.

Mankiw, N.Gregory. 1987. The optimal collection of seigniorage: Theory and evidence. *Journal of Monetary Economics* 20(2): 327–341.

Mayhew, Nicholas J. 2013. Prices in England, 1170–1750. *Past & Present* 219(1): 3–39.

McEvedy, Colin, and Richard. Jones. 1978. *Atlas of world population history.* Penguin Books Ltd.

Montiel Olea, José Luis, and Mikkel Plagborg-Møller. 2021. Local projection inference is simpler and more robust than you think. *Econometrica* 89(4): 1789–1823.

Newey, Whitney K, and Kenneth D West. 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation. *Econometrica* 55(3): 703–708.

North, Douglass Cecil. 1981. *Structure and change in economic history.* Norton.

Ober, J. 2015. Fiscal Policy in Classical Athens. In *Fiscal Regimes and the Political Economy of Premodern States*, edited by Scheidel, W., and A. Monson. Cambridge.

O'Brien, P. K. 2011. The History, Nature and Significance of an Exceptional Fiscal State for the Growth of the British Economy. *Economic History Review* 64(2): 408–446.

O'Brien, Patrick, and Nuno Palma. 2020. Not an ordinary bank but a great engine of

state: the Bank of England and the British economy, 1694-1844. *CEPR discussion paper 15400*.

Palma, Nuno. 2018a. Money and modernization in early modern England. *Financial History Review* 25: 231–261.

Palma, Nuno. 2018b. Reconstruction of Money Supply Over the Long Run: The Case of England, 1270-1870. *Economic History Review* 71(2): 373–392.

Palma, Nuno. 2019. American precious metals and their consequences for early modern Europe. *Handbook of the History of Money and Currency* 1–20.

Palma, Nuno. 2022. The real effects of monetary expansions: evidence from a large-scale historical experiment. *The Review of Economic Studies* 89(3).

Patterson, Clair C. 1972. Silver stocks and losses in ancient and medieval times. *The Economic History Review* 25(2): 205–235.

Pomeranz, Kenneth, and Steven Topik. 1999. *The World that Trade Created: Society, Culture, and the World Economy, 1400-the Present*. M.E. Sharpe.

Riley, J. C, and J. J McCusker. 1983. Money Supply, Economic Growth, and the Quantity Theory of Money: France, 1650-1788. *Explorations in Economic History* 20(3): 274–293.

Rosenthal, Jean-Laurent. 1990. The Development of Irrigation in Provence, 1700-1860: The French Revolution and Economic Growth. *The Journal of Economic History* 50(3): 615–638.

Santiago Caballero, Carlos, Leandro Prados de la Escosura, and Carlos Álvarez Nogal. 2020. Growth recurring in preindustrial Spain: half a millennium perspective. EHES Working Paper.

Sargent, Thomas. 1982. The Ends of Four Big Inflations. In *Inflation: Causes and Effects*, 41–98. National Bureau of Economic Research.

Sargent, Thomas, and Neil Wallace. 1981. Some unpleasant monetarist arithmetic. *Quarterly Review* (fall).

Seljak, Anton. The Russian Monetary System from the Kievan Empire to 189. *Money Museum Working Paper*, translated by Graham Pascoe.

Sims, Christopher. 1994. A Simple Model for Study of the Determination of the

Price Level and the Interaction of Monetary and Fiscal Policy. *Economic Theory* 4(3): 381–99.

Sousa, Rita Martins de. 2006. *Moeda e metais preciosos no Portugal Setecentista (1688-1797)*. Lisboa: Imprensa Nacional Casa da Moeda.

Spasskii, IG. 1970. *Russkaia monetnaia sistema. Istoriko-numizmaticheskii ocherk (Russian Monetary System. Historical and Numismatic Study)*. Leningrad, Avrora Publ. (in Russian).

Spufford, Peter. 1988. *Money and Its Use in Medieval Europe*. Cambridge University Press.

Stock, James H., and Mark W. Watson. 2018. Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal* 128(610): 917–948.

TePaske, John. 2010. *A New World of Gold and Silver*. Brill.

Tilly, Charles. 1990. *Coercion, Capital, and European States, Ad 990-1992*. Basil Blackwell.

Velde, François R., and Warren E. Weber. 2000. A Model of Bimetallism. *Journal of Political Economy* 108(6): 1210–1234.

Vries, Peer. 2015. *State, economy and the great divergence: Great Britain and China, 1680s-1850s*. Bloomsbury Publishing.

Weber, Warren E. 2000. *International Data, 1810-1995*. Research Department, Federal Reserve Bank of Minneapolis.

Wojtowicz, G, A Wojtowicz, and S Piekarczyk. 2005. *A Monetary History of Poland*. Twigger.

Wrigley, Edward Anthony, Ros S Davies, RS Schofield, and JE Oeppen. 1997. *English population history from family reconstitution 1580-1837*. Cambridge University Press.

A The model

In this section, we construct a simple model that investigates the relationship between monetary and fiscal capacity. In a two-period, perfect information setup, we establish that a positive shock to fiscal capacity increases monetary capacity, because economic agents expect a lower future inflation tax. In a symmetric fashion, a positive shock to monetary capacity spills over to fiscal capacity, because it induces the state to invest in building the capacity to tax the money economy. Below, we introduce the model, solve for the equilibrium and the comparative statistics, and interpret the results.

A.1 Setup

Consider an economy inhabited by a multitude of private agents, which lasts for two periods. Aggregate income is equal to 1 in both periods. Preferences are linear in consumption, and in a public good g_s provided by the government. While consumption receives weight one in preferences, the public good receives weight α_s, a random variable which takes value $\alpha > 0$ with probability π, and value zero with probability $1 - \pi$. The public good can be interpreted as expenditure in defence, and the case $\alpha_s = \alpha$ as one in which the country must fight an external war in period s.

Income can be seen as the sum of the value added generated in a multitude of transactions taking place in each period. A share m_s of these transactions is conducted using money, as opposed to barter. As explained below, the share for period one is exogenously given, while the share for period two is determined by agents at the end of period one. We will refer to m_s as to the economy's degree of monetization.

The public good is produced using a decreasing returns to scale technology. To keep things simple, we assume a stark technology: up to γm_s units of the public good (where $0 < \gamma < 1$) can be produced at a one-to-one transformation ratio with the consumption good, while the marginal cost of producing additional units is prohibitively high. This implies that optimal production is zero if $\alpha_s = 0$, and γm_s if $\alpha_s = \alpha$. The assumption that maximum production is proportional to the monetary

economy ensures that public expenditures are proportional to public revenues. It is a simplifying assumption which does not affect the results qualitatively.

In each period, the government can impose an inflation tax i_s, inflicting a loss $i_s m_s$ on private agents and yielding revenues $\delta i_s m_s$.[1] We assume $1 > \delta > 1/\alpha$. By the first inequality, the inflation tax is inefficient, since it generates a higher cost on private agents than it generates revenues for the government. The second inequality ensures that, despite its inefficiency, the inflation tax is desirable as a last-resort source of revenues in case of war.

The government can also impose a tax on transactions. The key modelling assumption for taxation is that it was easier to tax money transactions than barter transactions and non-market economic activity. This assumption reflects the insight that monetized transactions were easier to monitor and assess (Bonney, 1995; Tilly, 1990). Furthermore, once collected, money taxes could be transported to the central treasury and redistributed back across the country, making fiscal centralization feasible. In contrast, in-kind taxes were primarily spent locally, often leaving the central government little effective say on their use (Vries, 2015, p. 124).

To capture the insight that a monetized economy is more convenient to tax, it is assumed that the government can impose the tax, t_s, only on money transactions. This strong assumption is made to simplify the model and can be relaxed to allow for taxation of barter, as long as it is more difficult than taxing the money transactions. The tax inflicts a loss $t_s m_s$ on private agents, generating an equal amount of revenues for the government. To keep the model simple, we assume that the inflation and transaction taxes are additive: they inflict a total cost $(t_s + i_s) m_s$ on private agents.

We have assumed that the transaction tax is more efficient than the inflation tax: while the former yields one unit of revenues per each unit paid by agents, the latter only yields δ units. This assumption is crucial. As further clarified below, it implies that the government strictly prefers to raise revenues through the transaction tax, which in turn implies that greater fiscal capacity creates the expectation of a lower

[1] With fiat money, the government imposed the tax by printing more of it. With commodity silver or gold money, states reminted old coins that were delivered to the mint with less silver or gold, paid the agents that delivered the coins with these new depreciated coins, and appropriated the difference.

inflation tax. This assumption could be micro-founded in a model in which the two taxable bases were made endogenous. To the extent that agents can choose not to conduct transactions in money in order to avoid the inflation tax (or to conduct them using a foreign currency), but they cannot choose not to conduct transactions at all in order to avoid the transaction tax, then the taxable base of the inflation tax would be more elastic than that of the transaction tax, making the former less efficient than the latter.[2]

Central to the model is the notion of monetary capacity. It refers to a state's capacity to issue and circulate money that is accepted by the public and it determines the monetization level of the economy. Monetary capacity depends on both the demand and the supply for money. On the demand side, to the extent that a government can commit to not imposing an inflation tax, agents will choose to conduct transactions in money rather than in kind, and monetary capacity will increase. In our model, money demand is endogenous, as private agents optimally choose the degree to which they rely on money for transactions (m_s). On the supply side, to the extent that a government has access to precious metals and can organize the production and circulation of money, monetary capacity will increase. In our model, the supply-side determinants of monetary capacity are exogenous, and captured by the parameter ϕ_μ.[3]

Agents choose monetization one period in advance, in accordance with their expectations about monetary capacity. Since the supply-side determinants of monetary capacity are exogenous, the key determinant of these expectations is the government's capacity to commit to a low future inflation tax. When this capacity is high, agents expect a low future inflation tax, or, in other words, they expect the monetary capacity to be high. They thus choose a high future level of monetiza-

[2]The argument relies on all types of transactions being subject to the transaction tax to some extent, including transactions in kind. While we have ruled this out in the model, it is of course true in reality.

[3]We model these supply-side determinants as exogenous for simplicity, and to be consistent with the empirical part of the paper. There, the identification comes from an exogenous increase in the European states' capacity to supply money, due to an increasing availability of precious metals from the New World. In reality, a state's capacity to supply money was not shaped solely by the availability of precious metals, but also depended on the state's capacity to organize the production of money, which itself was endogenous to fiscal capacity. If included in the model, such feedback from fiscal to monetary capacity would strengthen the mutual dependence between the two that the model identifies.

tion. In equilibrium, the government is able to commit to a low future inflation tax by investing in fiscal capacity, since it is against the government's interest to raise revenues through the inflation tax when it can do so through the more efficient transaction tax. Given this lag in decisions, monetization in period one (m_1) is exogenously given. In contrast, monetization in period two (m_2) is determined at the end of period one, based on the expectations about the monetary capacity.

More formally, agents set monetization in period two as

$$m_2 = \mu_2 \left(\underset{-}{i_2^e + t_2^e}, \underset{+}{\phi_\mu} \right),$$

where the function μ_2 denotes monetary capacity.[4] This monetary capacity function is increasing in ϕ_μ, the parameter capturing exogenous supply factors. In addition, the function is decreasing in the expected value of taxes on the monetary economy, $i_2^e + t_2^e$. These variables capture the demand factors described above: when the government can commit to a low future tax on the monetary economy, expectations will reflect this, resulting in a high expected monetary capacity and the choice of a high degree of monetization in period one. Note that, by assumption, monetary transactions are hit by the transaction tax as well as by the inflation tax. However, because the former tax is more efficient, in equilibrium, higher fiscal capacity in period two will result not only in a lower i_2, but also in a lower i_2+t_2, thus increasing monetary capacity.

The function μ_2 is assumed to take a value between zero and one, to be twice continuously differentiable, and to satisfy the condition: $\partial^2\mu_2/\left[\partial(i_2^e + t_2^e)\partial\phi_\mu\right] \leq 0$.[5] In words, if monetization increases due to exogenous shocks, it remains at least as responsive to changes in tax rates as it was before. Intuitively, the higher the share of transactions conducted using money, the more responsive the share should be to changes in the cost of holding money.[6]

[4]We use a greek letter to denote monetary capacity, to be consistent with the literature on fiscal capacity. This uses a greek letter, τ_t, to denote fiscal capacity, even though this is an endogenous variable. In the rest of our notation, we stick to the more general convention of using greek letters to denote parameters.

[5]The function μ_2 could be micro-founded using a standard shopping-time model.

[6]This assumption could be relaxed, since it is only sufficient (but not necessary) for our results. That the assumption is only sufficient for the results can be appreciated by inspecting condition

In each period, the transaction tax cannot exceed the government's current fiscal capacity. Fiscal capacity in period one is exogenously given. We represent it by the function $\tau_1 \underset{+}{(\phi_\tau)} \in (0,1)$, where the parameter ϕ_τ captures all the exogenous supply factors that may increase the government's fiscal capacity. Increasing fiscal capacity to $\tau_2 > \tau_1(\phi_1)$ costs $F[\tau_2 - \tau_1(\phi_1)]$, where F is a continuously differentiable, increasing and convex function. There is a maximum attainable level of fiscal capacity, $\overline{\tau} \in [\tau_1(\phi_1), 1]$. To ensure that the government's problem has an interior solution, we also assume $F'(0) = 0$ and $F'[\overline{\tau} - \tau_1(\phi_1)] = \infty$.

Finally, we assume $\delta[1 - \tau_1(\phi_1)] + \tau_1(\phi_1) - F[\overline{\tau} - \tau_1(\phi_1)]/m_1 > \gamma > \overline{\tau}$. The first inequality ensures that the government can always generate enough revenues (net of investment in fiscal capacity) to provide the optimal amount of the public good. The second inequality ensures that in the "war scenario", the government must resort to the inflation tax. This assumption narrows down the analysis to the case of interest.

A.2 Timing and government objectives

The timing of the game is as follows:

- Period one:
  - Nature sets α_1.
  - The government sets i_1, t_1 and τ_2.
  - Private agents set m_2. The public good is produced, payoffs realise.
- Period two:
  - Nature sets α_2.
  - The government sets i_2 and t_2.
  - The public good is produced, payoffs realise.

(14) below.

Instantaneous indirect utility in period s is equal to

$$v_s = 1 - (i_s + t_s)\, m_s + \alpha_s g_s,$$

while V_s denotes inter-temporal utility. The government maximises $V_2 = v_2$ with respect to i_2 and t_2 in period two, and $V_1 = v_1 + v_2^e$ with respect to i_1, t_1 and τ_2 in period one.

A.3 Equilibrium

We now derive the equilibrium of the model. We begin by stating the key relationship between endogenous variables:

Result 1. *Greater investment in future fiscal capacity (a higher τ_2) leads to a higher monetary capacity and future monetization (higher μ_2 and m_2).*

Proof. We solve for the equilibrium using backward induction. The government's problem in period two is

$$\begin{aligned} \max_{i_2, t_2} \;\; & 1 - (i_2 + t_2)\, m_2 + \alpha_2 g_2 \;\; \text{s.t.} \\ & g_2 \leq (\delta i_2 + t_2)\, m_2 \\ & t_2 \leq \tau_2. \end{aligned}$$

If $\alpha_2 = 0$, then the optimal expenditure is zero, and the solution to the government's problem is $i_2 = t_2 = 0$. If $\alpha_2 = \alpha$, then the optimal expenditure is γm_2. Given that the inflation tax is less efficient a source of revenues than the transaction tax, the government first sets the transaction tax to its highest possible level, and then finances the remaining expenditure through the inflation tax:[7]

$$t_2^* = \tau_2 \tag{6}$$

$$i_2^* = \frac{\gamma - \tau_2}{\delta}. \tag{7}$$

[7]Note that the constraint $t_2^* + i_2^* < 1$ is always satisfied. To see this, recall assumption $\delta(1 - \tau_1) + \tau_1 - F\left[\tau_2 - \tau_1(\phi_\tau)\right]/m_1 > \gamma$. Since $\tau_2 \geq \tau_1$ and $F \geq 0$, this assumption implies $\delta(1 - \tau_2) + \tau_2 > \gamma$, which can be re-arranged as $\tau_2 + (\gamma - \tau_2)/\delta < 1$, or $t_2^* + i_2^* < 1$.

At the end of period one, forward-looking agents set $i_2^e = \pi i_2^*$ and $t_2^e = \pi t_2^*$. They then set

$$m_2^* = \mu_2 \left(\pi \frac{\gamma - \tau_2 (1-\delta)}{\underset{-}{\delta}}, \underset{+}{\phi_\mu} \right) \tag{8}$$

Thus, μ_2 is increasing in τ_2, proving the result. □

Investment in future fiscal capacity guarantees that the government will finance a greater portion of future expenditure through the transaction tax, rather than the inflation tax. Since the former is less distortionary, this amounts to a lower future taxation of the monetary economy, or greater monetary capacity. The greater monetary capacity induces agents to increase future monetization. Thus, Result 1 implies that investment in fiscal capacity serves two purposes: It increases public revenues and is a commitment device that paves the way for an increase in monetary capacity. Since μ_2 is in equilibrium a function of τ_2, it will be helpful to henceforth write the function as $\mu \left(\underset{+}{\tau_2}, \underset{+}{\phi_\mu} \right)$, which is increasing in both arguments.

Consider now the government's problem in period one. Once future choices (6)-(8) are taken into account, the expected period two instantaneous utility, v_2^e, is only a function of τ_2. Then, the government's problem is

$$\begin{aligned} \max_{i_1, t_1, \tau_2} \quad & 1 - (i_1 + t_1) m_1 + \alpha_1 g_1 + v_2^e (\tau_2) \quad \text{s.t.} \\ & g_1 + F [\tau_2 - \tau_1 (\phi_\tau)] \leq (\delta i_1 + t_1) m_1 \\ & t_1 \leq \tau_1 (\phi_\tau) . \end{aligned}$$

By an earlier assumption, the government can always raise enough revenues to pay for both the optimal expenditure on the public good and for the optimal investment in fiscal capacity. It follows that the optimal investment problem can be solved as the stand-alone problem,

$$\begin{aligned} \max_{\tau_2} V_1 (\tau_2) &= k - \lambda F [\tau_2 - \tau_1 (\phi_\tau)] + \pi \left[\alpha\gamma - \left(\frac{\gamma - \tau_2}{\delta} + \tau_2 \right) \right] \mu_2 (\tau_2, \phi_\mu) \\ &= k - \lambda F [\tau_2 - \tau_1 (\phi_\tau)] + \pi \frac{(\delta\alpha - 1)\gamma + (1-\delta)\tau_2}{\delta} \mu_2 (\tau_2, \phi_\mu) , \end{aligned}$$

where k is an expression which does not depend on τ_2, and $\lambda \in \{1, 1/\delta\}$ is the opportunity cost of public revenues in period one.

The solution to this problem satisfies the first-order condition

$$\frac{\partial V_1(\tau_2)}{\partial \tau_2} = 0, \tag{9}$$

which can also be written as

$$\lambda F'[\tau_2 - \tau_1(\phi_\tau)] = \overbrace{\pi \frac{1-\delta}{\delta} \mu_2(\tau_2, \phi_\mu)}^{\text{I}} + \overbrace{\pi \frac{(\delta\alpha - 1)\gamma + (1-\delta)\tau_2}{\delta} \left(\frac{\partial \mu_2(\tau_2, \phi_\mu)}{\partial \tau_2} \right)}^{\text{II}}. \tag{10}$$

The second-order condition is

$$\frac{\partial^2 V_1(\tau_2)}{\partial (\tau_2)^2} < 0. \tag{11}$$

Equation (10) clearly illustrates the trade-off faced by the government. It equalises the marginal cost of investment in fiscal capacity, $\lambda F'[\tau_2 - \tau_1(\phi_\tau)]$ to its marginal return. As anticipated, the latter can be decomposed into two parts. First, investment in fiscal capacity increases future revenues (term I). This term is increasing in monetary capacity, which increases the future taxable base. Second, it allows the government to commit to a low future inflation tax, thus increasing monetary capacity (term II).

Exploring (10), one finds that there is always an internal solution to (9) and (11). However, the solution is not necessarily unique.[8] Thus, the solution to the optimal investment problem, τ_2^*, may have to be chosen between alternative solutions to (9) and (11). These complications are of no consequence for our purposes, since we are

[8]To see that the solution might not be unique, note that the marginal cost on the leoft-hand side of (10) is a standard marginal cost curve, which continuously increases from zero to infinity as τ_2 increases from τ_1 to $\overline{\tau}$. However, the marginal benefit on the right-hand side might also be increasing in τ_2, because term I is increasing, and the shape of term II depends on the sign of the second derivative of $\mu_2(\tau_2, \phi_\tau)$, which we have not made any assumption on. To see that an internal solution exists, note that the curve on the right-hand side of (10) is a continuous function, and it is positive and finite for both $\tau_2 = \tau_1$ and $\tau_2 = \overline{\tau}$. Given the properties of the curve on the left-hand side, the latter must cross the former at least once from below. At that point, both (9) and (11) are satisfied.

not interested in the solution per se. We are instead interested in the comparative statics of the solution, which, as will be clear momentarily, is the same for all solutions.

For completeness, we conclude by solving the rest of the government problem. If $\alpha_1 = 0$, then optimal expenditure on the public good is zero, implying that taxes are set to the minimum required to pay for investment in fiscal capacity. If $\alpha_1 = \alpha$, then

$$t_1^* = \tau_1(\phi_\tau) \tag{12}$$

$$i_1^* = \frac{\gamma + F\left[\tau_2^* - \tau_1(\phi_\tau)\right]/m_1 - \tau_1(\phi_\tau)}{\delta}. \tag{13}$$

After solving for the equilibrium (t_1^*, i_1^*, τ_2^*, m_2^*, t_2^* and i_2^*), we next turn to the comparative statics.

A.4 Comparative statics

In the short run, that is, taking monetary and fiscal capacity as given, monetary expansion (higher inflation) and higher taxes are clearly substitutes since they are alternative tools to generate revenues (see equations 6-7 and 12-13).[9]

In the long run, however, monetary and fiscal capacity are endogenous. The central result of the paper establishes that, in this long-run setup, monetary and fiscal capacity, and thus, monetary expansion and higher taxes, are complements.

Proposition 1. *An exogenous positive shock to monetary capacity (ϕ_μ) increases both monetary capacity (μ_2^*) and fiscal capacity (τ_2^*) in period two. Likewise, an exogenous positive shock to fiscal capacity (ϕ_τ) increases both monetary capacity and fiscal capacity in period two.*

[9]For example, suppose there was a temporary shock which increased fiscal capacity from τ_s to $\tau_s + \epsilon$. In response, the government would increase t_s by ϵ, and reduce i_s (by ϵ/δ).

Proof. Note first that

$$\frac{\partial^2 V_1(\tau_2)}{\partial\tau_2\partial\phi_\mu} = \pi\frac{1-\delta}{\delta}\frac{\partial\mu_2(\tau_2,\phi_\mu)}{\partial\phi_\mu} + \pi\frac{(\delta\alpha-1)\gamma+(1-\delta)\tau_2}{\delta}\left(\frac{\partial^2\mu_2(\tau_2,\phi_\mu)}{\partial\tau_2\partial\phi_\mu}\right) > 0$$

$$\frac{\partial^2 V_1(\tau_2)}{\partial\tau_2\partial\phi_\tau} = \lambda F''\left[\tau_2 - \tau_1(\phi_\tau)\right]\frac{\partial\tau_1(\phi_\tau)}{\partial\phi_\tau} > 0, \tag{14}$$

where the first inequality follows from the fact that, by assumption, $\partial^2\mu_2/(\partial\tau_2\partial\phi_\mu) = -\partial^2\mu_2/(\partial i_2^e\partial\phi_\mu) \geq 0$.

Consider any solution τ_2^* to (9) and (11). The total differential of $\partial V_1(\tau_2^*)/\partial\tau_2^* = 0$ with respect to ϕ_μ is

$$\frac{\partial^2 V_1(\tau_2^*)}{\partial\tau_2^*\partial\phi_\mu}d\phi_\mu + \frac{\partial^2 V_1(\tau_2^*)}{\partial(\tau_2^*)^2}d\tau_2^* = 0$$

$$\frac{d\tau_2^*}{d\phi_\mu} = -\frac{\frac{\partial^2 V_1(\tau_2^*)}{\partial\tau_2^*\partial\phi_\mu}}{\frac{\partial^2 V_1(\tau_2^*)}{\partial(\tau_2^*)^2}} > 0. \tag{15}$$

Given (15), since $\mu_2^* = \mu_2(\tau_2^*,\phi_\mu)$ is an increasing function of both its arguments, it immediately follows that $d\mu_2^*/d\phi_\mu > 0$. The total differential of $\partial V_1(\tau_2^*)/\partial\tau_2^* = 0$ with respect to ϕ_τ is

$$\frac{\partial^2 V_1(\tau_2^*)}{\partial\tau_2^*\partial\phi_\tau}d\phi_\tau + \frac{\partial^2 V_1(\tau_2^*)}{\partial(\tau_2^*)^2}d\tau_2^* = 0$$

$$\frac{d\tau_2^*}{d\phi_\tau} = -\frac{\frac{\partial^2 V_1(\tau_2^*)}{\partial\tau_2^*\partial\phi_\tau}}{\frac{\partial^2 V_1(\tau_2^*)}{\partial(\tau_2^*)^2}} > 0. \tag{16}$$

Given (16), it again follows from the properties of $\mu_2^* = \mu_2(\tau_2^*,\phi_\mu)$ that $d\mu_2^*/d\phi_\tau > 0$. □

To interpret this result, note first that an exogenous positive shock to monetary capacity (ϕ_μ) not only increases monetary capacity (μ_2^*) and actual monetization levels in the second period (m_2^*), but also the second-period fiscal capacity (τ_2^*) and actual tax collection (t_2^*). This impact works by easing tax collection and hence incentivizing the government to invest in future fiscal capacity. In the empirical

section, we test this prediction, relying on the exogenous positive shock to monetary capacity due to the discovery of silver and gold in the New World.

Secondly, in a symmetric fashion, the model predicts that an exogenous positive shock to fiscal capacity increases not only the second-period fiscal capacity and actual tax collection, but also monetary capacity and actual monetization levels in the second period. This impact works by lowering the inflation expectations of the private agents, and making them hold more money.[10]

Lastly, these two effects do not work independently, but rather reinforce and magnify each other. This mutual dependence does not necessarily imply a virtuous cycle of monetary and fiscal capacity building, however. Just as a positive exogenous shock to one is predicted to spill over to the other, so does a negative exogenous shock. Hence, the model predicts that monetary and fiscal capacity move together, but depending on the particulars of the sequence of exogenous shocks to the system, the trajectory of their co-movement might change course.

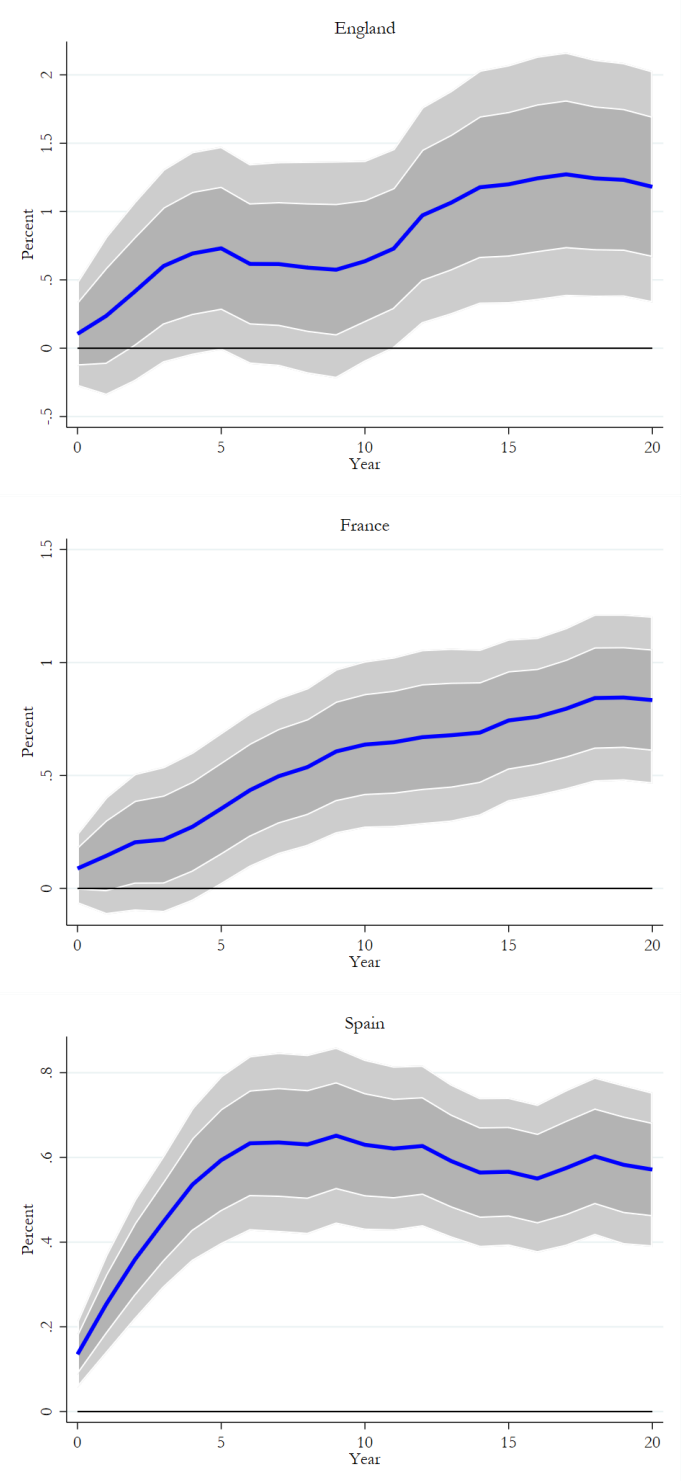
In the long run, since monetary and fiscal capacity are complements, so are monetary expansion and higher taxes. To see this, it is sufficient to note that an increase in monetary capacity implies a monetary expansion: It implies greater monetization, which will typically require a greater nominal money supply.[11]

[10]In this setup, the impact of fiscal capacity on monetary capacity works through the money demand channel. We do not formally model the potential impact through the money supply channel. If, as historical evidence suggests, a higher fiscal capacity affords states greater ability to produce and circulate money (Depeyrot, 2020; Karaman et al., 2020), this impact would complement the money demand channel, and result in a greater aggregate impact on monetization.

[11]In principle, an increase in the real money supply could also be obtained through a decrease in prices. In the early modern period, however, this was made difficult by a technological constraint: that it was difficult to put too little silver in coins.

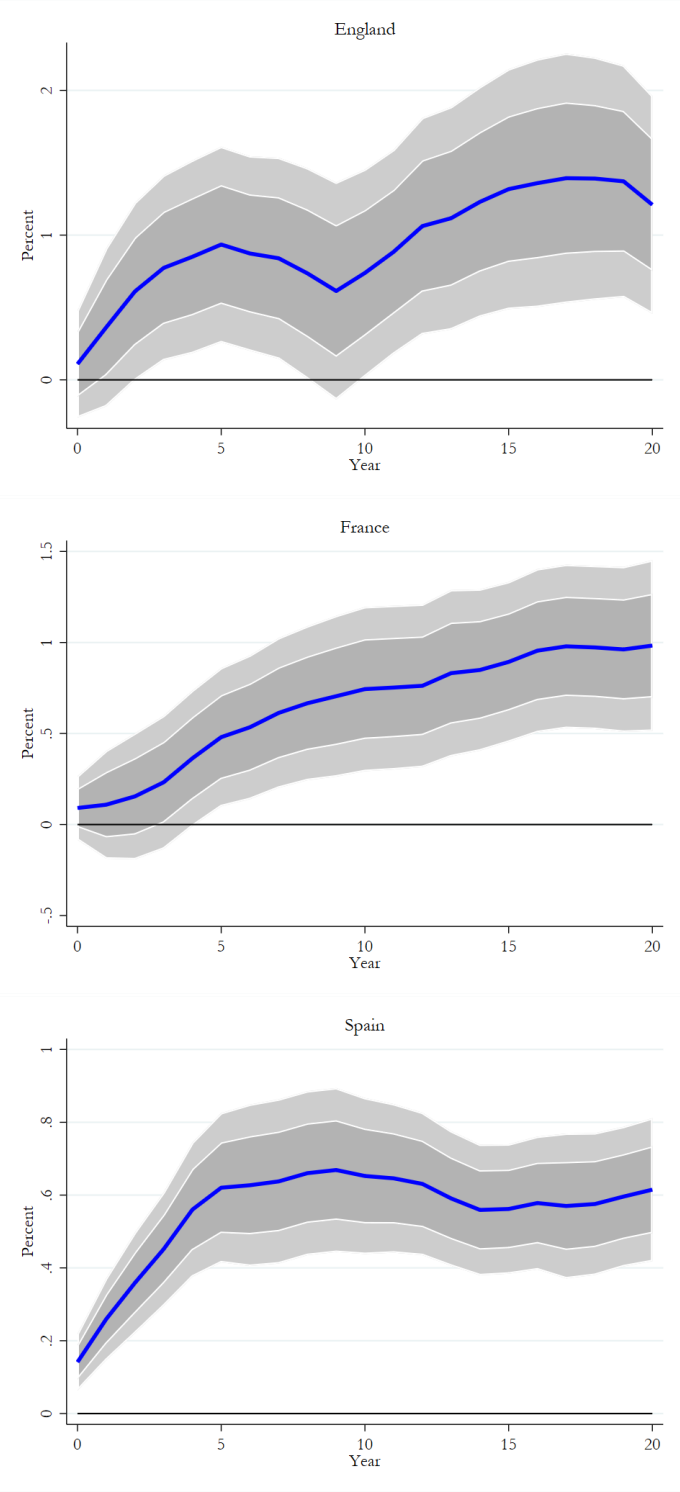
B Robustness

Figure A1: Robustness: contemporaneous lag of metals production



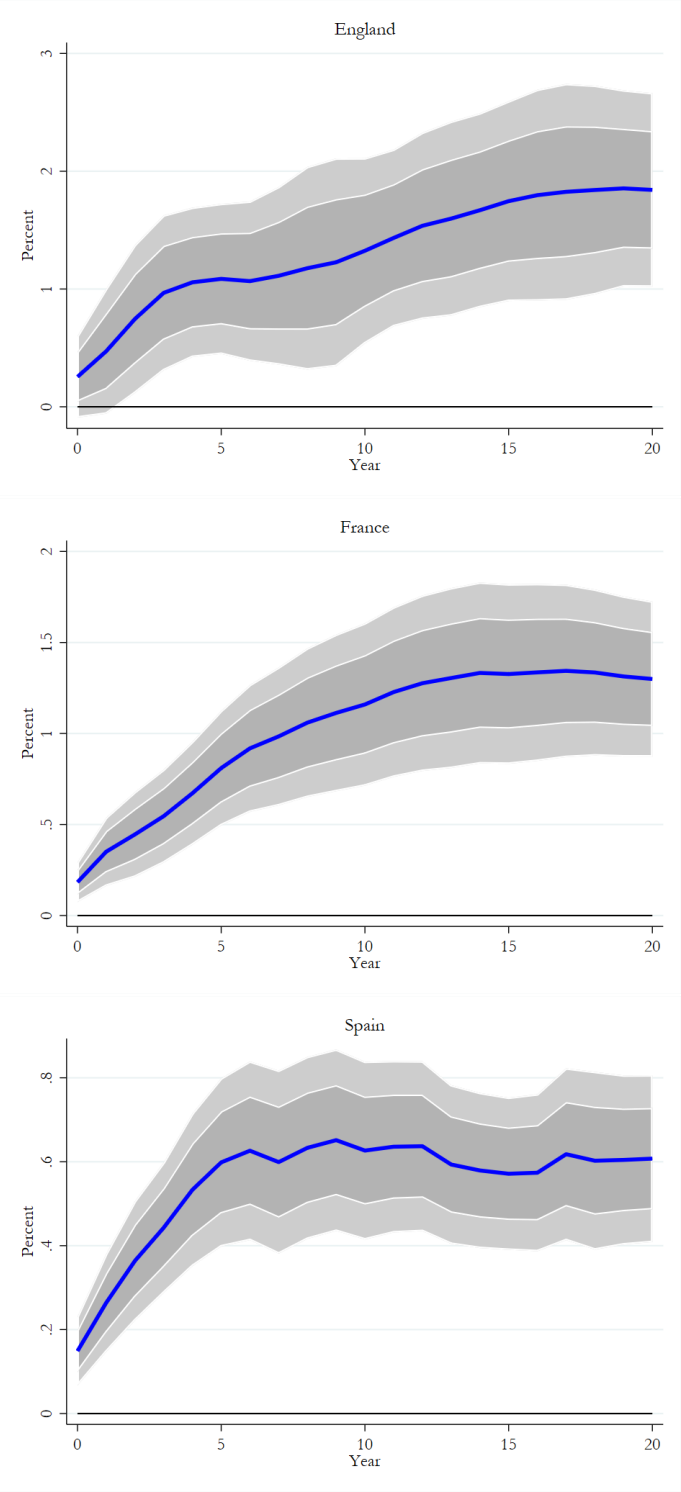
Note: LP-IV impulse responses, with contemporenous metals production (Column 2 of table 2).

Figure A2: Robustness: second lag of metals production



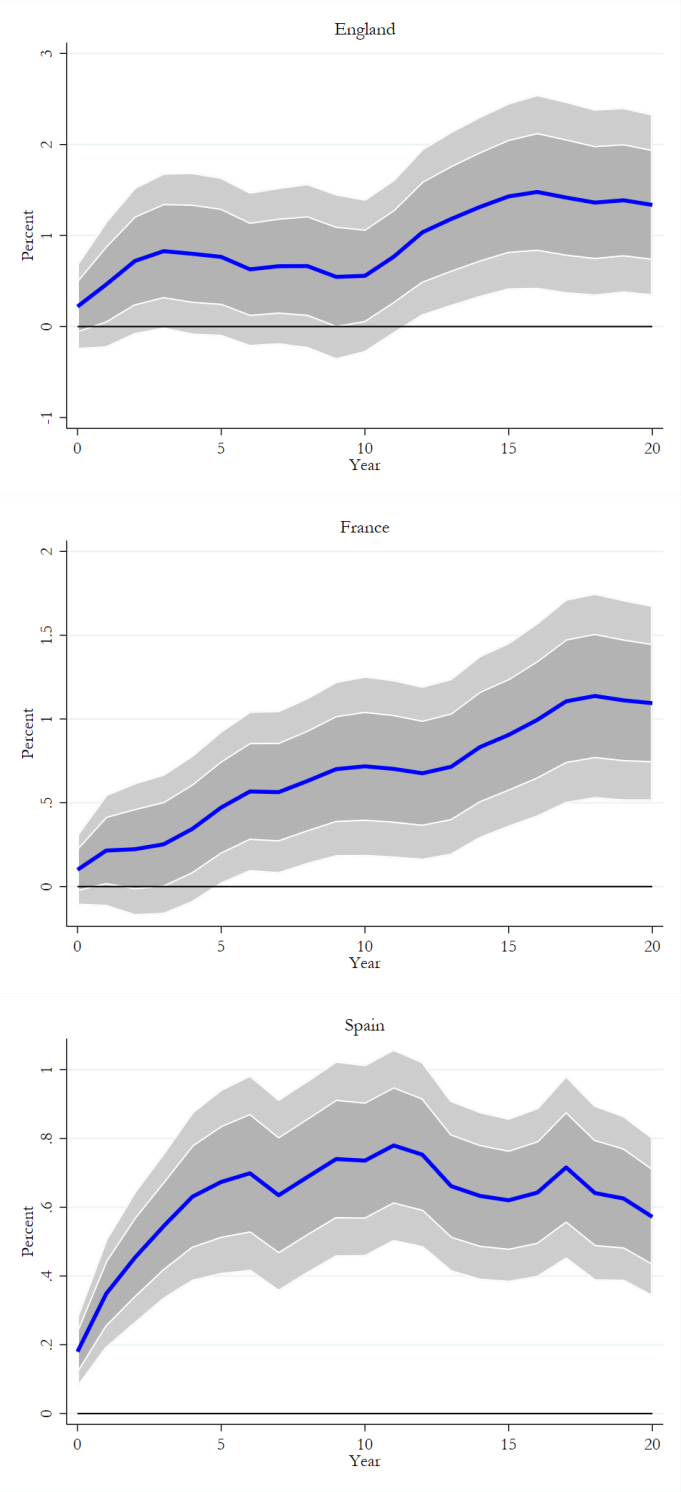
Note: LP-IV impulse responses, with second lag of metals production (Column 3 of table 2).

Figure A3: Robustness: no GDP control



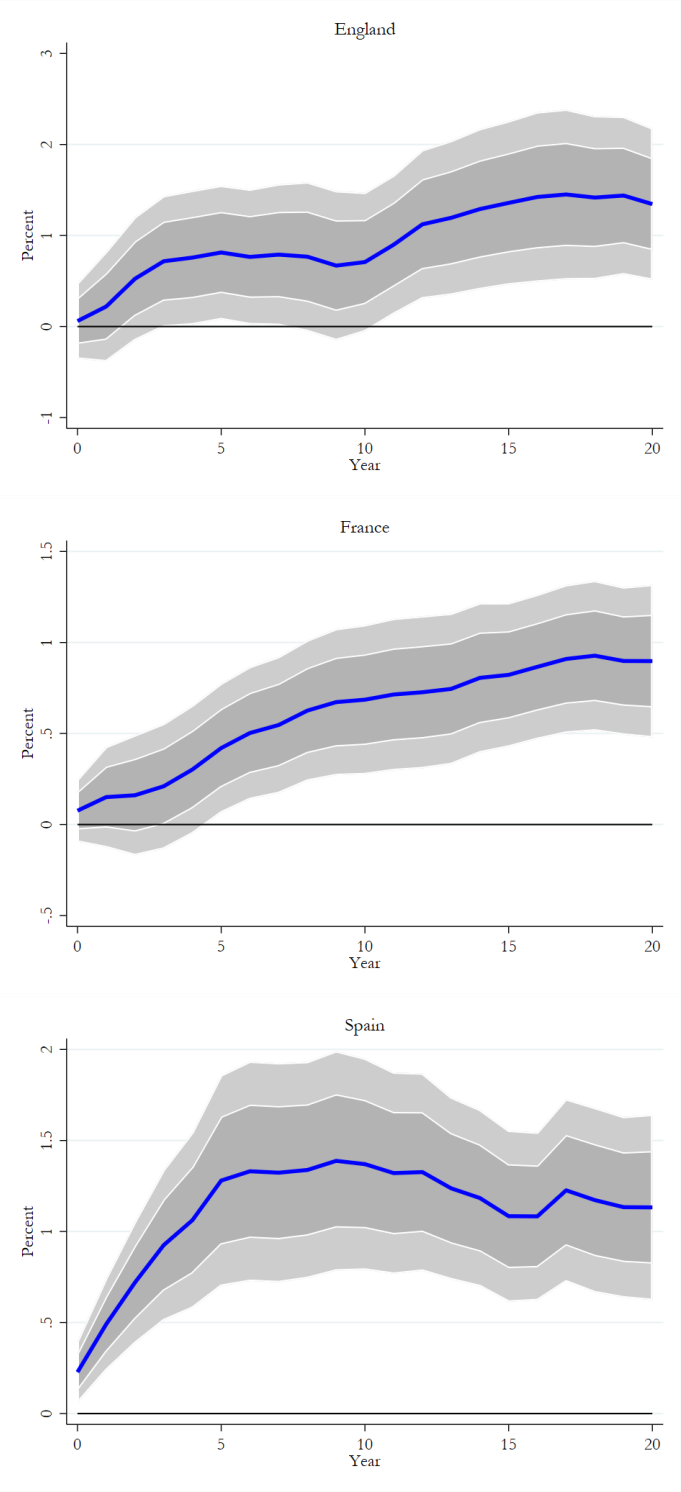
Note: LP-IV impulse responses, with no GDP control (Column 4 of table 2).

Figure A4: Robustness: war control



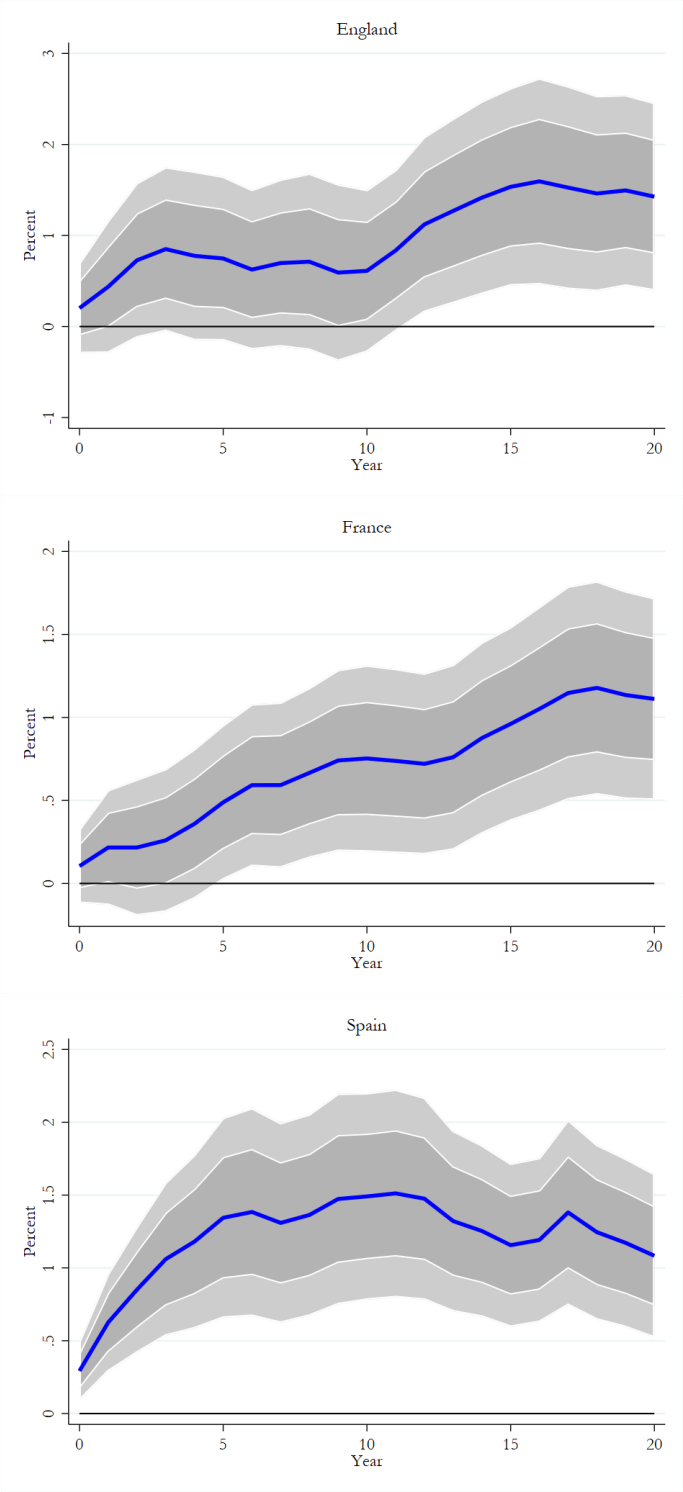
Note: LP-IV impulse responses, with additional war control (Column 5 of table 2).

Figure A5: Robustness: parliament control



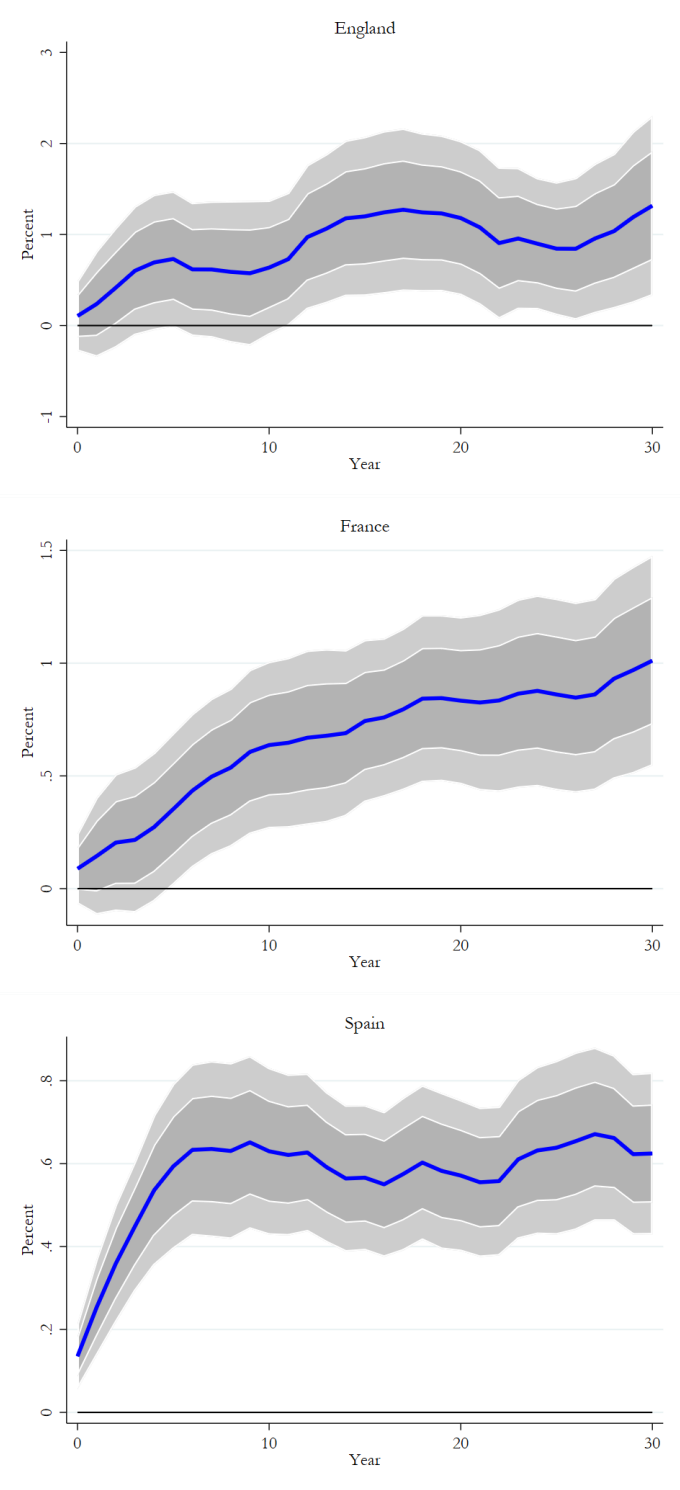
Note: LP-IV impulse responses, with additional parliament control (Column 6 of table 2).

Figure A6: Robustness: war and parliament control



Note: LP-IV impulse responses, with additional war and parliament controls (Column 7 of table 2).

Figure A7: Robustness: horizon up to 30



Note: Baseline LP-IV impulse responses, but with horizon up to $h = 30$.

The Vagaries of the Sea: evidence on the real effects of money from maritime disasters in the Spanish Empire*

Forthcoming in: *The Review of Economics and Statistics*†

Abstract

We estimate the effect of money supply changes on the real economy by exploiting a recurring natural experiment: maritime disasters in the Spanish Empire (1531-1810) which resulted in the loss of substantial amounts of silver money. We find that negative money supply shocks caused Spanish real output to decline. A transmission channel analysis highlights slow price adjustments and credit frictions as mechanisms through which money supply changes affected the real economy. Especially large output declines occurred in textile manufacturing against the backdrop of a credit crunch that impaired merchants' ability to supply their manufacturers with inputs.

Keywords: monetary shocks, credit channel, price rigidity, wage rigidity, local projection

JEL Codes: E43, E44, E52, N10, N13

*This chapter is co-authored with Yao Chen (Erasmus University Rotterdam), Nuno Palma (University of Manchester; Instituto de Ciências Sociais, Universidade de Lisboa; CEPR) and Felix Ward (Erasmus University Rotterdam).

†Brzezinski, Adam, Yao Chen, Nuno Palma, and Felix Ward. Forthcoming. The vagaries of the sea: evidence on the real effects of money from maritime disasters in the Spanish Empire. *Review of Economics and Statistics*. URL: https://doi.org/10.1162/rest_a_01223.

1. Introduction

The Columbian voyage of 1492 marked the beginning of three centuries in which vast amounts of monetary silver were shipped from America to Spain. During that time, Spain's money supply was subjected to the vagaries of the sea: maritime disasters that resulted in the loss of silver-laden ships gave rise to random contractions in the amount of money that arrived in Spain. We exploit this repeated natural experiment to obtain well-identified estimates of the causal effects of an exogenous money supply change.

To conduct the empirical analysis, we compile a novel dataset of maritime disasters in the Spanish Empire. For each maritime disaster, we collect data on the quantity of silver lost, the cause of the disaster, and the quantity of silver that was salvaged in the aftermath of the event. We assess the effect that these money supply shocks had on the real economy by estimating impulse response functions (IRFs) through local projections. The resulting IRFs describe the trajectory of macroeconomic variables in the aftermath of money supply shocks. We provide evidence that our estimates are not contaminated by correlated shocks using a series of robustness checks and diagnostic statistics.

We find that a negative shock to silver inflows amounting to 1% of the Spanish money supply led to a 1% drop in real output that persisted for several years. Our analysis highlights two important transmission channels through which money shocks affected economic activity: credit market frictions and nominal rigidities. The first is evidenced by a 1 to 2 percentage points increase in lending rates in response to the negative money shock. Contemporary merchant letters further indicate that the monetary contraction caused a credit crunch during which credit became severely rationed. With respect to the second channel, we find that aggregate prices responded only sluggishly to a monetary shock. Absent an instantaneous adjustment of prices, the scarcity of money weighed on the quantity of goods that were transacted. In contrast, nominal wages were generally more responsive, though

more so for unskilled than for skilled labor.

An analysis of disaggregated data indicates that the response of the Spanish economy was influenced by the *putting-out system*, in which raw materials and intermediate inputs were supplied to manufacturers by merchants. Merchants' ability to raise the working capital necessary to supply input goods to manufacturers was negatively affected by disaster-induced credit crunches. Consistent with this, input prices declined more markedly than output prices in the aftermath of maritime disasters. In addition, the textile industry, whose reliance on merchants for the provision of input goods was particularly pronounced, experienced particularly large output declines.

The first contribution of this paper is to provide well-identified, reduced-form estimates of the causal effects of money supply shocks. In doing so, we add to the body of evidence that sheds light on the interaction between money and the real economy based on historical monetary experiments (Velde, 2009; Palma, 2022). Beyond documenting a causal effect, our analysis draws on an extensive body of disaggregated data to indicate the transmission channels through which money affected the real economy. Our identification strategy is related to Koudijs (2015, 2016), who exploits weather-induced interruptions to shipping traffic across the English Channel in the early modern period to analyze how information flows affect stock market valuations. The strength of natural experiments lies in that they require few assumptions to identify causal effects. This luxury is seldom provided to studies of monetary policy, which usually are more demanding in this respect. Methodological examples include structural VARs (Christiano, Eichenbaum, and Evans, 1999; Uhlig, 2005; Bernanke, Boivin, and Eliasz, 2005; Coibion, 2012), estimated DSGE models (Ireland, 1997, 2003; Smets and Wouters, 2007), instrumental variable strategies (Jordà, Schularick, and Taylor, 2020) often applied in combination with high frequency data (Gertler and Karadi, 2015; Miranda-Agrippino and Ricco, 2018; Nakamura and Steinsson, 2018a), and narrative approaches to identifying monetary shocks (Friedman and Schwartz, 1963; Romer and Romer, 1989, 2004; Cloyne and Hürtgen, 2016). Comprehensive overviews of common identifica-

tion methods in macroeconomics are provided by Ramey (2016) and Nakamura and Steinsson (2018b).

The second contribution of this paper is to trace the real effects of money through its various transmission channels (e.g. Mishkin, 1995; Kuttner and Mosser, 2002; Auclert, 2019). In this regard, our findings underscore the relevance of nominal rigidities (Calvo, 1983; Christiano, Eichenbaum, and Evans, 2005; Nakamura and Steinsson, 2013; Gorodnichenko, Sheremirov, and Talavera, 2018), and credit frictions (Kiyotaki and Moore, 1997; Carlstrom and Fuerst, 1997; Bernanke, Gertler, and Gilchrist, 1999). That these two channels – so familiar to economists today – were already relevant in the early modern period points to their deep-rooted influence on macroeconomic fluctuations.

2. Money supply shocks and econometric method

This section works toward obtaining reduced-form estimates of the causal effects of money on the economy. It begins with a description of the Spanish monetary system in the early modern period and then presents our shock measure, data, and econometric strategy.

2.1. Money and precious metal inflows in early modern Spain

Money in early modern Spain consisted mainly of coins made of precious metals – above all silver (Palma, 2022).[1] In terms of functionality, early modern precious metal money is comparable to narrow money aggregates today. Other varieties of money existed, but precious metal coins were more widely accepted than their surrogates, such as banknotes (Nightingale, 1990). As late as 1875, gold and silver made up 85% of the Spanish money supply (Tortella et al., 2013, p.78). Our analysis therefore focuses on gold and silver coins, which we interchangeably name "money" in the following.

[1] Copper coins also played a role, but they were only practicable for small transactions (Sargent and Velde, 2002), and their prominence fluctuated over time (Motomura, 1994). Only for a few decades after 1617 did copper coins make up a significant share of Spain's overall money supply (Velde and Weber, 2000).

Spain's money supply was heavily influenced by the inflow of precious metals from America (Desaulty et al., 2011). Annual arrivals were large, ranging from 1% to 15% of the Spanish money stock (Chen et al., 2021). Precious metal inflows primarily constituted remittances and transfers of income from abroad.[2] Regulation required all arriving precious metals to be minted (Hamilton, 1934, pp.25,20). Over time, more and more of America's precious metal output was already minted in American mints (Céspedes del Castillo, 1996; Irigoin, 2020) and thus the vast majority of precious metals that arrived in Spain did so in coined form (Costa et al., 2013; de Paula Perez Sindreau, 2016, p.63).[3]

Public authorities at Imperial mints guaranteed the silver content of the coinage, but precious metal mines were owned and run by private entrepreneurs (Elliott, 2006, p.93). For much of the early modern period 85% to 95% of precious metal remittances from America were privately owned (García-Baquero González, 2003; Costa et al., 2013). Only in the late 18th century did the Crown's share begin to exceed 20%. The annual arrival of treasure ships was scheduled in advance (Chaunu et al., 1955), and publicly available prognoses of how much precious metals would arrive were on average correct (Palma, 2022).[4] Thus, maritime disasters led to the loss of money that Spain's private sector thought it possessed.[5]

[2]Only between one tenth and one third of precious metal inflows from America constituted payment for Spanish exports (based on export values from Phillips, 1990, p.82). The annual data on Spanish exports to America that becomes available from the late 18th century onward (Esteban, 1981) exhibits no correlation with Spain's precious metal inflows (Table B.1 in the Appendix). Also note that, because a typical round trip from Spain to the colonies took somewhat more than a year (Chaunu and Chaunu, 1977, pp. 229ff.), it is possible to account for any export correspondent of money inflows by controlling for lagged indicators of Spanish economic activity.

[3]Some silver shipments were insured (Baskes, 2013). To the extent that silver shipments were insured within the Spanish merchant community maritime disaster losses still constituted an aggregate shock to the Spanish money supply. Foreign insurers also existed, but no quantitative information exists about the extent to which they were involved in insuring Spanish silver shipments. Our finding that European lending rates beyond Spain's borders did not systematically increase in the aftermath of maritime disaster losses indicates that money losses were predominantly born by Spanish entities (see Figure B.14 in the Appendix).

[4]*Avisos* – small and fast dispatch boats – brought news about the amount of treasure a fleet carried (Martín, 1965, p.lxxxvii).

[5]By regulation, transatlantic business with Spanish colonies was restricted to Spanish merchants (Nogues-Marco, 2011, p.6). As a consequence, a Spanish entity initially owned the arriving precious metals, even if most of it eventually diffused across Europe.

2.2. Maritime disasters

We collect data on maritime disasters from the historical literature on Spanish precious metal shipments (Chaunu and Chaunu, 1977; Morineau, 1985; Mangas, 1989; Walton, 1994; Bonifacio, 2010), catalogs of sunken treasure ships (Potter, 1972; Marx, 1987), and `todoavante.es` which provides a description of treasure fleet voyages based on primary archival sources. The main variables we collect are the date of a disaster, its location, its cause, the amount of precious metals lost, and the amount of precious metals that was salvaged. We restrict our data collection to maritime disasters that resulted in the loss of monetary gold and silver destined for Spain. The resulting list of events is described in Table 1. A detailed listing of the sources for each individual disaster event is provided in Table A.1 in the Appendix.

In total we observe maritime disasters that produced money shocks to the Spanish economy in 33 out of 280 years.[6] The most frequent cause of maritime disasters was bad weather, such as hurricanes. Navigational errors rank second. A distant third reason for the loss of silver was capture by privateers. The most notable such event occurred in 1628, when the Spanish fleet, carrying 80 tonnes of silver, was captured by Dutch privateer Piet Heyn. Finally, in three instances silver-laden ships were destroyed in naval combat. In 1804, for example, the Spanish treasure ship *Nuestra Señora de las Mercedes* was engaged by British naval forces off the coast of Portugal. In the ensuing Battle of Cape Saint Mary the treasure ship exploded, resulting in the loss of more than 35 tonnes of silver.

We argue that these events constitute useful natural experiments. Bad weather as well as navigational errors were unrelated to economic conditions in Spain. Capture and combat admittedly were rooted in interstate conflicts that affected the Spanish economy in more ways than just through the influx of silver. Conditional on that, however, the capture and destruction of silver ships was driven by the random emergence of tactical opportunities, not the evolution of economic variables in Spain. Moreover, our results are robust to excluding conflict-based events. To test whether

[6] The accident rate in terms of ship numbers and ship tonnage was below 5% (Chaunu et al., 1955, Tome VI-1, pp. 869-871) – considerably lower than the fraction of disaster years.

Table 1: Maritime disaster events

Year	Silver loss	Cause of maritime disaster	Salvaged silver	Location	Month
1537	18,258 kg	capture	NA	Azores	Sep
1550	2,693 kg	navigational error	0 kg	Azores	NA
	5,386 kg	weather	0 kg	Havana	NA
1554	73,031 kg	weather	36,516 kg	Veracruz	Apr
1555	12,781 kg	weather	0 kg	Veracruz	Jan
1563	24,430 kg	weather	0 kg	Bermudas	Sep
1567	109,547 kg	weather	0 kg	Havana	Apr
1591	255,610 kg	weather	191,708 kg	Azores	Sep
1605	204,488 kg	weather	0 kg	Havana	Nov
1621	382 kg	weather	128 kg	Bermudas	Sep
1622	188,951 kg	weather	25,561 kg	Havana	Sep
1623	76,345 kg	weather	0 kg	Havana	Sep
1624	51,122 kg	weather	0 kg	Havana	Apr
1628	30,538 kg	navigational error	8,339 kg	Veracruz	Jul
	80,660 kg	capture	NA	Havana	Sep
1631	150,241 kg	weather	4,427 kg	Veracruz	Oct
	58,169 kg	navigational error	25,561 kg	Portobelo	Jun
1634	7,635 kg	navigational error	2,556 kg	Havana	Nov
1641	76,683 kg	weather	0 kg	Havana	Oct
1654	255,610 kg	navigational error	89,464 kg	Lima	Oct
1656	127,805 kg	navigational error	63,903 kg	Havana	Jan
	122,693 kg	combat	0 kg	Europe	Sep
	51,122 kg	capture	NA	Europe	Sep
1682	153,366 kg	weather	0 kg	Cartagena	May
1698	161,540 kg	navigational error	153,366 kg	Havana	Mar
1702	5,308 kg	combat	0 kg	Europe	Oct
	2,042 kg	capture	NA	Europe	Oct
1708	281,171 kg	combat	0 kg	Cartagena	Jun
	5,112 kg	capture	NA	Cartagena	Jun
1715	309,972 kg	weather	133,969 kg	Havana	Jul
1730	139,556 kg	weather	37,215 kg	Havana	Sep
1733	311,908 kg	weather	308,246 kg	Havana	Jul
1750	6,748 kg	weather	359 kg	Havana	Aug
	3,573 kg	capture	NA	Havana	Aug
1752	49,620 kg	weather	47,345 kg	Rio de plata	Jul
1753	38,134 kg	weather	0 kg	Lima	Jan
1762	62,895 kg	capture	NA	Europe	May
1786	185,716 kg	navigational error	178,847 kg	Europe	Feb
1800	51,264 kg	navigational error	24,641 kg	Lima	Nov
1802	13,742 kg	weather	0 kg	Havana	Oct
1804	36,645 kg	combat	0 kg	Europe	Oct
	76,439 kg	capture	NA	Europe	Oct

maritime disasters were in any way related to economic conditions in Spain we also look at the pre-disaster behavior of economic variables. The pre-event analysis

confirms that maritime disasters were not preceded by systematic fluctuations in the Spanish economy (Figure B.13 in the Appendix).

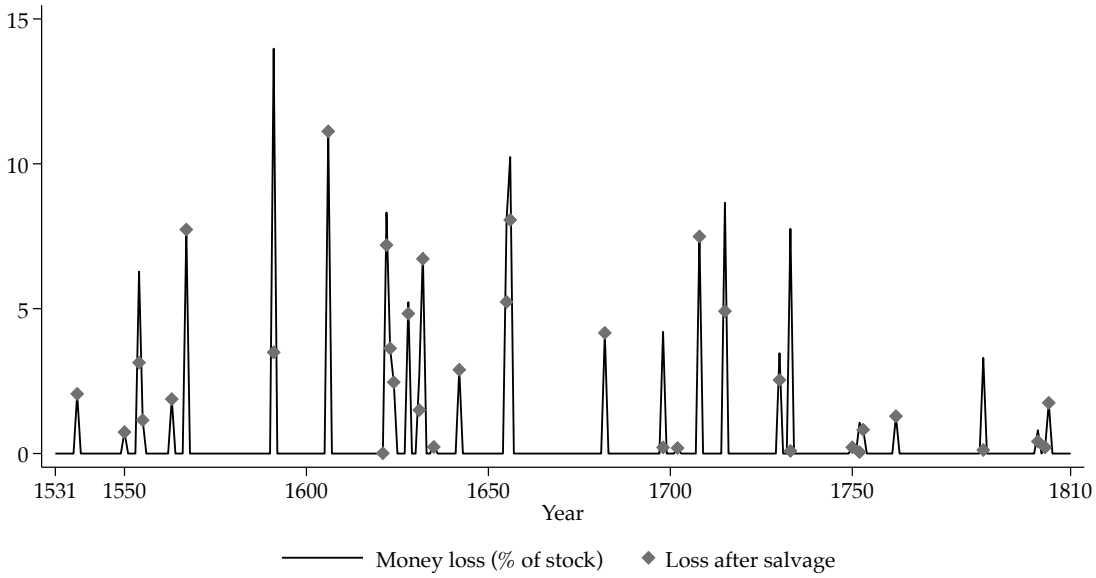
Clean identification requires that our IRF estimates are not contaminated by the influence of correlated shocks. Two potential concerns in this regard are the non-monetary losses and fiscal losses that were associated with maritime disasters. In Appendix A.2 we show that neither can account for the response of Spain's economy to maritime disaster events, suggesting that our IRF estimates indeed describe monetary effects. Importantly, we show that maritime disasters on outbound voyages, in which no monetary metals were lost, had no effect on Spanish economic outcomes. This finding moderates concerns that maritime disasters exerted their effect in non-monetary ways, such as through the loss of ships, lives, and non-money cargo.[7] We also show that our results are not driven by the loss of Crown revenues, which typically amounted to significantly less than one fifth of money arrivals (García-Baquero González, 2003; Costa et al., Forthcoming). While maritime disasters could trigger sovereign debt crises, Spain's early modern public sector was too small for this to generate noticeable spillover effects into private sector lending rates or aggregate economic outcomes.

To arrive at our money supply shock measure, we divide the absolute silver losses by the money stock level provided in Chen et al. (2021). More concretely, we express each silver loss as a fraction of the preceding year's money stock.[8] The timing of the shock to the Spanish economy furthermore depends on the location where the maritime disaster took place. For example, it took about 3 months for precious

[7] More generally, the value of non-monetary losses – ships and cargo – was small compared to the value of lost precious metals, and the number of lives lost in maritime disaster events never amounted to a significant fraction of the labor force in Spain. Table A.3 in the Appendix provides a quantitative overview of the non-monetary losses associated with maritime disasters.

[8] Absent annual data on Spanish money in- and outflows vis-à-vis other countries it is impossible to determine how shocks to Atlantic money inflows affected Spain's overall money growth rate. International monetary models, such as Hume's price-specie flow model (Hume, 1752), or the monetary approach to the balance of payments (Flynn, 1978; Frenkel and Johnson, 2013) predict that Spanish money losses eventually diffused across Spain's borders. Fragmentary trade data indicates that Spain indeed imported less from abroad in the aftermath of a contractionary money shock, but it did not succeed in increasing its exports (see section B.1 in the Appendix). The prolonged response of Spanish prices suggests that loss diffusion did not occur quickly (see section 3.2.2). In fact, for the majority of prices a sustained recovery only set in five years after the shock, indicating that Spain's money stock acted as a buffer between precious metal inflows from America and other European countries' money stocks. This notion is further supported by the lack of a systematic lending rate response beyond Spain's borders (see Figure B.14 in the Appendix).

Figure 1: Monetary shock measure



metals loaded in Portobelo (Panama) and Cartagena (Colombia) to arrive in Spain. Fourth quarter maritime disasters in this area thus constituted shocks to Spain's money supply one year ahead. We use the detailed information on return shipping times from Chaunu and Chaunu (1977, pp. 229ff.) to determine the timing of the shocks.[9] The resulting shock measure is depicted in Figure 1. Grey markers indicate the money loss excluding salvaged amounts. On average, maritime disaster shocks resulted in a loss of money that amounted to 4.1% of the Spanish money stock. The shock's size ranged from 0.02% to 14%, with a median value of 3.1%.[10]

2.3. Outcome variables

We analyze two types of outcome variables: First, variables that describe real economic output. Second, variables that describe channels through which money shocks might have affected real economic output in the early modern Spanish economy.

[9] We apply the following travel times to Spain: European coasts: 0 months, Azores: 0.5 months, Bermudas: 1.5 months, Havana: 2 months, Cartagena: 2.75 months, Portobelo: 3 months, Veracruz: 4 months, Lima: 4 months, Rio de plata: 4 months. We attribute each maritime disaster to its most recent place of departure. E.g. to a maritime disaster occurring on the Veracruz-Havana leg of the New Spain fleet we attribute a 4 month travel time. A robustness check in which we assume that travel times lasted one month longer produces very similar results (Figure B.10 in the Appendix). The exact months for the 1550 disasters are unknown. We set them to July and September according to the disaster locations and the typical return voyage schedules reported by Chaunu and Chaunu (1977, pp.229ff.).

[10] Our money shock measure resembles a (negative) helicopter drop in that it directly altered the level of monetary assets held by the private sector.

Real economic output. Economic historians have recently rebuilt early modern national accounts for many countries using large amounts of data from sources such as probate inventories and the account books of monasteries, universities, and hospitals. In this vein, Álvarez-Nogal and Prados de la Escosura (2013) have compiled an annual time series on Spanish real output that covers our sample period.[11] This series is based on a demand function approach that combines data on incomes (wage rates and land rents) and goods prices (agricultural and manufacturing output prices) into an estimate of real GDP.[12] We use this series to obtain an impression of how Spain's aggregate economic activity responded to money supply shocks.

While recent efforts by economic historians have produced much improved macro aggregates for the early modern period, their construction is based on stronger assumptions than that of their modern equivalents.[13] To address this concern our analysis includes a wide range of disaggregated data which originates from the account books of companies, notarial deeds, and the tax records of cities and guilds. This data is region-, sector-, and period-specific (e.g. 16th century Segovian textile production). We use these fragments to broaden the evidence base and throw additional light on the sectoral response of the Spanish economy.

First, Phillips and Phillips (1997, Appendix I) provide annual data on wool production approximated by the size of the flock of sheep. From the beginning of our sample up to 1563 the number of sheep under the jurisdiction of the *Mesta* – a Castilian livestock owners' association – was counted annually to assess membership fees.[14] After that, the annual count was replaced by a rough estimate, until in the early 1600s annual counts resumed for two decades (Klein, 1920, p.27).[15] Through-

[11] While the original publication presents 11-year moving averages the authors have kindly shared the underlying annual data with us.

[12] The baseline GDP estimate by Álvarez-Nogal and Prados de la Escosura (2013) also incorporates urbanization rate data that is interpolated over several decades. As this interpolated data is uninformative with respect to our analysis of short-run monetary effects we remove it from their GDP estimate. This leaves the shape and statistical significance of the real output response unaffected, but increases its absolute size by around one third.

[13] For details about the reconstruction procedures see de Jong and Palma (2018). For details about the typical data sources used see Palma (2020).

[14] While the Mesta flock size data is dominated by sheep, it also includes cows, horses, goats, and swines. The time series by Phillips and Phillips (1997) adds the latter animals to the overall Mesta flock size number according to the following key: 1 cow = 1 horse = 6 sheep; 1 goat = 1 swine = 1 sheep.

[15] The flock size was usually counted during hibernation in Extremadura. The annual flock size given by the sources thus represents beginning-of-year values. We shift this series one year forward

out, the *Mesta* accounted for a significant fraction of total Spanish wool production (20-50% according to Phillips and Phillips, 1997, p.291).

Second, García Sanz (1977) has compiled an annual time series on the production of fine cloths in Segovia. Segovia was Spain's most important textile manufacturing center (Massana, 1999, p.36). At times, close to 50% of its population worked in textile manufacturing (Le Flem, 1976, p.531). García Sanz' series spans almost a century, from 1699 to 1790, providing us with extensive insight into how one of early modern Spain's most prominent manufacturing industries responded to money supply shocks.

Third, the tax records of the city of Toledo provide another data source that is informative with respect to economic activity in Spain's textile manufacturing sector. Toledo was a key transit hub for the distribution of textiles (Montemayor, 1996, p.239), and each year it auctioned off the right to collect taxes on cloths entering the city. Most textiles came to Toledo for final finishing and subsequent re-export: Segovian textiles, for example, passed through Toledo before reaching their Southern customers. Similarly, many Valencian textiles first arrived in Toledo before continuing to their Northern customers. The Toledean tax data is therefore informative with respect to activity in Spain's textile manufacturing sector more generally. Montemayor (1981) has compiled the existing annual data on Toledean tax auction revenues for coarse and fine cloths (1540–1660).[16] The resulting series complement the textile output series by García Sanz, and enable us to evaluate the reaction of Spanish textile manufacturing for an earlier time period and a wider production region. Montemayor (1981) furthermore provides an annual tax revenue series that covers foodstuffs sold in the Toledean market.[17] We use this series to throw additional light on the behavior of primary goods consumption in the after-

to facilitate the interpretation of our impulse response function results.

[16] We deflate the two nominal auction revenue series with a Spanish cloth price index. We calculate this cloth price index as an equi-weighted average of the cotton, hemp, twine, and linen cloth price series provided by Losa and Zarauz (2021) and Allen (2001). The nominal tax receipt series produce almost identical IRF estimates, as the response of cloth prices to money supply shocks was moderate.

[17] We deflate the foodstuffs tax revenue series with a food-heavy consumer price index for New Castile from Reher and Ballesteros (1993). We obtain very similar IRF results when deflating the foodstuffs tax series with an equi-weighted index of all the Madrilenian food price series provided by Losa and Zarauz (2021).

math of money shocks.[18]

Finally, Odriozola Oyarbide (2002) has compiled an annual time series on Basque ship building that covers our entire sample period. This time series summarizes information from the purchase contracts and notarial deeds that have survived in Basque archives. It misses smaller ships that have been commissioned on a handshake basis without notarial help. The Basque region was an important shipbuilding hub in early modern Spain. The time series by Odriozola Oyarbide is therefore indicative of overall Spanish ship production. A caveat pertains to the analysis of this series given the nature of the money shock we analyze: to the extent that ship owners wanted to replace the ships they had lost in maritime disasters, the shipbuilding response mixes monetary effects with an increase in replacement demand.

Transmission channels. A multitude of price and wage series exist for early modern Spain. They allow us to trace the nominal adjustment of Spain's economy to money shocks in considerable detail. On the aggregate level, Álvarez-Nogal and Prados de la Escosura (2013) provide consumer price and wage series that cover our entire sample period. These series form the data basis for our headline price and wage IRFs. We further unpack Spain's nominal adjustment by drawing on disaggregated price and wage series.

There exist a large number of goods-level price series for early modern Spain – more than is practical to analyze individually (e.g. Hamilton, 1934, 1947; Losa and Zarauz, 2021). We therefore aggregate the available price series into equi-weighted price indices that cover various product categories of interest (e.g. animal products, threads, coarse and fine cloths). Table A.4 in the Appendix describes the sources, the series, and how we combine them into product-category price indices.

Disaggregated wage series are somewhat less plentiful than disaggregated price series (e.g. Hamilton, 1934; Feliu, 1991; Losa and Zarauz, 2021). However, enough data exists to distinguish between important occupational categories (e.g. skilled

[18] To the extent that tax farmers were risk averse and money shocks increased uncertainty about tax revenues, tax farmers' bids may incorporate an additional risk premium in the aftermath of maritime disasters. Part of any initial tax auction return decline may thus reflect a higher risk premium rather than lower expected tax revenues. Such a risk premium is likely to diffuse once the dust has settled one to two years after the shock, leaving the validity of the IRF estimate for all subsequent horizons unaffected.

versus unskilled labor, primary sector versus secondary sector). Table A.5 in the Appendix describes sources, series, and their aggregation into wage indices for different occupational groups. Besides throwing additional light on how money shocks worked their way through Spain's economy, the disaggregated price and wage series also allow us to draw some conclusions about the types of nominal frictions that were present.

Time series that convey information about credit market conditions are more difficult to obtain. Usury laws led to the hiding of interest payments. Additionally, many original sources were lost and no longer exist (Homer and Sylla, 1991; Pike, 1966, p. vi). Fortunately, lending rates left their mark in exchange rates – more particularly in the prices of bills of exchange that were systematically quoted in financial markets throughout Europe.

The exchange rate embodied in a bill of exchange differs from the spot exchange rate in that it describes the amount of currency to be delivered at one place today in exchange for another currency at another place at a later date. This time delay means that bills of exchange combined a spot exchange transaction with a lending transaction. In fact, against the background of the prohibition of many types of lending through usury laws, bills of exchange became Europe's dominant lending instrument. They allowed lending rates to be hidden within what on the surface was a foreign exchange contract (Bell et al., 2017; Flandreau et al., 2009; de Roover, 1967; de Malynes, 1601).

Extensive datasets of early modern bill of exchange quotations have been compiled (Da Silva, 1969; Schneider et al., 1992, 1994; Denzel, 2010), enabling us to throw light on the fluctuations in lending rates that Spanish merchants faced. Appendix A.4 provides a detailed description of how we use bill of exchange prices to infer the behavior of lending rates in Spain.[19] We complement the lending rate data with information contained in merchant letters that have been compiled by economic historians (da Silva, 1959a; Martín, 1965; Lockhart et al., 1976). In some of these

[19]The calculated interest rate fluctuations are best regarded as fluctuations in risky lending rates, rather than risk-free rates. Lenders were generally perceptive of the riskiness of borrowers' financial position, and high lending rates were charged to compensate for high default risk (Carrasco González, 1996, ch.2).

letters, contemporary merchants describe the credit market conditions they faced in the aftermath of maritime disaster events. Importantly, these letters describe episodes of credit rationing which the observable lending rate data does not reflect.

2.4. Econometric method

We use local projections (Jordà, 2005) to estimate impulse response functions (IRFs) over a five-year horizon, $h = 0, ..., 5$:

$$(Y_{t+h} - Y_{t-1})/Y_{t-1} = \alpha_h + \beta_h S_t + \gamma_h X_t + u_{t+h} \tag{1}$$

where Y_t is the outcome of interest (output, prices, wages, interest rates). u_{t+h} denotes the horizon-specific error term. S_t is the money supply shock. X_t comprises control variables, including the part of the initial money loss that is subsequently salvaged.[20] The estimated β_h coefficients describe the cumulative response of the outcome variable to a monetary shock: β_0 captures the cumulative response between period $t-1$ and t, β_1 captures the cumulative response between period $t-1$ and $t+1$, and so on.

The baseline specification includes up to two lags, five leads, and the contemporaneous values of the following exogenous control variables: the money supply shock S_t, the amount of money that was salvaged after a maritime disaster (% of stock), silver lost due to capture (% of stock), Spanish temperatures (Anderson et al., 2017), and indicators of the number of military conflicts that Spain was involved in (based on the historical conflict catalogue by Brecke, 1999). Finally we also include two lags and, with the exception of the dependent variable, the contemporaneous values of the following endogenous controls: price level growth, wage growth, real GDP growth, and changes in the money inflow-to-stock ratio.[21] As is common in IRF

[20] Alternatively, it is possible to account for salvaging by directly subtracting salvaged amounts from the shock measure. Figure B.6 in the Appendix shows very similar results for this approach.

[21] The first half of the 17th century witnessed the spread of copper coins in the Spanish economy. For a period of several decades, copper coins made up a substantial fraction of the overall Spanish money supply. We therefore allow the effects of money shocks to differ for the period 1617 to 1664. The beginning of that period is marked by the public authorities' decision to allow the minting of copper coins. Copper minting became largely inconsequential after 1664 when a decision was made to halt minting. While copper money existed before and after this period, its role was much

analyses based on local projections, we saturate our baseline specification with a rich set of control variables (Stock and Watson, 2018; Jordà et al., 2020). A more parsimonious specification that besides the lagged dependent variable's growth rate contains only the silver shock regressors results in similar IRF estimates (Figure B.7 in the Appendix).[22] The shortness of some of the disaggregated time series requires us to thin out the set of control variables to prevent the number of coefficients from exceeding the number of observations. In these cases we fall back onto the parsimonious specification.[23] For each estimated IRF we report point-wise confidence bands, based on Newey-West standard errors with the lag order of autocorrelation equalling five (Newey and West, 1987).

3. The causal effects of money supply shocks

3.1. Output effects

What was the real effect of exogenous money supply changes on the Spanish economy? Figure 2 displays the shock and its impact on aggregate economic activity. The money shock IRF describes a one-off reduction in money inflows amounting to 1% of the money stock. The IRF's quick return to zero one year after the impulse shows that subsequent Atlantic inflows did not compensate for the initial maritime disaster loss. In response to the shock, Spanish real output begins to fall. Output troughs one year later at around 1% below its trend level. Output stays below trend for several years before fully recovering four years after the shock.

We now turn to the analysis of the disaggregated output data. The top left panel

diminished. Only within the 1617 to 1664 time window did copper play an important role in the Spanish monetary system (Velde and Weber, 2000).

[22] The robustness of the results with respect to various other relevant alterations in specification and shock measure is documented in Appendix B.4. The results are robust to excluding money shocks that were rooted in international conflict (capture and combat events) from the analysis (Figure B.8). We obtain very similar results when we estimate IRFs through an autoregressive distributed lag (ARDL) model as in Romer and Romer (2004) and Cloyne and Hürtgen (2016) (Figure B.9). We also construct an alternative shock series under the assumption that the news about a disaster took one month longer than usual to arrive in Spain, which yields very similar results (Figure B.10). The calculation of placebo IRFs based on 500 random temporal reshufflings of the shock measure yield results centered around 0 (Figure B.11), confirming that our shock measure does not simply pick up a spurious relation between the variables.

[23] This applies to the following series: Mesta flock size, Segovian textile production, Toledean coarse and fine textiles.

of Figure 3 shows primary sector responses. Foodstuffs consumption responded only modestly.[24] By contrast, wool production displays a striking response. Sheep flock sizes drop by 7% within one year. Various pieces of evidence suggest that herd owners liquidated their livestock in an attempt to raise money. First, high lending rates and merchant letters reflect a scramble for liquidity occurring in the aftermath of money losses (see section 3.2.1). Second, consistent with an increase in the supply of meats, the price for mutton and beef fell quickly (see section 3.2.2). Finally, historical documents show that selling commodities to overcome a liquidity shortage was common practice in early modern Spain. Martín (1974, p.278) provides an account of wool merchants that lament having to sell part of their flock to raise money precisely when prices are lowest. Even those who did not own any commodities could participate in their sell-off (Pike, 1966, p.44): Contemporary theologian-economist de Mercado (1587, chapter XXII) describes a specific contract type – the *barata* – that allowed those in need of money to borrow commodities from wholesalers and immediately sell them at a substantial discount to raise the much needed cash. A fire sale of livestock is thus a compelling explanation for the sharp drops in flock size that we observe in the aftermath of contractionary money supply shocks.[25]

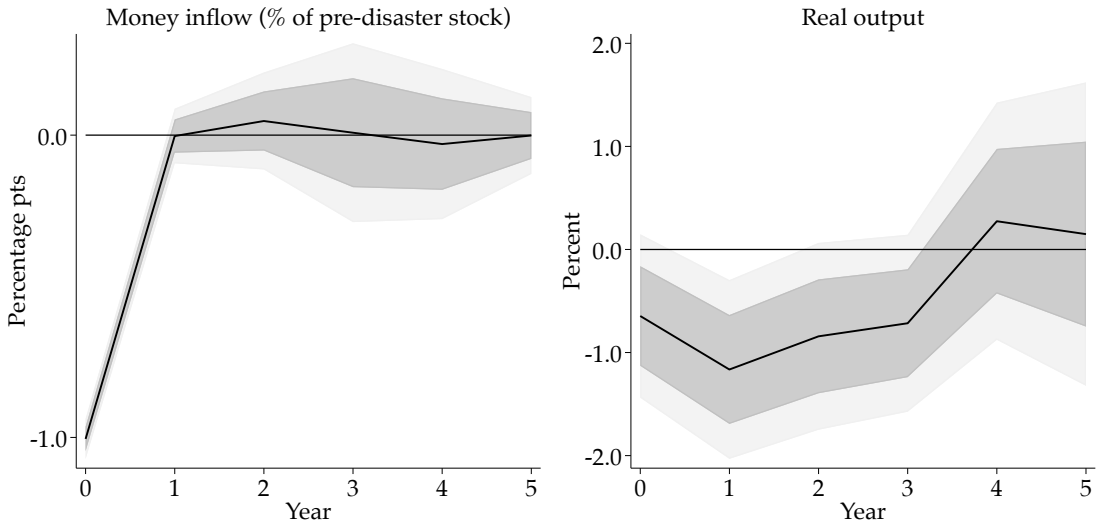
Further down the production chain, textile production in Segovia falls by 2% on impact (see top left panel of Figure 3). Production then remains 2% below trend for another two years before showing any signs of recovery. The lower left panel of Figure 3 confirms the textile industry's contraction based on the Toledean tax data. Cloth entries into Toledo decrease between 2% and 4%. Compared to the Segovian production data, Toledean cloth entries exhibit a delayed response, and the IRF for fine cloth entries is estimated less precisely than that for coarse cloths.[26] Generally,

[24] Separate Toledean tax revenue series for fish and cereals allow us to further distinguish between essential and luxury food items (Montemayor, 1981). Whereas the consumption of cereals responds similarly to all other foodstuffs, the consumption of fish – a luxury good – declines more markedly. This is indicative of belt-tightening behavior among households (Figure B.1 in the Appendix).

[25] Another potential explanation springs from the fact that shepherds obtained credit and cash from local landlords to supply their herds, e.g. with salt (Carrère, 1974, p.208). A scarcity of money and credit may thus have forced a reduction in flock size as shepherds became unable to sustain their herds. However, the drop in flock size is exceedingly large and short-lived when compared to production declines in other industries that relied on credit to purchase intermediate inputs (e.g. textile manufacturing).

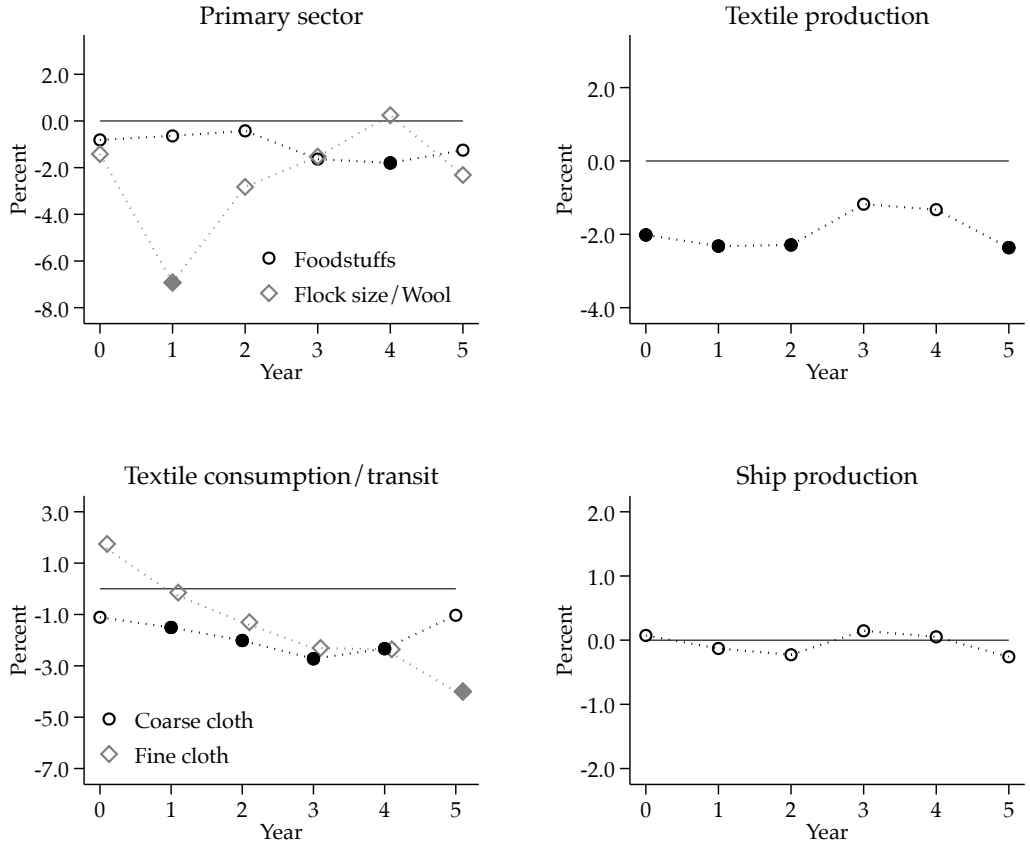
[26] One potential explanation for the delayed response is that tax auction revenues can reflect bidders' backward-looking expectations rather than the current volume of taxable goods.

Figure 2: Real output response to a negative 1 percentage point money inflow shock



Notes: Gray areas – 1 standard deviation and 90% confidence bands.

Figure 3: Real output responses to a negative 1 percentage point money inflow shock



Notes: Solid markers – significance at the 10% level. Hollow markers – no significance at the 10% level.

however, the two Toledean tax series corroborate the evidence for a contractionary effect of money supply shocks on textile manufacturing.[27]

We conclude our analysis of output quantities with a look at ship production. The lower right panel of Figure 3 shows that Basque shipbuilding did not react to the monetary shock. In our setting, the shipbuilding response has added relevance, because ships were lost in maritime disasters. While the value of ships lost was trivial compared to the value of money lost (see section A.2.2 in the Appendix), the replacement demand for new ships may nevertheless have contributed to stabilizing output in the shipbuilding industry. The lack of response of ship production indicates that if any such stabilization effect existed, it did not go beyond offsetting the negative effects springing from the maritime disaster shock.

In sum, the IRF estimates suggest that exogenous money supply changes have real effects. A 1% contractionary money shock is followed by a 1% decline in the aggregate output measure. Output declines are particularly large and persistent in textile manufacturing. The evidence presented in the subsequent sections highlights credit market frictions and slowly adjusting prices as important transmission channels through which money supply shocks exerted their influence.

3.2. Monetary transmission channels

This section analyzes the transmission channels through which money supply shocks affected the real economy. It does so by taking a closer look at the data for indications of whether certain transmission channels were active or not.

3.2.1 External financing and credit frictions

Despite the prohibition of many forms of interest rates, external financing played an important role in early modern Europe. Debt obligations commonly made up between half and two thirds of merchants' liabilities, implying asset-to-net-worth

[27] Montemayor (1981) provides additional tax auction returns for tax rights pertaining to Toledean theater and brothel revenues. The results for these series, appropriately deflated, are shown in Figure B.1 in the Appendix. Brothel visits decline gradually, whereas theater visits do not respond significantly.

ratios in the 2 to 3 range (Costa, 1997; Gelderblom and Jonker, 2004; Oldland, 2010).[28] For Spain, an extensive study of 149 merchant housholds' post-mortem inventories from 17th century Toledo reveals an average asset-to-net-worth ratio of 2.2 (Pérez, 1992).[29] A snapshot of a merchant company's balance sheet from Cadiz in 1747 shows an asset-to-net-worth ratio of 2.8 (Bernal and Ruiz, 1992, pp.369-370).[30] To which extent did credit market conditions deteriorate in response to negative money supply shocks?

The left panel of Figure 4 shows that a 1 percentage point reduction in money inflows led to a 1.5 percentage point increase in the lending rate. The lending rate remains elevated for several years after the shock.[31] However, the lending rate IRF does not exhibit an on impact response, despite the on-impact drop in textile production (see section 3.1) – a sector highly dependent on the input goods supplied by merchants. The conjunction of an immediate output contraction and a delayed lending rate increase can be indicative of an episode of credit rationing during which all but the most creditworthy borrowers lose access to credit (Stiglitz and Weiss, 1981). During such episodes observable lending rates fail to reflect the prohibitively high premiums faced by average borrowers. In contrast to the lending rates analysis, an analysis of merchant letters that were written in the aftermath of maritime disasters is indicative of credit rationing.

In a letter dated February 1568, ten months after the loss of monetary metals amounting to around 7% of Spain's money supply, a Spanish merchant complained about the lack of credit supply at the exchange fairs (da Silva, 1959a, p.12). Tight money and credit market conditions persisted into the second half of 1568: A mer-

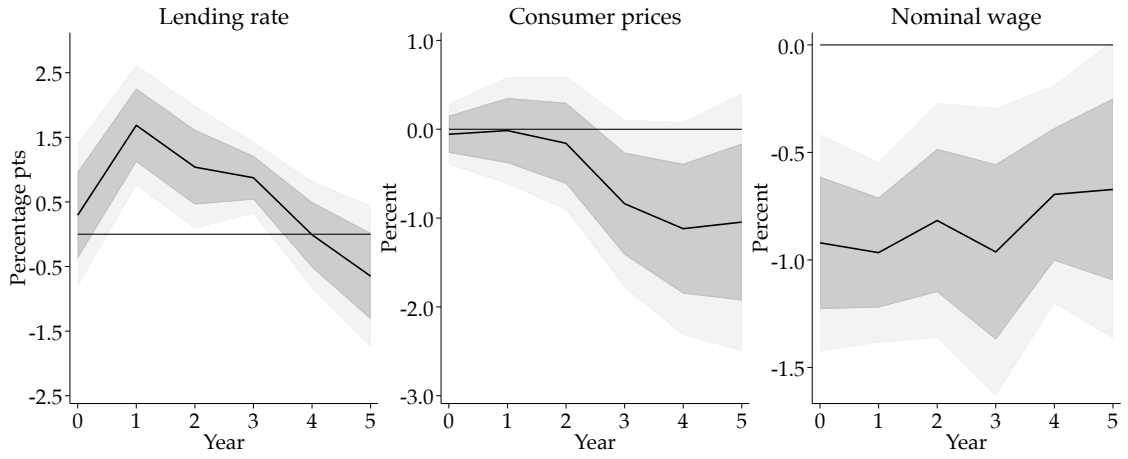
[28] The most important suppliers of credit were wealthy merchants, merchant-bankers, ecclesiastical institutions, and nobles (Pike, 1965; Milhaud, 2015).

[29] In the calculation of this asset-to-net-worth ratio equity-like liabilities include ex ante capital and dowries brought into the merchant family through marriage.

[30] This number results from treating credit-financed owner contributions as equity. Otherwise, the company was characterized by a higher asset-to-net-worth ratio.

[31] As a placebo test we also estimated IRFs for the lending rates in several other Western European financial centers. The results show that lending rate increases are prevalent across Spanish cities whereas this is not the case for other European cities (Figure B.14 in Appendix B). Appendix A.4.3 explores the concern that in the aftermath of maritime disasters, bills of exchange prices reflected merchants' fears of currency debasement, rather than Spanish lending rate fluctuations. When we calculate lending rates from bill prices in a way that eliminates the devaluation risk component, we find a similar lending rate response, with a peak of 3.5% one year after the shock, and a persistence of two years (Figure A.12).

Figure 4: Transmission variable responses to negative 1 percentage point money inflow shock



Notes: Gray areas – 1 standard deviation and 90% confidence bands.

chant letter dated August 25th 1568 continues to diagnose the Iberian peninsula with a deficiency in money and credit supply (da Silva, 1959b, p.22). One year after another disaster in 1591, Simón Ruiz – one of the wealthiest merchants of his time – wrote a letter to one of his factors instructing him to use money parsimoniously and instead use bills of exchange for payment whenever possible. This suggests that the wealthiest merchants still had access to credit instruments which they could use to bridge the money shortage (Martín, 1965, p.xiv). The sensitivity of Spanish financial markets to American silver arrivals is further evidenced by the severe effects that mere delays in silver shipments could develop (Pike, 1966, p.87).[32]

Merchants' access to credit had repercussions for the wider economy. This is because Spain's early modern manufacturing industry was characterized by the *putting-out system* in which producers received raw materials and intermediate inputs from merchants. Merchants, in turn, relied on smoothly functioning credit markets to finance their purchase of input goods (García Sanz, 1977; Montemayor, 1996, p.249). Thus, when the cost of external finance increased, the manufacturing industry's supply with input goods could dwindle. In this regard, it is difficult to overestimate the centrality of merchants in the early modern Spanish economy. Many weavers received threads, looms, and even lodging from their merchant (Vaquero,

[32] A Sevillian merchant's letter from November 25th 1553 laments such a delay, and states that the only credit still available to him came with a 60% annualized interest rate (Lockhart et al., 1976, pp.91ff.).

2020; Carande Thobar, 1976, p.471). For 17th century Toledo, Pérez (1992, p.92) quantifies the linkage between merchants and producers based on post-mortem inventory data found in notarial deeds. He finds that it was common for one merchant to supply between 100 and 200 manufacturers, e.g. master weavers. Montemayor (1996, p.230) provides another example of 100 master silk weavers, who ensured subsistence for a total of 8000 people. The entire enterprise depended on only 10 to 12 large merchants for its supply of silk. Merchants' access to credit was thus crucial for the functioning of the wider economy.

Maritime disasters quickly affected credit conditions in Spain, because money arrivals from America immediately fed into the central nodes of the money market – financial exchanges and exchange fairs. This is reflected in a unique set of records from 1570 and 1571, which registered the outflow of American precious metals from their port of arrival together with their destination. These records reveal that the money's most important destination were Spanish regions that hosted important financial exchanges and exchange fairs (da Silva, 1965, p.67). Maritime disasters thus brought about a scarcity of liquidity and they lowered the net worth of those individuals who were the owners of the lost money.[33] Both raised doubts about final settlement at the exchange fairs. The heightened default risk increased the cost of external finance as lenders became more hesitant to extent credit and sellers became more reluctant to accept credit instruments for payment – a credit crunch ensued.[34]

[33] A back-of-the-envelope calculation suggests that silver losses only amounted to a small fraction of the overall Spanish wealth level. The Spanish annual precious metal inflow-to-GDP ratio ranges from 0.4% to 6.8%, whereas the closest wealth-to-GDP ratios from Piketty (2014) are 7 for England in 1700, and 5 for Spain in 1901. Considering a wealth-to-GDP ratio in the 5 to 7 range, annual Spanish precious metal inflows only amounted to between 0.06% and 1.4% of the Spanish wealth level. Small wealth losses can nevertheless develop sizeable aggregate effects if they are amplified by a critical node in the economy, such as levered entrepreneurs or financial intermediaries whose decisions influence the economy's aggregate investment (Kiyotaki and Moore, 1997; Bernanke et al., 1999; Brunnermeier and Sannikov, 2014). Relatedly, the literature that traces the origins of aggregate fluctuations back to the amplification of idiosyncratic shocks along an economy's input-output linkages (Gabaix, 2011; Acemoglu et al., 2012; Carvalho, 2014) suggests that negative money shocks to Spanish merchants could develop such a powerful effect throughout the Spanish economy, in part because merchants were located at a central node in Spain's early modern production network. Our results are particularly reminiscent of research that highlights sector specific financial frictions in the the propagation of shocks along production chains (Bigio and La'o, 2020; Costello, 2020; Luo, 2020; Demir et al., 2020; Altinoglu, 2021; Alfaro et al., 2021).

[34] This picture is echoed in the historical literature. Morineau (1985, p.69) describes how the loss of the 1622 fleet sparked panic in Seville. More recently, the empirical evidence presented by Schularick, ter Steege, and Ward (2021) highlights the potency of contractionary monetary shocks in triggering financial crises.

3.2.2 Price adjustment

A leading explanation for money non-neutrality is that prices fail to adjust to money supply changes. To which extent did prices adjust in early modern Spain?[35] The middle panel of Figure 4 shows that the aggregate Spanish price level reacts sluggishly to money supply shocks. Only after three years does the consumer price response fall to -1%. Even then, however, the IRF estimate is not statistically significant. This renders price rigidity a suspect in our search for the monetary transmission channels through which money supply shocks exerted their influence on the real economy.[36]

The historical literature describes two ways in which reduced money arrivals first met rigid prices to bring about a reduction in output. First, consumption decreased as households went through a period of belt-tightening in the aftermath of money losses. Pike (1966, pp.30ff.) describes the consumption of silks and embroidery that the arrival of money from America triggered. To the extent that goods prices failed to adjust to negative money shocks, households were likely to cut back on this consumption after maritime disasters. Second, Spain's manufacturing regions were an important destination for American money arrivals.[37] Maritime disaster losses brought this money flow to a halt and thus disrupted the manufacturing industry's provision with money.[38] When prices for input goods failed to adjust sufficiently, liquidity constrained merchants and manufacturers had to reduce their purchase

[35] One source of nominal price rigidity in early modern Spain were guilds. They were prevalent in the urban manufacturing and service sectors, but they also could be found in the primary sector, and among rural artisans (Ogilvie, 2011, p.19). As producers of differentiated products, guilds could set collective monopoly prices for the output of their members. If large enough, or if in possession of a chartered right to be the exclusive buyer of a particular industry's output (Ogilvie, 2011, p.33), guilds also could set price ceilings on the raw materials they purchased from their suppliers (Ogilvie, 2014).

[36] Appendix B.2 explores whether Spain's nominal adjustment was aided by a decrease in the silver content of Spanish coins. We find that such debasements at most played a small auxiliary role over the medium term.

[37] Sevillian records from 1570 and 1571, which registered the outflow of American precious metals from their port of arrival together with their destination, identify Spain's manufacturing regions the second most important destination for American money arrivals. Only regions hosting important financial exchanges received more money (da Silva, 1965, p.67). Relatedly, Phillips (1987, p.540) states that part of the investment in the Spanish industrial sector was directly financed out of profits that were generated in the colonies.

[38] A merchant letter from September 5th, 1577 exemplifies this link between the arrival of American money and textile manufacturing investments. In this letter, Simón Ruiz instructs one of his factors to send the money arriving from America to him as soon as it arrives, so he can invest it in the production of woollen cloth (da Silva, 1959b, p.130).

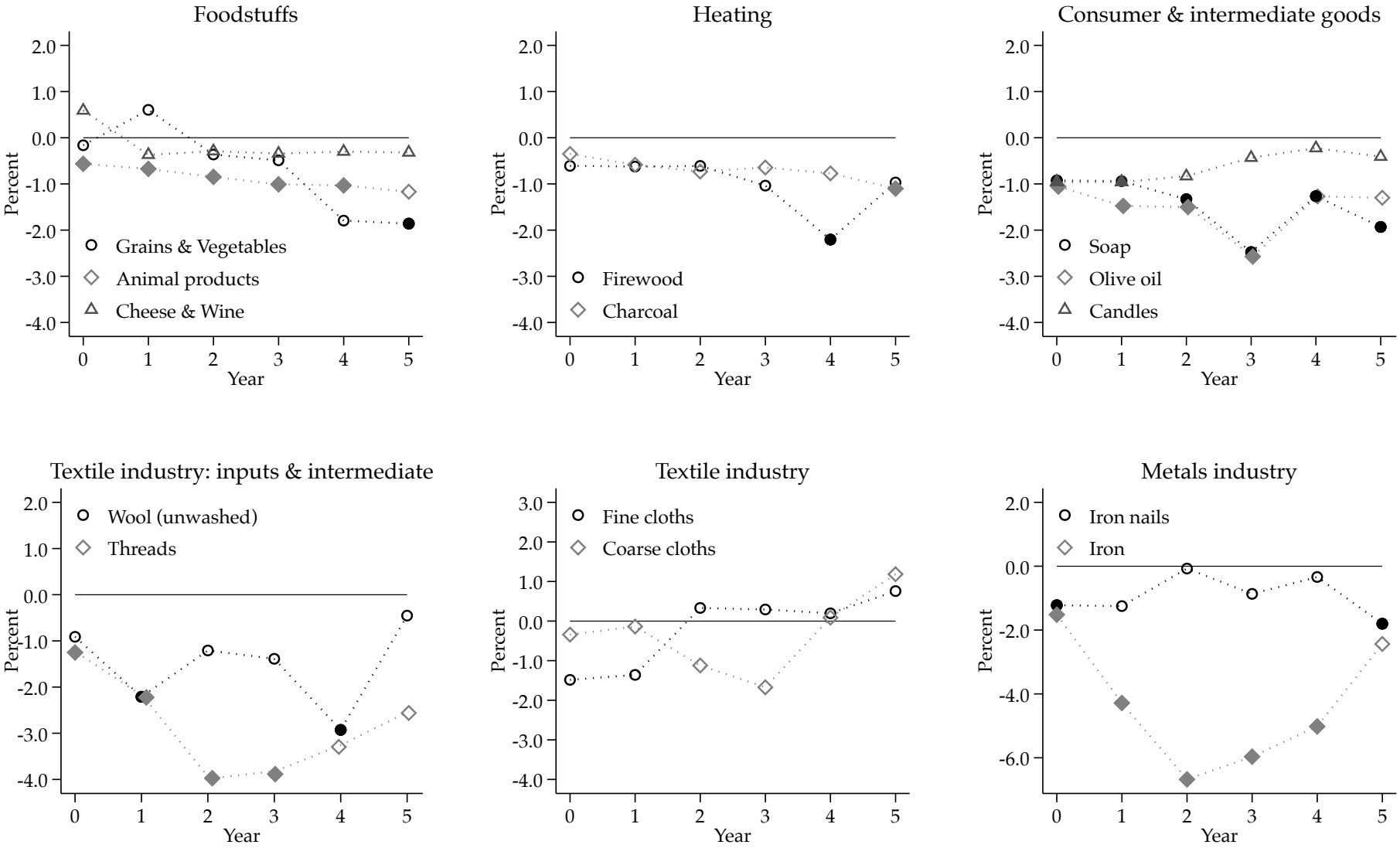
volumes.

To better understand the nominal adjustment of Spain's early modern economy we turn to the analysis of the disaggregated price data. We begin with the prices for essential consumer goods – food and heating. The top left panel in Figure 5 shows that most foodstuff prices decline slowly in response to a money shock. This is the case for essentials, such as grains and vegetables, as well as processed items, such as cheese and wine. The price reaction of these food items is furthermore statistically insignificant over much of the five-year horizon we analyze. Firewood and charcoal prices follow the same pattern (see top middle panel). By contrast, prices for animal products – meats, eggs, and honey – fall immediately (see top left panel). The liquidation of livestock in response to a liquidity shortage described earlier presents one economic rationale for this. Even more disaggregated data reveals that the significant on-impact response is shared by mutton and beef prices (Figure B.1 in the Appendix).

The prices for nonessential consumer goods (soap, olive oil, and candles) fall quickly within the year of impact (see top right panel). The prices for soap and olive oil continue to fall for several more years. Both goods are not just consumer items, but also important intermediate inputs for the textile industry. They are required in large quantities to wash the wool, and to lubricate textile fibres for processing (García, 2007, p.183). The lower left panel of Figure 5 displays the price responses of two more textile industry inputs – unwashed wool and threads. Both prices fall substantially, but the wool price decline is somewhat less pronounced.[39]

[39]Wool was an important export item in early modern Spain. The relative stability of wool prices may thus reflect an initial increase in wool exports (see section B.1 in the Appendix). In addition, the fire sale in livestock documented earlier reduced wool supply during the time it took the flock size to recover (see section 3.1).

Figure 5: Price responses to a negative 1 percentage point money inflow shock



Notes: Solid markers indicate significance at the 10% level.

In contrast to the textile industry's input prices, its output prices for coarse and fine cloths do not fall significantly. A similar divergence between input and output prices can be observed in metallurgy.[40] Whereas the iron price declines by a cumulative 6%, the price for nails only briefly falls by around 1%.[41] This input-output price divergence ties in with the notion that a credit crunch paralyzed merchant procurement of raw materials and intermediate inputs (see section 3.2.1). The imperfect pass-through of input prices to output prices is furthermore indicative of price rigidities in the manufacturing sector.[42]

3.2.3 Wage adjustment

Wages are another nominal variable, whose failure to adjust can give rise to money non-neutrality. The IRF estimate depicted in the right panel of Figure 4 shows that Spain's aggregate wage index immediately falls by around 1% in response to the negative money supply shock. Given the delayed response of consumer prices this implies an immediate real wage drop of 1%. This on-impact adjustment defies the common presumption that wages were more rigid than goods prices.[43] The following analysis of the disaggregated wage data traces the origins of early modern Spain's rapid nominal wage adjustment.

The upper left panel of Figure 6 isolates one occupational group that saw its wage drop markedly on impact: Unskilled agricultural laborers, such as harvesters. The rapidity of the nominal adjustment for this occupational group is perhaps not surprising given that unskilled agricultural labor was hired on a seasonal or even daily basis. Nevertheless, the magnitude of the wage drop stands out at 4%.[44]

[40] Gautier and Bihan (2020) and Pasten et al. (2020) make the related point that sectoral heterogeneity in price rigidity can amplify the real effects of monetary shocks.

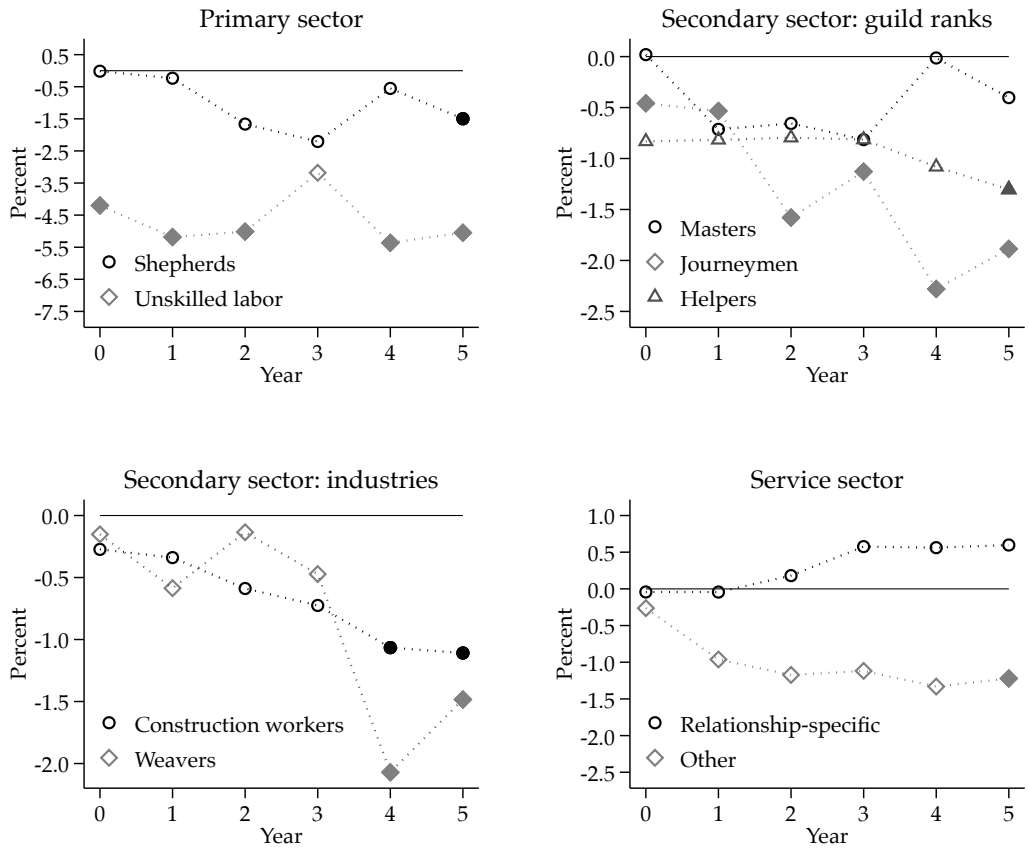
[41] See de Gracia (2002) for a description of the putting-out system in Spain's early modern iron manufacturing sector.

[42] Note, however, that such imperfect pass-through of input prices to output prices can also result from endogenous markup fluctuations (Rotemberg and Woodford, 1999) and the repeated amplification of an initial liquidity shock along a production chain that is characterized by financial frictions (Bigio and La'o, 2020; Altinoglu, 2021).

[43] As our identification strategy exploits large shocks, our findings may reflect optimally timed wage changes in response to sizeable shocks as described by Golosov and Lucas (2007).

[44] Absent a contraction in the consumption of agricultural goods (see section 3.1), agricultural labor demand presumably remained relatively stable in the face of money supply shocks. This suggests that the marked decline in the wage for unskilled agricultural workers reflects an increase in their supply. The analysis of available employment flow data fragments supports this interpretation (Appendix B.3).

Figure 6: Nominal wage responses to negative 1 ppt money inflow shock



Notes: Solid markers indicate significance at the 10% level.

In contrast to the immediate wage drop among unskilled agricultural workers, the wages of skilled shepherds declined only gradually by around 1%.[45]

To a lesser extent, the contrast between skilled and unskilled wages is also echoed in the secondary sector. The upper right panel of Figure 6 displays the response of various occupations according to their guild rank. Skilled masters only experienced an insignificant wage decrease, whereas helpers, and especially journeymen saw their wages decline more markedly. Reminiscent of more recent findings, monetary contractions in early modern Spain thus aggravated inequality by widening the wage gap between skilled and unskilled workers (Carpenter and Rodgers III, 2004; Coibion et al., 2017). The guild rank results suggest that differences in wage responses were

[45]Shepherding is categorized as a skilled occupation because it involved important decision making. For example, the timing of herd movements affected the quality of the sheeps' wool. Knowledge of the terrain and meteorology were advantageous. The occupation also involved financial decision making connected with the continued supply of the transhumant herd with necessary inputs, such as salt.

in part driven by institutional factors. Masters, for example, typically shared in the wage setting power of their guilds, whereas journeymen and helpers were more likely to be wage takers with little collective bargaining power.

We find no evidence indicating that secondary sector wage responses differed across industries: the gradual wage decline among weavers resembles the gradual wage decline among construction workers (see lower left panel of Figure 6). Thus, secondary sector wages as a whole exhibit a rather gradual adjustment to money supply shocks. Given that labor is a key input into the secondary sector's production process, the failure of nominal wages to immediately adjust implies that liquidity constrained merchants could employ less labor in their production networks.

One group of jobs whose nominal wage rate remains entirely unaffected by the monetary contraction are relationship-specific service sector jobs. This covers occupations that build on mutual trust (e.g. doormen, stewards) and occupations that are characterized by the accumulation of relationship-specific knowledge (e.g. bookkeepers). By contrast, other service sector occupations (e.g. basket emptiers, cart-loaders) see their wage rates decline gradually by around 1%.

4. Conclusion

In this paper we use a series of natural experiments to identify the causal effects that run from the monetary side of the economy to the real side. Maritime disasters in the Spanish Empire repeatedly gave rise to large losses of monetary metals. The causes of these disasters had nothing to do with the economy in Spain, and the corresponding precious metal losses therefore gave rise to exogenous variation in the influx of American precious metals into Spain.

This paper contributes to a long-standing debate in economics that already took place among contemporaries witnessing the arrival of the silver fleets. In 16th century Spain, theologians expressed the idea that an increase in money is absorbed by an equivalent increase in prices (de Azpilcueta, 1556; de Molina, 1597). Only a few decades later, English mercantilists hypothesized about the non-neutrality of money. They argued that an increase in money not only increases prices, but stimulates real

economic activity (Misselden, 1622; de Malynes, 1623). Our findings lend support to the latters' hypothesis.

Negative shocks to Spain's money supply caused real output to decline. Our analysis highlights credit market frictions and nominal rigidities as important monetary transmission mechanisms. Disruptions to the influx of American precious metals into Spain triggered credit crunches during which lending rates rose and credit became rationed. Consumer prices adjusted only sluggishly, forcing transaction quantities to decline in response to the diminished money supply. Disaggregated data indicate that the putting-out system influenced how the Spanish economy reacted to monetary shocks: by restricting money supply and credit access, maritime disasters impaired merchants' ability to purchase input goods for their manufacturers.

References

Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. 2012. The network origins of aggregate fluctuations. *Econometrica* 80(5): 1977–2016.

Alfaro, Laura, Manuel García-Santana, and Enrique Moral-Benito. 2021. On the direct and indirect real effects of credit supply shocks. *Journal of Financial Economics* 139(3): 895–921.

Allen, Robert C. 2001. The great divergence in European wages and prices from the Middle Ages to the First World War. *Explorations in Economic History* 38(4): 411–447.

Altinoglu, Levent. 2021. The origins of aggregate fluctuations in a credit network economy. *Journal of Monetary Economics* 117: 316–334.

Álvarez-Nogal, Carlos, and Leandro Prados de la Escosura. 2013. The rise and fall of Spain (1270–1850). *The Economic History Review* 66(1): 1–37.

Anderson, Robert Warren, Noel D. Johnson, and Mark Koyama. 2017. Jewish persecutions and weather shocks: 1100–1800. *The Economic Journal* 127(602): 924–958.

Auclert, Adrien. 2019. Monetary policy and the redistribution channel. *The American Economic Review* 109(6): 2333–67.

Baskes, Jeremy. 2013. *Staying afloat: Risk and uncertainty in Spanish atlantic world trade, 1760-1820*. Stanford University Press.

Bell, Adrian R., Chris Brooks, and Tony K. Moore. 2017. Cambium non est mutuum: Exchange and interest rates in medieval Europe. *The Economic History Review* 70(2): 373–396.

Bernal, Antonio-Miguel, and Isabel Martínez Ruiz. 1992. *La financiación de la Carrera de Indias (1492-1824): Dinero y crédito en el comercio colonial español con América*. Fundación El Monte.

Bernanke, Ben S., Jean Boivin, and Piotr Eliasz. 2005. Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics* 387–422.

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist. 1999. The financial accelerator in a quantitative business cycle framework. *Handbook of Macroeconomics* 1: 1341–1393.

Bigio, Saki, and Jennifer La'o. 2020. Distortions in production networks. *The Quarterly Journal of Economics* 135(4): 2187–2253.

Bonifacio, Claudio. 2010. *Galleons and sunken treasure.* Signum Ops.

Brecke, Peter. 1999. Violent conflicts 1400 AD to the present in different regions of the world. 1999 Meeting of the Peace Science Society.

Brunnermeier, Markus K., and Yuliy Sannikov. 2014. A macroeconomic model with a financial sector. *American Economic Review* 104(2): 379–421.

Calvo, Guillermo A. 1983. Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics* 12(3): 383–398.

Carande Thobar, Ramón. 1976. Telares y los paños en el mercado de lanas en Segovia. In *Produzione, commercio e consumo dei panni di lana*, 469–73. Olschki.

Carlstrom, Charles T., and Timothy S. Fuerst. 1997. Agency costs, net worth, and business fluctuations: A computable general equilibrium analysis. *The American Economic Review* 893–910.

Carpenter, Seth B., and William M. Rodgers III. 2004. The disparate labor market impacts of monetary policy. *Journal of Policy Analysis and Management* 23(4): 813–830.

Carrasco González, María Guadalupe. 1996. Los instrumentos del comercio colonial en el Cádiz del siglo XVII (1650-1700). *Estudios de Historia Económica* 35.

Carrère, Claude. 1974. Aspects de la production et du commerce de la laine en Aragon au milieu du XVe siècle. In *La lana come materia prima: I fenomeni della sua produzione e circolazione nei secoli XIII-XVII*, edited by Spallanzani, Marco, 205–19. Olschki.

Carvalho, Vasco M. 2014. From micro to macro via production networks. *Journal of Economic Perspectives* 28(4): 23–48.

Céspedes del Castillo, Guillermo. 1996. *Las casas de moneda en los reinos de Indias: Las cecas indianas en 1536-1825.* Museo Casa de la Moneda.

Chaunu, Huguette, Pierre Chaunu, and Guy Arbellot. 1955. *Séville et l'Atlantique, 1504-1650*. Travaux & mémoires de l'Institut des hautes études de l'Amérique latine ; 4. Paris: Institut des hautes études de l'Amérique latine.

Chaunu, Pierre, and Huguette Chaunu. 1977. *Séville et l'Amérique aux XVIe et XVIIe siècles*. Flammarion.

Chen, Yao, Nuno Palma, and Felix Ward. 2021. Reconstruction of the Spanish money supply, 1492–1810. *Explorations in Economic History* 81.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 1999. Monetary policy shocks: What have we learned and to what end? *Handbook of Macroeconomics* 1: 65–148.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 2005. Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy* 113(1): 1–45.

Cloyne, James, and Patrick Hürtgen. 2016. The macroeconomic effects of monetary policy: A new measure for the United Kingdom. *American Economic Journal: Macroeconomics* 8(4): 75–102.

Coibion, Olivier. 2012. Are the effects of monetary policy shocks big or small? *American Economic Journal: Macroeconomics* 4(2): 1–32.

Coibion, Olivier, Yuriy Gorodnichenko, Lorenz Kueng, and John Silvia. 2017. Innocent bystanders? Monetary policy and inequality. *Journal of Monetary Economics* 88: 70–89.

Costa, Leonor Freire. 1997. *Naus e galeões na Ribeira de Lisboa: A construção naval no século XVI para a Rota do Cabo*. Patrimonia historica: Dissertações.

Costa, Leonor Freire, Susana Münch Miranda, and Pilar Nogues-Marco. Forthcoming. An economic history of the Iberian Peninsula, 700-2000. In *Globalization and the Enlightenment, 1500-1800*, edited by Costa, Leonor Freire, and Regina Grafel. Cambridge University Press.

Costa, Leonor Freire, Maria Manuela Rocha, and Rita Martins de Sousa. 2013. *Ouro do Brasil*. Imprensa Nacional-Casa da Moeda.

Costello, Anna M. 2020. Credit market disruptions and liquidity spillover effects in

the supply chain. *Journal of Political Economy* 128(9): 3434–3468.

da Silva, José Gentil. 1959a. *Marchandises et finances II: Lettres de Lisbonne, 1563-1578*. S.E.V.P.E.N.

da Silva, José Gentil. 1959b. *Marchandises et finances III: Lettres de Lisbonne, 1563-1578*. S.E.V.P.E.N.

da Silva, José Gentil. 1965. *En Espagne: développement économique, subsistance, déclin*. Mouton.

Da Silva, José-Gentil. 1969. *Banque et crédit en Italie au XVIIe siècle*, volume 2. Paris: Klincksieck.

de Azpilcueta, Martín. 1556. Commentary on the resolution of money. In Journal of Markets & Morality, 7(1), 2004.

de Gracia, Hilario Rodríguez. 2002. Relaciones comerciales en el siglo XVII: hierro guipuzcoano, cerrajería de Mora (Toledo) y seda toledana. *Boletín de la Real Sociedad Bascongada de Amigos del País* 58(2).

de Jong, Herman J., and Nuno Palma. 2018. Historical National Accounting. In *An Economist's Guide to Economic History*, edited by Blum, M., and C. L. Colvin. Palgrave Macmillan.

de Malynes, Gerard. 1601. *A Treatise on the Canker of England's Commonwealth*. London: Printed by Richard Field for William Iohnes printer.

de Malynes, Gerard. 1623. *The center of the circle of commerce*. London: Printed by William Jones, and are to be sold by Nicholas Bourne at the Royal Exchange.

de Mercado, Fray Tomás. 1587. *Summa de tratos y contratos*. Seville: Fernando Díaz.

de Molina, Luis. 1597. A treatise on money. In CLP Academic 2015.

de Paula Perez Sindreau, Francisco. 2016. *La casa de moneda de Sevilla, su historia*. Universidad de Sevilla.

de Roover, Raymond. 1967. The scholastics, usury, and foreign exchange. *Business History Review* 41(3): 257–271.

Demir, Banu, Beata Javorcik, Tomasz K. Michalski, and Evren Ors. 2020. Financial constraints and propagation of shocks in production networks. *The Review of*

Economics and Statistics 1–46.

Denzel, Markus A. 2010. *Handbook of world exchange rates, 1590–1914*. Routledge.

Desaulty, Anne-Marie, Philippe Telouk, Emmanuelle Albalat, and Francis Albarède. 2011. Isotopic Ag–Cu–Pb record of silver circulation through 16th–18th century Spain. *Proceedings of the National Academy of Sciences* 108(22): 9002–9007.

Elliott, John H. 2006. *Empires of the Atlantic world: Britain and Spain in America, 1492-1830*. Yale University Press.

Esteban, Javier Cuenca. 1981. Statistics of Spain's colonial trade, 1792-1820: Consular duties, cargo inventories, and balances of trade. *The Hispanic American Historical Review* 61(3): 381–428.

Feliu, Gaspar. 1991. *Precios y salarios en la Cataluña moderna*, volume II. Banco de España.

Flandreau, Marc, Christophe Galimard, Clemens Jobst, and Pilar Nogués-Marco. 2009. The bell jar: Commercial interest rates between two revolutions, 1688–1789. In *The Origins and Development of Financial Markets and Institutions: From the Seventeenth Century to the Present*, edited by Atack, Jeremy, and Larry Neal, chapter 6, 161–208. Cambridge University Press.

Flynn, Dennis O. 1978. A new perspective on the Spanish Price Revolution: The monetary approach to the balance of payments. *Explorations in Economic History* 15(4): 388–406.

Frenkel, Jacob A., and Harry G. Johnson, editors. 2013. *The Monetary Approach to the Balance of Payments (Collected Works of Harry Johnson)*. Routledge.

Friedman, Milton, and Anna J. Schwartz. 1963. *A monetary history of the United States, 1867-1960*. Princeton University Press.

Gabaix, Xavier. 2011. The granular origins of aggregate fluctuations. *Econometrica* 79(3): 733–772.

García, Ricardo Hernandez. 2007. *La industria textil en Palencia durante los siglos XVI y XVII*. Ediciones Universidad de Valladolid.

García-Baquero González, Antonio. 2003. American gold and silver in the eighteenth century: From fascination to accounting. In *Global connections and monetary his-*

tory, 1470-1800, edited by Dennis O. Flynn, Richard von Glahn, Arturo Giráldez, 107–121. Ashgate.

García Sanz, Ángel. 1977. *Desarrollo y crisis del Antiguo Régimen en Castilla la Vieja: Economía y sociedad en tierras de Segovia, 1500-1814*. Akal.

Gautier, Erwan, and Hervé Le Bihan. 2020. Shocks vs menu costs: patterns of price rigidity in an estimated multi-sector menu-cost model. *The Review of Economics and Statistics* 1–45.

Gelderblom, Oscar, and Joost Jonker. 2004. Completing a financial revolution: The finance of the Dutch East India trade and the rise of the Amsterdam capital market, 1595–1612. *The Journal of Economic History* 64(3): 641–672.

Gertler, Mark, and Peter Karadi. 2015. Monetary Policy Surprises, Credit Costs, and Economic Activity. *American Economic Journal: Macroeconomics* 7(1): 44–76.

Golosov, Mikhail, and Robert E. Lucas. 2007. Menu costs and Phillips curves. *Journal of Political Economy* 115(2): 171–199.

Gorodnichenko, Yuriy, Viacheslav Sheremirov, and Oleksandr Talavera. 2018. Price setting in online markets: Does IT click? *Journal of the European Economic Association* 16(6): 1764–1811.

Hamilton, Earl Jefferson. 1934. *American treasure and the price revolution in Spain, 1501-1650*, volume 1125. Harvard University Press Cambridge, MA.

Hamilton, Earl Jefferson. 1947. *War and prices in Spain: 1651-1800*, volume 81. Harvard University Press Cambridge, MA.

Homer, Sidney, and Richard E. Sylla. 1991. *A history of interest rates*. Rutgers University Press, 3 edition.

Hume, David. 1752. *Political Discourses*. A. Kincaid & A. Donaldson.

Ireland, Peter N. 1997. A small, structural, quarterly model for monetary policy evaluation. In *Carnegie-Rochester Conference Series on Public Policy*, volume 47, 83–108. Elsevier.

Ireland, Peter N. 2003. Endogenous money or sticky prices? *Journal of Monetary Economics* 50(8): 1623–1648.

Irigoin, Alejandra. 2020. Rise and demise of the global silver standard. In *Handbook of the History of Money and Currency*, 383–410. Springer, Singapore.

Jordà, Òscar. 2005. Estimation and inference of impulse responses by local projections. *The American Economic Review* 95(1): 161–182.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2020. The effects of quasi-random monetary experiments. *Journal of Monetary Economics* .

Kiyotaki, Nobuhiro, and John Moore. 1997. Credit Cycles. *The Journal of Political Economy* 105(2): 211–248.

Klein, Julius. 1920. *The Mesta: A study in Spanish economic history 1273–1836*. Harvard University Press.

Koudijs, Peter. 2015. Those who know most: Insider trading in eighteenth-century Amsterdam. *Journal of Political Economy* 123(6): 1356–1409.

Koudijs, Peter. 2016. The boats that did not sail: Asset price volatility in a natural experiment. *The Journal of Finance* 71(3): 1185–1226.

Kuttner, Kenneth N., and Patricia C. Mosser. 2002. The monetary transmission mechanism: Some answers and further questions. *Federal Reserve Bank of New York Economic Policy Review* 8(1): 15–26.

Le Flem, Jean-Paul. 1976. Vraies et fausses splendeurs de l'industrie textile ségovienne (vers 1460 – vers 1650). In *Produzione, commercio e consumo dei panni di lana*, 525–536. Olschki.

Lockhart, James, Enrique Otte, Alan Knight, and James Marvin Lockhart. 1976. *Letters and people of the Spanish Indies: sixteenth century*, volume 22. Cambridge University Press.

Losa, Ernesto López, and Santiago Piquero Zarauz. 2021. Spanish subsistence wages and the Little Divergence in Europe, 1500–1800. *European Review of Economic History* 25(1): 59–84.

Luo, Shaowen. 2020. Propagation of financial shocks in an input-output economy with trade and financial linkages of firms. *Review of Economic Dynamics* 36: 246–269.

Mangas, Fernando Serrano. 1989. *Armadas y flotas de la plata 1620-1640*. Madrid:

Banco de España.

Martín, F. Ruiz. 1965. *Lettres marchandes échangées entre Florence et Medina del Campo*, volume 124. S.E.V.P.E.N.

Martín, Felipe Ruiz. 1974. Pastos y Ganaderos en Castilla: La Mesta (1450–1600). In *La lana come materia prima: I fenomeni della sua produzione e circolazione nei secoli XIII-XVII*, edited by Spallanzani, Marco, 271–85. Olschki.

Marx, Robert F. 1987. *Shipwrecks in the Americas*. Courier Corporation.

Massana, Rosa Ros. 1999. *La industria textil lanera de Béjar (1680-1850): la formación de un enclave industrial*. Junta de Castilla y Leon Consejeria.

Milhaud, Cyril. 2015. Priests or bankers? Ecclesiastical credit in early modern Spain. Paris School of Economics Working Paper.

Miranda-Agrippino, Silvia, and Giovanni Ricco. 2018. Identification with external instruments in structural VARs under partial invertibility. OFCE Sciences Po, Working Paper n. 24 06/2018.

Mishkin, Frederic S. 1995. Symposium on the monetary transmission mechanism. *Journal of Economic Perspectives* 9(4): 3–10.

Misselden, Edward. 1622. *Free trade or, the meanes to make trade florish*. Printed by John Legatt.

Montemayor, Julián. 1981. Une conjoncture municipale: les propios de Tolède (1540-1600). *Mélanges de la Casa de Velázquez* 17(1): 183–204.

Montemayor, Julián. 1996. *Tolède entre fortune et déclin (1530-1640)*. Presses Univ. Limoges.

Morineau, Michel. 1985. *Incroyables gazettes et fabuleux métaux: les retours des trésors américains d'après les gazettes hollandaises (XVIe-XVIIIe siècles)*, volume 7. Les Editions de la MSH.

Motomura, Akira. 1994. The best and worst of currencies: Seigniorage and currency policy in Spain, 1597–1650. *The Journal of Economic History* 54(1): 104–127.

Nakamura, Emi, and Jón Steinsson. 2013. Price rigidity: Microeconomic evidence and macroeconomic implications. *Annual Review of Economics* 5(1): 133–163.

Nakamura, Emi, and Jón Steinsson. 2018a. High-frequency identification of mone-

tary non-neutrality: The information effect. *The Quarterly Journal of Economics* 133(3): 1283–1330.

Nakamura, Emi, and Jón Steinsson. 2018b. Identification in macroeconomics. *Journal of Economic Perspectives* 32(3): 59–86.

Newey, Whitney K., and Kenneth D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 703–708.

Nightingale, Pamela. 1990. Monetary contraction and mercantile credit in later medieval England. *The Economic History Review* 43(4): 560–575.

Nogues-Marco, Pilar. 2011. The microeconomics of bullionism: Arbitrage, smuggling and silver outflows in Spain in the early 18th century. Universidad Carlos III de Madrid. Working Papers in Economic History, WP 11-05.

Odriozola Oyarbide, María Lourdes. 2002. *Construcción naval en el País Vasco, siglos XVI-XIX. Evolución y análisis comparativo.* San Sebastián: Diputación Foral de Gipuzkoa, Departamento de Economía y Turismo, 2002.

Ogilvie, Sheilagh. 2011. *Institutions and European trade: Merchant guilds, 1000–1800.* Cambridge University Press.

Ogilvie, Sheilagh. 2014. The economics of guilds. *Journal of Economic Perspectives* 28(4): 169–92.

Oldland, John. 2010. The allocation of merchant capital in early Tudor London. *The Economic History Review* 63(4): 1058–1080.

Palma, Nuno. 2020. Historical account books as a source for quantitative history. In *History and economic life: A student's guide to approaching economic and social history*, edited by Christ, G., and P. Roessner. Routledge.

Palma, Nuno. 2022. The real effects of monetary expansions: evidence from a large-scale historical experiment. *The Review of Economic Studies* 89(3).

Pasten, Ernesto, Raphael Schoenle, and Michael Weber. 2020. The propagation of monetary policy shocks in a heterogeneous production economy. *Journal of Monetary Economics* 116: 1–22.

Pérez, Francisco José Aranda. 1992. Los mercaderes de Toledo en el Seiscientos:

bases económicas y status sociopolítico. *Investigaciones históricas: Época moderna y contemporánea* (12): 71–96.

Phillips, Carla Rahn. 1987. Time and duration: A model for the economy of early modern Spain. *The American Historical Review* 92(3): 531–562.

Phillips, Carla Rahn. 1990. The growth and composition of trade in the Iberian empires, 1450-1750. In *The rise of merchant empires: Long-distance trade in the early modern world 1350–1750*, edited by Tracy, James D., 34–101. Cambridge University Press.

Phillips, Carla Rahn, and William D. Phillips. 1997. *Spain's golden fleece: Wool production and the wool trade from the Middle Ages to the nineteenth century.* Johns Hopkins University Press Baltimore.

Pike, Ruth. 1965. The Sevillian nobility and trade with the new world in the sixteenth century. *Business History Review* 39(4): 439–465.

Pike, Ruth. 1966. *Enterprise and adventure.* Cornell University Press.

Piketty, Thomas. 2014. *Capital in the twenty-first century.* Harvard University Press.

Potter, John Stauffer. 1972. *The treasure diver's guide.* New York: Doubleday.

Ramey, Valerie A. 2016. Macroeconomic shocks and their propagation. In *Handbook of Macroeconomics*, volume 2, 71–162. Elsevier.

Reher, David S., and Esmeralda Ballesteros. 1993. Precios y salarios en Castilla la Nueva: la construccion de un indice de salarios reales, 1501–1991. *Revista de Historia Económica-Journal of Iberian and Latin American Economic History* 11(1): 101–151.

Romer, Christina D., and David H. Romer. 1989. Does monetary policy matter? A new test in the spirit of Friedman and Schwartz. *NBER Macroeconomics Annual* 4: 121–184.

Romer, Christina D., and David H. Romer. 2004. A new measure of monetary shocks: derivation and implications. *The American Economic Review* 94(4): 1055–1084.

Rotemberg, Julio J., and Michael Woodford. 1999. The cyclical behavior of prices and costs. *Handbook of Macroeconomics* 1: 1051–1135.

Sargent, Thomas J., and François R. Velde. 2002. *The big problem of small change*. Princeton University Press.

Schneider, Jürgen, Oskar Schwarzer, and Markus A. Denzel. 1994. *Währungen der Welt III: Europäische Wechselkurse im 17. Jahrhundert*. Franz Steiner Verlag Stuttgart.

Schneider, Jürgen, Oskar Schwarzer, Friedrich Zellfelder, and Markus A. Denzel. 1992. *Währungen der Welt VI: Geld und Währungen in Europa im 18. Jahrhundert*. Franz Steiner Verlag Stuttgart.

Schularick, Moritz, Lucas ter Steege, and Felix Ward. 2021. Leaning against the wind and crisis risk. *American Economic Review: Insights* 3(2): 199–214.

Smets, Frank, and Rafael Wouters. 2007. Shocks and frictions in US business cycles: a Bayesian DSGE approach. *The American Economic Review* 97(3): 586–606.

Stiglitz, Joseph E., and Andrew Weiss. 1981. Credit rationing in markets with imperfect information. *The American Economic Review* 71(3): 393–410.

Stock, James H., and Mark W. Watson. 2018. Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal* 128(610): 917–948.

Tortella, Gabriel, J. García Ruiz, and José Luis García Ruiz. 2013. *Spanish money and banking: A history*. Springer.

Uhlig, Harald. 2005. What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics* 52(2): 381–419.

Vaquero, Ángel Santos. 2020. Complejidad del Putting-Out-System en Toledo y su jurisdicción (siglos XVII-XVIII). *Revista electrónica de Historia Moderna* 10(40): 204–227.

Velde, François R. 2009. Chronicle of a deflation unforetold. *Journal of Political Economy* 117(4): 591–634.

Velde, François R., and Warren E. Weber. 2000. Fiat money in 17th century Castile. Federal Reserve Bank of Minneapolis and University of Minnesota Minneapolis.

Walton, Timothy R. 1994. *The Spanish treasure fleets*. Pineapple Press Inc.

Appendix contents

A Data **226**

A.1 Precious metal losses 226

A.2 Correlated shocks 234

A.3 Price and wage indices 241

A.4 Lending Rates 244

B Additional results **258**

B.1 International trade 260

B.2 Changes in the silver value of the unit of account 263

B.3 Labor flows 265

B.4 Robustness checks 267

A. Data

A.1. Precious metal losses

Selection of maritime disasters, losses, and salvaging

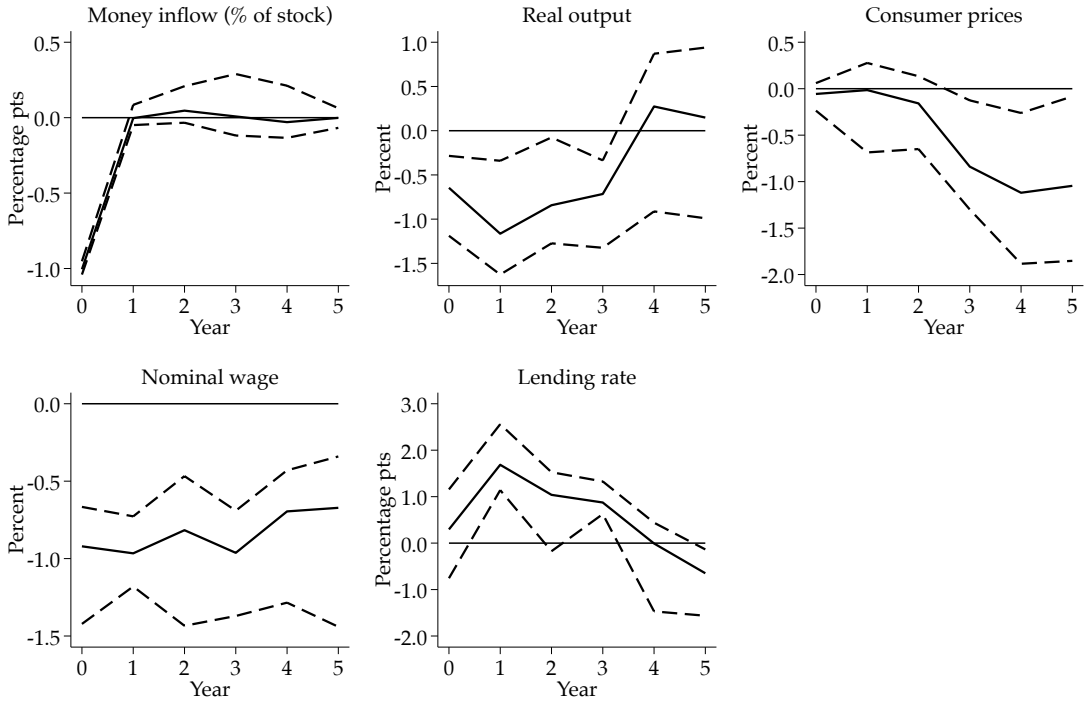
Our selection of precious metal losses is based on an extensive review of the existing literature on maritime disasters in the Spanish Empire. Table A.1 presents an overview of source materials. We include only precious metal shipments that were bound for Spain. As a consequence, all but one of the maritime disasters listed in Table A.1 took place in the Atlantic Ocean. The only exception is the sinking of the Santa Leocadia, which in 1800 carried precious metals from Peru to Panama. While 130,000 pesos of its treasure constituted a fiscal transfer between Spanish American colonies, the remainder was to be transported over the Isthmus of Panama and loaded onto ships destined for Spain.

While our sample of maritime disasters includes all known major events, we are likely to miss some incidents in which small individual merchant vessels sank with an unregistered amount of coins. The precious metal loss associated with these must have been small to escape widespread public attention. Our sample contains some such small disaster events which illustrate the type of losses that could have escaped widespread notice. For example, the 1550 wreck of the merchant caravel *Santa Maria de la Piedad* near Terceira Island (Azores) added around 100,000 pesos ($\approx$ 2.5 tonnes) to the loss for that year.

To address the sensitivity of our results with respect to the selection of maritime disaster events, we calculate impulse response functions based on subsets of our maritime disaster list. Each subset leaves out three random disaster events. The resulting range of IRF estimates depicted in Figure A.1 shows that the baseline results' characteristics are robust to the exact selection of disaster events.

Our analysis accounts for salvaging. How quickly precious metals could be salvaged depended on the accessibility of the sunken ship. When a ship sank in shallow waters, its treasure was accessible to divers and could thus be quickly salvaged. When sunken treasure was salvaged within weeks after the disaster event we do

Figure A.1: Disaster subsets: negative 1 percentage point money inflow shock

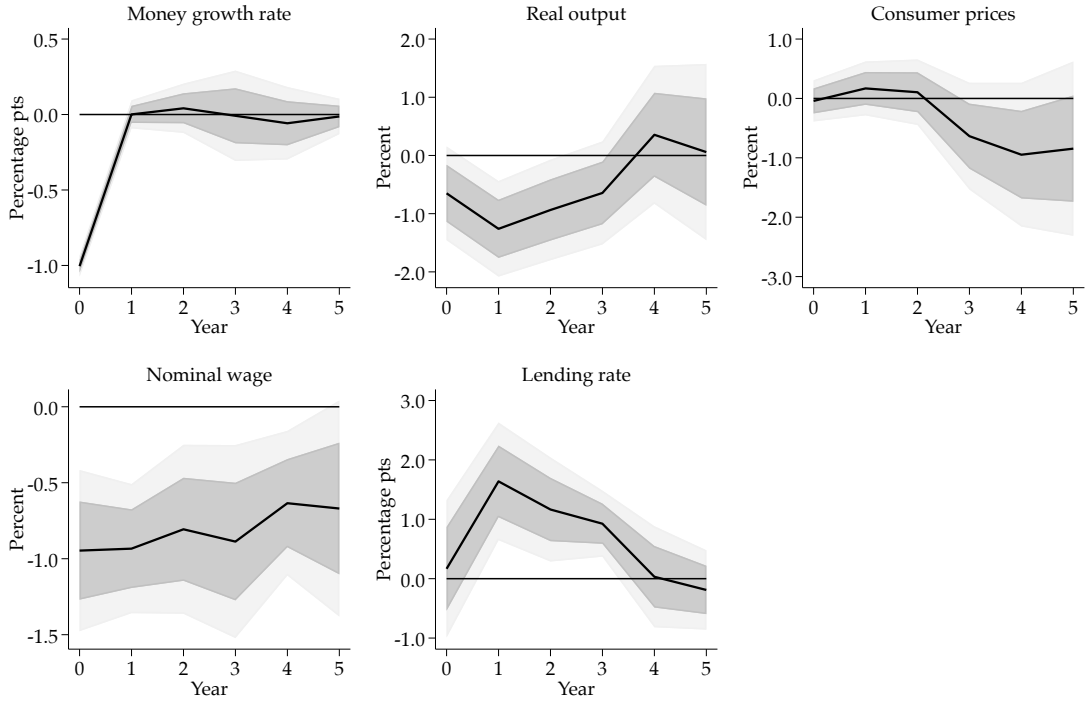


Notes: Solid line – full sample response. Gray area – 1st to 99th percentile (pointwise) IRF range based on 100 disaster subsets, each leaving out three disasters.

not include it in our loss series (e.g. the 1711 Nueva España fleet), nor do we include such immediate salvaging into our salvage series. Many salvaging operations, however, lasted longer, because shipwrecks first had to be dragged to more shallow waters, or because bad weather denied the salvaging party access to the wreck site. As soon as salvaging operations dragged on for several months we include the associated disasters' losses into the loss series, and the subsequently salvaged amounts into the salvage series. Few salvaging operations lasted more than a year, because shifting sands commonly hindered further salvaging efforts (Chen et al., 2021, p.4). Salvaged precious metals were typically transported to Spain on board of the following year's treasure fleet (Walton, 1994, p.166). In our baseline analysis the first lag of the salvage variable thus controls for the effect of the arrival of salvaged precious metals on the Spanish outcome variables of interest.[1]

[1]To the extent that actually salvaged amounts did not coincide with merchants' expectations thereof, the year after a maritime disaster shock could give rise to a secondary salvaging news shock. This expectation error is not accounted for in the baseline specification, whose disaster loss and salvaging regressors only account for the actually realized net loss. In principle, this could give rise to an omitted variables problem (the expectation error is omitted) and a measurement error problem (ex ante expected net losses are confounded with ex post realized net losses). On

Figure A.2: Impulse responses for negative 1 ppt money inflow shock (long salvaging)

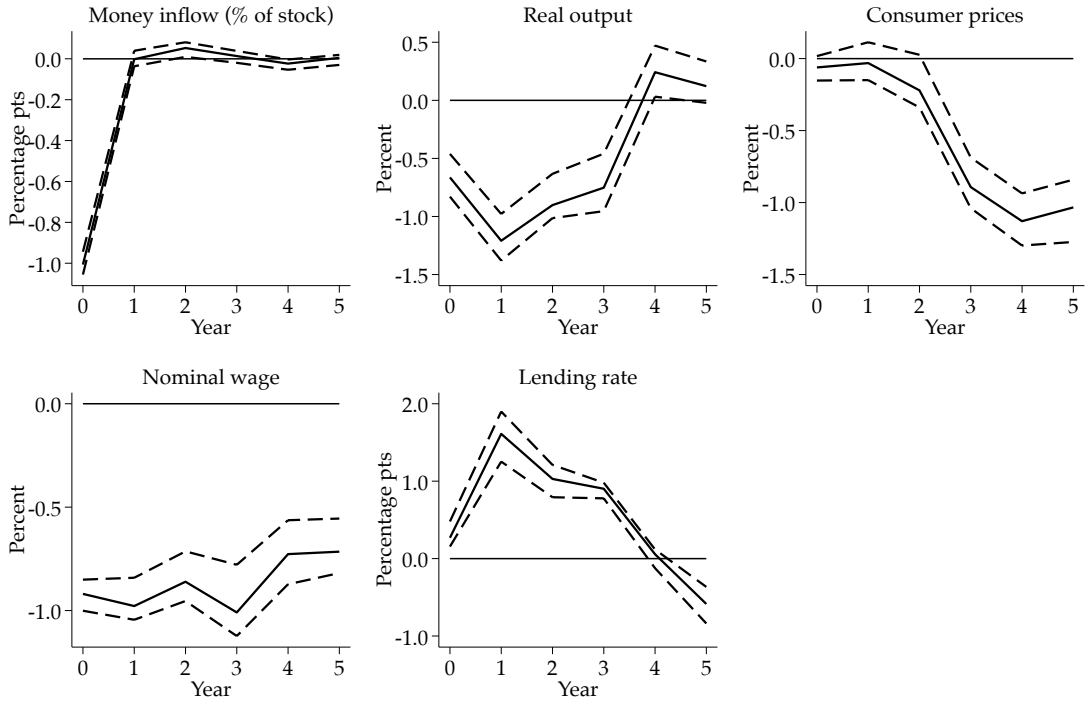


Notes: Results based on specification that separately controls for the amount of salvaged precious metals from the 1656 and 1715 wreck sites. In both cases, salvaging operations extended beyond one year.

We are aware of two maritime disaster events for which salvaging operations extended beyond one year. First, the salvaging of the 1715 loss lasted until 1718. The return of successive salvaging operations, however, declined rapidly (Peterson, 1975, p.369). Second, Potter (1972, p.219) describes the repeated salvaging operations that took place after the 1656 disaster until shifting sands eventually also frustrated this effort (Marx, 1987, p.316). Our baseline analysis does not account for the additional delay in the arrival of the salvaged precious metals associated with the maritime disasters of 1656 and 1715. Figure A.2, however, shows that the baseline results are robust to adding separate controls for the amounts salvaged in the aftermath of these two more protracted events.

both accounts, however, the large variance of the loss measure limits bias (see the next section for quantitative measurement error examples). Also, the salvaging expectation error was likely to be moderate because uncertainty about the amount of treasure that could be salvaged was largely determined by a shipwreck's location. The location in turn was either known, in which case salvaging operations were undertaken, or not, in which case no salvaging occurred.

Figure A.3: Smuggling error: negative 1 percentage point money inflow shock



Notes: Solid line – baseline IRFs. Dashed lines – min-max IRF range calculated from 100 shock series with random error term added to non-zero shock values. Error term $\sim \mathcal{N}(0, 0.63^2)$.

Smuggling and measurement error

Because private treasure flows were taxed upon arrival in Spain there existed an incentive to not register them. The data collected by Mangas (1989, p.316) and Morineau (1985, pp.242 and 375) suggests that, on average, smuggling amounted to 30% of registered shipments in the 16th century, 67% in the 17th century, and 47% in the 18th century. We apply these smuggling shares whenever no better information on treasure size is available (e.g. due to a complete salvaging of the treasure at a later time). The smuggling rate's standard deviation equals 7 percentage points (according to 27 observations for the 17th century (Mangas, 1989)). Given an average inflow-to-stock ratio of 9% this translates into a 0.63 percentage point standard error for the inflow shock measure.

While the appropriate non-classical measurement error case is reported in the next paragraph, the magnitudes involved can be more transparently laid out with classical measurement error calculus. The variance in the shock measure equals 13.56 ($= \sigma_x^2$). The standard deviation of the measurement error that derives from

uncertain smuggling rates amounts to 0.63 percentage points. Given that only 33 out of our 280 sample years (12%) had maritime disasters and thus were affected by the measurement error, this implies a measurement error variance, σ_u^2, of 0.0057 ($= 0.12^2 \cdot 0.63^2$). For these quantities, the classical attenuation factor, $-\frac{\sigma_u^2}{\sigma_u^2+\sigma_x^2}$, amounts to just -0.04%. Put differently, because identification largely rests on the difference between large losses in disaster years and zero losses in non-disaster years, the 0.63 percentage point standard deviation error in the shock measure does not matter much.[2]

To cover the non-classical measurement error case, we add a random error term to each non-zero shock measure. The random term is centered around 0 and has a standard deviation of 0.63 percentage points – corresponding to the 0.63 percentage point standard deviation in the money inflow-to-stock ratio implied by the smuggling rates reported in Mangas (1989). We then calculate 100 different IRFs based on 100 shock series with random errors added. The distance between the resulting IRF range and the baseline IRF estimate conveys an idea of the magnitude of the bias that this measurement error can generate. The resulting IRF range depicted in Figure A.3 confirms that any bias resulting from the smuggling derived measurement error is small.

The maritime disasters in 1591 does not lend itself to the standard treatment described above. This is because it is not associated with the loss of any officially registered precious metals. However, according to Walton (1994) the multiple shipwrecks of this year led to the loss of 10 million pesos. This is the figure we use in our baseline analysis. The size of this loss may in part be explained by the hibernation of the previous year's fleet in Havanna. The 1591 fleet thus likely carried more than one year's worth of unregistered treasure. Nevertheless, because of the uncertainty which the 1591 losses are shrouded in, we also conduct a robustness check in which we exclude this event from the analysis. Figure A.4 shows that our results are robust to the exclusion of this event.

[2]This assumes that measurement error is zero for non-disaster years. In light of the above discussion of the incomplete coverage of small disaster events this might be an exaggeration. Assuming non-disaster years are afflicted by the same measurement error as disaster years results in a still small attenuation factor of -4%.

Figure A.4: Excluding 1591 disaster: negative 1 percentage point money inflow shock

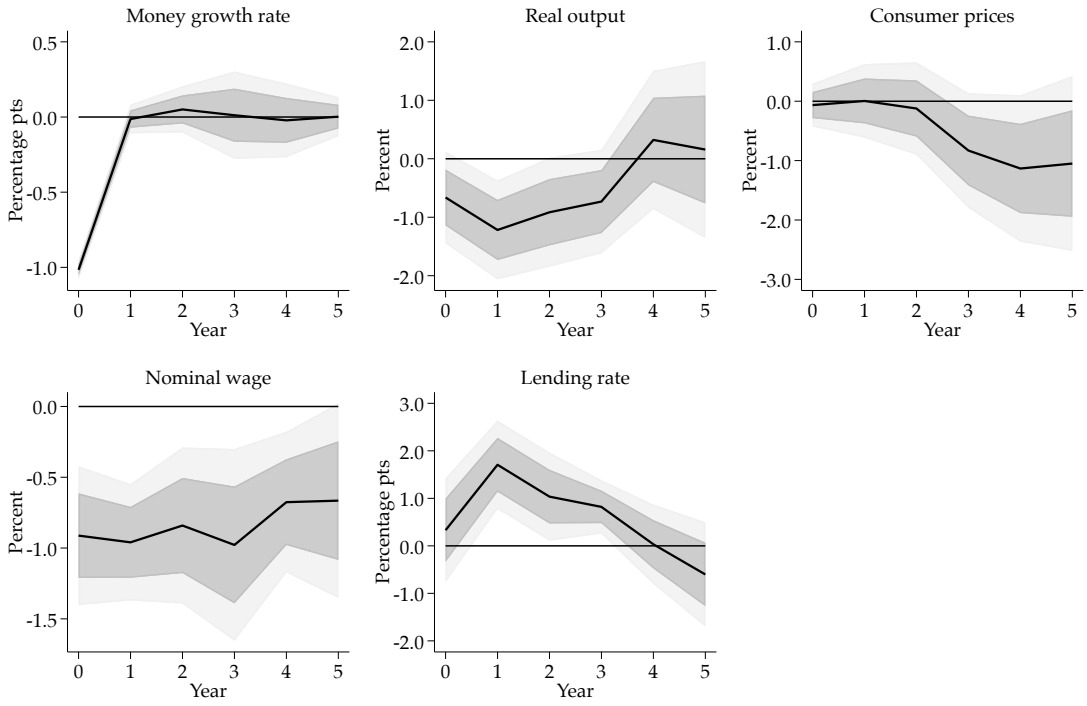


Table A.1: Maritime disaster event sources

Year	Source	Silver equivalent	Notes
1537	Walton (1994, p.24)[1]	18,258 kg	around 500,000 pesos captured
1550	Potter (1972, pp.215,299)[2]	8,079 kg	more than 300,000 pesos sunken
1554	Walton (1994, p.61)	73,031 kg	almost 3 million pesos in treasure sunken, about half salvaged
1555	Potter (1972, p.160), Bonifacio (2010)	12,780 kg	500,000 pesos sunken
1563	Earle (2007, pp.9ff.), Bonifacio (2010)	24,430 kg	around 1 million pesos sunken
1567	Walton (1994, p.61)	109,547 kg	more than 4 million pesos sunken; salvaging failed
1591	Walton (1994, p.83)	255,610 kg	10 million pesos sunken; about 3/4 salvaged
1605	Walton (1994, pp.83-84)[3]	204,488 kg	8 million pesos sunken; salvaging failed
1621	Marx (1987, p.302.)	382 kg	around 15,000 pesos in treasure sunken; most of it salvaged
1622	Potter (1972, pp.215ff.), Marx (1987, pp.200ff.)	188,951 kg	more than 7 million pesos in treasure sunken; partly salvaged
1623	Marx (1987, p.202)	76,345 kg	about 3 million pesos sunken
1624	Mangas (1989, p.318)	51,122 kg	2 million pesos sunken
1628	Potter (1972, p.160), Marx (1987, p.248)	30,538 kg	around 1.2 million pesos sunken, largely salvaged
1628	Venema (2010, p.213)	80,660 kg	177,000 pounds of silver and 66 pounds of gold captured
1631	Marx (1987, p.424)	58,169 kg	more than 2 million pesos sunken; 1 million pesos salvaged
1631	Marx (1987, p.249)	150,241 kg	more than 5.5 million pesos sunken; very little salvaged
1634	Sandz and Marx (2001, p.129)	7,635 kg	around 300,000 pesos in treasure sunken, partly salvaged
1641	Mangas (1989, p.318)	76,683 kg	3 million pesos sunken
1654	Earle (2007, p.83)	255,610 kg	10 million pesos sunken; 3.5 million pesos recovered
1656	Potter (1972, p.432)	173,815 kg	2 million pesos captured; around 5 million pesos sunken

1656	Walton (1994, pp.128, 140)	127,805 kg	5 million pesos sunken, 2.5 million pesos salvaged
1682	Bueno (1996, p.84)	153,366 kg	6 million pesos sunken
1698	www.todoavante.es	161,540 kg	around 6.5 million pesos sunken, 6 million pesos salvaged
1702	Kamen (1966)	7,350 kg	around 80,000 pesos captured and sunken by British and Dutch
1708	Phillips (2007b, pp.46,181), Sedgwick (1970)[4]	286,283 kg	11 million pesos sunken and 200,000 pesos captured
1715	Marx (1987, p.431)	309,972 kg	12 million pesos sunken; around 5 million salvaged
1730	Walton (1994, p.166)	139,556 kg	more than 5.5 million pesos sunken; partly salvaged
1733	Fine (2006, p.153)	311,908 kg	around 12.5 million pesos sunken; almost all salvaged
1750	Putley (2000), Amrhein (2007, ch.1)[5]	10,321 kg	272,000 pesos sunken; 14,467 pesos salvaged; 144,000 pesos captured
1752	Marx (1987, p.443)	49,620 kg	2 million pesos sunken, mostly salvaged
1753	Marx (1987, p.443)	38,134 kg	1.5 million pesos sunken
1762	The Gentleman's and London Magazine (1763, p.528)	62,895 kg	Around 2.5 million pesos captured
1786	Potter (1972, pp.349ff.)	185,716 kg	7.5 million pesos sunken, mostly salvaged
1800	Bravo (2010)	51,264 kg	Around 2 million pesos sunken; partly salvaged
1802	Sandz and Marx (2001, p.218)	13,742 kg	Around 0.5 million pesos sunken
1804	Cobbett (1804, p.663)	113,084 kg	1.5 million pesos sunken; 3 million pesos captured

Notes: The loss figures listed in the table account for the loss of unregistered silver (see text). [1]Loss calculated as American production destined for Spain from TePaske (2010), minus 10% American retention rate, minus the Spanish arrival figure provided in the source. [2]Exact disaster months unknown. Disaster months for the two 1550 events are set to July and September according to the disaster locations and the return voyage schedule by Chaunu and Chaunu (1977, pp.229ff.). [3]Also lists other disaster events which are not included here, because they did not result in the loss of money; e.g. because the disaster occurred on outbound voyages or because an accessible wreck site allowed for rapid salvaging. [4]URL: http://www.historyofparliamentonline.org/volume/1715-1754/member/wager-sir-charles-1666-1743#footnoteref3_g4iwhgx. [5]Loss associated with the ship "El Salvador" corrected to 240,000 pesos; correction confirmed with the author.

A.2. Correlated shocks

In this section we discuss the non-monetary losses and fiscal losses that were associated with maritime disasters. Both can be viewed as correlated shocks that could interfere with our identification strategy. In this section we show that neither can account for the response of Spain's economy to maritime disaster events, suggesting that our IRF estimates indeed describe monetary effects.

A.2.1 The Crown's finances and sovereign risk

Although 85 to 95% of precious metal inflows from the colonies were privately owned (García-Baquero González, 2003; Costa et al., Forthcoming), the remainder nevertheless could constitute an important source of revenue for the Crown.[3] In the first half of the 16th century only 4% to 10% of crown revenues derived from American precious metals. Under Phillip II this figure rose up to 20%. In the 17th century, the share of American precious metals in crown revenues wanes again to around 11% (Ulloa, 1977; Drelichman and Voth, 2010; Comín and Yun-Casalilla, 2012; Álvarez-Nogal and Chamley, 2014). Thus, the loss of silver in maritime disasters could put stress on the Crown's finances, making a sovereign debt crisis more likely.

Table A.2: Sovereign debt crises and maritime disasters

Debt crisis year	1557	1560	1575	1596	1607	1627	1647	1652	1662	1686	1700
Nearest disaster year	1555	1555	1567	1591	1605	1624	1641	1641	1656	1682	1698
Distance (in years)	2*	5*	8	5*	2*	3*	6	11	6	4*	2*

Notes: * indicates sovereign debt crises that occurred within five years after a maritime disaster. Sovereign debt crisis years from Pike (1966), Artola (1982), Homer and Sylla (1991), Álvarez-Nogal (2003), Reinhart and Rogoff (2009), and Álvarez-Nogal and Chamley (2014).

The Royal Treasury's revenues and expenditures were small compared to the Spanish economy – on average 3% between 1555 and 1596, and around 5% at the end of our sample (Ulloa, 1977; Drelichman and Voth, 2010; Barbier and Klein, 1981). Furthermore, a substantial share of this was not spent in Spain, but for military purposes abroad. Together with the fact that the Crown's silver revenues

[3] Only in the late 18th century does the Royal Treasuries' share of precious metal remittances increase to above 20%. Occasionally, however, the Crown sequestered part of the privately owned treasure through forced loans that were reimbursed later with additional interest (Sardone, 2019).

amounted to no more than one fifth of all government revenues, this suggests that maritime disasters only had a small direct effect on government spending within Spain.

Maritime disasters, however, could indirectly affect government spending, by raising the spectre of default and spooking the Spanish Crown's creditors. Despite its small public sector share, the Spanish public debt-to-GDP ratio was large – at times exceeding 50% according to Álvarez-Nogal and Chamley (2014).[4] Sovereign debt crises therefore might have negatively affected the Spanish economy through a reduction in public debt holder wealth. It is important to note, however, that during sovereign debt crises not all debt was written down. In fact, actual debt write-downs may have been quite small (Álvarez-Nogal and Chamley, 2014).

We analyze whether the real effects of money losses were the consequence of a monetary transmission through the Crown's finances in two steps. First, to see whether maritime disasters provoked sovereign debt crises, Table A.2 lists the dates of sovereign defaults, together with the closest preceding maritime disaster date. Indeed, seven out of the eleven sovereign debt crises in our sample occurred within five years after a maritime disaster. However, this also implies that 26 maritime disaster years were not followed by a sovereign debt crisis.

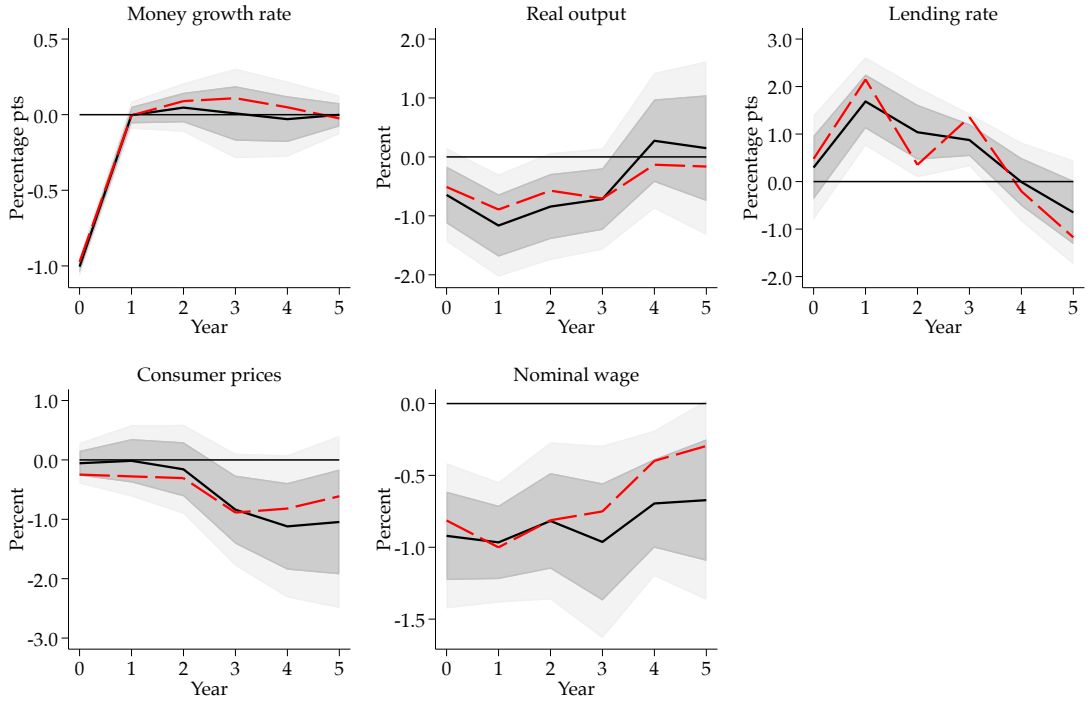
Second, to analyze the extent to which the money non-neutrality result is driven by these seven sovereign debt crises we re-estimate our baseline IRFs with an adjusted specification, which allows money losses to develop different effects on the economy if they are followed by a sovereign debt crisis. In particular, we amend the baseline specification from section 2.4 in the following way:

$$(Y_{t+h} - Y_{t-1})/Y_{t-1} = \alpha_h + \beta_h S_t + \delta_h S_t * C_t + \eta_h C_t + \gamma_h X_t + u_{t+h},$$

where C_t is a binary indicator that equals 0 except in sovereign debt crisis years and the preceding five years. This purges sovereign debt crisis effects over the full five year horizon over which the IRF extends. Figure A.5 shows that the adjusted IRFs closely resemble the baseline IRFs, indicating that the real effects of money supply

[4] Accordingly, debt servicing costs used up a large part of government revenues.

Figure A.5: Debt crises: Responses to negative 1 percentage point money inflow shock

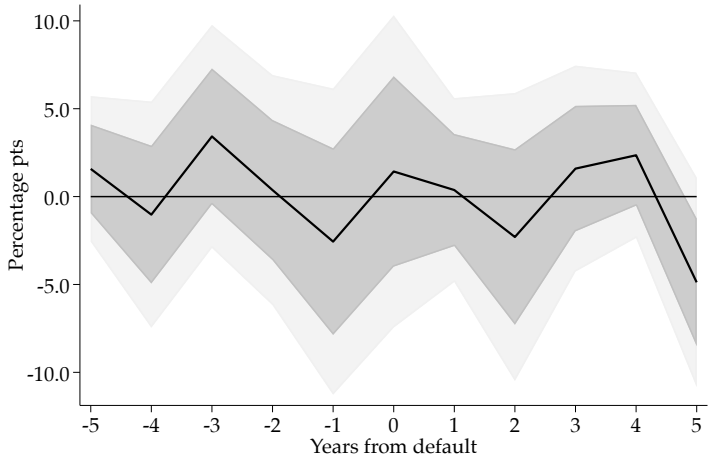


Notes: Solid black line – baseline specification. Dashed red line – specification controlling for separate effect of sovereign debt crises. Gray areas – 1 standard deviation and 90% confidence bands.

shocks did not significantly differ according to whether or not they were associated with sovereign debt crises. Thus, the Crown's finances do not appear to have played a major role in the transmission of money supply shocks.

Another way in which the Crown's finances could drive our non-neutrality result is if private sector lending rates were influenced by sovereign risk. This is often the case today, where private sector credit conditions are tied to sovereign credit worthiness, and sovereign bond rates act as a baseline for private sector interest rates (CGFS, 2011; IMF, 2013; Bocola, 2016). In pre-modern economies, however, private interest rates typically were lower than public rates, as lending to monarchs was considered to be more risky (de Mercado, 1587; Henriques and Palma, 2020, p.307). To see whether private sector lending rates in our sample were affected by sovereign risk, we analyze the behavior of the lending rate series around sovereign default dates. To this end, Figure A.6 shows average annual lending rate changes in the -5 to +5 year window around sovereign default years. We find little indication that lending rates budged either in the up-run or the aftermath of a sovereign default.

Figure A.6: Lending rates around sovereign defaults



Notes: Solid black line – mean estimate of annual lending rate change. Gray areas – 1 standard deviation and 90% confidence bands. Average lending rate changes from 0 to +5 are based on a forward projection equivalent to baseline specification (eq. 1), but replacing the money loss regressors with a binary sovereign default indicator. Lending rate changes from -1 to -5 are based on a symmetric backward projection.

Overall, we find no evidence to suggest that the Crown's public finances were an important mediator along the causal chain that translated money losses into real output effects.

A.2.2 Non-monetary wealth

The maritime disasters that resulted in money losses also entailed the loss of non-monetary wealth, such as ships. In this section we document that non-monetary losses were small compared to the monetary ones. Table A.3 provides estimates of the value of the non-monetary wealth loss associated with maritime disasters. In each of the 33 maritime disaster years between 1 and 30 ships were lost. The resale value for ships lay in the 1600 to 8000 peso range (Carrasco González, 1996, p.156).[5] Using the lower value, this implies that even in the maritime disaster with the largest number of ships lost, the overall ship wealth loss amounted to no more than 0.9% of the precious metal loss. For the high re-sale value of 8000 pesos, the corresponding fraction is 4.5%. The 4.5% value, however, corresponds to an outlier

[5]Few ships made more than four transatlantic return voyages. After that, they often were sold and used in calmer waters (Carrasco González, 1996, p.156). Repairing and outfitting worn out ships after a long-distance voyage cost nearly as much as buying a new ship (Gelderblom et al., 2013).

event in which 30 ships were lost, whereas the median ship value loss in our sample amounts to 0.8% of the precious metal loss – even for the high-ship value scenario.[6]

Table A.3: Value of non-precious metal losses

Precious metal share (% of import value)	Ship loss (per disaster)	Ship value		Life loss	
		Low	High	Low	High
Until late C18th:	1 to 30	*Pesos per ship:*		*Crew per 100t ship:*	
80%	(median=3)	1600	8000	30	100
From late C18th:		*Percent of metal loss:*		*Percent of labor:*	
> 55%		<0.9%	<4.5%	<0.09%	<0.3%

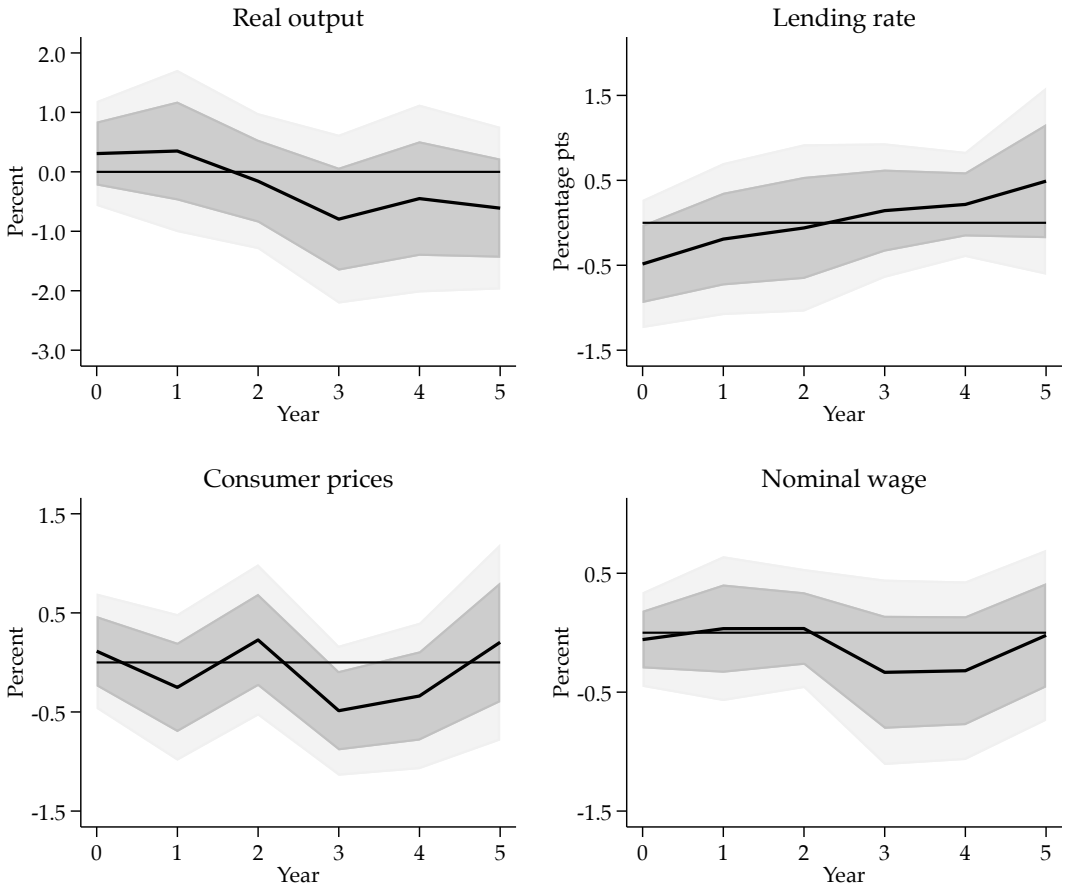
Sources: Precious metal import share from Fisher (1985), García-Baquero González (2003), and Cuenca-Esteban (2008). Ship values from Carrasco González (1996, p.156). The ship values are resale values that reflect ship depreciation. *Crew per ship*: Crew per 100 tonnes from de Vries (2003), quinquennial data on average ship tonnage from Phillips (1990). *Percent of labor*: Population data from Álvarez-Nogal and Prados de la Escosura (2013). Working-age (16-50) fraction (49.6%) from 1787 census described in Martín (2005).

The amount of lives lost as a fraction of the Spanish working population was small, never amounting to more than 0.3%, even when large numbers of ships sank, and even when assuming very large crews. For more moderate crew size estimates and median ship loss numbers this fraction is negligibly small.

Precious metals comprised 80% of the total value of imports up to the late 18th century (García-Baquero González, 2003; Fisher, 2003). Only in the late 18th century did the share of non-metal colonial goods – such as tobacco, cacao, and sugar – increase significantly above 20%. Fisher (2003) documents that between 1782 and 1796 43.6% of the value of the commodities imported into Barcelona and Cádiz pertained to non-metal colonial goods. To address the concern that at the end of our sample the monetary shock measure becomes diluted with a non-monetary wealth shock, we conduct a subsample analysis in which we end our sample in 1780. The results are very similar (Figure B.12 in the Appendix).

[6]The majority of vessels in the Atlantic had less than 400 tonnes capacity (Carrasco González, 1996, p.157). However, ships carrying the treasure often were from the military squadron that protected the fleets – two to six military ships with a larger tonnage. Such large ships were more expensive to build. Phillips (2007a) documents the construction costs for two such 1000 tonne ships in the late 17th century. Building them anew cost 63,419 pesos. Assuming a resale value of 20,000 for such ships, and assuming all sinking ships are of this type, the median value of ship losses across disaster events would add up to 1.9% of the value of the precious metal loss.

Figure A.7: Economic effects of maritime disasters on outbound voyage, per loss of 1 ship



Notes: Solid black line – mean IRF. Gray areas – 1 standard deviation and 90% confidence bands. The graph shows IRFs estimated on the basis of baseline specification 1 augmented by an indicator of the number of ships lost on outbound voyages.

Finally, we use maritime disasters on outbound voyages from Spain to America to assess the economic impact of non-monetary maritime disasters on the Spanish economy.[7] Maritime disasters on outbound voyages led to similar non-monetary losses as maritime disasters on inbound voyages. The key difference is that maritime disasters on the outbound voyage did not result in the loss of large quantities of money. Between 1531 and 1810, we record 14 instances of outbound fleet voyages that ended in disaster. The number of ships lost in such events ranged from 1 to 16. In the large events more than a thousand lives and cargo worth more than 2 million

[7]Most sources that provide historical accounts of maritime disasters on inbound voyages also do so for outbound voyages. In particular we use Marx (1987), Walton (1994), and the primary archival data from `todoavante.es`. We apply the same dating convention as for the inbound disaster shocks, i.e. we assume news about a maritime disaster takes time to travel to Europe.

pesos was lost (equivalent in value to around 51 tonnes of silver). Despite this, Figure A.7 shows that these events did not leave their mark on aggregate Spanish economic variables. Neither output, prices, wages, nor interest rates responded to the loss of lives, ships, and cargo on outbound shipments. This supports the interpretation of inbound maritime disaster IRFs as monetary effects.[8]

[8]This finding also constitutes prima facie evidence against an independent effect of several psychological forces that could have been triggered by maritime disaster events (e.g. a downturn in economic sentiment driven by superstition or the mourning about relatives lost at sea).

A.3. Price and wage indices

Table A.4: Price indices

Index	Input variables	Coverage	Source
		Foodstuffs	
Grains & Vegetables	barley, wheat, rice, legumes	1531–1800; Madrid, Seville, Valladolid, Barcelona, Valencia	Losa and Zarauz (2021)
Animal products	mutton, beef, eggs, fish, honey	—"—	—"—, Global price and income database `http://gpih.ucdavis.edu/`; compiled from Hamilton (1934, 1936, 1947); Feliu (1991)
Cheese & Wine	wine, cheese	—"—	—"—
		Heating	
Firewood	firewood	—"—	Losa and Zarauz (2021)
Charcoal	charcoal	—"—	—"—
		Consumer & intermediate goods	
Soap	soap	—"—	—"—
Olive oil	olive oil	—"—	—"—
Candles	candles	—"—	—"—
		Textile industry: inputs & intermediate	
Wool (unwashed)	unwashed wool	1558–1808 (with gaps); Casla church, Guadalupe monestary, Segovia Cathedral, Andalusia/Seville, Old Castile	Phillips and Phillips (1997); Drelichman (2009)
Threads	twine, black thread, white thread, yarn	1651–1800; New Castile	Hamilton (1947)
		Textile industry	
Fine cloths	white cloth, serge, linen	—"—	—"—
Coarse cloths	burlap, sackcloth, grogram	1651–1800 (with gaps); New Castile	—"—
		Metals industry	
Iron	iron	1655–1779; Basque region	Arregui (1991)

Iron nails	iron nails	1533–1605 (with gaps); Old Castile-León, New Castile	Global price and income database http://gpih.ucdavis.edu/; compiled from Hamilton (1934, 1936, 1947)

Table A.5: Wage indices

Index	Input variables	Coverage	Source
		Primary sector	
Shepherds	shepherds	1545–1800 (with gaps); Barcelona & surroundings	Feliu (1991)
Unskilled labor	agricultural day laborers, shepherd boys	1546–1809 (with gaps); Andalusia, Catalonia, Old Castile/Palencia	Feliu (1991); Lázaro (2001); Carrión (2002)
		Secondary sector: guild rank	
Masters	carpenters, masons, stonecutters	1531–1798 (with gaps); Andalusia, New Castile, Barcelona, Valencia	Hamilton (1934); Feliu (1991)
Journeymen	carpenters, masons, stonecutter, blacksmiths, cabinetmakers, wheelwrights	1531–1800 (with gaps); Andalusia, New Castile, Valencia	Hamilton (1934, 1947)
Helpers	masons, blacksmiths, cabinetmakers, carpenters	1531–1800 (with gaps); New Castile, Barcelona, Valencia	Hamilton (1934, 1947); Feliu (1991)
		Secondary sector: industries	
Construction workers	unskilled builders	1531–1800; Madrid, Seville, Valladolid, Barcelona, Valencia	Losa and Zarauz (2021)
Weavers	weavers	1531–1650 (with gaps); Valencia	Hamilton (1934)
		Service sector	
Relationship-specific	sacristans, bookkeepers, wardrobe keepers, doormen, stewards	1531–1650; Andalusia, New Castile, Valencia	Hamilton (1934)

Other	nurses, wetnurses, laundress, draymen, basket emptiers, cart loaders, gardeners, cooks, maids	1531–1650; Andalusia, Old Castile-León, Valencia	Hamilton (1934)

A.4. Lending Rates

Extensive datasets on early modern bill prices have been compiled by Da Silva (1969), Schneider et al. (1992, 1994), and Denzel (2010). It is possible to estimate (unobserved) lending rates from (observed) bills of exchange prices (Flandreau et al., 2009a; Nogues-Marco, 2011). In this section we describe two methods of doing so.

A.4.1 Accounting for nominal exchange rate fluctuations

Consider the following non-arbitrage condition: Suppose a London merchant possesses Pound Sterling (£), but wants to obtain 1 Spanish Peso (P) in Seville in one month's time. The merchant can do this in two ways. First, he can buy a bill of exchange on Seville with one month maturity for U_t^L Pounds in London. This bill of exchange entitles the merchant to receive 1 Spanish Peso in Seville in one month. Alternatively, the merchant can purchase Pesos on the spot exchange market at the spot exchange rate $E_t^{P/£}$ (Pesos per Pound Sterling). The merchant can then lend out the obtained Pesos in Seville and after one month reclaim the principal times the Sevillian monthly gross lending rate R_t^S. Thus, to receive 1 Peso in one months time, the second approach requires the merchant to initially buy $1/R_t^S$ Pesos for $1/[E_t^{P/£} R_t^S]$ Pounds.

This example clarifies how the price of a bill of exchange can be interpreted as containing a spot exchange rate component and a lending rate component:

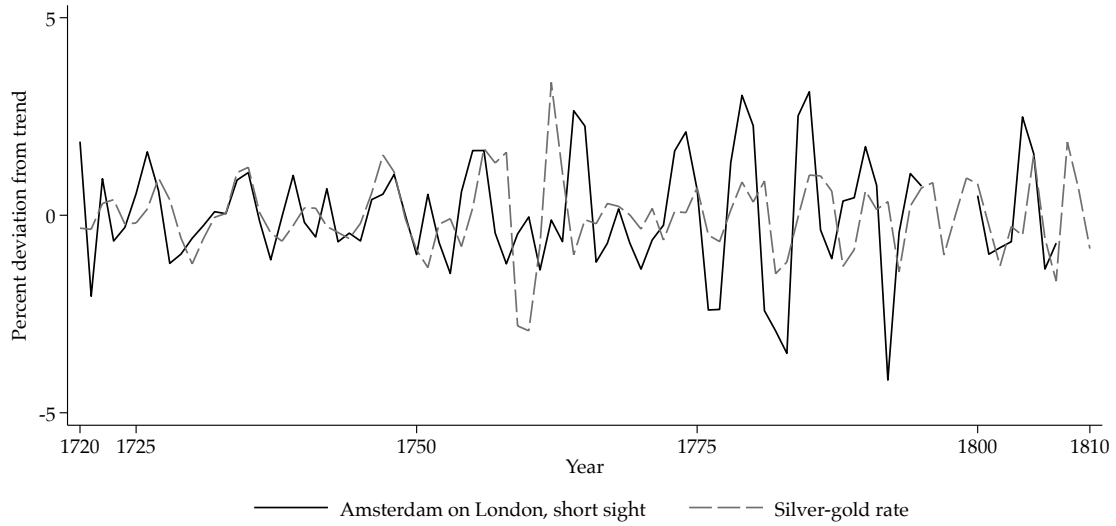
$$U_t^L = 1/[E_t^{P/£} R_t^S]. \tag{A.1}$$

Taking logs and detrending we obtain

$$\hat{r}_t^S = -\hat{u}_t^L - \hat{e}_t^{P/£}, \tag{A.2}$$

where small letters denote logs, and detrended variables are denoted with hats. Bill of exchange prices inform us about $\hat{u}_t^L$. To obtain an estimate for fluctuations in the Sevillian loan rate, $\hat{r}_t^S$, we need information about spot exchange rate fluctuations,

Figure A.8: Exchange rate vs. gold-silver rate



Notes: Exchange rate data based on Amsterdam price of short sight bill on London. Cyclical fluctuations based on HP detrended series ($\lambda = 6.25$).

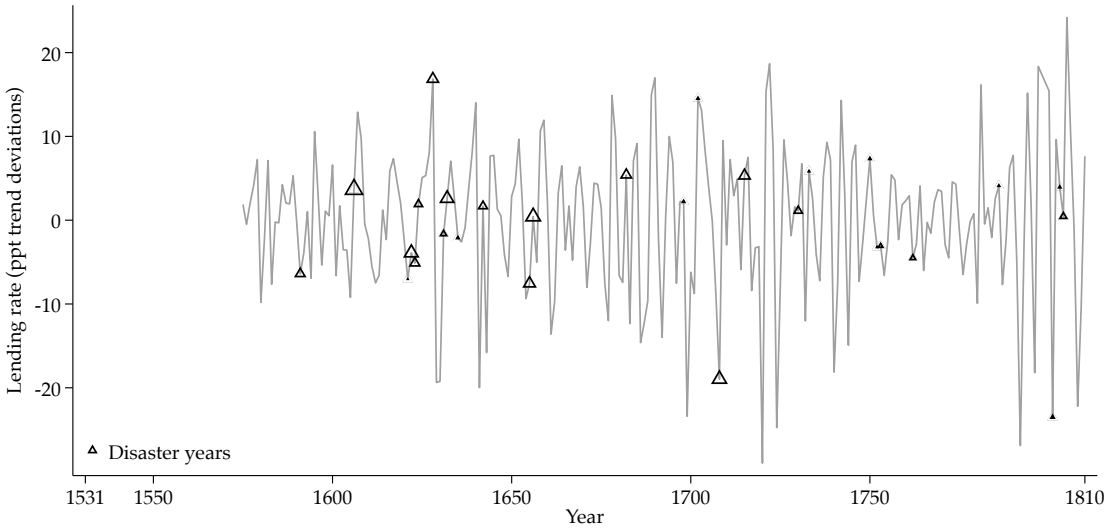
$\hat{e}_t^{P/£}$.

In our sample, currencies were commodity based. Spot exchange rates thus depended on the relative price of different monetary metals – typically gold or silver. This can be seen by comparing fluctuations in silver-gold rates with fluctuations in the price of short sight bills. Short sight bills were redeemable immediately upon presentation to the payer; their price thus resembles a spot exchange rate. Figure A.8 compares fluctuations in the silver-gold rate (Soetbeer, 1879; Spooner, 1972) with fluctuations in the Amsterdam price for a short sight bill on London. Although fluctuations in the silver-gold rate are not perfectly described by fluctuations in the Amsterdam bill price, the two series are highly correlated.

Absent short sight bill prices for Spanish cities, we therefore use silver-gold rates as an indicator for spot exchange rates to calculate Spanish lending rates according to equation A.2. For two silver-based currencies, e.g. Dutch Guilders and Spanish Pesos, this assumes that bill of exchange price fluctuations represent lending rate fluctuations. By contrast, for a gold-based currency, e.g. Pound Sterling after 1717, we subtract silver-gold rate fluctuations from bill of exchange price fluctuations to arrive at Spanish lending rates.

We use equation A.2 to calculate fluctuations in the lending rates for seven

Figure A.9: Lending rate



Spanish cities: Barcelona, Cadiz, Madrid, Medina del Campo, Saragossa, Sevilla, Valencia. We then calculate aggregate Spanish lending rate fluctuations as the average across these seven cities.[9] Figure A.9 depicts the resulting series. The series' standard deviation is 8.8 ppts. For comparison, the standard deviations for analogously detrended long-run returns data for advanced economies in modern times are as follows: 18.3 ppts for the total returns on equity shares, 7.1 ppts for total housing returns, and 6.9 ppts for long-term government bonds (based on data from Jordà et al., 2019, for 17 developed economies from 1870 to 2016).

A.4.2 Transportation times and transaction costs

The non-arbitrage condition described above abstracts from transportation times and transaction costs. International arbitrage in the early modern period, however, required the international transport of letters and money. Travelling from London to Seville by boat took about two weeks. Travelling from Genoa to Seville took somewhat less. Incorporating such frictions into a more realistic non-arbitrage condition gives rise to a non-arbitrage band that implies an upper and lower bound lending rate estimate.

[9] We also account for occasional de- and re-basements by adding the log detrended values of the silver content of Spain's currency to equation A.2 (Allen, 2001; Karaman et al., 2020). We do not account for foreign currency debasements as they are unlikely to be correlated with our shock measure. Even within Spain, debasements did not systematically occur in the immediate aftermath of money shipment losses (section B.2).

15 days constituted a smooth transit from Seville to London, based on an a distance of 1300 nautical miles and an average shipping speed of 4 knots – the typical speed of Spanish mail boats (Kelly and Ó Gráda, 2019). The respective journey time from Genoa to Seville is 11 days for 1100 nautical miles. In the following analysis of transportation times and their implication for lending rate estimates we therefore consider trips of two weeks.[10]

Consider a bill of exchange issued in London at time t that delivers 1 Peso in Seville at time $t+1$. Let the price of this bill of exchange be $U^L_{t,t+1}$ Pounds. First, to derive the upper bound lending rate estimate, consider the following arbitrage strategy:

1. Sell the bill of exchange in London at time t and exchange the obtained amount ($U^L_{t,t+1}$ Pounds) into Pesos at the spot exchange rate $E^{P/£}_t$ (Pesos per Pound Sterling).
2. Travel to Seville with the Peso receipt from step 1 ($U^L_{t,t+1}E^{P/£}_t$), and lend the Pesos out at the Sevillian interest rate for a period lasting from $t+\tilde{t}$ to $t+1$, where $\tilde{t}$ denotes the travel time between London and Seville. We denote the associated Sevillian gross interest rate by $R^S_{t+\tilde{t},t+1}$. We also allow the transportation and exchange of money to be associated with a proportional transportation and transaction cost $\tilde{\theta}$. Thus, the remaining Peso amount that can be lent out in Seville after transport and transaction costs is $U^L_{t,t+1}E^{P/£}_t(1-\tilde{\theta})$.
3. At $t+1$ in Seville, the bill of exchange and the loan mature, implying an expenditure of 1 Peso and a receipt of $U^L_{t,t+1}E^{P/£}_t(1-\tilde{\theta})R^S_{t+\tilde{t},t+1}$ Pesos.

The bill of exchange is too expensive, i.e. $U^L_{t,t+1}$ is too high, if $U^L_{t,t+1}E^{P/£}_t(1-\tilde{\theta})\mathbb{E}_t R^S_{t+\tilde{t},t+1} > 1$. In this case, the above strategy implies an expected profit. Arbitrage activity thus works towards establishing the following inequality:

$$\begin{aligned} & U^L_{t,t+1}E^{P/£}_t(1-\tilde{\theta})\mathbb{E}_t R^S_{t+\tilde{t},t+1} \leq 1 \\ \Leftrightarrow\ & \mathbb{E}_t R^S_{t+\tilde{t},t+1} \leq \frac{1}{U^L_{t,t+1}E^{P/£}_t(1-\tilde{\theta})} \end{aligned}$$

[10] Overland transportation times were similar. With an average speed of 100 miles per day letters could travel faster than anything else in early modern Europe (Parker, 2014, p.299).

With $\Omega_t^U \equiv \frac{\mathbb{E}_t R^S_{t+\tilde{t},t+1}}{R^S_{t,t+1-\tilde{t}}}$, we can write the above inequality as

$$R^S_{t,t+1-\tilde{t}} \leq \frac{1}{U^L_{t,t+1} E_t^{P/£} (1-\tilde{\theta}) \Omega_t^U}.$$

After taking natural logarithms, detrending, and periodizing the lending rate, fluctuations in the upper bound lending rate, $\hat{r}_t^U$, are defined as:

$$\hat{r}^S_{t,t+1} \leq \frac{1}{1-\tilde{t}} \left(-\hat{u}^L_{t,t+1} - \hat{e}_t^{P/£} - \hat{\omega}_t^U \right) \equiv \hat{r}_t^U, \tag{A.3}$$

where small letters denote logs, and detrended variables are indicated with hats. The periodization term $1/(1-\tilde{t})$ is necessary to translate a lending rate that pertains to the sub-period $t, t+1-\tilde{t}$ to a lending rate that covers one whole period, $t, t+1$. Besides the periodization term, the only other term that this equation adds to the non-arbitrage condition without transportation time is ω_t^U – log-detrended Ω_t^U. Ω_t^U can be decomposed into two economically meaningful terms: $\Omega_t^U = \frac{\mathbb{E}_t R^S_{t+\tilde{t},t+1}}{R^S_{t,t+1-\tilde{t}}} = \frac{\mathbb{E}_t R^S_{t+\tilde{t},t+1}}{R^S_{t+\tilde{t},t+1}} \frac{R^S_{t+\tilde{t},t+1}}{R^S_{t,t+1-\tilde{t}}}$. The two terms reflect an expectation error and a travel lag, both of which are inherent in non-instantaneous arbitrage activity. The expectation error stems from merchants forming an expectation at time t about $t+\tilde{t}$ lending rates in Seville. The travel lag term describes how the Sevillian lending rate changes during the journey from London to Seville, i.e. the period during which the arbitrage activity has commenced through the sale of a bill in London, but during which the lending rate for the Sevillian loan has not yet been locked in. These expectation error and travel lag terms reflect the two ways in which London bill prices and spot exchange rates become informationally insufficient as a description of Sevillian lending rates in a non-zero travel time environment.

Unfortunately, Ω_t^U is not observable. However, it is possible to assess how the omission of this term in the calculation of the upper bound lending rate series $\hat{r}_t^U$ affects our IRF estimate for this variable. First, consider the expectation error term. In our setting, the expectation error pertains to a two week window after the money shock hits Seville, and during which news about the disaster has not yet arrived in London. The informational insufficiency of London bill prices with respect

to Sevillian credit market conditions thus only pertains to one out of 12 monthly bill price observations, whose average forms the annual bill price which enters the lending rate series calculation. This annual averaging dilutes the influence of this one monthly bill price. For example, when lending rates in Seville unexpectedly increase by 2 ppts, from 7 to 9%, in response to a maritime disaster in t, and merchants in London continue to expect a 7% Sevillian lending rate until the bad news arrives, then the annual expectation error term will be $0.9984 = (1.07^{1/12} \cdot 1.09^{11/12})/1.09$. Omitting this term implies an underestimate of the actual upper bound lending rate increase in disaster years of around 0.16 ppts ($= 1/0.9984 - 1$). An IRF estimate based on the upper bound lending rate series that neglects the expectation error would thus modestly underestimate the actual on-impact response of the upper bound, thus somewhat narrowing the upper-lower bound IRF range. However, to the extent that credit rationing initially led to a reduction in the quantity of credit supplied, rather than an increase in lending rates, the topic is moot. This is because no systematic expectation error with respect to Sevillian lending rates emerges.

Second, consider how the omission of the travel lag term $\frac{R^S_{t+\tilde{t},t+1}}{R^S_{t,t+1-\tilde{t}}}$ affects the upper bound lending rate IRF estimate. In our setting, when a money shock hits at t the travel lag term turns less than 1 to the extent that the Sevillian lending rate immediately after the shock $t, t+1-\tilde{t}$ exceeds the lending rate two weeks further down the road. Omitting this term in the calculation of the upper bound lending rate thus leads us to underestimate the on-impact increase in the upper bound lending rate series during maritime disaster years. The informational deficiency of London bills and spot exchange rates that the travel lag term compensates for is again restricted to a two week period and thus a single monthly bill price observations. Assuming Seville's lending rate increases by 2 ppts in t, from 7 to 9%, and then falls back to 7% after $t+1-\tilde{t}$, then the annual travel lag term again amounts to 0.9984. As for the expectation error term, credit rationing renders the issue moot, to the extent that the Sevillian lending rate does not systematically change between t and $t+1$.[11]

[11] Beyond the behavior of $\hat{\omega}^U_t$ in response to money shocks, there exist other shocks that will push $\hat{\omega}^U_t$ up and down. More generally, omitting this term in the calculation of the upper bound lending rate thus introduces a measurement error, which likely increases the standard error of any

Next, we derive a lower bound lending rate estimate. If the London-issued bill of exchange on Seville is too cheap, the following strategy is profitable:

1. At time $t-\tilde{t}$, borrow $\mathbb{E}_{t-\tilde{t}}\left[U^L_{t,t+1}E^{P/£}_t\right]/(1-\tilde{\theta})$ Pesos in Seville for the period $t-\tilde{t}$ to $t+1$ at the local interest rate $R^S_{t-\tilde{t},t+1}$.
2. Travel to London, deduct transportation and transaction costs from the amount borrowed in step one[12], and exchange the remainder, $\mathbb{E}_{t-\tilde{t}}\left[U^L_{t,t+1}E^{P/£}_t\right]$, into Pounds at the spot exchange rate $E^{P/£}_t$. Use the obtained Pounds to buy bills of exchange on Seville at the price $U^L_{t,t+1}$. The total number of bills of exchange that can be bought in this way is $\mathbb{E}_{t-\tilde{t}}\left[U^L_{t,t+1}E^{P/£}_t\right]/(U^L_{t,t+1}E^{P/£}_t)$.
3. Travel back to Seville with the bills of exchange. At time $t+1$, both the bill of exchange and the loan mature, leading to the receipt of $\mathbb{E}_{t-\tilde{t}}\left[U^L_{t,t+1}E^{P/£}_t\right]/(U^L_{t,t+1}E^{P/£}_t)$ Pesos and an expenditure of $\mathbb{E}_{t-\tilde{t}}\left[U^L_{t,t+1}E^{P/£}_t\right]/(1-\tilde{\theta})R^S_{t-\tilde{t},t+1}$ Pesos.

The bill of exchange is too cheap if $1/(U^L_{t,t+1}E^{P/£}_t) > 1/(1-\tilde{\theta})R^S_{t-\tilde{t},t+1}$. Arbitrage leads to

$$R^S_{t-\tilde{t},t+1} \geq \frac{1-\tilde{\theta}}{U^L_{t,t+1}E^{P/£}_t}$$

Define $\Omega^L_t = \frac{R^S_{t,t+1+\tilde{t}}}{R^S_{t-\tilde{t},t+1}}$, and periodize the interest rate to get

$$R^S_{t,t+1} \geq \left(\frac{(1-\tilde{\theta})\Omega^L_t}{U^L_{t,t+1}E^{P/£}_t}\right)^{1/(1+\tilde{t})}$$

After taking natural logarithms and detrending we arrive at the lower bound lending rate estimate

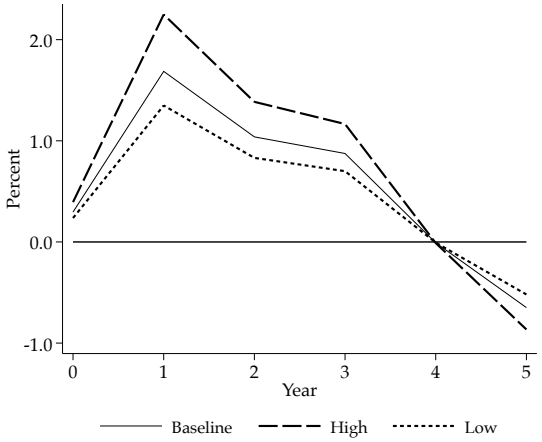
$$\hat{r}^S_{t,t+1} \geq \frac{1}{1+\tilde{t}}\left(-\hat{u}^L_{t,t+1} - \hat{e}^{P/£}_t + \hat{\omega}^L_t,\right) \equiv \hat{r}^L_t,$$

where small letters denote logarithms, and hats indicate detrended variables. Besides the periodization term $1/(1+\tilde{t})$ the lower bound lending rate thus includes a travel lag term – log-detrended $\Omega^L_t = \frac{R^S_{t,t+1+\tilde{t}}}{R^S_{t-\tilde{t},t+1}}$. This term is unobservable, and

IRF estimate.

[12] This should include an amount set aside for the return journey to Seville.

Figure A.10: Lending rate IRF range implied by a two-week travel time



Notes: Solid line – baseline response. Upper and lower bound IRFs for a two-week travel time period.

we thus calculate the lower bound lending rate series as $\frac{1}{1+\bar{t}}\left(-\hat{u}^{L}_{t,t+1} - \hat{e}^{P/£}_{t}\right)$. This affects the lower bound lending rate series in a similar way as the omission of Ω^{U}_{t} affects the upper bound series. Again, the travel lag term compensates for the informational deficiency of London bill prices and spot exchange rates with respect to contemporaneous Sevillian lending rates.

In our setting, when a money shock hits at t the travel lag term turns above 1 to the extent that the Sevillian lending rate increases in response to the shock. Omitting this term in the calculation of the lower bound lending rate thus leads us to underestimate the on-impact increase in the lower bound lending rate series during maritime disaster years. The informational deficiency of London bill prices and spot exchange rates is again restricted to a two week period and thus a single monthly bill price observations. Assuming Seville's lending rate increases by 2 ppts in t, from 7 to 9%, and remains at 9% thereafter, then the annual travel lag term amounts to $1.0015 = 1.09/(1.07^{1/12} \cdot 1.09^{11/12})$, implying a 0.15 ppt underestimation of the actual lower bound lending rate resulting from omitting the travel lag term. As for the expectation error term, credit rationing renders the issue moot, to the extent that the Sevillian lending rate has no on-impact response to the money shock.

Assuming a transportation time of two weeks and a bill of exchange maturity of two months, the scaling factors $\frac{1}{1+\bar{t}}$ and $\frac{1}{1-\bar{t}}$ amount to 0.8 and 1.33 respectively. Figure A.10 shows the upper and lower bound rate responses we obtain for the two-

week travel time scenario. The upper bound rate response suggests that lending rates increased around 2 percentage points in response to a negative 1 ppt money supply shock. According to the lower bound rate response lending rates increased by around 1.2 percentage points. The omission of expectation error and travel lag terms in the calculation of the upper and lower bound lending rate series implies that the on-impact IRFs are potentially afflicted by a downward bias. However, no bias is present to the extent that Spain's on-impact response to maritime disasters was characterized by credit rationing instead of an increase in lending rates. In sum, while transportation times add to the uncertainty in the lending rate response the evidence suggests that lending rates increased between one and two ppts after a 1 ppt money inflow shock.

A.4.3 Cancelling nominal exchange rate fluctuations

Gold-silver rate fluctuations can be an inaccurate proxy for nominal exchange rate fluctuations. This is because transportation costs can impede the arbitrage that would align nominal exchange rates with (bi-)metallic exchange rates (Bernholz and Kugler, 2011; Nogues-Marco, 2013). This opens the door for deviations between nominal exchange rates and (bi-)metallic exchange rates. For example, in the aftermath of a maritime disaster an acute shortage of Spanish silver coins may lead to an appreciation of Spanish peso coins in spot exchange markets that is not reflected in the gold-silver rate. Relatedly, falling prices for tradable goods in Spain may contribute to upward pressure on the Spanish exchange rate as merchants bid up the price of bills of exchange that provide them with the Spanish currency they require to buy the cheaper Spanish tradable goods.[13] Another concern is that merchants may have deemed Spanish coin debasements more likely after maritime disasters. As a consequence, post-disaster bill price changes might reflect debasement fears rather than Spanish lending rate fluctuations. This section describes an approach to calculating lending rates from bill prices that is robust to such concerns.

The currency market exchange rate of Spanish coins can be taken entirely out of

[13]Both these scenarios would imply that our baseline approach to calculating lending rates from bill prices would disguise some of the actual lending rate increase in the aftermath of maritime disasters.

the picture by using the prices of bills of exchange of different maturity. Consider an Amsterdam bill of exchange on Seville with 1-month maturity and a London bill of exchange on Seville (or another Spanish city) with 2-month maturity. The price for the Amsterdam bill is

$$U_t^{A,1m} = 1/[E_t^{P/G} R_t^{S,1m}], \tag{A.4}$$

where $E_t^{P/G}$ is the spot exchange rate between Dutch Guilders and Spanish Pesos (Pesos per Guilder), and $R_t^{S,1m}$ is the one-month lending rate in Seville. Analogously, the price for the London bill of exchange is

$$U_t^{L,2m} = 1/[E_t^{P/£} R_t^{S,2m}], \tag{A.5}$$

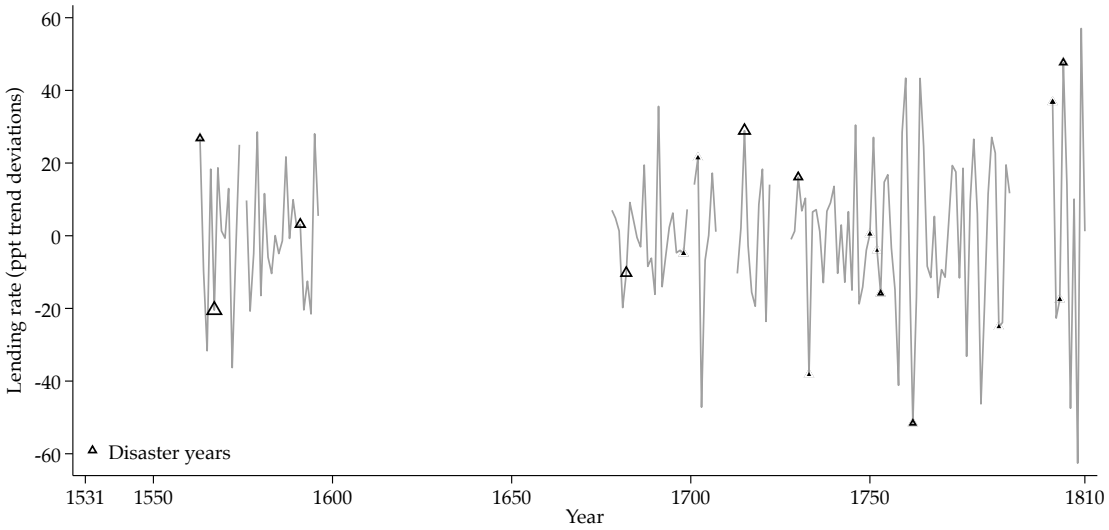
where $R_t^{S,2m}$ is the two-month lending rate in Seville. Dividing the two prices, we arrive at

$$\frac{U_t^{A,1m}}{U_t^{L,2m}} = \frac{E_t^{P/£}}{E_t^{P/G}} \frac{R_t^{S,2m}}{R_t^{S,1m}} = E_t^{G/£} R_t^{S,1m}. \tag{A.6}$$

$E_t^{G/£}$ is the spot exchange rate between Pounds Sterling and Dutch Guilders (Guilders per Pound). In other words, fluctuations of the nominal exchange rate of Spanish currency are cancelled out of expression A.6. What remains is the nominal exchange rate between two non-Spanish currencies that is unlikely to be affected by a scarcity or Spanish currency or Spanish debasement risk in the aftermath of maritime disasters. The log-detrended relative price of two bills of exchange of different maturity thus can serve as a proxy of the variation in the Spanish monthly interest rate. The main drawback of this alternative approach is the scarcity of suitable bill of exchange pairs of different maturity. As a consequence, this approach produces one third fewer observations than the (bi-)metallic ratio-based approach.

Figure A.11 shows the lending rate series that results from canceling the Spanish currency's exchange rate in this way (*NER cancelation rate*). With a standard deviation of 22, it exhibits more variation than the lending rate based on using the gold-silver rate as an indicator for the spot exchange rate (*NER proxy rate*). Partly this is due to the extremely high volatility of the NER cancelation rate at the

Figure A.11: Lending rate, NER cancelation approach



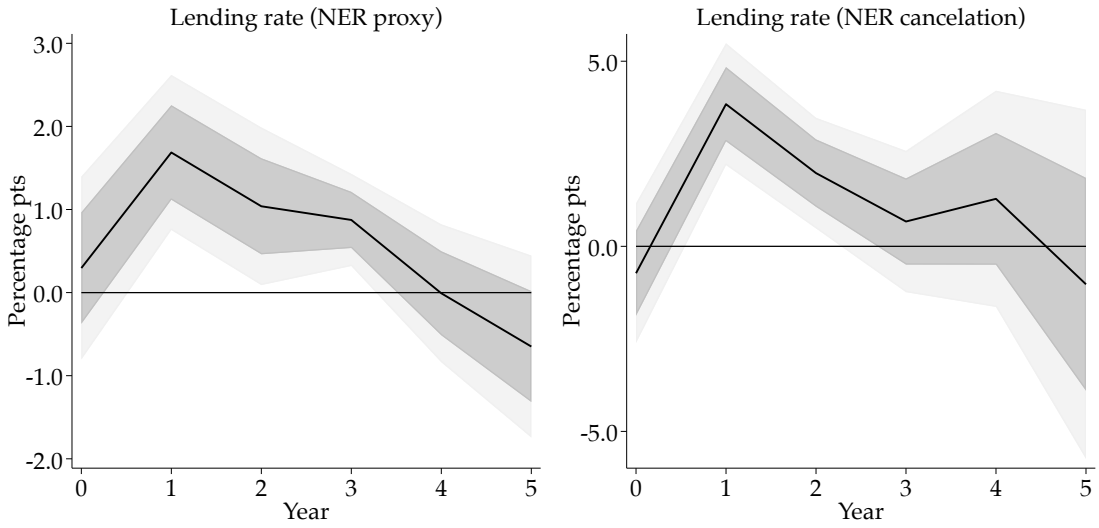
end of the sample during the Napoleonic Wars, the exclusion of which lowers the series' standard deviation by 4.5 ppts to 17.5. The prime suspect behind the NER cancelation series' higher volatility, however, is that in contrast to the NER proxy series it is not an average of lending rates across various Spanish cities. Instead, at each point in time, it reflects price fluctuations of bills of exchange on only one or two Spanish cities. As a consequence, the NER cancelation series lacks the variance reducing effect of the diversification implicit in summing across less than perfectly correlated Spanish market places. This is confirmed by the data: Three of the lending rates series for the individual Spanish cities that underlie the NER proxy rate have standard deviations above 19 ppts. One upshot of this is that the NER cancelation series is less representative of lending rate fluctuations in Spain as a whole, and hence we prefer the NER proxy lending rate series.

Figure A.12 compares the IRFs of both lending rate measures. According to both measures lending rates increased significantly one year after the shock. The NER cancelation rate's peak response, however, is about 2.5 times as large as the NER proxy rate's response. It is also somewhat less persistent.

A.4.4 Data on bill of exchange prices

Europe's early modern financial centers quoted bills of exchange prices for several Spanish market places. Our sources for the bills of exchange prices are Da Silva

Figure A.12: Alternative lending rate responses (negative 1 ppt money inflow shock)



(1969), Schneider et al. (1992, 1994), and Denzel (2010). In our data collection we focus on quotations from the largest financial centers (Amsterdam, London, Paris), as well as several other important nodes in the European financial network (Antwerp, Bisenzone fairs, Genoa, Hamburg, Venice) (Flandreau et al., 2009b). Since the bill of exchange price data for specific financial hub-destination pairs are often limited in length, we combine price quotations from different financial hubs to construct long-run time series (Table A.6). Based on these, we use equation A.2 to calculate fluctuations in the average Spanish lending rate as $\hat{r}_t^{Spain} = \frac{1}{I}\sum_i \hat{r}_t^i$, where $i \in$ {Barcelona, Cadiz, Madrid, Medina del Campo, Saragossa, Sevilla, Valencia}.[14]

The maturity of a typical bill of exchange ranges between 1 and 3 months. We use bills of M months maturity to calculate annual interest rates as follows: $\hat{r}_t^{annual} = \frac{12}{M}\hat{r}_t^{Mmonths} = -\frac{12}{M}\hat{u}_t^{Mmonths} - \frac{12}{M}\hat{e}_t^{P/£}$. We obtain detrended logarithms of the prices of bills of exchange, $\hat{u}_t^{Mmonths}$, by applying the HP filter with the smoothing parameter, λ, set to 6.25. Note that this approach provides us with estimates of trend deviations in loan rates, not estimates of loan rate levels. This is an important difference to the approaches pioneered in Flandreau et al. (2009a)

[14]Bisenzone was the Genoese exchange fair. As the Genoese used the Bisenzone fair to raise money for the Spanish Crown, one concern is that the bill prices at the Bisenzone fair reflect the financial standing of the Spanish Crown, rather than the Spanish private sector (Pezzolo and Tattara, 2008). However, the NER cancelation rate response (Figure A.12) which is not based on bill prices from the Bisenzone fair, suggests that the positive lending rate response is not driven by the behavior of bill prices on this fair.

Table A.6: Bills of exchange, NER proxy

Time period	Destination city	Financial center
1590 - 1796	Amsterdam	London
1575 - 1685	Antwerp	Bisenzone
1686 - 1697		London, Bisenzone
1756 - 1764		London, Venice
1678 - 1699	Cadiz	Amsterdam (1700: Antwerp)
1598 - 1722	Barcelona	Bisenzone
1591 – 1722	Frankfurt	Bisenzone
1776 – 1785		Leghorn
1805 – 1811		Paris
1552 - 1674	Genoa	Bisenzone
1678 - 1698		Amsterdam
1699 - 1811		London
1626 - 1666	Hamburg	Venice
1667 - 1811		London
1593 - 1654	London	Amsterdam
1657 - 1686		Amsterdam, Hamburg
1687 - 1806		Hamburg
1580 - 1607	Madrid	Bisenzone
1661 - 1734		Amsterdam
1735 - 1811		London
1579 - 1722	Medina del Campo	Bisenzone
1575 - 1722	Milan	Bisenzone
1740 - 1796		Venice
1619 - 1714	Paris	Amsterdam
1715 - 1811		Hamburg
1598 - 1722	Saragossa	Bisenzone
1575 - 1681	Seville	Bisenzone
1682 - 1712		Amsterdam, Bisenzone
1713 - 1804		Amsterdam
1585 - 1660	Venice	Bisenzone
1661 - 1696		Amsterdam
1698 - 1807		London

and Nogues-Marco (2011), which result in loan rate level estimates.

For the alternative method to derive Spanish interest rates, pairs of bills of exchange with different maturities are required. When data of such pairs are not available for the same Spanish city, we use bills of exchange on different Spanish cities (Table A.7). The Spanish interest rate is then calculated according to equation A.6, and then log-detrended. The maturity difference between the bills of exchange is always one month. Thus the resulting interest rate is monthly. This monthly rate is then multiplied by 12 to arrive at the annual rate.

Table A.7: Bills of exchange, NER cancelation

Time period	Destination city	Financial center	Maturity (month)
1563 - 1574	Valencia	Lyon	1
	Seville	Antwerp	2
1576 - 1596	Valencia	Lyon	1
	Seville	Lyon	2
1678 - 1699	Cadiz	Amsterdam	1
	Madrid	Amsterdam	2
1700 - 1710	Cadiz	Antwerp	1
	Cadiz	Amsterdam	2
1713 - 1777	Madrid	Amsterdam	2
	Madrid	London	3
1778 - 1810	Madrid	Amsterdam	2
	Cadiz	London	3

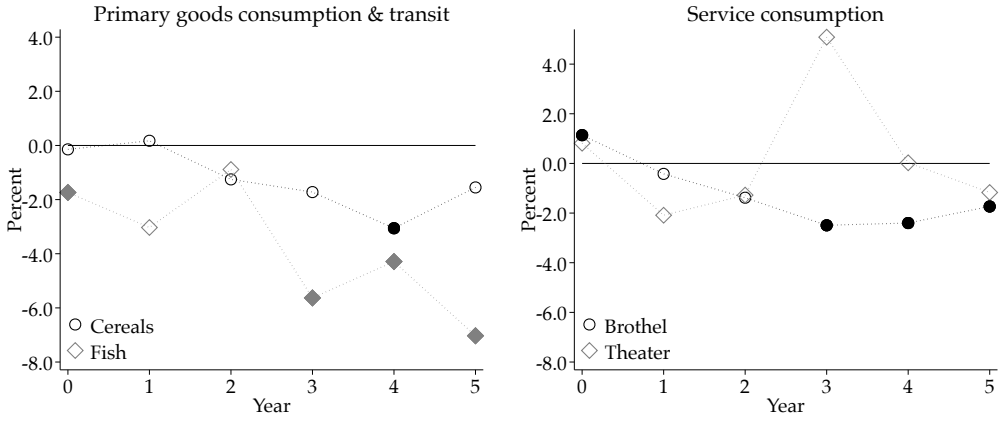
B. Additional results

Table B.1: Silver inflows and Spanish exports

Spanish export (growth rate)	-0.03		-0.03
	(0.02)		(0.03)
L.Spanish export (growth rate)		-0.02	-0.02
		(0.03)	(0.03)
F-statistic (p-value)	0.20	0.52	0.37
R squared	0.10	0.03	0.13
Observations	18	17	17

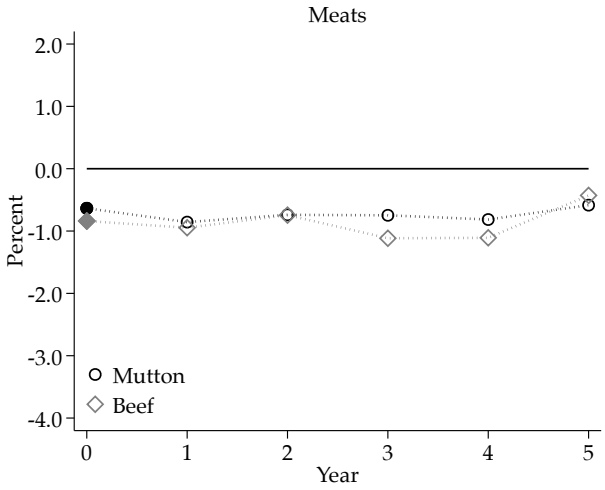
Notes: * 90%, ** 95%, *** 99%. Standard errors in parentheses. Dependent variable – silver inflow growth rate. Independent variables – Spanish export value growth rate. Export data from Esteban (1981).

Figure B.1: Responses to negative 1 percentage point money inflow shock



Notes: Nominal cereals tax revenues deflated with wheat and barley price index from nearby Madrid (Losa and Zarauz, 2021). Fish tax revenues are deflated with fish prices from Madrid from Allen (2001). The theater and brothel tax revenue series are deflated with the annual wage series for New Castile from Reher and Ballesteros (1993).

Figure B.2: Response to negative 1 percentage point money inflow shock



B.1. International trade

Price falls in tradable goods, such as iron and wool, raise the question whether Spanish exports increased in the aftermath of contractionary money shocks. Such an increase in exports, when paid for with precious metals, can counteract the monetary contraction at home. Relatedly, a Spanish recession can also lead to a reduction in Spanish import demand, thereby increasing the amount of precious metals that are retained in Spain.[15] The predictions that Spain's monetary loss will thus eventually diffuse across borders is shared by international monetary models, such as the price-specie flow model (Hume, 1752) and the monetary approach to the balance of payments (Flynn, 1978; Frenkel and Johnson, 2013).[16]

Wool and iron were two important export commodities for early modern Spain (Bilbao, 1987; Drelichman, 2009). While trade data for our sample period is generally scarce, available time series on Spanish wool and iron exports are long enough to straddle several maritime disaster events. The iron export data by Bilbao and Fernández de Pinedo (1982) covers exports from Biscayan ports from 1656 to 1805.[17] The wool export data by Alonso (1994) covers exports through Spain's Northern ports from the beginning of our sample until 1585.[18]

[15] Relatedly, blanket bans on precious metals exports were in place before 1551, from 1560 to 1566, from 1583 to 1586, from 1590 to 1593, and in 1600 (Martín, 1965, p.xxxii). After 1785, Spanish officials also began to rein in unregistered silver outflows with increasing success (Stein and Stein, 2003, p.310). These dates are not correlated with the occurrence of maritime disasters, and the blanket bans thus did little to increase Spanish money retention in the aftermath of the shocks we analyze.

[16] By modern standards, the early modern Spanish economy was relatively closed. Available export data for the late 18th century, after a period of rapid trade expansion, show that Spain's exports to its colonies averaged 1.5% of Spanish GDP (data from Esteban, 1981; Álvarez-Nogal and Prados de la Escosura, 2013). Lacking export success in other European economies, Spain's colonial exports were an important part of its total exports (Prados de la Escosura and Tortella Casares, 1983; Prados de la Escosura, 1993, pp.276-277). Spain's total export-to-GDP ratio was thus unlikely to exceed 5%. In addition, Spanish money outflows provide an insight into the size of Spain's chronic trade deficit with the rest of Europe. On average, Spanish money outflows amounted to around 94% of its money inflows (Chen et al., 2021). Money inflows, in turn, amounted to 3.7% of Spanish GDP on average. Putting both numbers together suggests that Spanish net-imports amounted to 3.5% of Spanish GDP – a large deficit given the small gross flows.

[17] This export series is calculated from tax auction returns. The tax right entitled the owner to a fixed amount of Maravedis per weight unit of iron exports. The tax auction returns thus reflects real export quantities. As for other tax auction series, an expectation error caveat applies (see section 3.1): To the extent that tax farmers' revenue expectations were disappointed due to the effects of the unexpected monetary shock, the on-impact response calculated from this data fails to reflect the actual response of iron exports.

[18] Wool exports are expressed in "sacks". While the exact content of a "sack" of wool is uncertain, the discussion in the historical literature has settled on a weight estimate of about 80 kilograms

The left panel in Figure B.3 describes how Spanish wool and iron exports reacted to the contractionary money shock. Real wool exports initially increased by around 4%.[19] Within the following two years, Spanish wool exports decline by close to 10%, which ties in with Spain's diminished capacity to produce wool after its flock size has diminished (see section 3.1). Real iron exports from Biscay exhibit no reaction. For the attraction of foreign coins, nominal rather than real exports are crucial. The considerable fall in Spanish iron and wool prices suggests that nominal exports might have fallen considerably (see section 3.2.2). However, we do not observe export prices at Spanish ports, which might have been more stable than Spain's domestic prices.

How did Spanish imports react to the monetary contraction? Spain's iron exports turn out to be informative in this regard. Iron exports were used to pay for non-food imports, because regulation discouraged the use of precious metals for international payment except for the import of foodstuffs (Lonbide and Ruano, 2007). As such, the iron export response can also be interpreted as indicative of Spanish non-food imports arriving at Biscayan ports. Assuming iron's export price fell by no more than iron's domestic price this implies an upper bound for Spain's nominal import reduction through Biscayan ports of 6% within two years (see section 3.2.2).[20]

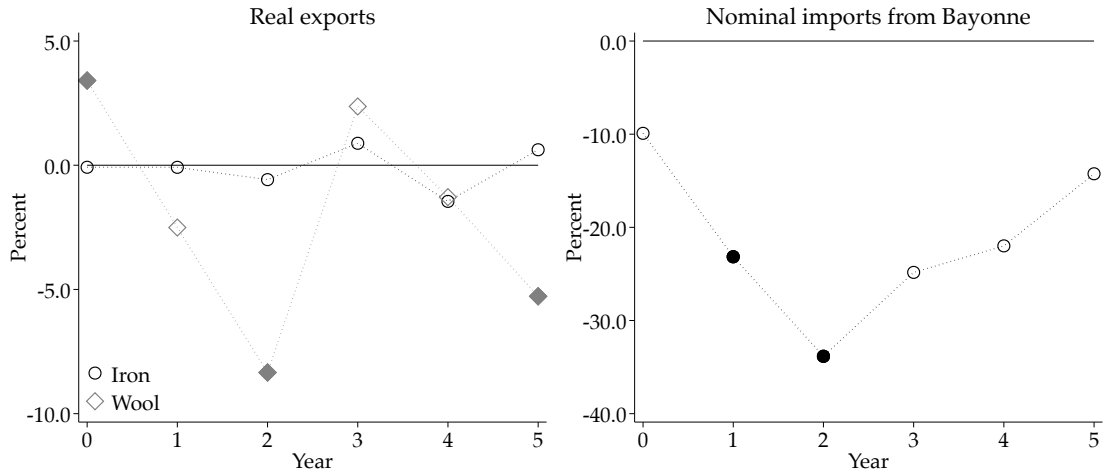
Another snippet of information on Spain's import response comes from a short time series on exports from Bayonne to Spain between 1746 and 1780 (Jaupart, 1966). Bayonne, a French city that is located just across the French border, was a key node in the Spanish-French overland trade (Stein and Stein, 2003, p.308). Much of the American silver that left Spain through the Basque region and Navarre entered France through Bayonne. The right panel in Figure B.3 displays how Spain's imports from Bayonne behaved in the aftermath of maritime disasters. Imports collapsed by 30% within two years, demonstrating that Spanish import demand was not

(Grafe, 2002; Drelichman, 2009).

[19] Contemporary merchant letters suggest that any wool that could not be sold in Spain was quickly prepared for export (Martín, 1965, p.xc).

[20] The Spanish government took various measures to discourage the outflow of silver money from Spain. As a consequence, wool and iron were frequently used as money substitutes in international trade. To transfer money from Spain to Italy, for example, Genoese merchants bought wool in Spain, exported the wool to Italy, and sold it there. Thus, the decrease in Spanish wool exports may also reflect a reduction in the amounts of funds that merchants desired to transfer from Spain to the rest of Europe.

Figure B.3: International trade response to a -1 percentage point money inflow shock



Notes: Solid markers indicate significance at the 10% level. IRF estimates for imports from Bayonne and wool exports based on parsimonious specification.

insulated from monetary contractions.[21] Spanish imports from Bayonne, however, are hardly representative. The close geographic proximity and the high weight of luxury goods among French exports probably rendered imports from Bayonne especially vulnerable to maritime disaster losses and the ensuing belt-tightening among Spanish households (Stein and Stein, 2003, p.313).

Overall, the wool and iron export IRFs suggest that the diffusion posited by international monetary models did not occur through a short-run increase in exports. Sporadic import evidence, however, is consistent with the notion that Spanish households and merchants curtailed their import demand, and thereby contributed to the diffusion of Spain's money loss to the rest of the Europe.

[21] Note that the four maritime disaster losses that the Bayonne import series straddles are among the smallest, ranging from 0.05% to 1.2% of Spain's money stock (after salvaging).

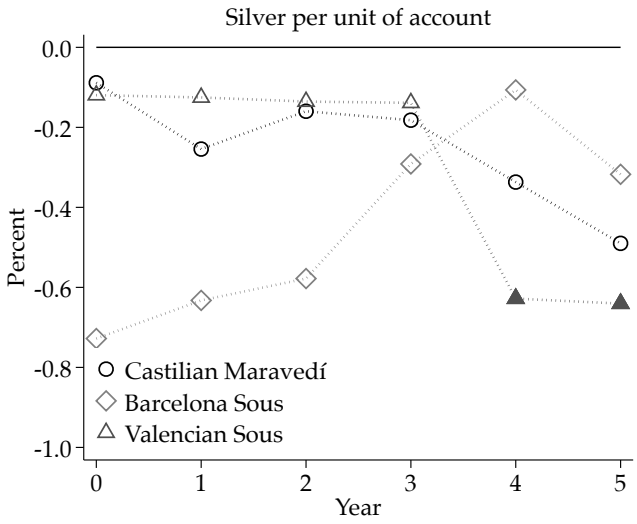
B.2. Changes in the silver value of the unit of account

In early modern Spain nominal adjustments could take place not only through a fall in wages and prices, but also through a change in the silver value of the unit of account (UOA) (Sargent and Velde, 2002; Velde, 2009; Karaman et al., 2018). Wages and prices in Castile were commonly expressed in *Maravedí* – the Castilian UOA – and Castilian coins possessed a *Maravedí* face value. Within the borders of present-day Spain, different UOAs existed. Thus, wages and prices in Barcelona and Valencia were commonly expressed in *Barcelona sous* and *Valencian sous*, respectively. The most direct way in which the silver value of one UOA could change was through the monetary authority's decision to change the silver content of coins. This, however, was an infrequent occurrence. More frequent were fluctuations in the exchange rate between silver coins and copper coins, which within our sample led to fluctuations in silver prices while the same prices expressed in UOAs stayed constant. Here, we analyze whether changes in the silver value of Spanish UOAs contributed to the nominal adjustment of the Spanish economy in the aftermath of maritime disasters.

The silver value of Spanish UOAs was stable for most of the 16th and 18th centuries, but in the 17th century it declined and especially in Castile at times behaved erratically (see Hamilton, 1934, ch. 4 for a detailed description of this period). To see whether such fluctuations in the silver value of Spanish UOAs contributed to nominal adjustments in the Spanish economy we analyze the time series on the silver content of the Castilian *Maravedí*, the *Barcelona sous* and the *Valencian sous* (Allen, 2001; Losa and Zarauz, 2021). We estimate IRFs for these UOAs' silver content based on the baseline specification 1.

Figure B.4 shows the results. For Castile, the silver content of the Maravedí does not move significantly in the five years following a maritime disaster shock. However, a downward trend after year three may be indicative of an increase in debasement risk over the medium term. The silver content of the Barcelona sous falls by 0.7% on impact, but the IRF is statistically insignificant throughout. Only the Valencian sous exhibits a significant debasement, but this only takes place four years after the shock. These IRFs suggest that debasements at best played a small

Figure B.4: Currency debasements in response to a -1 percentage point money inflow shock



Notes: Solid markers indicate significance at the 10% level.

auxiliary role for the short-run adjustment of prices and wages.

B.3. Labor flows

The marked decline in the wage for unskilled agricultural workers in the aftermath of maritime disasters could be a consequence of an increase in the supply of this type of workers. One snippet of data that is informative in this regard comes from the renovation records of the Royal Palace in Madrid. Hamilton (1947, p.208) provides an annual time series (1737–1800) on the number of different occupation types that were employed in this renovation project. This data offers us a rare glimpse into how employment flows responded to money supply shocks.

Our IRF estimate suggests that this particular construction project right away released between 10% and 20% of the occupation types that were employed in it (Figure B.5 in the Appendix). Given that many of Madrid's construction workers originated form its rural surroundings, one plausible scenario is that the labor that was released from urban construction flowed back to the pool of unskilled agricultural laborers (Andrés Ucendo and Lanza García, 2014). Any release of labor from textile manufacturing was similarly prone to increase the supply of unskilled agricultural laborers. This is because many weavers were rural peasants who supplemented their incomes through manufacturing jobs (García Sanz, 1977, p.224). In fact, many of the early production stages of textile manufacturing were deliberately outsourced to rural areas by merchants intent on avoiding the guild regulations prevalent in cities (Montañés, 2012; Montemayor, 1996, p.228).

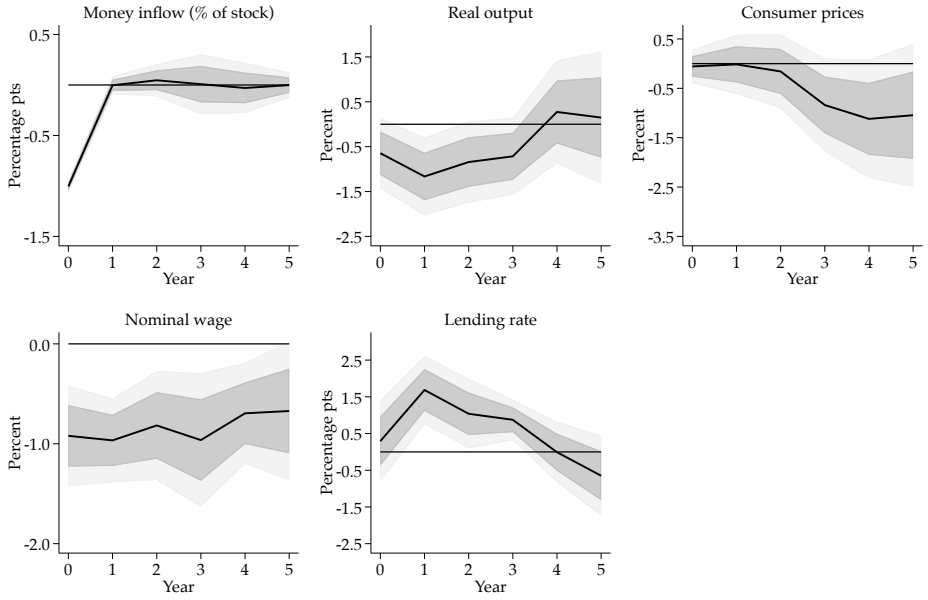
Figure B.5: Response to negative 1 percentage point money inflow shock



Notes: Parsimonious specification. Local projections including only two lags of the dependent variable's growth rate, together with leads, lags, and the contemporaneous value of the net money shock variable among the regressors.

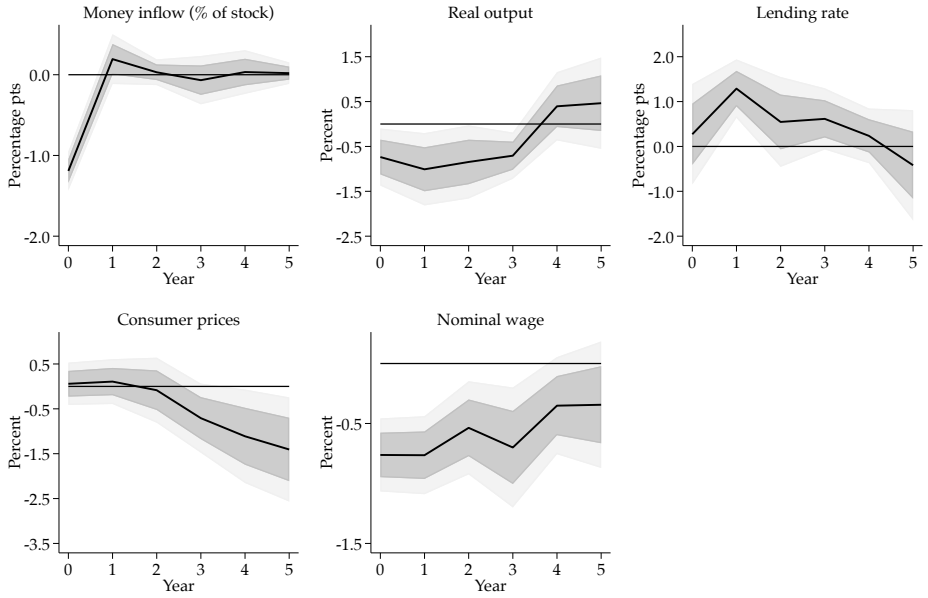
B.4. Robustness checks

Figure B.6: Impulse responses for negative 1 ppt money inflow shock (excl. salvaged)



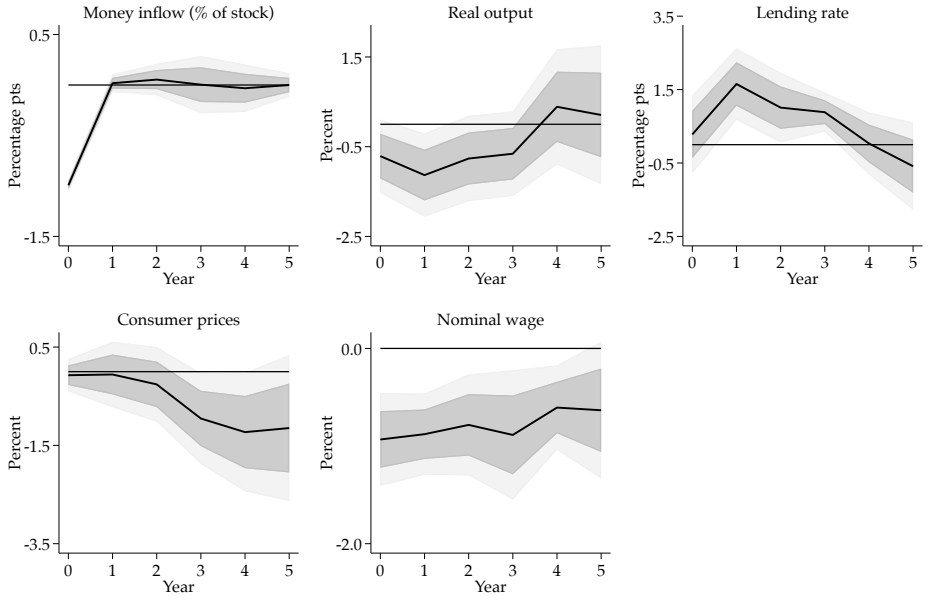
Notes: This figure shows IRF results based on a money shock measure that excludes subsequently salvaged money losses.

Figure B.7: Impulse responses for negative 1 ppt money inflow shock (parsimonious)



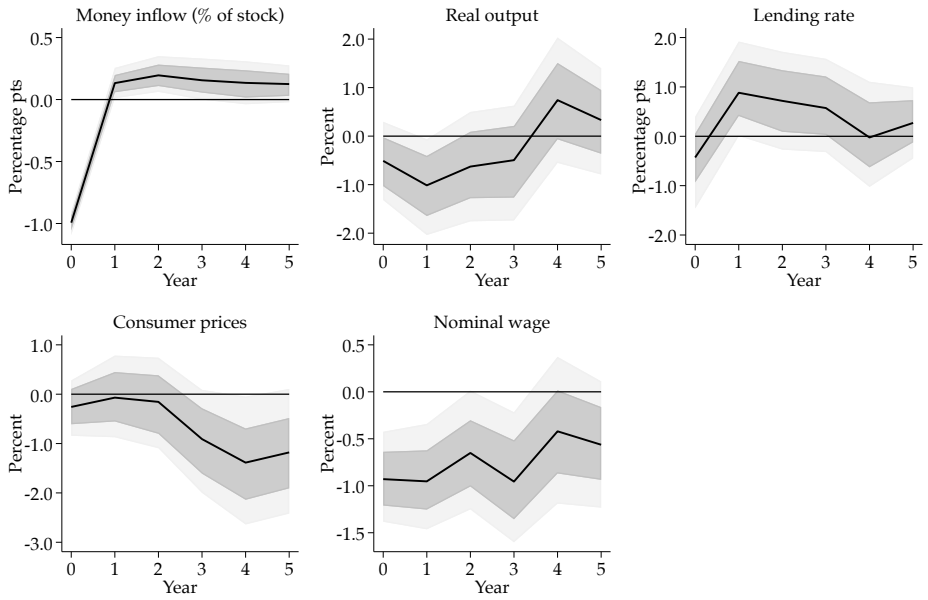
Notes: Parsimonious specification. Local projections including only two lags of the dependent variable's growth rate, and the money shock indicators among the regressors.

Figure B.8: Impulse responses for negative 1 ppt money inflow shock (excl. conflicts)



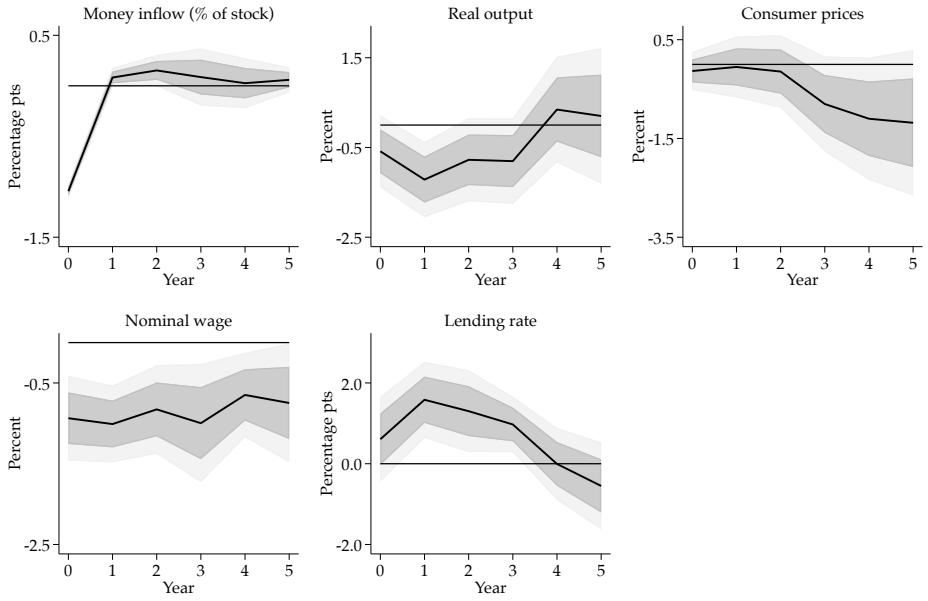
Notes: This figure shows IRF results based on a specification that allows money shocks caused by combat or capture events to develop different effects, by including an interaction term between the money shock and a conflict dummy.

Figure B.9: ARDL model results (negative 1 ppt money inflow shock)



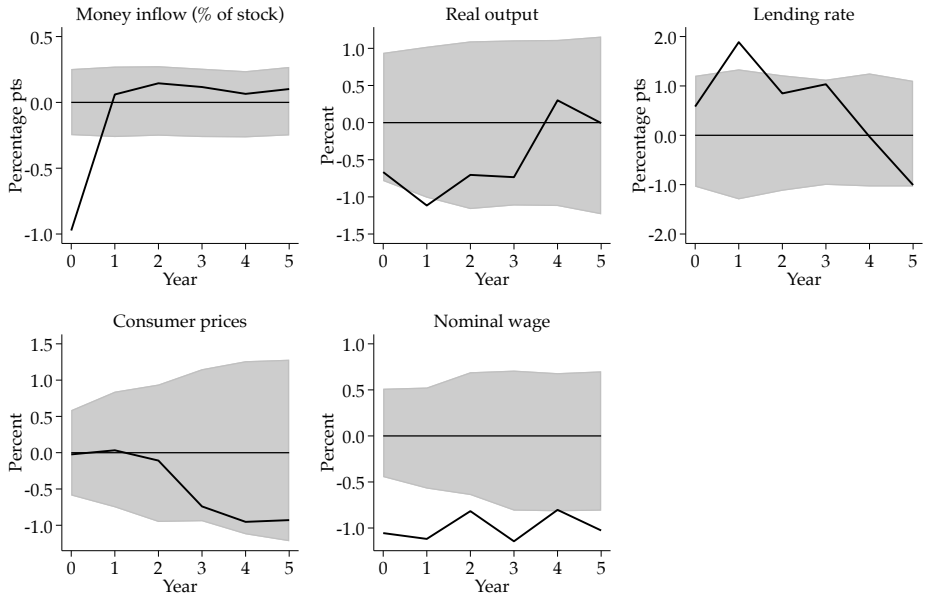
Notes: Autoregressive distributed lag (ARDL) model including up to 4 lags of the endogenous regressors, and up to 4 leads of the exogenous regressors. Money inflows are expresses relative to the current money stock.

Figure B.10: Impulse responses for negative 1 ppt money inflow shock (one month delay)



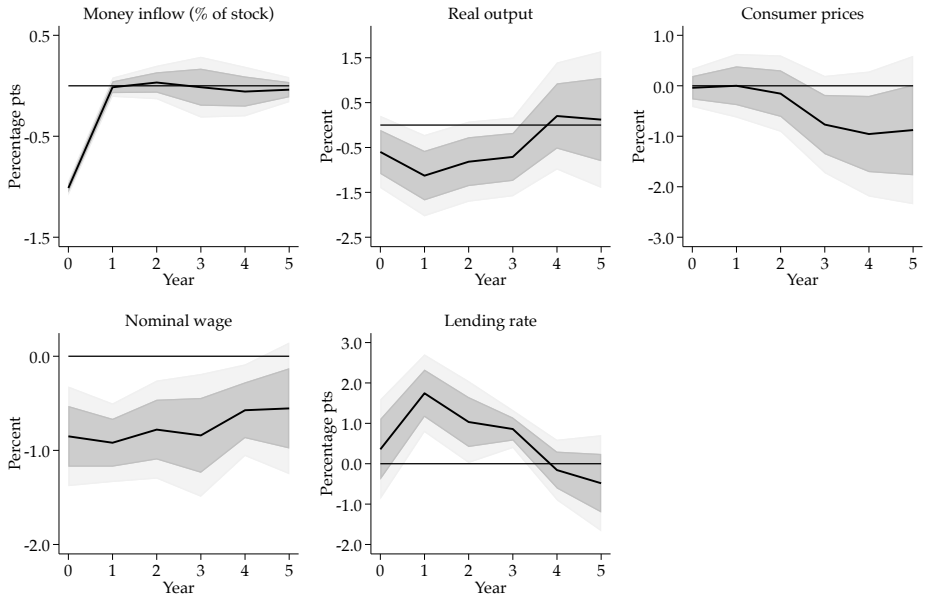
Notes: Results based on an alternative shock series that assumes that the news of maritime disasters arrived in Spain one month later than assumed in the baseline series.

Figure B.11: Placebo test



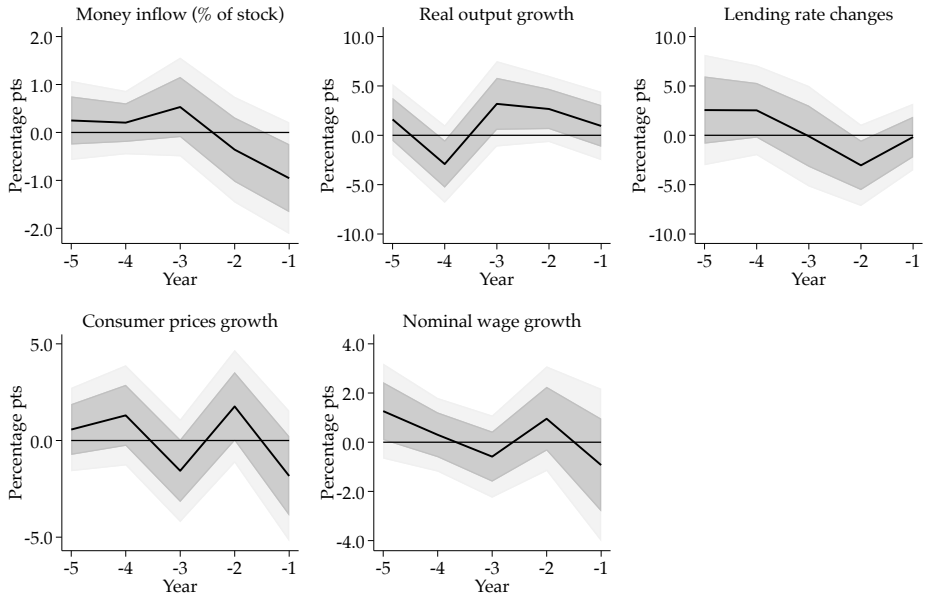
Notes: Black line – baseline IRF. Grey area – 10% to 90% percentile range based on 500 IRFs for temporally randomly reshuffled money shocks.

Figure B.12: Impulse responses for negative 1 ppt money inflow shock (pre-1780 sample)



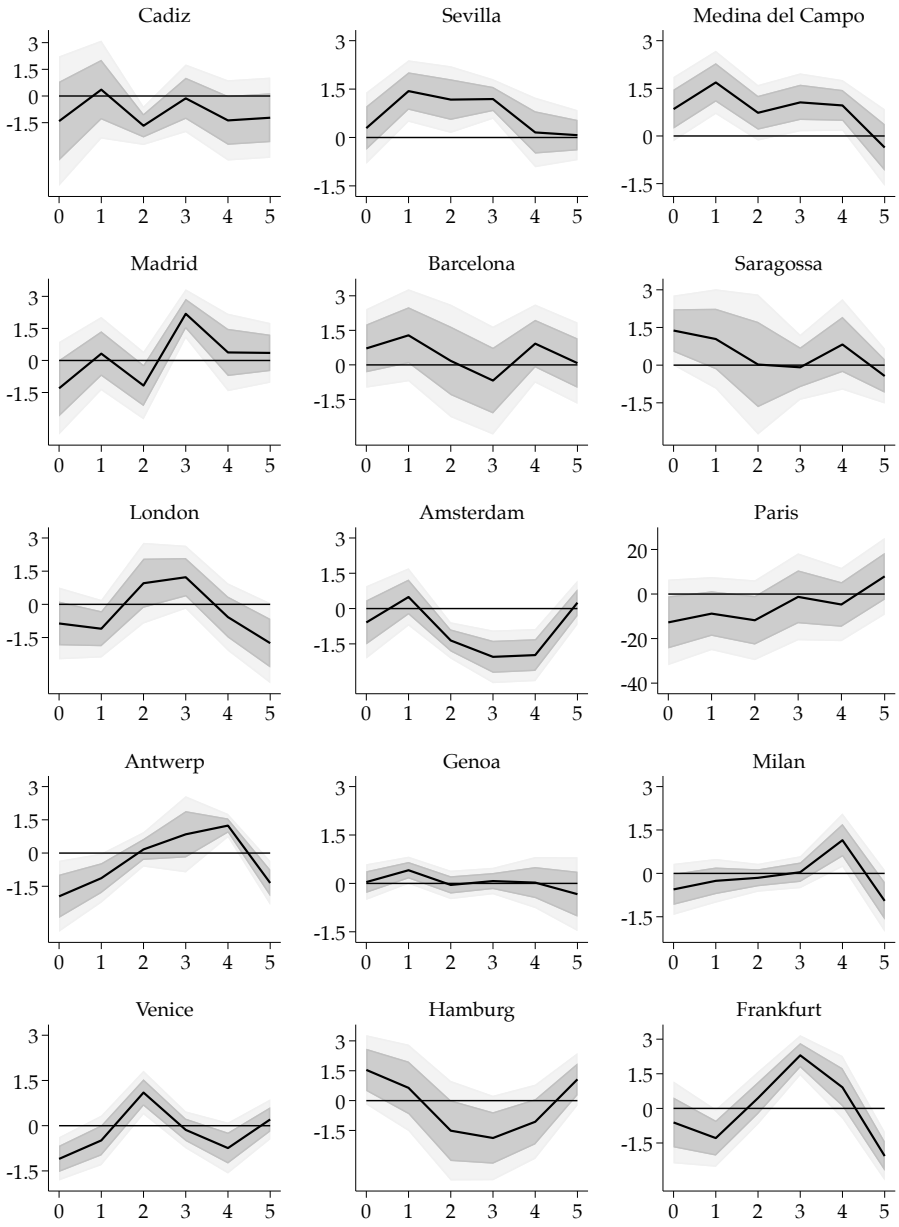
Notes: This figure shows IRF results based on a pre-1780 sample to remove any influence of structural changes occurring at the end of the 18th century.

Figure B.13: Pre-event analysis



Notes: Black line – average growth rate of variables prior to maritime disaster years compared to non-disaster years.

Figure B.14: European interest rate responses (negative 1 ppt money inflow shock)



Notes: y-axis – percentage points. x-axis – years. The figure shows the IRFs for those cities where the available data allows for the construction of long-run lending rate series. The lending rate series for each of these cities straddles at least 15 maritime disaster events. Only the series for Cadiz straddles fewer disaster events, because its observations are concentrated in the 18th century. Its IRF is nevertheless shown here, because the Cadiz series goes into the average Spanish lending rate series. Lending rates in some non-Spanish cities actually decreased (e.g. Amsterdam), possibly indicating a "run to safety" response, in which lending retracted from Spain and moved to safer destinations in the aftermath of maritime disasters.

References

Allen, Robert C. 2001. The great divergence in European wages and prices from the Middle Ages to the First World War. *Explorations in Economic History* 38(4): 411–447.

Alonso, Hilario Casado. 1994. El comercio internacional burgalés en los ss. XV y XVI. In *Actas del V Centenario del Consulado de Burgos:(1494-1994)*, 175–248. Diputación Provincial de Burgos.

Álvarez-Nogal, Carlos. 2003. The role of institutions to solve sovereing debt problems: the Spanish monarchy's credit (1516-1665). Universidad Carlos III de Madrid. Economic History and Institution Series 04. Working Paper 03-08.

Álvarez-Nogal, Carlos, and Christopher Chamley. 2014. Debt policy under constraints: Philip II, the Cortes, and Genoese bankers. *The Economic History Review* 67(1): 192–213.

Álvarez-Nogal, Carlos, and Leandro Prados de la Escosura. 2013. The rise and fall of Spain (1270–1850). *The Economic History Review* 66(1): 1–37.

Amrhein, John Jr. 2007. *The hidden galleon*. New Maritima Press.

Andrés Ucendo, José Ignacio, and Ramón Lanza García. 2014. Prices and real wages in seventeenth-century Madrid. *The Economic History Review* 67(3): 607–626.

Arregui, Ignacio M.^{a} Carrion. 1991. Los precios del hierro en Guipuzcoa durante los siglos XVII y XVIII: un buen indicador de la coyuntura del sector siderurgico. *Revista de Historia Economica-Journal of Iberian and Latin American Economic History* 9(3): 481–498.

Artola, Miguel. 1982. *La hacienda del antiguo régimen*. Madrid: Alianza Editorial.

Barbier, Jacques A., and Herbert S. Klein. 1981. Revolutionary wars and public finances: The Madrid Treasury, 1784–1807. *The Journal of Economic History* 41(2): 315–339.

Bernholz, Peter, and Peter Kugler. 2011. Financial market integration in the early modern period in Spain: Results from a threshold error correction model. *Economics Letters* 110(2): 93–96.

Bilbao, Luis María, and Emiliano Fernández de Pinedo. 1982. Auge y crisis de

la siderometalurgia tradicional en el País Vasco (1700-1850). In *La economía española al final del Antiguo Régimen. II. Manufacturas,*, 133–228. Madrid: Alianza.

Bilbao, Luis María Bilbao. 1987. *La industria siderometalúrgica tradicional en el País Vasco (1450-1720)*. Instituto de Estudios Fiscales.

Bocola, Luigi. 2016. The pass-through of sovereign risk. *Journal of Political Economy* 124(4): 879–926.

Bonifacio, Claudio. 2010. *Galleons and sunken treasure*. Signum Ops.

Bravo, Mariano Sánchez. 2010. Fragata "Santa Leocadia": Episodios de su naufragio y rescate del tesoro. *Revista del Instituto de Historica Marítima* (46).

Bueno, Pablo Emilio Pérez-Mallaína. 1996. *El hombre frente al mar: Naufragios en la carrera de Indias durante los siglos XVI y XVII*. 15. Universidad de Sevilla.

Carrasco González, María Guadalupe. 1996. Los instrumentos del comercio colonial en el Cádiz del siglo XVII (1650-1700). *Estudios de Historia Económica* 35.

Carrión, José-Miguel Martínez. 2002. *El nivel de vida en la España rural, siglos XVIII-XX*. Universidad de Alicante.

CGFS. 2011. The impact of sovereign credit risk on bank funding conditions. Committee on the Global Financial System, CGFS Paper No 43.

Chaunu, Pierre, and Huguette Chaunu. 1977. *Séville et l'Amérique aux XVIe et XVIIe siècles*. Flammarion.

Chen, Yao, Nuno Palma, and Felix Ward. 2021. Reconstruction of the Spanish money supply, 1492–1810. *Explorations in Economic History* 81.

Cobbett, William. 1804. Capture of Spanish frigates. Cobbett's Weekly Political Register.

Comín, Francisco Comín, and Bartolomé Yun-Casalilla. 2012. Spain: From composite monarchy to nation-state, 1492-1914. An exceptional case? In *The rise of fiscal states: A global history 1500-1914*, edited by Yun-Casalilla, Bartolomé, Patrick K. O'Brien, and Francisco Comín Comín. Cambridge University Press.

Costa, Leonor Freire, Susana Münch Miranda, and Pilar Nogues-Marco. Forthcoming. An economic history of the Iberian Peninsula, 700-2000. In *Globalization and the Enlightenment, 1500-1800*, edited by Costa, Leonor Freire, and Regina

Grafel. Cambridge University Press.

Cuenca-Esteban, Javier. 2008. Statistics of Spain's colonial trade, 1747–1820: New estimates and comparisons with Great Britain. *Revista de Historia Económica-Journal of Iberian and Latin American Economic History* 26(3): 323–354.

Da Silva, José-Gentil. 1969. *Banque et crédit en Italie au XVIIe siècle*, volume 2. Paris: Klincksieck.

de Mercado, Fray Tomás. 1587. *Summa de tratos y contratos*. Seville: Fernando Díaz.

de Vries, Jan. 2003. Connecting Europe and Asia: a quantitative analysis of the cape-route trade, 1497-1795. In *Global Connections and Monetary History, 1470-1800*, edited by ane Richard von Glahn, Dennis O. Flynn, and Artúro Giraldez. Ashgate.

Denzel, Markus A. 2010. *Handbook of world exchange rates, 1590–1914*. Routledge.

Drelichman, Mauricio. 2009. License to till: The privileges of the Spanish Mesta as a case of second-best institutions. *Explorations in Economic History* 46(2): 220–240.

Drelichman, Mauricio, and Hans-Joachim Voth. 2010. The sustainable debts of Philip II: A reconstruction of Castile's fiscal position, 1566–1596. *The Journal of Economic History* 70(4): 813–842.

Earle, Peter. 2007. *Treasure hunt: Shipwreck, diving, and the quest for treasure in an age of heroes*. Macmillan.

Esteban, Javier Cuenca. 1981. Statistics of Spain's colonial trade, 1792-1820: Consular duties, cargo inventories, and balances of trade. *The Hispanic American Historical Review* 61(3): 381–428.

Feliu, Gaspar. 1991. *Precios y salarios en la Cataluña moderna*, volume II. Banco de España.

Fine, John Christopher. 2006. *Treasures of the Spanish main: Shipwrecked galleons in the New World*. Rowman & Littlefield.

Fisher, John Robert. 1985. *Commercial relations between Spain and Spanish America in the era of free trade, 1778-1796*. 13. Centre for Latin-American Studies,

University of Liverpool.

Fisher, John Robert. 2003. Mining and imperial trade in eighteenth-century Spanish America. In *Global connections and monetary history, 1470-1800*, edited by Dennis O. Flynn, Richard von Glahn, Arturo Giráldez, 107–121. Ashgate.

Flandreau, Marc, Christophe Galimard, Clemens Jobst, and Pilar Nogués-Marco. 2009a. The bell jar: Commercial interest rates between two revolutions, 1688–1789. In *The Origins and Development of Financial Markets and Institutions: From the Seventeenth Century to the Present*, edited by Atack, Jeremy, and Larry Neal, chapter 6, 161–208. Cambridge University Press.

Flandreau, Marc, Christophe Galimard, Clemens Jobst, and Pilar Nogués-Marco. 2009b. Monetary geography before the Industrial Revolution. *Cambridge Journal of Regions, Economy and Society* 2(2): 149–171.

Flynn, Dennis O. 1978. A new perspective on the Spanish Price Revolution: The monetary approach to the balance of payments. *Explorations in Economic History* 15(4): 388–406.

Frenkel, Jacob A., and Harry G. Johnson, editors. 2013. *The Monetary Approach to the Balance of Payments (Collected Works of Harry Johnson)*. Routledge.

García-Baquero González, Antonio. 2003. American gold and silver in the eighteenth century: From fascination to accounting. In *Global connections and monetary history, 1470-1800*, edited by Dennis O. Flynn, Richard von Glahn, Arturo Giráldez, 107–121. Ashgate.

García Sanz, Ángel. 1977. *Desarrollo y crisis del Antiguo Régimen en Castilla la Vieja: Economía y sociedad en tierras de Segovia, 1500-1814*. Akal.

Gelderblom, Oscar, Abe de Jong, and Joost Jonker. 2013. The formative years of the modern corporation: The Dutch East India Company VOC, 1602–1623. *The Journal of Economic History* 73(4): 1050–1076.

Grafe, Regina. 2002. Northern Spain between the Iberian and the Atlantic worlds: trade and regional specialisation, 1550–1650. *European Review of Economic History* 6(2): 269–275.

Hamilton, Earl Jefferson. 1934. *American treasure and the price revolution in Spain,*

1501-1650, volume 1125. Harvard University Press Cambridge, MA.

Hamilton, Earl Jefferson. 1936. *Money, prices, and wages in Valencia, Aragon, and Navarre, 1351-1500*. Harvard University Press.

Hamilton, Earl Jefferson. 1947. *War and prices in Spain: 1651-1800*, volume 81. Harvard University Press Cambridge, MA.

Henriques, Antonio, and Nuno Palma. 2020. Comparative European institutions and the little divergence, 1385-1800. CEPR Discussion Paper No. DP14124 (v.3).

Homer, Sidney, and Richard E. Sylla. 1991. *A history of interest rates*. Rutgers University Press, 3 edition.

Hume, David. 1752. *Political Discourses*. A. Kincaid & A. Donaldson.

IMF. 2013. Global financial stability report. April 2013, International Monetary Fund, Washington, DC.

Jaupart, Fernand. 1966. *L'activité commerciale de Bayonne au XVIIIe siècle*. Société des Sciences, Lettres et arts de Bayonne.

Jordà, Òscar, Katharina Knoll, Dmitry Kuvshinov, Moritz Schularick, and Alan M Taylor. 2019. The rate of return on everything, 1870–2015. *The Quarterly Journal of Economics* 134(3): 1225–1298.

Kamen, Henry. 1966. The Destruction of the Spanish silver Fleet at Vigo in 1702. *Historical Research* 39(100): 165–173.

Karaman, K. Kıvanç, Sevket Pamuk, and Seçil Yıldırım. 2018. Money and monetary stability in Europe, 1300-1914. CEPR Discussion Paper No. DP12583.

Karaman, K Kıvanç, Şevket Pamuk, and Seçil Yıldırım-Karaman. 2020. Money and monetary stability in Europe, 1300–1914. *Journal of Monetary Economics* 115: 279–300.

Kelly, Morgan, and Cormac Ó Gráda. 2019. Speed under sail during the early industrial revolution (c. 1750–1830). *The Economic History Review* 72(2): 459–480.

Lázaro, Javier Moreno. 2001. Precios de las subsistencias, salarios nominales y niveles de vida en Castilla la Vieja.: Palencia, 1751-1861. *Documentos de trabajo de la Asociación Española de Historia Económica* 2.

Lonbide, Xabier Alberdi, and Álvaro Aragón Ruano. 2007. Relaciones comerciales entre puertos franceses y guipuzcoanos en las décadas finales del siglo XVI y la primera mitad del siglo XVII: síntoma de vitalidad económica. *Pedralbes: Revista d'historia moderna* (27): 219–262.

Losa, Ernesto López, and Santiago Piquero Zarauz. 2021. Spanish subsistence wages and the Little Divergence in Europe, 1500–1800. *European Review of Economic History* 25(1): 59–84.

Mangas, Fernando Serrano. 1989. *Armadas y flotas de la plata 1620-1640*. Madrid: Banco de España.

Martín, Alberto Marcos. 2005. Viejos en la ciudad. La estructura de edad de la población en los núcleos urbanos españoles del Antiguo Régimen. In *Vejez, envejecimiento y sociedad en España, siglos XVI-XXI*, volume 83, 67–100. Univ de Castilla La Mancha.

Martín, F. Ruiz. 1965. *Lettres marchandes échangées entre Florence et Medina del Campo*, volume 124. S.E.V.P.E.N.

Marx, Robert F. 1987. *Shipwrecks in the Americas*. Courier Corporation.

Montañés, María Ángela Jiménez. 2012. La industria textil y su regulación en el siglo XVI: caso particular de Toledo. *Pecunia: revista de la Facultad de Ciencias Económicas y Empresariales* (14): 107–132.

Montemayor, Julián. 1996. *Tolède entre fortune et déclin (1530-1640)*. Presses Univ. Limoges.

Morineau, Michel. 1985. *Incroyables gazettes et fabuleux métaux: les retours des trésors américains d'après les gazettes hollandaises (XVIe-XVIIIe siècles)*, volume 7. Les Editions de la MSH.

Nogues-Marco, Pilar. 2011. Tipos de cambio y tipos de interés en Cádiz en el siglo XVIII (1729-1788). Banco de España Estudios de Historia Económica N. 58.

Nogues-Marco, Pilar. 2013. Competing bimetallic ratios: Amsterdam, London, and bullion arbitrage in mid-eighteenth century. *The Journal of Economic History* 73(2): 445–476.

Parker, Geoffrey. 2014. *Imprudent king: A new life of Philip II*. Yale University

Press.

Peterson, Mendel. 1975. *The funnel of gold*. Little, Brown, & Company.

Pezzolo, Luciano, and Giuseppe Tattara. 2008. "Una fiera senza luogo": Was Bisenzone an international capital market in sixteenth-century Italy? *The Journal of Economic History* 68(4): 1098–1122.

Phillips, Carla Rahn. 1990. The growth and composition of trade in the Iberian empires, 1450-1750. In *The rise of merchant empires: Long-distance trade in the early modern world 1350–1750*, edited by Tracy, James D., 34–101. Cambridge University Press.

Phillips, Carla Rahn. 2007a. Spanish shipbuilding in the late 17th century: Art, science and bureaucracy. Cuadernos de estudios borjanos 50.

Phillips, Carla Rahn. 2007b. *The treasure of the San José: Death at sea in the War of the Spanish Succession*. JHU Press.

Phillips, Carla Rahn, and William D. Phillips. 1997. *Spain's golden fleece: Wool production and the wool trade from the Middle Ages to the nineteenth century*. Johns Hopkins University Press Baltimore.

Pike, Ruth. 1966. *Enterprise and adventure*. Cornell University Press.

Potter, John Stauffer. 1972. *The treasure diver's guide*. New York: Doubleday.

Prados de la Escosura, Leandro. 1993. La pérdida del imperio y sus consecuencias económicas en España. In *La independencia americana: consecuencias económicas*, edited by Prados de La Escosura, Leandro, and Samuel Amaral, 253–300. Madrid: Alianza Editorial.

Prados de la Escosura, Leandro, and Gabriel Tortella Casares. 1983. Tendencias a largo plazo del comercio exterior Español, 1714-1913. *Revista de Historia Economica - Journal of Iberian and Latin American Economic History* 1(2).

Putley, Julian. 2000. *The virgins' treasure isle*. Virgin Island Books.

Reher, David S., and Esmeralda Ballesteros. 1993. Precios y salarios en Castilla la Nueva: la construccion de un indice de salarios reales, 1501–1991. *Revista de Historia Económica-Journal of Iberian and Latin American Economic History* 11(1): 101–151.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2009. *This time is different: Eight centuries of financial folly.* Princeton University Press.

Sandz, Victoria, and Robert F. Marx. 2001. *Encyclopedia of western Atlantic shipwrecks and sunken treasure.* London: McFarland.

Sardone, Sergio. 2019. Forced loans in the Spanish Empire: The first requisition of American treasures in 1523. *The Economic History Review* 72(1): 57–87.

Sargent, Thomas J., and François R. Velde. 2002. *The big problem of small change.* Princeton University Press.

Schneider, Jürgen, Oskar Schwarzer, and Markus A. Denzel. 1994. *Währungen der Welt III: Europäische Wechselkurse im 17. Jahrhundert.* Franz Steiner Verlag Stuttgart.

Schneider, Jürgen, Oskar Schwarzer, Friedrich Zellfelder, and Markus A. Denzel. 1992. *Währungen der Welt VI: Geld und Währungen in Europa im 18. Jahrhundert.* Franz Steiner Verlag Stuttgart.

Sedgwick, Romney. 1970. *The History of Parliament: The House of Commons, 1715-1754.* Published for the History of Parliament Trust, by Oxford University Press.

Soetbeer, Adolf. 1879. *Edelmetall-Produktion und Werthverhältniss zwischen Gold and Silber seit der Entdeckung Amerika's bis zur Gegenwart.* Justus Perthes.

Spooner, Frank C. 1972. *The international economy and monetary movements in France, 1439-1725*, volume 138 of *Harvard Economic Studies.* Harvard University Press.

Stein, Stanley J., and Barbara H. Stein. 2003. *Apogee of empire: Spain and New Spain in the age of Charles III, 1759–1789.* Johns Hopkins University Press.

TePaske, John J. 2010. *A new world of gold and silver.* Brill.

The Gentleman's and London Magazine. 1763. Stated Account of the Sale of the Hermione, and her Cargo. Dublin: Printed by John Exshaw.

Ulloa, Modesto. 1977. *La hacienda real de Castilla en el reinado de Felipe II.*, volume 2. Madrid: Fundación Universitaria Española.

Velde, François R. 2009. Chronicle of a deflation unforetold. *Journal of Political*

Economy 117(4): 591–634.

Venema, Janny. 2010. *Kiliaen van Rensselaer (1586-1643) Designing a New World.* Uitgeverij Verloren.

Walton, Timothy R. 1994. *The Spanish treasure fleets.* Pineapple Press Inc.

Concluding remarks

My thesis considered a diverse and non-exhaustive set of ways in which subjective beliefs are determined, and in which they in turn affect outcomes. It analysed how populist forces are shaped if political narratives rule freely, unconstrained by data (Chapter 1). It considered how prevailing science scepticism can inhibit the effectiveness of shelter-in-place policies in the case of COVID-19 (Chapter 2). It reasoned that subjective beliefs are a key mechanism that explains the co-evolution of a state's monetary and fiscal capacity (Chapter 3), and showed that shocks to expectations have real effects in the economy (Chapter 4).

Brief reflection on the chapters that constitute this thesis reveals that beliefs can be formed in very different ways, with different implications. On the one hand, they may be shaped by self-motivated reasoning and largely ignore empirical evidence (Chapters 1 and 2). On the other, beliefs may be formed based on observations and more rational forms of inference (Chapters 3 and 4). These ways of characterising belief formation are not mutually exclusive, of course, and in most cases both of these considerations will play an important role. For simplicity, the frameworks presented in this thesis did not go into considering the trade-offs agents may face between the first and second mode of belief formation, which is an interesting and ongoing area of research inquiry (Little, 2019).

The thesis brings up several other important follow-up questions that merit further consideration. How do political parties come up with new narratives, and what prevents them from doing so? Going beyond scapegoating narratives, how can governments leverage narratives together with policies to influence behaviour—and when will such attempts fail? How could one empirically measure the co-evolution of political narratives, voter beliefs, and party policies? In an ongoing research agenda that considers the role of narratives, I am currently following up on some of these questions. I will briefly touch on two ongoing projects to show the potential future research that can build on this thesis.

In ongoing work with Tim Besley, we analyse how a government can use narra-

tives alongside more conventional policies to guide behaviour. In a stylised model, we illustrate the possibility of policy ineffectiveness, where prevailing narratives in society deter the government from pursuing its preferred policy, due to limited buy-in from citizens and firms. Such a policy ineffectiveness trap cannot be overcome by simply devoting more resources towards the policy. Indeed, increasing the budget for the government's preferred policy may backfire: in the face of public opposition, these additional resources will be used for alternative policies because of an inability of the government to precommit. In the future, we plan to apply the general framework to analysing how science-sceptic narratives prevent climate action, and what can change this impasse; when narratives about the effectiveness of tax policy changes will fail to convince citizens; and under which conditions place-based policies ("Levelling Up") will succeed to convince a critical mass of supporters in order to be effective.

In complementary work, I consider ways in which political narratives can be captured empirically, in order to study their co-evolution with voter beliefs and implemented policies. For this purpose, I am applying natural language processing tools to parliamentary speeches. Speeches are accessible for 9 countries through the *ParlSpeech V2* dataset (Rauh and Schwalbach, 2020). To classify these speeches across topics of interest, I use a supervised machine learning tool, the naïve Bayes classifier. The first step of the methodology consists in screening speeches through wordstem searches (e.g. for "climate change") and then manually labelling them on a five point scale (e.g. for versus against climate action). In a second step, the naïve Bayes classifier classifies the unlabelled speeches on the same topic based on the training set. This classifier can be applied to a diverse set of topics, including political narratives on climate change, redistributive policies, and immigration. These measures will allow one to analyse the co-evolution of party positions, voter beliefs, and other topics of policy relevance.

I hope that this thesis has convinced some sceptical readers of the importance of considering how subjective beliefs are formed when studying topics in political economy. Much remains to be done to further the insights of this growing area of

enquiry, both in terms of theory and empirics. Perhaps this thesis can act as a small contribution to this endeavour.

References

Little, Andrew T. 2019. The Distortion of Related Beliefs. *American Journal of Political Science* 63(3): 675–689.

Rauh, Christian, and Jan Schwalbach. 2020. The ParlSpeech V2 data set: Full-text corpora of 6.3 million parliamentary speeches in the key legislative chambers of nine representative democracies. Available at https://doi.org/10.7910/DVN/L4OAKN.