Essays on Weak Identification in High-Dimensional Models with Applications in Macroeconomics



Max-Sebastian Dovì

Brasenose College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2022

Word count: 462 words on p. 59 × 151 pages = 69,762

Acknowledgements

My deepest gratitude goes to my supervisor Prof. Sophocles Mavroeidis. Without working as a research assistant for him, I would not have been introduced to the fascinating literature of weak identification. I also could not have hoped to get even a foothold in this literature without his patient support and guidance. His comments and feedback on all of my work were crucial. I often came out of meetings grappling with a new insight he provided, realising how fuzzy my understanding of a given issue was prior to the meeting. This allowed my work to reach a level that I would not have been able to achieve on my own. I am also very grateful to Prof. Anders Kock. By working with him on the first chapter of my thesis, I learned an incredible amount on econometric theory. In particular, this collaboration taught me the importance of establishing rigorous proofs using methodical and transparent (ultimately, simple) steps. I will continue to treasure this way of thinking, in econometrics and beyond.

My work also benefitted from comments and feedback from: Dr. Elena Bobeica, Prof. Xu Cheng, Prof. Andrea Ferrero, Dr. Benny Hartwig, Prof. Jan Kiviet, Prof. Frank Kleibergen, Dr. Gerrit Koester, Prof. Damian Kozbur, Dr. Michele Lenza, Prof. Anna Mikusheva, Dr. Christiane Nickel, Prof. Bent Nielsen, Dr. Chiara Osbat, Prof. Mikkel Plagborg-Møller, and Dr. Riccardo Trezzi.

My doctoral studies would not have been possible without funding from the German National Merit Foundation, and the Department of Economics at the University of Oxford. This support is gratefully acknowledged.

Felix, Luke, Madeleine, Moritz, Sudarshana, and Valentin: thank you for all these years of friendship, and for keeping me sane throughout the past five. Thanks also to the chapel community at Brasenose College, and in particular to Julia.

I dedicate this thesis to my parents, Konstanze and Vincenzo. Thank you for not giving up on a recalcitrant child, and your unconditional support throughout my undergraduate and graduate studies. None of this would have been possible without you.

Abstract

This thesis consists of four self-contained chapters.

Chapter 2 (co-authored with Prof. Sophocles Mavroeidis and Prof. Anders Kock) considers hypothesis testing in instrumental variable (IV) regression models with few included exogenous covariates but many instruments–possibly more than the number of observations. We look for a method of inference that controls asymptotic size when there is heteroskedasticity and the instruments may be arbitrarily weak. We show that a ridge-regularised version of the jackknifed Anderson and Rubin (1949, henceforth AR) test achieves this objective. This test weakens the assumptions needed for recently proposed jackknifed AR tests, and extends their scope to situations in which there are more instruments than observations. Monte-Carlo simulations indicate that our method has favourable finite-sample size and power properties compared to recently proposed alternative approaches in the literature. An empirical application on the elasticity of substitution between immigrants and natives in the US illustrates the usefulness of the proposed method for practitioners.

Chapter 3 considers limited-information inference on New Keynesian Phillips Curves (NKPCs) in the presence of weak and high-dimensional IVs. Beyond the efficiency concerns previously raised in the literature, I show by simulation that ad-hoc selection procedures can lead to substantial biases in post-selection inference. I propose a Sup Score test that remains valid under dependent data, arbitrarily weak identification, and a number of IVs that increases exponentially with the sample size. Conducting inference on a standard NKPC with 361 IVs and 179 observations, I find substantially wider confidence sets than those commonly found.

Chapter 4 (co-authored with Dr. Gerrit Koester and Dr. Christiane Nickel) addresses the endogeneity of slack in Phillips Curves in reduced-form contexts. Endogeneity of the labour market slack in reduced-form Phillips Curves is usually addressed either by including proxies for omitted supply shocks, or by using instrumental variables. Using the Kiviet (2022) Kinky Least Squares estimator, we find evidence that supply-shock proxies should not be omitted from PCs, and that many popular instrumental variables seem to be invalid. We estimate a standard backward-looking wage Phillips Curve by Kinky Least Squares and find that unless a large *negative* correlation between the slack variable and the error term is assumed, the coefficient of the slack variable is significantly negative.

Chapter 5 provides a weak-identification robust IV method for identifying the structural covariance matrix in structural vector autoregressions (SVARs) where data for the reduced-form VAR are available over a longer horizon than for the IV. I apply this method to analyse the effect of US monetary shocks on real and financial variables in the US and other countries. While the effects on US real and financial variables are qualitatively similar to those reported previously in the literature, I find little evidence for US monetary policy spillovers in countries other than the US, in line with the theoretical predictions of the Mundellian Trilemma.

Statement of Co-Authorship

Some chapters were written as co-authored chapters. The following section provides an overview of co-authorship.

Chapter 1

This chapter is single-authored.

Chapter 2

This chapter is coauthored with Prof. Anders Kock and Prof. Sophocles Mavroeidis. Prof. Sophocles Mavroeidis had the original idea of developing a high-dimensional weak-identification robust test using regularisation techniques. After trying out several different statistics, I proposed the RJAR test used in this chapter. I also devised the broad strategy to deriving its limiting distribution. Prof. Anders Kock and Prof. Sophocles Mavroeidis helped rework several parts of the proofs. Prof. Anders Kock derived the proof showing that our assumptions nest the ones of the two approaches closest to the RJAR test. Prof. Sophocles Mavroeidis crucially reworked the framing of our proposed approach. I was responsible for most of the writing, the simulations, and the empirical applications.

Chapter 3

This chapter is single-authored.

Chapter 4

This chapter is coauthored with Dr. Gerrit Koester and Dr. Christiane Nickel, following a traineeship I completed at the European Central Bank in 2019. Dr. Gerrit Koester and Dr. Christiane Nickel suggested I apply new econometric methods to analyse Euro Area wage Phillips Curves. I suggested the use of the Kiviet (2022) estimator. I was subsequently responsible for the implementation of the analysis, and the drafting.

Chapter 5

This chapter is single-authored.

Chapter 6

This chapter is single-authored.

Contents

1 Introduction **1**

2 A Ridge-Regularised Jackknifed Anderson-Rubin Test **5**

2.1 Introduction 5

2.2 Model 9

2.3 The RJAR test 10

2.3.1 Definition of the RJAR test 10

2.3.2 Asymptotic properties of the RJAR test 12

2.3.3 Closest alternatives in the literature 14

2.4 Simulations 16

2.4.1 Size 17

2.4.2 Power 19

2.5 Empirical Application 26

2.6 Conclusion 31

3 Inference on the New Keynesian Phillips Curve with Very Many Instrumental Variables **33**

3.1 Introduction 33

3.2 Model 36

3.3 Methodology 37

3.3.1 Post-Selection Low-Dimensional Inference 38

3.3.2 A High-Dimensional Sup Score Test for Dependent Data . . 41

3.4 Simulations 44

3.5 Empirical Application 51

3.6 Conclusion 55

4 Addressing the Endogeneity of Slack in Phillips Curves **57**

4.1 Introduction 57

4.2 Model 60

4.3 Methodology and Data 61

4.3.1 Single-Equation Estimation 62

4.3.2 Instrumental Variable Estimation 62

4.3.3 Kinky Least Squares Estimation . . . 63
4.3.4 Data . . . 65
4.4 Results . . . 67
4.4.1 Single-Equation Estimation Results . . . 67
4.4.2 Instrumental Variable Estimation Results . . . 68
4.4.3 Using Kinky Least Squares to Test Phillips Curves Specifications 69
4.4.4 Kinky Least Squares Estimation Results . . . 78
4.5 Conclusion . . . 81

5 Robust Inference with Unbalanced Proxy SVARs and the Mundellian Trilemma 83
5.1 Introduction . . . 83
5.2 Methodology . . . 85
5.3 The Effect of US Monetary Policy Shocks . . . 89
5.4 Conclusion . . . 97

6 Conclusion 99

Appendices

A Appendix for Chapter 2 103
A.1 Proofs . . . 104
A.1.1 Lemmas . . . 104
A.1.2 Proof of Proposition 1 . . . 114
A.1.3 Proof of Theorem 1 . . . 116

B Appendix for Chapter 3 123
B.1 Proofs . . . 124
B.1.1 Proof of Theorem 2 . . . 124
B.1.2 Derivation of Concentration Parameters for Simulations . . 124
B.2 Data . . . 127

C Appendix for Chapter 4 131
C.1 Data . . . 131

D Appendix for Chapter 5 133
D.1 Proofs . . . 133
D.1.1 Proof of Proposition 2 . . . 133

References 137

1 Introduction

Instrumental Variables (IVs) are central for statistical inference in Economics. By extracting the exogenous variation of an endogenous variable, they make it possible to establish causal relations based on observational data. One well-known pitfall of using IVs is that they may contain little information on the causal effect the researcher is interested in studying. This is known as weak identification. Another pitfall that has recently received scholarly attention is the problem of (very) many IVs, where the number of IVs available is not negligible with respect to (and, in the case of 'very' many IVs, may in fact exceed) the number of observations. This is sometimes known as the (very) many IV problem. Although these two strands of the literature have individually received substantial attention in the past, less attention has been dedicated to studying the joint problem of (very) many weak IVs.

This thesis contributes to closing this gap in the literature by developing new econometric techniques to (jointly) address the problems of weak identification and the presence of (very) many IVs. In addition to providing new methodological tools for inference, this thesis also seeks to employ these (and other) econometric tools to help answer important macroeconomic questions, paying particular attention to how the results obtained through these new and more robust methods, compare to those previously reported in the literature. Therefore, this thesis also provides a contribution to the empirical Macroeconomics literature.

Chapter 2 introduces the ridge-regularised jackknifed Anderson and Rubin (1949) (RJAR) test. The RJAR test provides an inference procedure that remains valid under heteroskedastic data and very many weak IVs. The methodological contribution of this chapter is twofold. First, the RJAR test remains valid under conditions that are strictly weaker than the recently proposed tests of Crudu et al. (2021) and Mikusheva and Sun (2022). In particular, by deriving the limiting distribution of the test statistic from the bottom-up, we are able to show that the so-called balanced-design assumption made by these authors is stronger than needed. This is far from a uniquely theoretical consideration. In our empirical application taken from Card (2009), we are able to show that while the balanced-design assumption does not appear to hold, the assumption needed for the RJAR test to be valid appears to hold more plausibly. Second, the RJAR tests allows for the case where the number of IVs exceeds the number of observations (i.e., it allows for very many IVs). The only other approach that allows for more IVs than observations is the Sup Score test of Belloni et al. (2012). Furthermore, extensive Monte-Carlo evidence shows the very favourable power properties of the RJAR test compared to the alternatives of Belloni et al. (2012), Crudu et al. (2021), and Mikusheva and Sun (2022). One drawback of the RJAR test proposed in Chapter 2 is that it treats the IVs as fixed. While this assumption is reasonable in cross-sectional contexts, it is hardly ever appropriate in time-series contexts. This issue is considered in Chapter 3, where I propose a Sup Score test that remains valid under dependent data, while continuing to be robust to arbitrarily weak identification and very many IVs. This Sup Score is one of the first weak-identification robust inference procedures that remains valid under very many IVs and time-series data. Chapter 5 provides a weak-identification robust IV method for identifying the structural covariance matrix in 'unbalanced' structural vector autoregressions (SVARs), where data for the reduced-form VAR are available over a longer horizon than for the IV. A comparison of the confidence sets implied by this approach with the approach that estimates the reduced-form VAR using only data taken from the sample for which data for the IV is available suggests that the proposed approach has favourable power properties.

The second purpose of this thesis is to use these (and other existing) techniques to contribute to different strands of the empirical Macroeconomics literature.

Chapter 3 contributes to the empirical Macroeconomics literature in a twofold way. First, it uses extensive Monte-Carlo evidence based on a standard New Keynesian Phillips Curve (NKPCs) to show how endogenous selection of IVs can lead to invalid inference, even if the IVs are valid pre-selection. Second, it uses the Sup Score proposed in this thesis to obtain confidence intervals for a US NKPC that are robust to invalid inference due to endogenous selection. Substantially larger confidence sets are found than the ones reported in the literature. In particular, short-term inflation expectations are not found to significantly affect inflation today. Furthermore, comparing the IVs identified by the selection procedure that yielded the worst bias in the Monte-Carlo simulations to the IVs commonly used in previous studies provides suggestive evidence that previous results may have suffered from an IV selection bias. Although the Sup Score proposed in Chapter 3 is robust to IV selection biases, it still suffers from specification uncertainty. This is the inferential uncertainty that stems from the fact that it is impossible (even with the methods proposed in this thesis) to consider all the available IVs. This means that any inferential decision (even though it may not be biased) will be contingent on the particular choice of IVs. Chapter 4 illustrates this problem with reference to a standard wage Phillips Curve (PC) for the Euro Area. It is shown that, depending on the set of IVs chosen, the coefficient of the slack variable is sometimes found to be significantly different from zero, while other times it is not. Rather than arguing for a 'preferred specification' (which could increase the risk of selection bias), Chapter 4 proposes to turn the problem of endogeneity-robust inference on PCs on its head: using the recent econometric estimator of Kiviet (2022), it is argued that so long as the correlation between the slack variable and error term of the PC is not negative and large in absolute value, the coefficient on the slack variable is significantly negative. By arguing that the correlation between the slack variable and the PC's error term is positive, Chapter 4 hence provides evidence that the crucial transmission mechanism from the real economy to nominal wages is active in

the Euro Area. Finally, Chapter 5 uses the robust proxy SVAR method proposed in this thesis to test the predictions of Mundellian Trilemma by studying the spillover effects of US monetary policy. While the effects on US real and financial variables are qualitatively similar to those reported previously in the literature, little evidence for US monetary policy spillovers in countries other than the US is found, in line with the theoretical predictions of the Mundellian Trilemma.

Though related by a unifying theme, each of the following four chapters are presented as self-contained articles. All chapters include a detailed discussion of the particular contributions achieved in each of them, as well as a detailed survey of the relevant literatures.

2

A Ridge-Regularised Jackknifed Anderson-Rubin Test

2.1 Introduction

Instrumental variables (IVs) are commonly employed in economics and related fields to estimate causal effects from observational data. Interest in inference with (very) many IVs that are potentially weak has recently received increased attention for at least two reasons. First, when identification is weak, researchers may attempt to obtain more precise inference by using a large number of IVs to capture the limited exogenous variation in the endogenous covariates. Second, recent econometric approaches and theoretical results make a number of IVs that approaches or even exceeds the number of observations more common. For instance, the so-called granular IV approach proposed by Gabaix and Koijen (2020) leads to a number of IVs equal to the size of the cross-section of the panel considered. This number can approach, and potentially outstrip, the number of available observations. Another example is the so-called saturation approach to the identification of (weighted) local average treatment effects as in Blandhol et al. (2022), which involves considering as IVs a sufficient number of interactions between the original IVs and the covariates. The number of IVs generated in this way can easily outstrip the number of observations available. However, there is no guarantee that a vast

number of IVs will be jointly informative about the causal effects of interest.[1] Hence, it is important to have methods of inference that remain reliable when using (very) many weak IVs. The present chapter contributes to the literature on the development of such methods.

We propose a ridge-regularised jackknifed Anderson-Rubin (RJAR) test to construct confidence sets for the coefficients of endogenous variables in potentially weakly-identified and heteroskedastic IV models when the number of IVs is large. Jackknife-based methods have recently been used in this context because they provide an asymptotic framework where the number of IVs diverges with the number of observations. However, by relying on existing central limit theorems developed for standard projection matrices, these methods require that the number of IVs be less than the number of observations, and often perform poorly when the number of IVs is close to (but still less than) the number of observations. Recently-proposed regularisation approaches for inference under many IVs require strong identification or a sparse relationship between the endogenous variables and the IVs to work well. By combining jackknifing with ridge regularisation, we provide a test that remains valid under heteroskedasticity, arbitrarily weak identification, and more IVs than observations, all while achieving good power both when the relationship between the endogenous variables and the IVs is sparse and when it is dense.

In the context of fully weak-identification robust inference with an increasing number of IVs, it is helpful to distinguish three different asymptotic regimes. The first 'moderately many' IVs regime allows the number of IVs, k_n, to grow with the sample size, n, but still requires it to be asymptotically negligible with respect to the sample size. Examples include Andrews and Stock (2007), who require $k_n/n^{1/3} \to 0$, and Phillips and Gao (2017), who require $k_n/n \to 0$.

The second 'many' IVs regime allows the number of IVs to be of the same order of magnitude as the number of observations. Anatolyev and Gospodinov (2011) provide an Anderson and Rubin (1949, henceforth AR) test that remains valid for

[1] Pretesting the strength of IVs, though practical, could be rather inefficient as a way of controlling the size of subsequent inference on causal effects. Moreover, weak-IV pretests are not available when the number of IVs exceeds the number of observations.

$k_n/n \to \lambda$, $0 \leq \lambda < 1$, $n \to \infty$, provided that the error terms are homoskedastic and a strong balanced-design assumption of the IVs is satisfied.[2,3] Bun et al. (2020) provide analogous results for the GMM version of the AR statistic under the assumption of independently and identically distributed (i.i.d.) data. These results were recently extended in the form of a jackknifed AR statistic by Crudu et al. (2021, henceforth CMS) and Mikusheva and Sun (2022, henceforth MS) to the case where errors are allowed to display arbitrary heteroskedasticity, and the only assumption on the IVs is that the diagonal entries of the projection matrix of the IVs are bounded away from unity from above.

The third 'very many' IVs regime allows the number of IVs to grow with n, and further allows the number of IVs to be greater than the number of observations. Belloni et al. (2012, henceforth BCCH) propose a Sup Score test that remains valid under mild conditions, and allows the number of IVs to increase exponentially with the sample size. Carrasco and Tchuente (2016) propose a ridge-regularised AR statistic that allows for more IVs than observations under the assumption of i.i.d. data. Kapetanios et al. (2015) extend Bai and Ng (2010) by proposing weak-identification robust factor-based statistics that in principle allow for the number of IVs to be larger than the number of observations, provided that the factor structure of the IVs is sufficiently strong.

Using the notation of the model introduced in the next section, Table 2.1 provides a schematic overview of the main assumptions and results in the literature on inference that is robust to many weak IVs.

Our test provides a twofold extension of the existing literature. First, it allows for valid inference under many IVs while further weakening the assumptions of similar tests proposed by CMS and MS. This is made possible by deriving the limiting behaviour of the RJAR statistic from the bottom up, without relying on the existing results in Chao et al. (2012) or Hansen and Kozbur (2014). Second, this test

[2]See Anatolyev (2012), Anatolyev and Yaskov (2017), and Crudu et al. (2021) for a discussion on (the restrictiveness of) the strong balanced-design assumption.

[3]See Kaffo and Wang (2017) for a bootstrapped version of the Anatolyev and Gospodinov (2011) AR test.

allows for more IVs than observations. The only other approach currently available in the literature that is robust to heteroskedastic error terms and more IVs than observations is the Sup Score test of BCCH. Simulations show that the RJAR test has power comparable to the Sup Score test of BCCH whenever the signal in the first stage is sparse (i.e., only a few of the IVs are informative), and substantially more power when the signal in the first stage is dense (i.e., not sparse). Finally, we provide a comparison of the most recently proposed approaches to conducting inference in possibly heteroskedastic linear IV models when the number of IVs is not negligible with respect to the sample size. Indeed, using extensive simulation evidence and an empirical application based on Card (2009), we provide a comparison between these existing 'state-of-the-art' methods in a controlled and comparable setting.

Notation. I_p denotes the $p \times p$ identity matrix. The entry (i, j) for an $m \times p$ matrix A is denoted as A_{ij} for $1 \leq i \leq m$, $1 \leq j \leq p$. $||A||_F := \sqrt{\text{tr}(A'A)}$ denotes the Frobenius norm of A, and $\text{tr}(B) := \sum_{i=1}^{m} B_{ii}$ for any $m \times m$ matrix B with entries given by B_{ij} for $1 \leq i, j \leq m$. The remaining notation follows standard conventions.

Organisation of the chapter. Section 2.2 specifies the linear IV model considered throughout. Section 2.3 introduces the RJAR test, and provides the main asymptotic results. Section 2.4 provides simulation evidence on the size and power of our RJAR test and compares it with the tests proposed by BCCH, CMS and MS. Section 2.5 considers an empirical application based on Card (2009). Section 2.6 concludes. All proofs are given in the Appendix.

Table 2.1: Weak-identification robust inference with many IVs: schematic comparison of main assumptions and results in the literature.

	k_n	Z	ε
Anatolyev and Gospodinov (2011) and Kaffo and Wang (2017)	$k_n/n \to \lambda$, $0 \leq \lambda < 1$, $n \to \infty$	Z is fixed, obeys strong balanced design assumption	i.i.d., $\mathbb{E}[\varepsilon_i] = 0$, $\mathbb{E}[\varepsilon_i^4] < \infty$
Belloni et al. (2012)	$\log(\max(k_n, n)) = o(n^{1/3})$, $n \to \infty$	Z is fixed	independent, $\mathbb{E}[\varepsilon_i] = 0$, $\mathbb{E}[\varepsilon_i^3] < \infty$
Kapetanios et al. (2015)	Factor-dependent	Potentially dependent data, $\mathbb{E}[Z_{il}^4] < \infty$, $l = 1, \ldots, k_n$, $\mathbb{E}[\varepsilon_i^4] < \infty$	
Carrasco and Tchuente (2016)	$n \to \infty$	$Z_i\varepsilon_i$ is i.i.d., $\mathbb{E}[Z_i\varepsilon_i] = 0$, $\mathbb{E}[X_{ig}^2] < \infty$	
Bun et al. (2020)	$k_n/n \to \lambda$, $0 \leq \lambda < 1$, $n \to \infty$	$Z_i\varepsilon_i$ is i.i.d., Z obeys strong balanced design assumption, $\mathbb{E}[Z_i\varepsilon_i] = 0$, $\mathbb{E}[Z_i^8\varepsilon_i^8] < \infty$	
Crudu et al. (2021)	$k_n/n \to \lambda$, $0 \leq \lambda < 1$, $k_n \to \infty$ as $n \to \infty$	Z is fixed, $P_{ii} \leq 1 - \delta$, $0 < \delta < 1$	independent, $\mathbb{E}[\varepsilon_i] = 0$, $\mathbb{E}[\varepsilon_i^4] < \infty$
Mikusheva and Sun (2022)	$k_n/n \to \lambda$, $0 \leq \lambda < 1$, $k_n \to \infty$ as $n \to \infty$	Z is fixed, $P_{ii} \leq 1 - \delta$, $0 < \delta < 1$	independent, $\mathbb{E}[\varepsilon_i] = 0$, $\mathbb{E}[\varepsilon_i^6] < \infty$
RJAR	$r_n \to \infty$ as $n \to \infty$	Z is fixed, $\sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n})^2 \geq c r_n^{\omega}$, $c > 0$, $\omega \geq 1$	independent, $\mathbb{E}[\varepsilon_i] = 0$, $\mathbb{E}[\varepsilon_i^4] < \infty$

Notes:
n is the number of observations, k_n is the number of IVs, and g is the number of endogenous variables. Z is the $n \times k_n$ matrix of IVs. X is the $n \times g$ matrix of endogenous variables. ε is the $n \times 1$ vector of structural error terms. $r_n := \text{rank}(Z)$. See also Equation (2.1) below.
$P := Z(Z'Z)^{-1}Z'$, $P^{\gamma_n} := Z(Z'Z + \gamma_n I_{k_n})^{-1}Z'$ for $\gamma_n \geq 0$ if $r_n = k_n$ and $\gamma_n \geq \gamma_- > 0$ if $r_n < k_n$.

2.2 Model

We consider the heteroskedastic linear IV model given by

$$y = X\beta + \varepsilon \tag{2.1a}$$

$$X = Z\Pi + V, \tag{2.1b}$$

where y is an $n \times 1$ vector containing the dependent variable, X is an $n \times g$ matrix containing the endogenous variables, β is a $g \times 1$ coefficient vector, ε is an $n \times 1$ vector containing the structural error terms, Z is an $n \times k_n$ matrix containing the IVs, Π is a $k_n \times g$ coefficient matrix, and V is an $n \times g$ matrix containing the first-stage errors. k_n can diverge with n, but g is fixed. Also, let y_i, X_i, ε_i, Z_i, and V_i denote the i^{th} row of y, X, ε, Z, and V, respectively. As in BCCH, CMS and MS,

we treat Z as fixed (non-stochastic), and assume that any exogenous covariates have been partialled out (see the discussion of Assumption 3 below). We *exclusively* consider methods that allow for arbitrarily weak identification, i.e., methods that control asymptotic size irrespective of the value of Π.

Inference is conducted on the coefficient vector β by testing hypotheses of the following type for a prespecified $\beta_0 \in \Re^g$:

$$H_0 : \beta = \beta_0 \text{ vs. } H_1 : \beta \neq \beta_0. \tag{2.2}$$

For a given non-randomised test of asymptotic size $\alpha \in (0, 1)$, a confidence set of asymptotic coverage $1 - \alpha$ can be constructed as the collection of those β_0 for which H_0 in Equation (2.2) is not rejected. For convenience, define

$$e(\beta_0) \coloneqq [e_1(\beta_0), \ldots, e_n(\beta_0)]', \qquad e_i(\beta_0) \coloneqq y_i - X_i\beta_0, \tag{2.3}$$

which we refer to as the structural error under the null hypothesis.

It should be noted that so long as the error term remains additive, the linearity in the structural equation (2.1a) and the first-stage equation (2.1b) can be relaxed to allow for any (known) real-valued function, without affecting the asymptotic properties of the RJAR test.

2.3 The RJAR test

This section introduces our proposed RJAR test, derives its large sample properties under the null hypothesis in Equation (2.2), and discusses its relationship to the most closely related tests in the literature: the jackknifed AR tests of CMS and MS, and the Sup Score test of BCCH.

2.3.1 Definition of the RJAR test

The original AR test of the null hypothesis in Equation (2.2) can be thought of as a test that the IVs are exogenous using the implied structural errors in Equation (2.3). More specifically, the AR test is a Wald test of the significance of the IVs in the auxiliary regression of the structural errors under the null, $e_i(\beta_0)$, on the

IVs, Z_i.[4] When the number of IVs k_n grows with n, i.e., when there are many (potentially weak) IVs, the χ^2 approximation of the original AR statistic breaks down. The recent chapters by CMS and MS that we shall briefly review in the next subsection propose jackknifed versions of the AR test that remain valid when k_n grows with n, but $k_n < n$. Our proposed method combines jackknifing with ridge regularisation of the auxiliary regression of $e_i(\beta_0)$ on Z_i.

We first standardise the IVs in-sample (after partialling out any covariates) as in BCCH (p. 2393) so that

$$\frac{1}{n}\sum_{i=1}^{n} Z_{ij}^2 = 1 \quad \text{for} \quad j = 1, \ldots, k_n. \tag{2.4}$$

The RJAR test for the testing problem in Equation (2.2) is then based on the following statistic:

$$RJAR_{\gamma_n}(\beta_0) := \frac{1}{\sqrt{r_n}\sqrt{\hat{\Phi}_{\gamma_n}(\beta_0)}} \sum_{i=1}^{n}\sum_{j\neq i} P_{ij}^{\gamma_n} e_i(\beta_0) e_j(\beta_0), \tag{2.5}$$

where $r_n := \text{rank}(Z)$ (assumed to be positive without loss of generality),

$$\hat{\Phi}_{\gamma_n}(\beta_0) := \frac{2}{r_n}\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^2 e_i^2(\beta_0) e_j^2(\beta_0), \tag{2.6}$$

and $P^{\gamma_n} := Z(Z'Z + \gamma_n I_{k_n})^{-1}Z'$ for $\gamma_n \geq 0$ if $r_n = k_n$ and $\gamma_n > 0$ if $r_n < k_n$. γ_n is a (sequence of) regularisation parameter(s) whose choice we discuss next.

We set the penalty parameter γ_n to

$$\gamma_n^* := \max \arg\max_{\gamma_n \in \Gamma_n} \sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^2, \tag{2.7}$$

where $\Gamma_n := \{\gamma_n \in \Re : \gamma_n \geq 0 \text{ if } r_n = k_n, \text{ and } \gamma_n \geq \gamma_- > 0 \text{ if } r_n < k_n\}$ for some constant γ_- not depending on n. The existence of γ_n^* is shown in Lemma 3 in the Appendix. We let γ_n^* be an element of $\arg\max_{\gamma_n \in \Gamma_n} \sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^2$ out of conservativeness, i.e., to make Assumption 3 below as plausible as possible given the IVs. Furthermore, we let γ_n^* be the maiximal element of this set because

[4]The AR statistic is defined as $e(\beta_0)'Z\widehat{\Sigma}Z'e(\beta_0)/n$, where $\widehat{\Sigma}$ is a consistent estimator of $\text{Var}[Z'e(\beta_0)/\sqrt{n}]$. The weak-IV-robust AR test of the null hypothesis in Equation (2.2) with asymptotic size α rejects the null if and only if the AR statistic exceeds the $1-\alpha$ quantile of a χ^2 distribution with k_n degrees of freedom.

the maximiser of the first argument is not necessarily unique without imposing additional assumptions on the singular values and left-singular vectors of the IVs (although in practice we only found unique maximisers). We choose to take the maximum of the maximisers to make the smallest eigenvalues of the ridge-regularised Gram matrix, $Z'Z + \gamma_n I_{k_n}$, as far away from zero as possible when $r_n < k_n$.

We now have all the ingredients to define our new test.

Definition 1. *The RJAR test rejects* $H_0 : \beta = \beta_0$ *in Equation* (2.2) *at significance level* $\alpha \in (0, 1)$ *if and only if*

$$RJAR_{\gamma_n^*}(\beta_0) > \mathcal{Q}(1-\alpha), \tag{2.8}$$

where $RJAR_{\gamma_n}(\beta_0)$ *is defined in Equation* (2.5), γ_n^* *is defined in Equation* (2.7), *and* $\mathcal{Q}(1-\alpha)$ *is the* $(1-\alpha)$ *quantile of the Standard Normal distribution.*

2.3.2 Asymptotic properties of the RJAR test

We make the following assumptions to derive the limiting distribution of $RJAR_{\gamma_n^*}(\beta_0)$ under the null hypothesis in Equation (2.2).

Assumption 1. $\{\varepsilon_i\}_{i\in\mathbb{N}}$ is a sequence of independent random variables satisfying $\mathbb{E}[\varepsilon_i] = 0$, $\inf_{i\in\mathbb{N}} \mathrm{Var}[\varepsilon_i] > 0$, and $\sup_{i\in\mathbb{N}} \mathbb{E}[\varepsilon_i^4] < \infty$.

Assumption 2. $r_n = \mathrm{rank}(Z) \to \infty$ as $n \to \infty$.

Assumption 3. There exist $\gamma_- > 0$, $c > 0$, $\omega \geq 1$, $N \geq 2$ such that for all $n \geq N$

1. If $r_n = k_n$, then there exists a $\gamma_n \in [0, \infty)$ such that $\sum_{i=1}^n \sum_{j\neq i} (P_{ij}^{\gamma_n})^2 \geq c r_n^\omega$
2. If $r_n < k_n$, then there exists a $\gamma_n \in [\gamma_-, \infty)$ such that $\sum_{i=1}^n \sum_{j\neq i} (P_{ij}^{\gamma_n})^2 \geq c r_n^\omega$

Assumption 1 is a mild condition on the structural error terms, and allows for conditional heteroskedasticity. It is the same as in CMS. It is slightly less restrictive than the one in MS (who require finite sixth moments on the structural error terms), and slightly more restrictive than the one in BCCH (who require finite third moments).

Assumption 2 is a weak technical assumption. It implies that both k_n and n diverge. It also allows for the sum of the number of IVs and the number of exogenous covariates to be larger than the number of observations, provided that the number of exogenous covariates that have been partialled out be sufficiently small (so that the rank of the matrix of partialled IVs continues to diverge). This assumption is weaker than the restriction on the dimensionality in MS and CMS, who require $r_n = k_n$, $k_n < n$ for each $n \in \mathbb{N}$, and $k_n \to \infty$ as $n \to \infty$. BCCH prove asymptotic size control of their Sup Score test under the assumption that $\log k_n = o(n^{1/3})$.

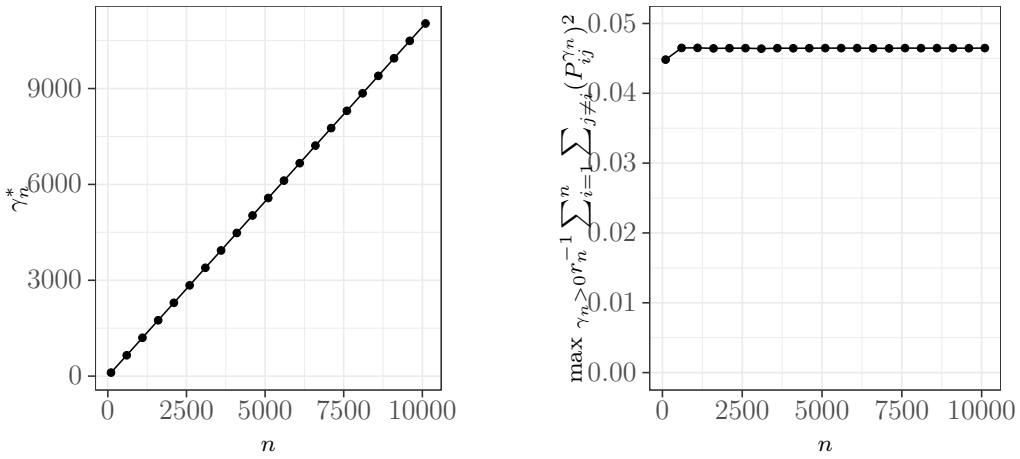
Assumption 3 is a high-level assumption on the number of IVs and their correlation structure. Assumption 3 implies that γ_n^* satisfies $\sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n^*})^2 \geq c r_n^\omega$. When $k_n < n$ (and there are no exogenous covariates, as in CMS and MS), Assumption 3 is weaker than the *balanced-design* assumption in CMS and MS which requires for $P := Z(Z'Z)^{-1}Z'$

$$\max_{1 \leq i \leq n} P_{ii} \leq 1 - \delta, \tag{2.9}$$

for some $0 < \delta < 1$. The balanced-design assumption implies Assumption 3.

Proposition 1. Suppose (as in CMS and MS) that $\text{rank}(P) = k_n$ and that there exists a $0 < \delta < 1$ such that $\max_{1 \leq i \leq n} P_{ii} \leq 1 - \delta$. Then Assumption 3 is satisfied.

It is difficult to give more primitive conditions for Assumption 3 that remain both easily-verifiable and sufficiently general when $k_n > n$. Figure 2.1 shows that for the Gaussian data-generating process (DGP) design used in the simulations in Section 2.4, the ratio of $\sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n})^2$ to r_n is constant, lending plausibility to Assumption 3 in this context. Figure 2.1 also shows that γ_n^* grows with the number of observations, which implies that our penalty parameter may grow with the sample size (in contrast to the one proposed in Carrasco and Tchuente (2016)). In practice, and similarly to the heuristics recommended in CMS, MS, and Hansen and Kozbur (2014), we recommend that in a given application, practitioners check the implied minimum value of c in Assumption 3. If very small values are observed, then Assumption 3 may be questionable.



(a) **(b)**

Figure 2.1: Numerical Evidence on Assumption 3 for $k_n = 1.9n$. Gaussian IVs drawn with mean 0 and $\text{Var}[Z_{il}] = 0.3$, $\text{Corr}[Z_{il}, Z_{im}] = 0.5^{|l-m|}$, and independently across i.

We are now in a position to derive the asymptotic distribution of the RJAR test under the null hypothesis.

Theorem 1. Suppose Assumptions 1–3 hold. Then, under the null hypothesis in Equation (2.2), for any sequence γ_n such that, for all n sufficiently large, $\sum_{i=1}^{n} \sum_{j \neq i} (P_{ij}^{\gamma_n})^2 \geq c r_n^{\omega}$ for some $c > 0$ and $\omega \geq 1$, the statistic $RJAR_{\gamma_n}(\beta_0)$ defined in Equation (2.5) satisfies

$$RJAR_{\gamma_n}(\beta_0) \xrightarrow{d} \mathcal{N}[0, 1].$$

Corollary 1. Under Assumptions 1–3 and the null hypothesis in Equation (2.2), the RJAR test given in Definition 1 has asymptotic size α.

Notice that we did not impose any assumption on the coefficients of the instruments Π in the first-stage regression in Equation (2.1). Therefore, the RJAR test is robust to arbitrarily weak identification.

2.3.3 Closest alternatives in the literature

The RJAR test is similar to the jackknifed AR tests proposed in CMS and MS. The distinguishing feature is the use of a ridge-regularised projection matrix P^{γ_n}, which

makes the RJAR test applicable also when $r_n < k_n$, unlike the aforementioned two tests that use the normal projection matrix P, and cannot be computed when $k_n > n$.

CMS assume $r_n = k_n < n$, and propose the jackknifed AR statistic given by

$$AR_{CMS}(\beta_0) := \frac{1}{\sqrt{k_n}\sqrt{\hat{\Phi}_{CMS}(\beta_0)}} \sum_{i=1}^{n} \sum_{j \neq i} C_{ij} e_i(\beta_0) e_j(\beta_0),$$

where $C := A - B$, $A := P + \Delta$, $B := (I_n - P)D(I_n - D)^{-1}(I_n - P)$, $\Delta := PD(I_n - D)^{-1}P - \frac{1}{2}PD(I_n - D)^{-1} - \frac{1}{2}D(I_n - D)^{-1}P$, and D is the diagonal matrix containing the diagonal elements of P. $\hat{\Phi}_{CMS}(\beta_0) := \frac{2}{k_n}\sum_{i=1}^{n}\sum_{j\neq i} C_{ij}^2 e_i^2(\beta_0) \times e_j^2(\beta_0)$. Under the null hypothesis in Equation (2.2), CMS show that $AR_{CMS}(\beta_0)$ converges to a Standard Normal distribution. The CLT underlying this result is a modified version of Lemma A2 of Chao et al. (2012) proposed in Bekker and Crudu (2015, Appendix A.4).

MS also assume that $r_n = k_n < n$, and propose a different jackknifed AR statistic that can be obtained from $RJAR_{\gamma_n}(\beta_0)$ in Equation (2.5) by setting $\gamma_n = 0$, and replacing $\hat{\Phi}_{\gamma_n}(\beta_0)$ with

$$\hat{\Phi}_{MS}(\beta_0) := \frac{2}{k_n} \sum_{i=1}^{n} \sum_{j \neq i} \frac{P_{ij}^2}{M_{ii}M_{jj} + M_{ij}^2} \left[e_i(\beta_0) M_i e(\beta_0)\right] \left[e_j(\beta_0) M_j e(\beta_0)\right], \qquad (2.10)$$

where $M = I_n - P$, and M_i is the i^{th} row of M. The reason why the unregularised jackknifed AR test in MS uses $\hat{\Phi}_{MS}(\beta_0)$ instead of the variance estimator given in Equation (2.6) evaluated at $\gamma_n = 0$, is because, according to their Theorem 4, the former yields higher power than the latter. It follows that the unregularised version of our RJAR test, which arises when $\gamma_n^* = 0$ in Equation (2.7), will be dominated by the jackknifed AR test of MS in terms of power, and practitioners may prefer the latter in those cases. We note, however, that the asymptotic size control and superior power of the jackknifed AR test of MS are proven under the balanced design assumption given in Equation (2.9) which is strictly stronger than our Assumption 3. This can have implications for the finite-sample performance of the jackknifed AR test of MS compared to the RJAR, which we investigate in Section 2.4.

When k_n is larger than n, the AR tests in CMS and MS are not applicable, since $Z'Z$ does not have full rank. In this case, the Sup Score test of BCCH

remains valid. BCCH first standardise the IVs as in Equation (2.4), and then propose the Sup Score statistic given by

$$S(\beta_0) = \max_{1 \leq j \leq k_n} \frac{\sqrt{n} \left| e(\beta_0)' z_j \right|}{\sqrt{(e(\beta_0) \otimes e(\beta_0))'(z_j \otimes z_j)}},$$

where z_j is the $n \times 1$ vector containing the j^{th} column of Z, i.e., the j^{th} instrument.

BCCH propose to use the critical value $\nu_\alpha = c_{BCCH} \sqrt{n} \mathcal{Q}\left(1 - \alpha/(2k_n)\right)$, $c_{BCCH} > 1$ and $\alpha \in (0, 1)$. They show that comparing their Sup Score statistic to this critical value yields a test of the null hypothesis in Equation (2.2) that has asymptotic size less than or equal to α. Being a supremum-norm test suggests that the BCCH Sup Score test will work well with a sparse first stage (i.e. where only a few elements of Π are zero), but may have lower power than the RJAR test when the first stage is dense. This is verified in the simulations in Section 2.4.

2.4 Simulations

We now investigate the size and power properties of the RJAR test and compare them to those of the tests proposed in BCCH, CMS and MS. We take our simulation setup from Hansen and Kozbur (2014) who in turn take theirs from BCCH. The DGP is given by

$$y_i = X_i \beta + \varepsilon_i \tag{2.11a}$$

$$X_i = Z_i' \pi + v_i, \tag{2.11b}$$

$$\begin{bmatrix} \varepsilon_i \\ v_i \end{bmatrix} \sim \mathcal{N}\left[0, \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon v} \\ \sigma_{\varepsilon v} & \sigma_v^2 \end{bmatrix}\right],$$

for $i = 1, \ldots, n = 100$, with $\sigma_\varepsilon^2 = 2$, $\sigma_v^2 = 1$ and $\sigma_{\varepsilon v} = 0.6\sigma_\varepsilon \sigma_v$.[5] The IVs are independent Gaussian with mean 0 and $\text{Var}[Z_{il}] = 0.3$ and $\text{Corr}[Z_{il}, Z_{im}] = 0.5^{|l-m|}$.[6] $\pi = \varrho\kappa$, where κ is a vector of zeros and ones that varies with the type of DGP

[5] Heteroskedasticity could be incorporated in standard ways, e.g., by scaling the structural error by the value of a randomly selected IV for every observation. As is to be expected given that all the approaches considered here are robust to heteroskedasticity of arbitrary form, such modifications do not qualitatively affect the simulations.

[6] Unreported simulations show that the results remain qualitatively unchanged if uncorrelated Gaussian IVs are considered instead.

considered (sparse or dense, as modelled below), and ϱ is some scalar that ensures that for a given concentration parameter, μ^2, the following relationship is satisfied:

$$\mu^2 = \frac{n\pi'\mathbb{E}\left[Z_i Z_i'\right]\pi}{\sigma_v^2}.$$

This implies that

$$\varrho = \sqrt{\frac{\sigma_v^2\mu^2}{n\kappa'\mathbb{E}\left[Z_i Z_i'\right]\kappa}}.$$

To illustrate how the sparsity structure of the first stage in Equation (2.11b) can affect the size and power of the studied tests, we consider both a sparse first stage and a dense first stage. Sparsity in the first stage is modelled by setting $\kappa = [\iota_5', 0_{k_n-5}']'$, where ι_q is a $q \times 1$ vector of ones, and 0_q is a $q \times 1$ vector of zeros. Density in the first stage is modelled by setting $\kappa = [\iota_{0.4k_n}', 0_{0.6k_n}']'$.

We consider $k_n = 30, 90, 190$. Throughout, γ_- in Assumption 3 is set equal to 1. For the case of 30 IVs, the RJAR test does not impose any regularisation ($\gamma_n^* = 0$). For the case of 90 IVs, $\gamma_n^* = 12.048$. We note that $\sum_{i=1}^n \sum_{j\neq i}(P_{ij}^{\gamma_n^*})^2 = 11.123 > 8.845 = \sum_{i=1}^n \sum_{j\neq i}(P_{ij})^2$. This shows that even in the case where $r_n < n$, ridge regularisation can make Assumption 3 strictly more plausible. For the case of 190 IVs, $\gamma_n^* = 109.187$. The variance estimator of MS occasionally yields a negative value. These cases are conservatively interpreted as a failure to reject the null hypothesis. As recommended by BCCH, $c_{BCCH} = 1.1$. The number of Monte Carlo replications is 10,000.

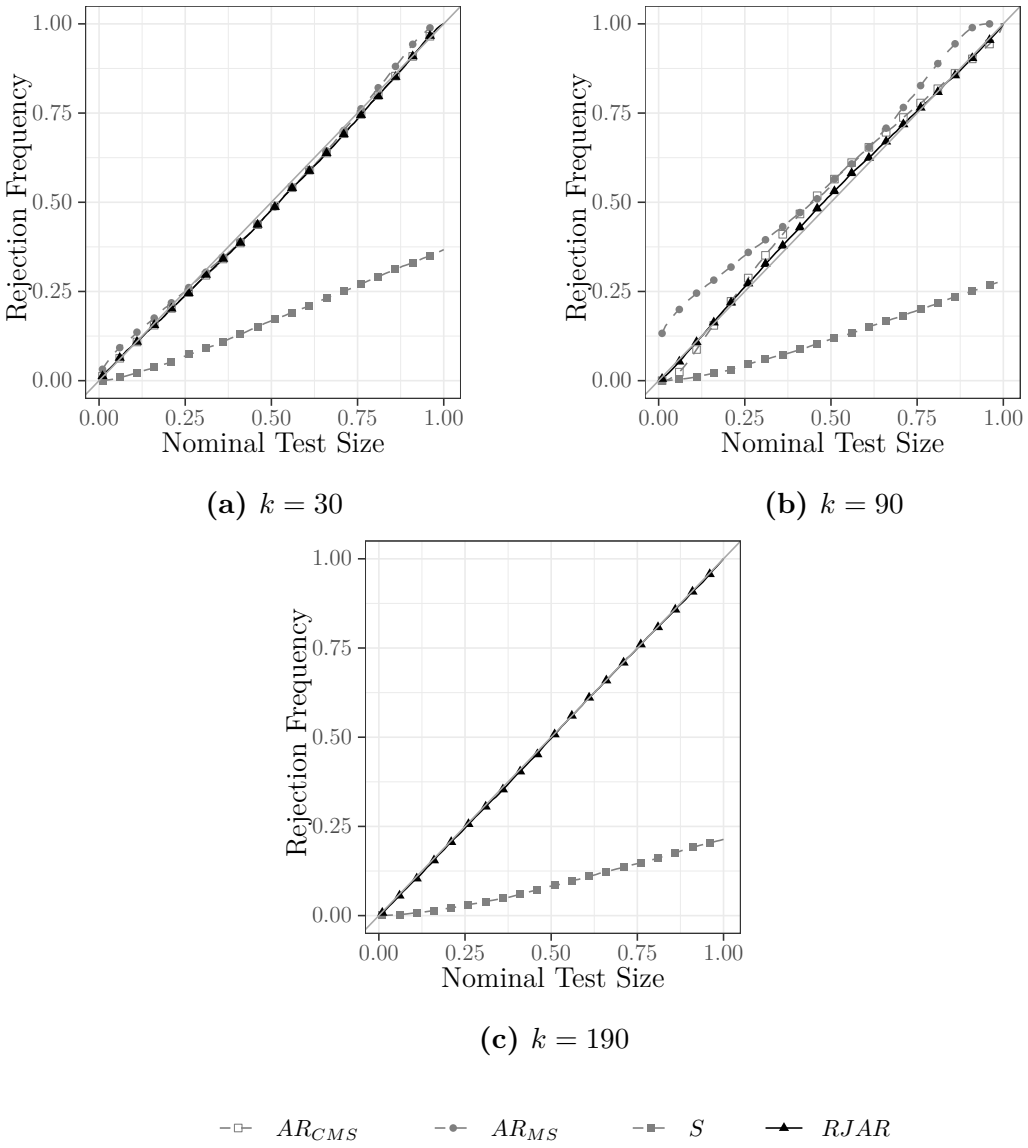
2.4.1 Size

Figure 2.2 shows the simulation results for the sparse first stage for tests of size 0.01 to 0.99, that is the rejection frequency under $H_0 : \beta_0 = 1$. As far as the illustration of the tests' size properties is concerned, the dense first stage yields virtually the same results, and is hence omitted. Since all tests are robust to weak IVs, the rejection frequencies of the tests are not affected by the value of the concentration parameter. For the case of 30 IVs, the AR tests of CMS and MS and the RJAR test have correct size, while the BCCH Sup Score test is undersized.

For the case of 90 IVs, the AR test of CMS appears to control size for common small nominal test sizes (e.g., 0.05 or 0.1). The AR test of MS appears to be generally oversized. For example, at nominal level 0.05, the rejection frequency of the test is 0.189.[7] The BCCH Sup Score test continues to be undersized, while the RJAR test continues to have correct size irrespective of the value of μ^2.

For the case of 190 IVs, only the BCCH Sup Score test and the RJAR test are feasible. As before, the BCCH Sup Score test is undersized, while the RJAR test has correct size.

[7] Unreported simulations confirm that if n and k_n are increased to about 1,000 and 900, respectively, there is no longer any size distortion.



(a) $k = 30$ **(b)** $k = 90$

(c) $k = 190$

Figure 2.2: PP Plots for Sparse IVs, $\mu^2 = 0$, $H_0 : \beta_0 = 1$.

2.4.2 Power

Figures 2.3–2.8 show the power results when the number of IVs and the sparsity pattern of the first stage is varied. For the case of 30 sparse IVs (Figure 2.3), all tests have similar power for the different degrees of identification considered. For the case of 30 dense IVs (Figure 2.4), the AR tests of CMS and MS have similar power to the RJAR test. In this case, the BCCH Sup Score test is noticeably less powerful. For the case of 90 sparse IVs (Figure 2.5), the BCCH Sup Score test and

the RJAR test exhibit very similar power properties. As shown in the subsection above, the AR test of MS fails to control the size of the test, while the AR test of CMS exhibits power properties substantially worse than those of the BCCH Sup Score test and the RJAR test. For the case of 90 dense IVs (Figure 2.6), the AR tests of CMS and MS continue to exhibt the same lower power and failure to control the test size as in the case of 90 sparse IVs. In contrast to the case of 90 sparse IVs, however, the RJAR test is substantially more powerful than the BCCH Sup Score test. For the case of 190 IVs, the results suggest that the RJAR test is more powerful than the BCCH Sup Score test for both the sparse and the dense first stage. This recommends the RJAR test vis-à-vis the BCCH Sup Score test.

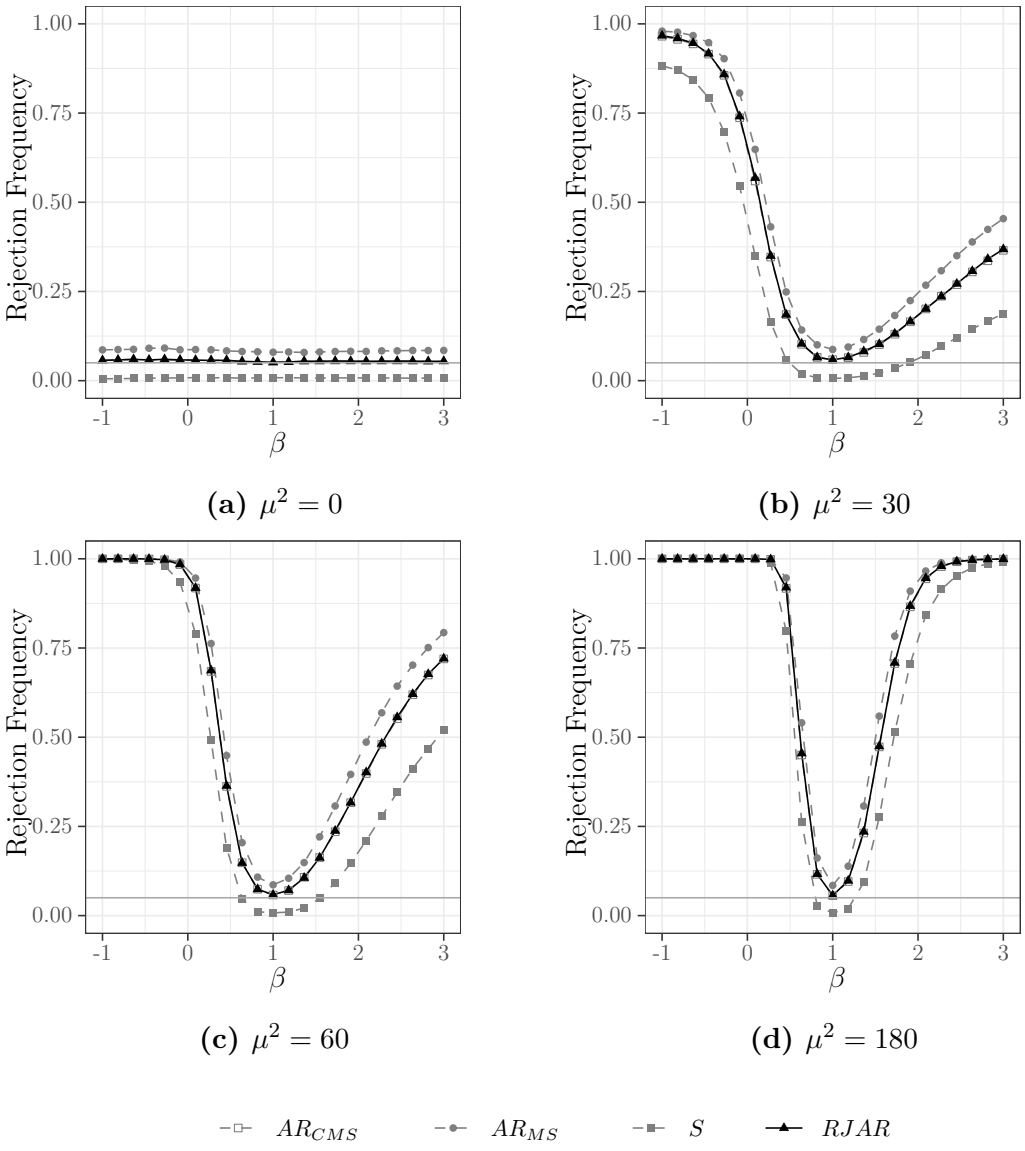


Figure 2.3: Power Curves for 30 Sparse IVs. Nominal test size: 5%.

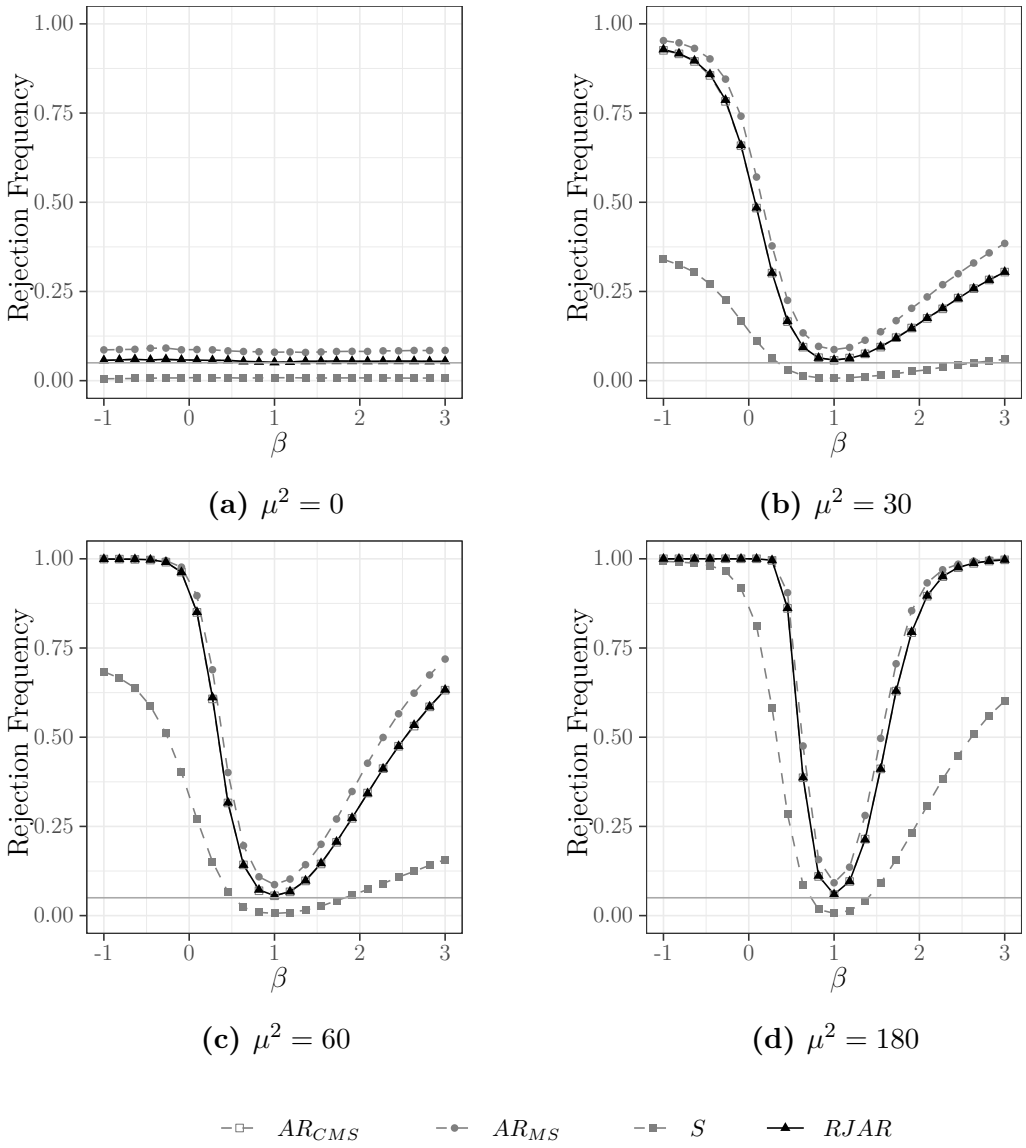


Figure 2.4: Power Curves for 30 Dense IVs. Nominal test size: 5%.

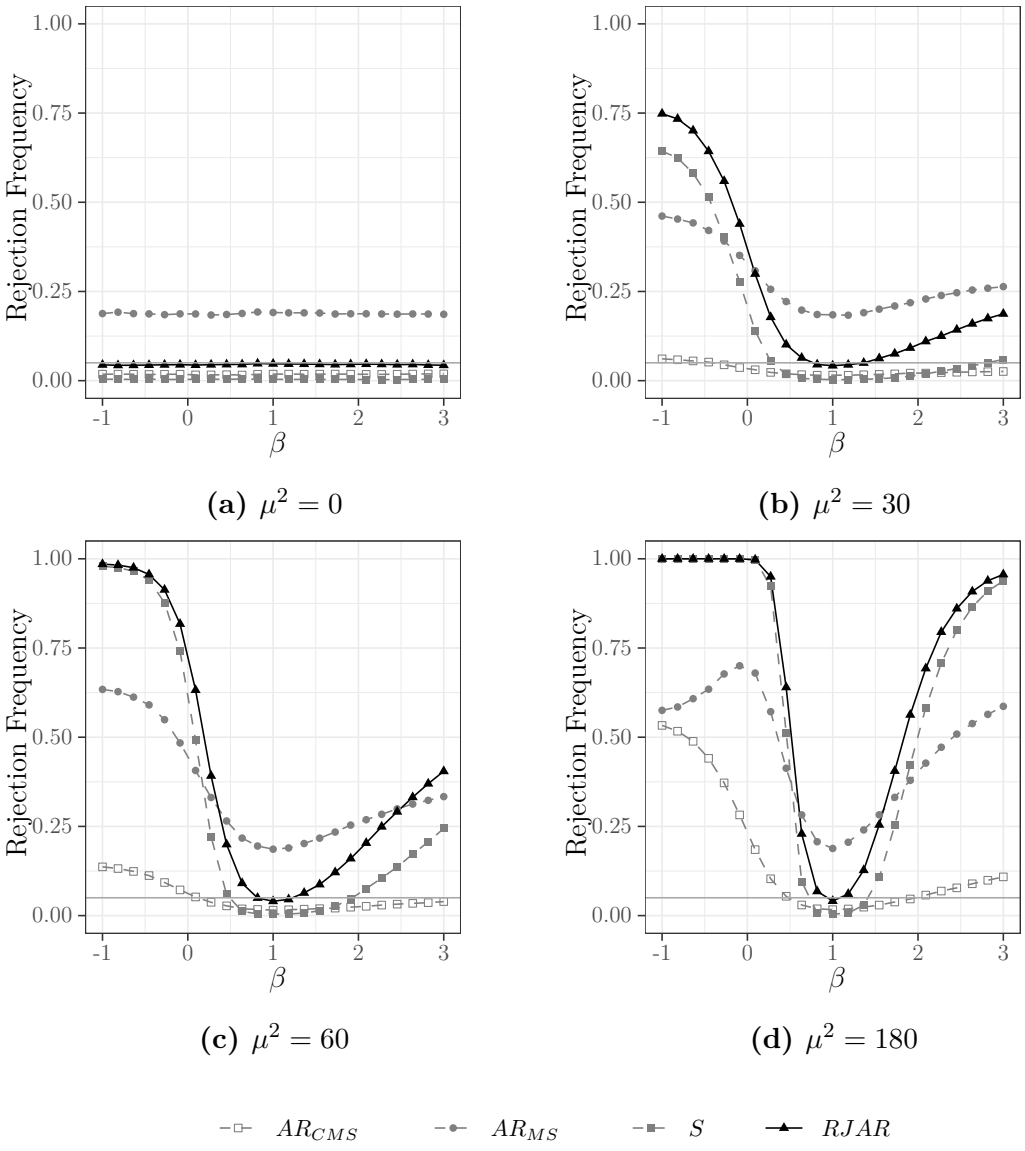
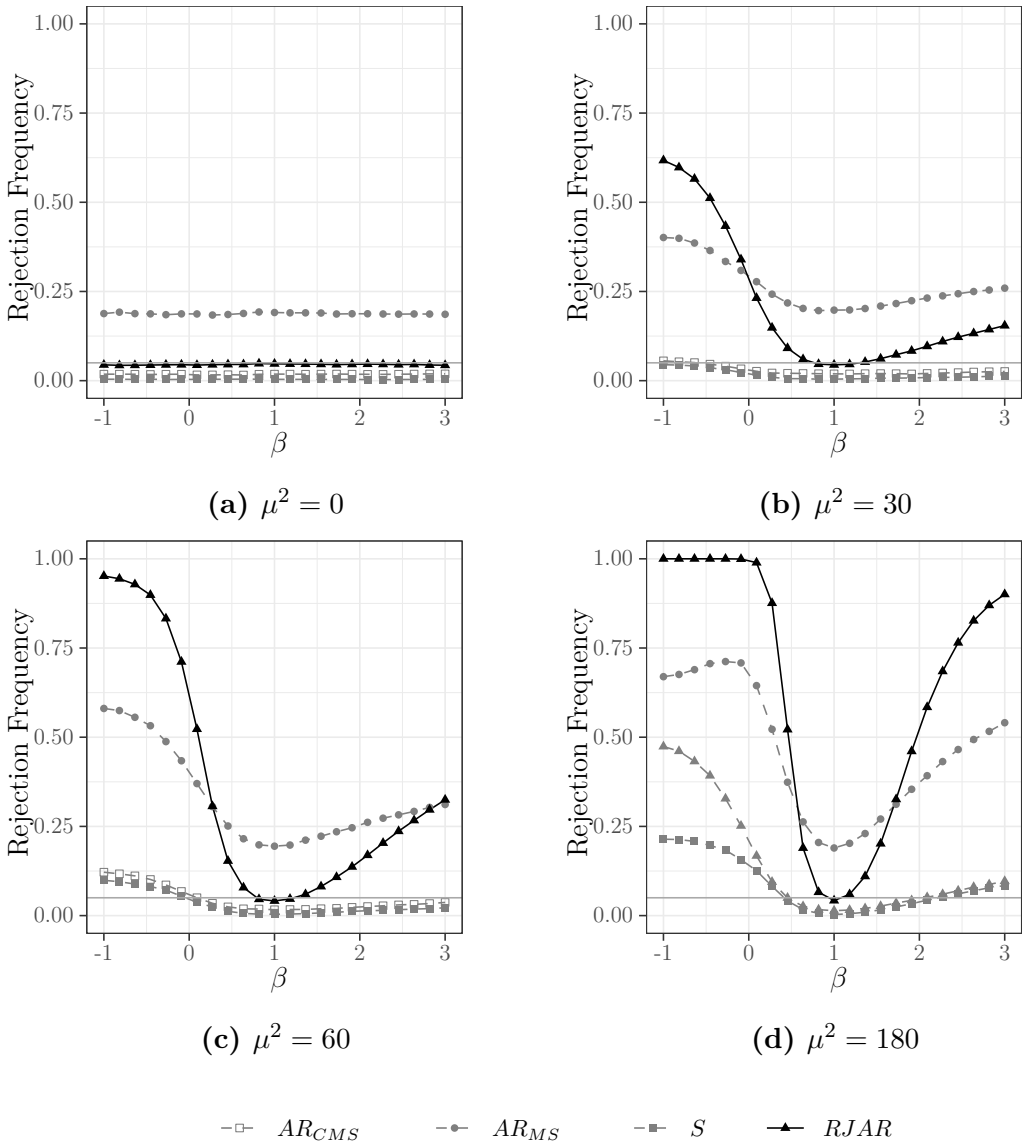


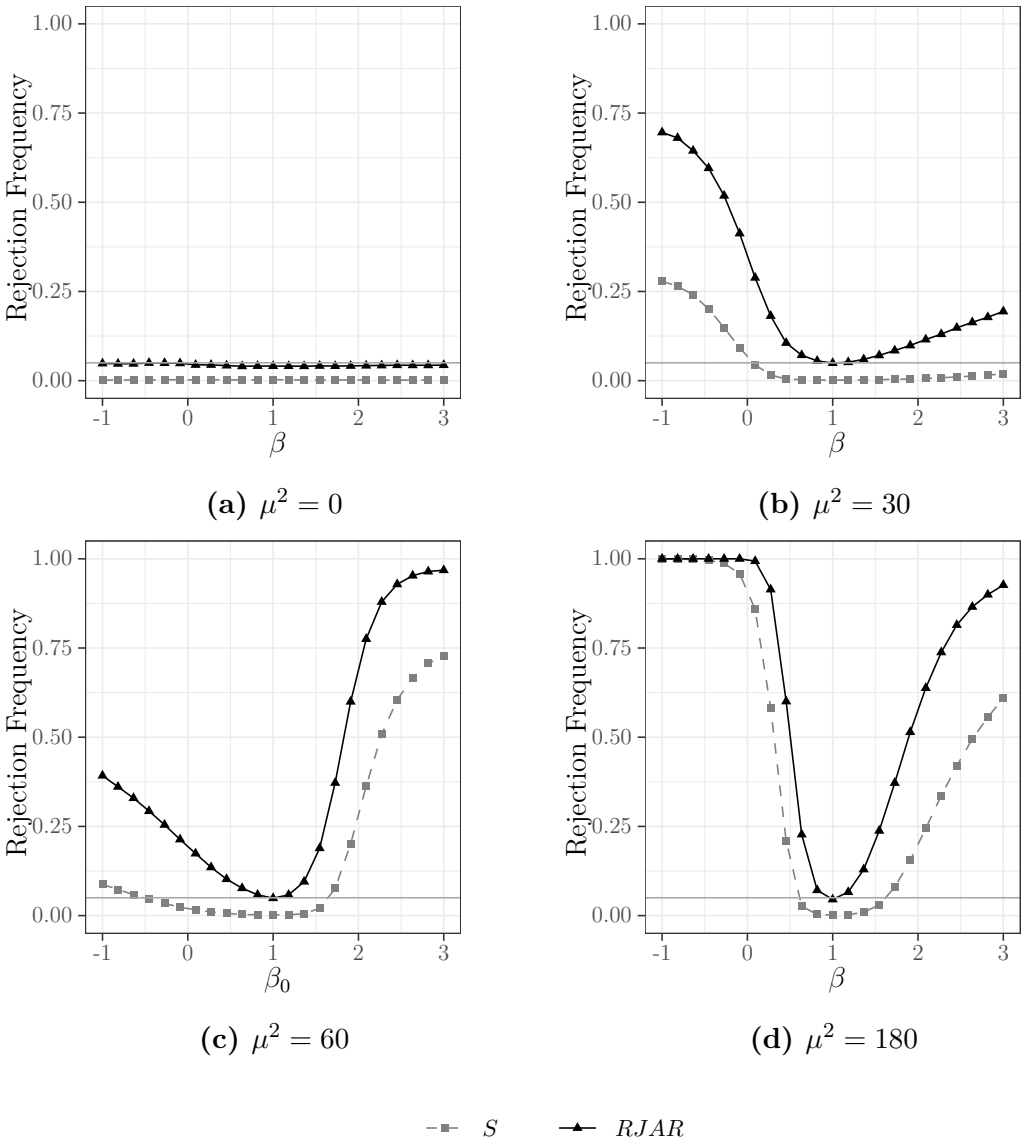
Figure 2.5: Power Curves for 90 Sparse IVs. Nominal test size: 5%.



(a) $\mu^2 = 0$ **(b)** $\mu^2 = 30$

(c) $\mu^2 = 60$ **(d)** $\mu^2 = 180$

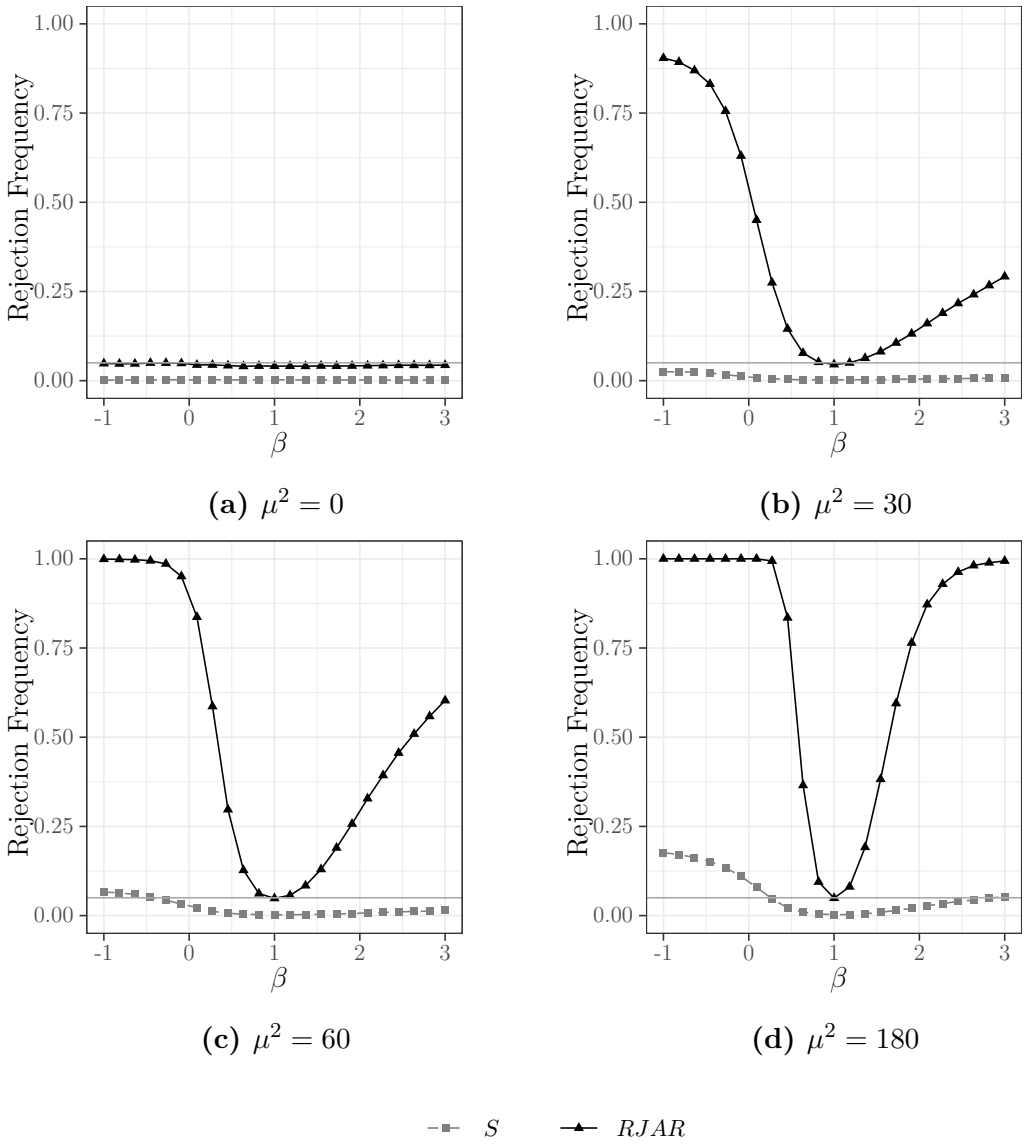
Figure 2.6: Power Curves for 90 Dense IVs. Nominal test size: 5%.



(a) $\mu^2 = 0$ **(b)** $\mu^2 = 30$

(c) $\mu^2 = 60$ **(d)** $\mu^2 = 180$

Figure 2.7: Power Curves for 190 Sparse IVs. Nominal test size: 5%.



(a) $\mu^2 = 0$ **(b)** $\mu^2 = 30$

(c) $\mu^2 = 60$ **(d)** $\mu^2 = 180$

Figure 2.8: Power Curves for 190 Dense IVs. Nominal test size: 5%.

2.5 Empirical Application

We consider an empirical application from Card (2009), who estimates the (negative) inverse elasticity of substitution between immigrants and natives in the US. The coefficient of interest is given by β_s in the following model:

$$y_{is} = \beta_s X_{is} + \delta_s' W_i + \varepsilon_{is}, \tag{2.12}$$

where y_{is} is the residual log wage gap between immigrant and native men in skill group s in city i, X_{is} is the log ratio of immigrant to native hours in skill group s of both men and women in city i, and W_i is a vector of city-level controls with coefficient vector δ_s, and ε_{is} is the structural error.

Following Card (2009), we identify the coefficient β_s by using as IVs the ratio of the number of immigrants from country l in city i to the total number of immigrants from foreign country l. The rationale for these IVs is that existing immigrant enclaves are likely to attract additional immigrant labour through social and cultural channels unrelated to labour market outcomes. As in Card (2009), we consider two skill groups (high school or college equivalent) separately. We consider two sets of IVs. First, we consider the original setup of Card (2009), using as IVs the 38 different countries from which immigrants emigrated. Second, motivated by the saturation approach of Blandhol et al. (2022), we consider the setup where these 38 original IVs are interacted with the nine available controls (including a constant). This yields 342 IVs. In both cases, the number of observations (i.e., the number of cities) is 124. Data is taken from a single cross section, as made available by Goldsmith-Pinkham et al. (2020).

As in the simulation exercise in Section 2.4, $\gamma_- = 1$, and $c_{BCCH} = 1$. A grid of 100 values for β_0 is used.

Figure 2.9 and Figure 2.10 show the confidence sets when $k_n = 38$ for high school workers and college workers, respectively. We find that $\gamma_n^* = 0$, implying that no regularisation is needed. This is in line with our simulations in Section 2.4, where we found that regularisation was not needed to maximise the sum in Assumption 3 when $k_n/n = 0.3$. The crucial balanced design assumption of CMS and MS appears questionable, since $\max_i P_{ii} = 0.944$. However, the term $r_n^{-1} \sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n^*})^2 = 0.513$ suggests that Assumption 3 is reasonably satisfied. This illustrates the practical use of the weaker assumptions we provide in this chapter. The 95% confidence set for each test is given by the values of β_0 that are below the grey horizontal line. The results broadly confirm the analysis in Card (2009), in that we find that less-educated immigrants to a city are closer to perfect substitutes

(the inverse elasticity of substitution is smaller in absolute magnitude) than more-educated immigrants (the inverse elasticity of substitutions is larger in absolute magnitude). We find that the confidence sets for high school workers is smallest for the jackknifed AR statistic of CMS, whereas the BCCH Sup Score test yields the smallest confidence interval for the application to college workers. This latter result suggests a very sparse first stage for college workers, i.e., a few nationalities being highly predictive of inflows of immigrant labour. In line with the simulation evidence on power for $k_n/n = 0.3$ in Section 2.4, the difference in confidence sets across the four different tests considered appears to be reasonably small.

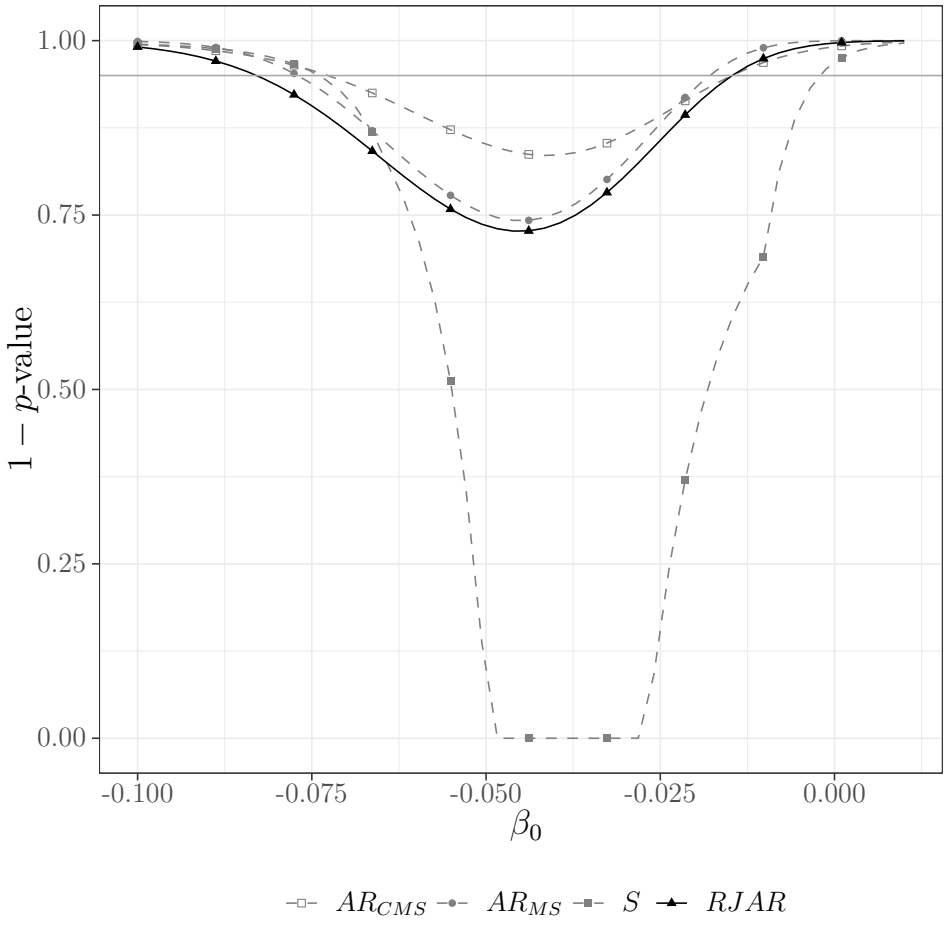


Figure 2.9: 95 % confidence sets for β_s for the application in Equation (2.12) for high school workers with 38 IVs. $\max_i P_{ii} = 0.944$. $\gamma_n^* = 0$, $r_n^{-1} \sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n^*})^2 = 0.513$.

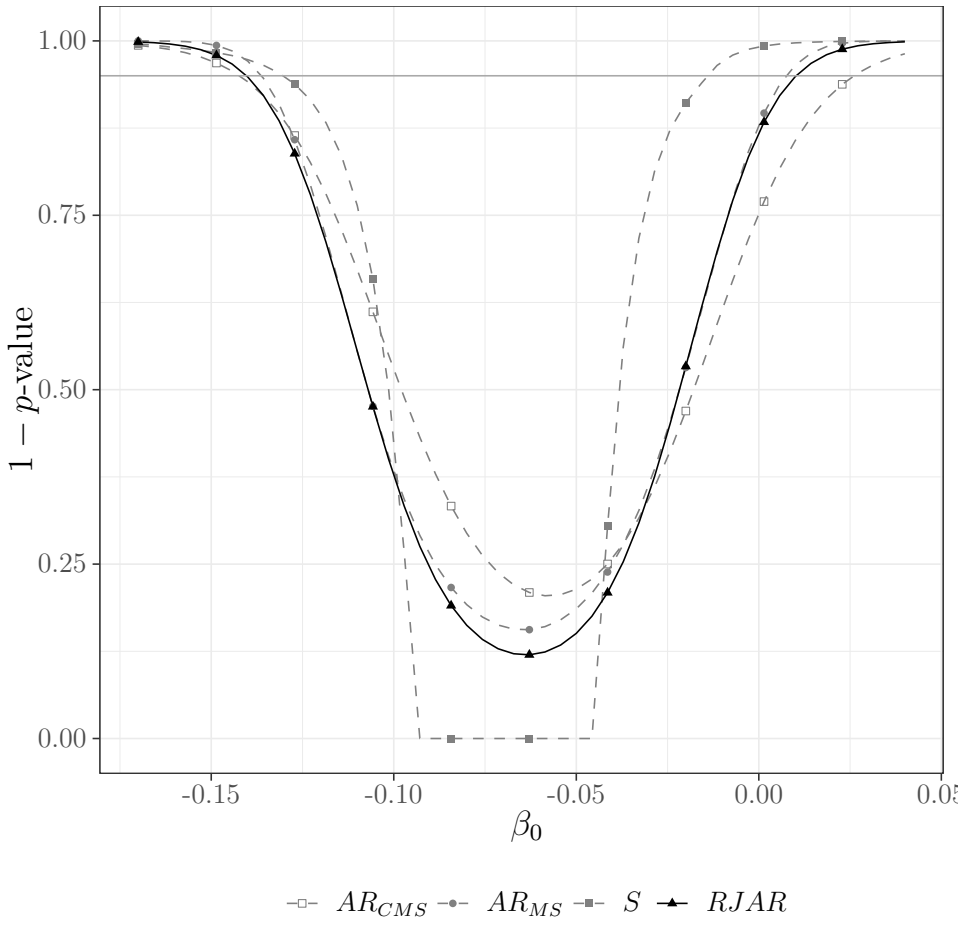


Figure 2.10: 95 % confidence sets for β_s for the application in Equation (2.12) for college workers with 38 IVs. $\max_i P_{ii} = 0.944$. $\gamma_n^* = 0$, $r_n^{-1} \sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n^*})^2 = 0.513$.

Figure 2.11 and Figure 2.12 show the confidence sets when $k_n = 342$ for high school workers and college workers, respectively. Since $r_n = 124 - 9 = 115 < 342$, the jackknifed AR statistics of CMS and MS are not applicable. We find that $\gamma_n^* = 5.299$ and $r_n^{-1} \sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n^*})^2 = 0.106$. In line with the simulation results on power reported in Section 2.4, the RJAR test yields smaller confidence intervals than the BCCH sup score test. The qualitative conclusions with respect to the case of 38 IVs remain unchanged. The results from the RJAR test in the high-dimensional case confirm the results of the low-dimensional case.

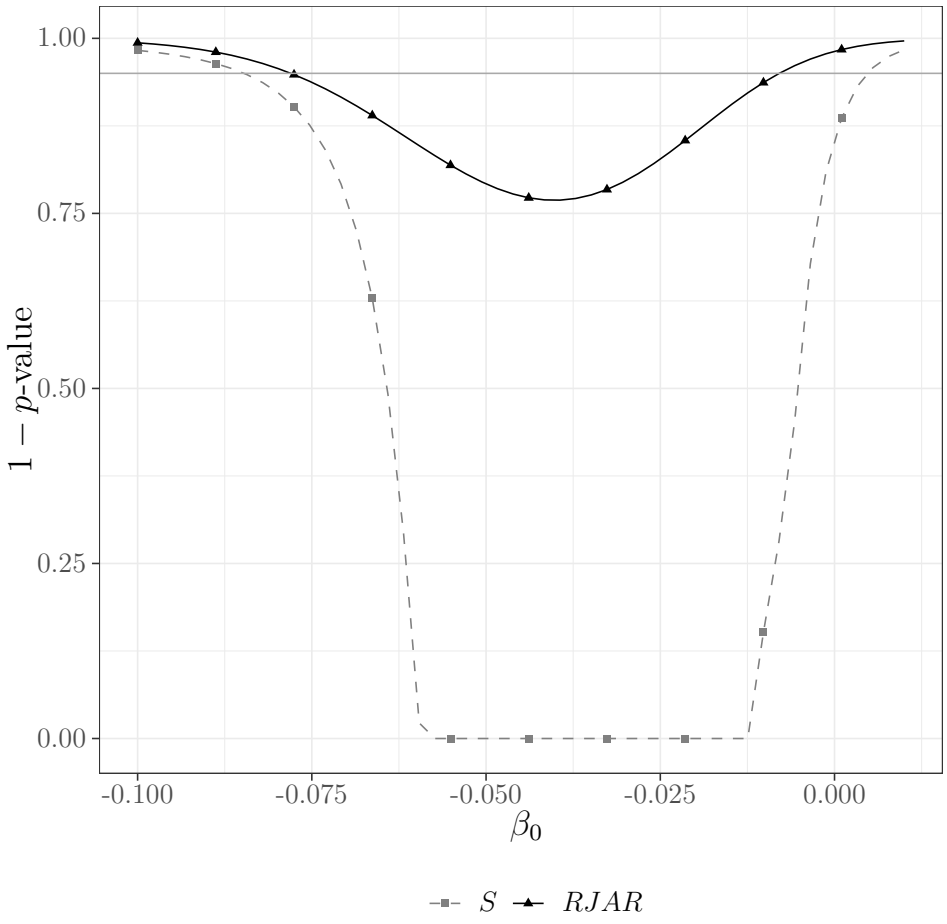


Figure 2.11: 95 % confidence sets for β_s for the application in Equation (2.12) for high school workers with 342 IVs. $\gamma_n^* = 5.299$, $r_n^{-1} \sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n^*})^2 = 0.106$.

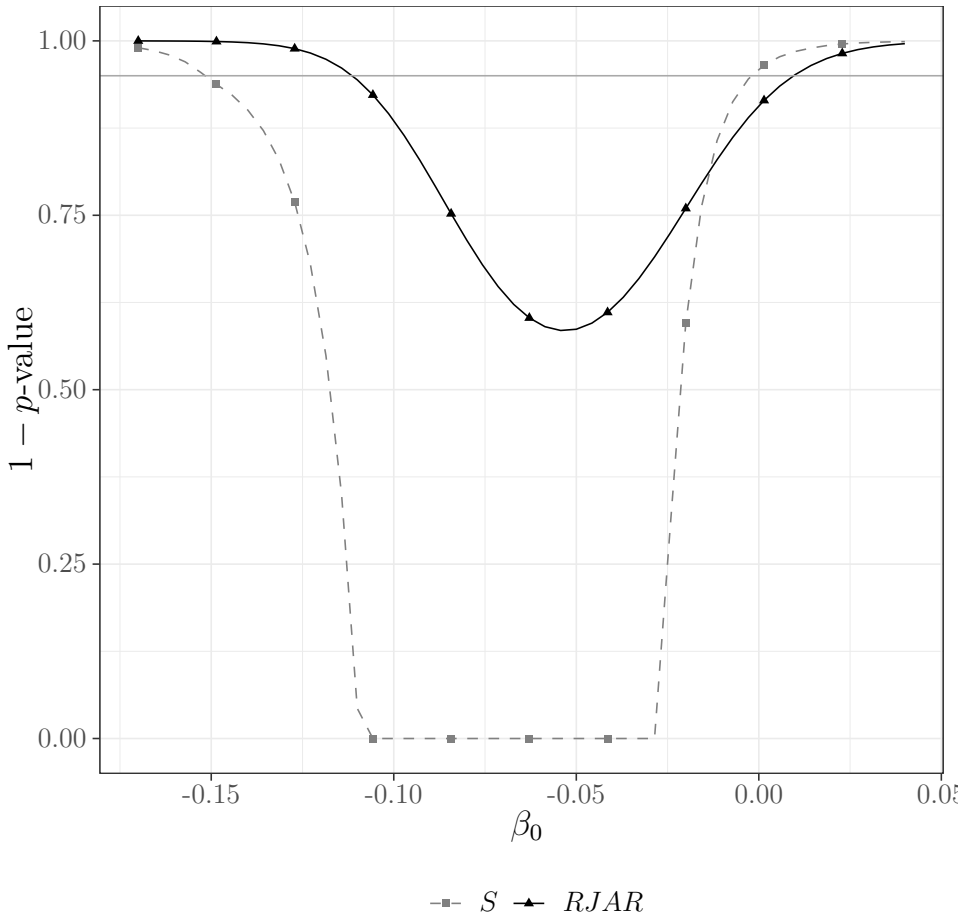


Figure 2.12: 95 % confidence sets for β_s for the application in Equation (2.12) for college workers with 342 IVs. $\gamma_n^* = 5.299$, $r_n^{-1} \sum_{i=1}^n \sum_{j \neq i} (P_{ij}^{\gamma_n^*})^2 = 0.106$.

2.6 Conclusion

We contributed to the literature on (very) many IVs in the cross-sectional linear IV model by proposing a new, ridge-regularised jackknifed AR test. Our test compares favourably with existing methods in the literature both theoretically, by allowing for high-dimensional IVs and weakening a common assumption on the IVs' projection matrix, and practically, by having correct size and displaying favourable power properties even when the number of IVs approaches or exceeds the number of observations.

3

Inference on the New Keynesian Phillips Curve with Very Many Instrumental Variables

3.1 Introduction

Instrumental variable (IV) methods are often used to conduct limited-information inference on (structural) single-equation macroeconomic relations that describe the dependence of a scalar variable on a set of covariates. Examples of such macroeconomic relations include New Keynesian Phillips Curves (NKPCs), Euler equations, and Taylor rules. IV-based limited-information inference on such macroeconomic relations has arguably proven popular because there is no requirement that parts of the model other than the specified relation itself be necessarily true to conduct valid inference. In virtually all applications, the relation is assumed to contain an additive error term that is shown (e.g., by the assumption of Rational Expectations (RE)) or primitively assumed to be uncorrelated with predetermined variables excluded from the specified relation. This makes any predetermined variable a valid IV.

As documented extensively in the existing literature, using IVs to conduct limited-information inference on such macroeconomic relations often runs into issues related to weak identification. This occurs when the variation in the IVs is only

able to explain a small portion of the variation of the endogenous variables.[1] This problem is especially pronounced when the analysis is restricted to using only a few variables to forecast the endogenous variables, a restriction that arises when using IV methods that treat the number of IVs as fixed relative to the sample size. Since any predetermined variable is a valid (if not very informative) IV, this naturally raises the question of which IVs to choose out of the very many available ones.

The limited literature that seeks to formally address the high dimensionality of the available IVs in such macroeconomic settings is primarily motivated by the potential inefficiency of using IVs selected in an ad-hoc way (Bayar, 2018; Berriel et al., 2016, 2019; Kapetanios et al., 2015; Mirza and Storjohann, 2014). Through simulations and/or empirical applications, these studies find smaller confidence sets than the ones implied by IVs traditionally used in the past. Although this evidence is certainly suggestive, it should be noted that formal efficiency claims rely on conditions that are not easily verifiable in practice.[2]

Rather than being motivated by such efficiency concerns, this chapter revisits the question of high-dimensional limited-information inference because some types of formal or intuitive regularisation can lead to invalid inference, even if weak-IV robust methods are used after regularisation. This is due to what Chernozhukov et al. (2015) call the 'endogeneity bias', which arises when variables are selected on the basis of their in-sample correlation with a model's error terms.

The first contribution of this chapter consists in illustrating how IV selection can lead to invalid inference in the context of limited-information inference on a standard NKPC. I do this by extending the simulations in Mavroeidis et al. (2014) to the more realistic case where the econometrician does not have oracle knowledge

[1] For the case of NKPCs, see Dufour et al. (2006), Kapetanios et al. (2015), Kleibergen and Mavroeidis (2009), Ma (2002), Mavroeidis et al. (2014), and Mirza and Storjohann (2014), for the case of Euler equations see Ascari et al. (2021), Kleibergen (2005), Stock and Wright (2000), and Yogo (2004), for the case of Taylor rules see Mavroeidis (2010) and Mirza and Storjohann (2014).

[2] For instance, factor-based approaches to reduce the dimensionality of the IVs likely work well if there is a factor structure, and if whatever explains most of the variation in the IVs, also explains (a good portion of) the variation of the endogenous variables. While the former may be made plausible through certain tests, the latter remains an assumption the researcher has to make. Similarly, a LASSO-based selection of IVs works well only under the assumption that the relation between the endogenous variables and the candidate IVs is sufficiently sparse.

on which IVs are the relevant ones, but rather has to choose amongst the very many available IVs. I consider different IV selection techniques, and show that several of them result in substantially invalid inference. The example of the NKPC is chosen for the sake of concreteness and due to its popularity in the literature. The same concerns extend to any of the many cases in Macroeconomics where a given structural equation can be estimated with very many valid IVs.

As a second contribution, I propose a Sup Score test to conduct IV-based limited-information inference on single-equation Macroeconomic relations. Contrarily to other approaches in the literature, this statistic requires no assumption on the factor structure of the IVs, nor does it make any sparsity-type assumption that requires only a few of the very many IVs to be relevant, while allowing for a number of IVs that increases exponentially with the sample size. This test directly contributes to the (very) many weak IVs literature predominantly restricted to the cross-sectional case (see Anatolyev and Gospodinov (2011), Belloni et al. (2012), Crudu et al. (2021), Mikusheva (2021), and Mikusheva and Sun (2022)), and can find application well beyond the example of NKPCs considered in this chapter.

The third contribution consists in applying the selection procedures considered in the simulation section and the Sup Score test to conduct IV-based limited-information inference on a standard hybrid NKPC with 361 IVs on a sample of 179 observations. I find that both the IVs selected and the confidence set implied by selection procedures that yield size distortions in the simulations are similar to the IVs selected and the confidence set implied by the IVs traditionally used in the past. This suggests that the results previously reported in the literature may suffer from endogeneity bias, and that they hence may undercover the true parameter values. By contrast, the confidence sets implied by the Sup Score test are considerably wider. In particular, and contrarily to the previous literature, I find no evidence that short-term inflation expectations matter in explaining inflation today.

Notation. For any real number a, $\lfloor a \rfloor$ indicates the largest integer b such that $b \leq a$. For any two real numbers a and b, $a \asymp b$ denotes $a = cb$ for some constant

$c > 0$, and $a \lesssim b$ denotes $a \leq cb$ for some constant $c > 0$, $a \vee b \equiv \max\{a, b\}$, $a \wedge b \equiv \min\{a, b\}$. The remaining notation follows standard conventions.

Organisation of the chapter. Section 3.2 introduces the model considered in this chapter. Section 3.3 outlines the methods used in this chapter to conduct inference in the context of very many IVs. Section 3.4 provides simulation-based evidence on the size and power of these methods. Section 3.5 revisits inference on the US NKPC using very many IVs. Section 3.6 concludes.

3.2 Model

The structural equation I consider is the hybrid NKPC of Galì and Gertler (1999),

$$(1 - \gamma)\Delta\pi_t = \lambda s_t + \gamma \mathbb{E}_t[\Delta\pi_{t+1}] + u_t, \tag{3.1}$$

where Δ is the difference operator, π_t is the inflation rate, s_t is the forcing variable, and λ, and γ are parameters of the model. u_t is an unobserved disturbance term, which can be interpreted as a measurement error, or as a shock to inflation, such as a cost-push shock.

The identifying moment conditions can be derived within the framework of Generalised Instrumental Variable (GIV) estimation. In this approach, realised one-period-ahead inflation is substituted in for expected inflation. This means that Equation (3.1) can be re-written as

$$(1 - \gamma)\Delta\pi_t = \lambda s_t + \gamma\Delta\pi_{t+1} + \underbrace{u_t - \gamma\left[\pi_{t+1} - \mathbb{E}_t[\pi_{t+1}]\right]}_{\epsilon_t}.$$

If it is further assumed that $\mathbb{E}_{t-1}[u_t] = 0$, the assumption of RE gives rise to the moment conditions

$$\mathbb{E}[Z_t\epsilon_t] = 0,$$

for any $k \times 1$ vector of predetermined variables Z_t. Due to the very large number of predetermined time series available, the dimension of Z_t is comparable to or larger than the number of observations, T. Combining the above two displays

gives the estimation problem that constitutes the particular example considered throughout this chapter:

$$\begin{aligned} \Delta\pi_t &= \lambda s_t + \gamma(\pi_{t+1} - \pi_{t-1}) + \epsilon_t, \\ \mathbb{E}[Z_t\epsilon_t] &= 0, \end{aligned} \tag{3.2}$$

where λ and γ are the coefficients on which the econometrician conducts (joint) inference.

It should be noted that the example of NKPCs (including the particular specification chosen), and the assumption of RE are not central to two of the contributions of this chapter. The same concerns relating to the endogeneity bias persist, and the same Sup Score test proposed below remains valid for the broad class of models defined by single-equation relations of the type

$$y = g(Y, \theta) + \varepsilon, \tag{3.3}$$

and moment equations given by

$$\mathbb{E}[Z_t\varepsilon_t] = 0, \tag{3.4}$$

where y is a $T \times 1$ vector, g is a known real-valued function, Y is a $T \times p$ matrix of (potentially) endogenous covariates, Z is a $T \times k$ matrix of variables such that $k \geq p$, θ is a $p \times 1$ vector of coefficients, p is fixed, and ε is a $T \times 1$ vector of error terms. This setup encompasses many popular applications in Macroeconomics, where k is of the same magnitude or even larger than T, such as limited-information inference on NKPCs, Euler equations, and Taylor rules. In particular, the NKPC considered in Equation (3.2) can be mapped into the more general model in Equation (3.3) as follows: $y = \pi$, $Y = [s\ \pi_{+1}]$, $\theta = [\lambda\ \gamma]'$, $g(Y, \theta) = Y\theta$, $\varepsilon = \epsilon$, and s, π are the $T \times 1$ stacked vectors of s_t, π_t, respectively.

3.3 Methodology

For all methods considered in this chapter, confidence sets are constructed by inverting statistics that test the hypothesis

$$H_0 : \theta = \theta_0 \text{ vs } H_1 : \theta \neq \theta_0. \tag{3.5}$$

The $(1-\alpha)$ confidence set can be constructed by collecting the values of θ_0 for which the null hypothesis in Equation (3.5) is not rejected at the α level of significance. For convenience, define $\varepsilon_0 \equiv y - g(Y, \theta_0)$.

3.3.1 Post-Selection Low-Dimensional Inference

Most of the existing literature that conducts inference on relations of the form presented in Equation (3.3) using moment conditions of the type shown in Equation (3.4) has employed methods that require the IVs to be low-dimensional. In the presence of very many IVs, these approaches can be seen as a two-step procedure. First, the IVs are selected. Second, a low-dimensional (weak-identification robust) method is applied with the selected IVs. The first step is usually not made explicit, and is often not given any attention, which makes it impossible to model this step accurately. In Section 3.3.1.2, I consider three different selection procedures that reasonably cover (in terms of their deleterious effect on subsequent inference) the range of selection procedures used in the previous literature. These are: random selection, 'crude thresholding', and LASSO. In Section 3.3.1.1, I outline the Anderson and Rubin (1949) (AR) test, which forms the post-selection inferential method common to all three selection procedures considered in this chapter.

Before proceeding, it is helpful to gain some intuition as to why IV selection may lead to invalid IVs. Consider the following projection ('first stage')

$$Y = Z\zeta + v,$$

where ζ is a $k \times 1$ vector of coefficients and v is a $T \times 1$ vector of error terms. Consider the case of no identification at all, $\zeta = 0$, and a selection procedure that selects the IVs that are most highly correlated with the endogenous variables, Y. This amounts to selecting those IVs that are most highly correlated in-sample with the first-stage error term. By the endogeneity of the system, this means that those IVs most correlated with the error term, ε, will be selected, so that *conditional on selection*, the IVs are no longer valid. This phenomenon carries over more broadly to cases of weak (but non-zero) identification as discussed in Hansen and Kozbur (2014).

3.3.1.1 The Anderson and Rubin (1949) Statistic

In this chapter, the AR statistic will be used for low-dimensional post-selection inference. The AR test is robust to arbitrarily weak identification, and has been used in Bayar (2018) and Dufour et al. (2006) for inference on NKPCs, and has attractive robustness properties vis-à-vis more sophisticated weak-identification robust methods (see Dufour (2003) for a discussion). This provides additional assurance that any bias found in the simulations is in fact due to the bias introduced by selection. It also recommends itself in the current context because it is computationally inexpensive. Letting $k_s \geq p$ denote the number of IVs selected, the AR statistic is given by

$$AR = \frac{\frac{1}{k_s}\varepsilon_0' P_{Z^s}\varepsilon_0}{\frac{1}{T-k_s}\varepsilon_0' M_{Z^s}\varepsilon_0}.$$

where $M_{Z^s} = I - P_{Z^s}$, $P_{Z^s} = Z^s(Z^{s\prime}Z^s)^{-1}Z^s$, and Z^s is the $T \times k_s$ matrix containing the selected IVs. Under the null hypothesis, the AR statistic converges in distribution to $1/k_s$ times a $\chi^2_{k_s}$ distributed random variable.

3.3.1.2 Ad-Hoc Selection of Instrumental Variables

Conducting inference with the AR statistic requires selecting a sufficiently small subset of k_s IVs from the available k IVs.[3] In most of the empirical studies on IV-based limited-information inference on macroeconomic relations, no explicit reason is given for choosing the k_s IVs that are subsequently used for analysis. Often, the choice of IVs is simply motivated with reference to previous studies that used those IVs. It is hence impossible to model the choice of IVs of the previous literature accurately in a simulation exercise. As an (imperfect) approximation, I consider the following three selection procedures.

[3] An often-used rule of thumb is to select k_s to be of the order of magnitude of $T^{1/3}$. This rate result is motivated by the results in Andrews and Stock (2007), and Newey and Windmeijer (2009), who show that this rate condition is sufficient for the case of independent data. Recently, fully weak-identification robust AR-type statistics have been developed that allow for the number of IVs to be of the order of magnitude of T (Anatolyev and Gospodinov, 2011; Crudu et al., 2021; Mikusheva and Sun, 2022). However, all of these approaches treat the IVs as fixed, and are hence not applicable in the context of time series.

The first selection procedure involves randomly selecting k_s IVs out of the k available IVs. Since the selection of IVs is not informed by the data itself, this selection procedure is guaranteed to not violate the identifying moment conditions.

The second selection procedure I consider will be referred to as crude thresholding. This involves first computing p separate $k \times 1$ vectors containing the sample correlations between the endogenous variables and all the candidate IVs, sorting the IVs in descending order of correlation, and constructing the vector of IVs for post-selection inference by taking the union of the first $\lfloor k_s/p \rfloor$ entries in each of the vectors. By selecting the variables based on in-sample correlations, this selection procedure is likely to break the exclusion restriction of the IVs selected. Although (to my knowledge) this crude thresholding has not been applied to IV-based limited-information inference, more sophisticated versions of thresholding have been considered in the past (e.g., Mirza and Storjohann (2014, Appendix B) and Bayar (2018)).[4]

The first two selection procedures (random selection and crude thresholding) arguably cover the extremes in terms of the effects IV selection can have on the validity of the IVs. Random selection provides the selection ideal, since it leaves the identifying moment conditions completely unaffected. However, particularly with reference to the traditional IVs often considered in the literature, it seems unlikely that random selection (over the very many available predetermined macroeconomic time series) led to choosing proximate lags of the endogenous variables as IVs. Indeed, given the persistence of most macroeconomic time series (and hence of the endogenous variables in any given application), it seems plausible that at least part of the motivation for considering proximate lags of the endogenous variables as IVs stems from their ability to usefully explain some of their in-sample variation. Suggestive evidence for this type of selection is also given by the fact that the IVs selected by crude thresholding in the empirical application in Section 3.5

[4]The hard thresholding in Mirza and Storjohann (2014, Appendix B) and Bayar (2018) is not applicable in high-dimensional contexts, since OLS is infeasible when there are more variables than observations.

show substantial overlap with these traditional IVs.[5] Therefore, it seems likely that random selection and crude thresholding provide a suggestive lower and upper bound on the selection-induced bias that could underlie existing empirical applications.

The third selection procedure I consider is a LASSO-based selection of IVs. This is motivated by the recent increase in popularity of penalisation-based approaches to the (very) many IV problem (see Belloni et al. (2012), Hansen and Kozbur (2014), and Ng and Bai (2009)). Furthermore, LASSO-based approaches to IV selection have also been applied to the case of NKPCs in Berriel et al. (2016, 2019). Here, IVs are selected by solving a LASSO optimisation problem of the following form for each endogenous variable

$$\hat{\zeta}_r = \underset{\zeta_r \in \mathbb{R}^k}{\arg\min} \sum_{t=1}^{T} (Y_{rt} - \zeta_r' Z_t)^2 + \Lambda_r |\zeta_r|,$$

where Y_{rt} is the element in position t of the $T \times 1$ vector Y_r given by the r^{th} column of Y, ζ_r for $r = 1, \ldots, p$ is a $k \times 1$ vector, and $\Lambda_r > 0$ for $r = 1, \ldots, p$ are scalar penalty parameters that are set such that $\hat{\zeta}_r$ has $\lfloor k_s/p \rfloor$ elements. The IVs selected are given by the IVs that have at least one corresponding non-zero entry in at least one of $\hat{\zeta}_r$ for $r = 1, \ldots, p$.

3.3.2 A High-Dimensional Sup Score Test for Dependent Data

Although the interplay between weak identification and high-dimensional IVs has recently received some attention (see Hansen and Kozbur (2014)), none of the currently available approaches are both robust to arbitrarily weak identifcation and applicable in a time-series context. Indeed, to the best of my knowledge, the only approach that is formally robust to arbitrarily weak identification in the presence of very many IVs is the Sup Score test of Belloni et al. (2012). The Sup Score test of Belloni et al. (2012), however, treats the IVs as fixed, and is hence not applicable in time-series contexts (see also the discussion in Mikusheva (2021)). In this section, I propose an IV inference procedure (the Sup Score test) that remains valid under high-dimensional dependent data.

[5] See also the ranking of IVs based on t-values in Mirza and Storjohann (2014, Table B1).

The Sup Score statistic I propose is given by

$$\mathcal{R} = \max_{1 \leq j \leq k} \left| \frac{1}{\sqrt{T}} Z_j' \varepsilon_0 \right|. \tag{3.6}$$

This can be seen as a non-studentised version of the Belloni et al. (2012) Sup Score statistic, which in turn can be interpreted as an extension to high dimensions of the AR statistic. It also bears some resemblance to the non-studentised AR statistic proposed by Horowitz (2018).

The critical values for the test statistic in Equation (3.6) are computed using a block bootstrap as in Chernozhukov et al. (2018, Appendix B). Decompose the sequence $\{1, \ldots, T\}$ into large blocks of length q, and small blocks of length r, and define the number of blocks as $m = \lfloor T/(q+r) \rfloor$. Furthermore, let $I = \{1, \ldots, q\}, \ldots, I_m = \{(m-1)(q+r)+1, \ldots, (m-1)(q+r)+q\}$. Define the $2k \times 1$ vector $\mathcal{Z}_t \equiv [Z_{t1}\varepsilon_{0t}, \ldots, Z_{tk}\varepsilon_{0t}, -Z_{t1}\varepsilon_{0t}, \ldots, -Z_{tk}\varepsilon_{0t}]'$, $\tilde{\mu}_j = \frac{1}{T}\sum_{t=1}^{T} \mathcal{Z}_{tj}$, for $j = 1, \ldots, 2k$ and construct

$$L_b = \max_{1 \leq j \leq 2k} \frac{1}{\sqrt{mq}} \sum_{l=1}^{m} e_l \sum_{t \in I_l} (\mathcal{Z}_{tj} - \tilde{\mu}_j),$$

where $\{e_l\}$ is a sequence of i.i.d. $\mathcal{N}[0,1]$ random variables. This is repeated for B replications, yielding $\{L_b\}_{b=1}^{B}$. The critical value for a test of size α of Equation (3.5) is given by

$$c(\alpha) = \text{conditional } (1-\alpha)\text{-quantile of } \{L_b\}_{b=1}^{B} \text{ given } \{Z_t\varepsilon_{0t}\}_{t=1}^{T}.$$

The decision rule for testing the null hypothesis in Equation (3.5) at the α level of significance is given by

$$\text{Reject } H_0 \iff \mathcal{R} > c(\alpha).$$

I now turn to sufficient conditions for the test to have correct size.

Assumption 4.

i. $Z_1\varepsilon_1, \ldots, Z_T\varepsilon_T$ are stationary and identically distributed, with β-mixing coefficient b_r $mb_r \lesssim T^{-a}$, for some constant $0 < a < 1/4$.[6]

ii. $\mathbb{E}[Z_t\varepsilon_t] = 0$, $0 < \mathbb{E}[Z_{tj}^2\varepsilon_t^2] < \infty$, $|Z_{tj}\varepsilon_t| \leq D_T$ almost surely for all $t = 1, \ldots, T$, $j = 1, \ldots, k$ and some constant $D_T \geq 1$.

iii. $q \asymp T^{\tilde{a}}$, $r \asymp T^{\breve{a}}$, $q + r \leq \frac{T}{2}$, where $\breve{a} + a < \tilde{a} < \frac{1}{2} - a$ for constants $\tilde{a}, \breve{a} > 0$.

iv. $k \lesssim \exp(T^{\bar{a}})$, $\bar{a} = \frac{2}{5}(\frac{1}{2} - a - \tilde{a}) \wedge \frac{1}{2}(\tilde{a} - \breve{a} - a)$.

v. $\max_I \frac{1}{q} \sum_{t \in I} \sum_{l \in I} \mathbb{E}[Z_{tj}\varepsilon_i Z_{lj}\varepsilon_l] < \infty$ for all $j = 1, \ldots, k$, where $\max_I$ are taken over all $I \subset \{1, \ldots, T\}$ of the form $I = \{t + 1, \ldots t + q\}$.

Assumption 4.i. states that the data is stationary, and identically but not necessarily independently distributed, provided a β-mixing condition (that need not be exponentially decaying) holds. I refer to Adamek et al. (2020) and Wong et al. (2020) for a recent collection of (high-dimensional) DGPs that fall into this category. For the case of NKPCs under RE considered in this chapter, this assumption is satisfied for any $r \geq 2$, since RE imply that the IVs are pre-determined at every point in time and that the structural error term is an MA(1) process. Assumption 4.ii. includes the identifying moment condition, and makes weak technical assumptions. Assumption 4.iii. specifies the size of the large and small blocks. Since $\bar{a} > 0$, the number of IVs is allowed to increase exponentially with the sample size. However, contrarily to other recent approaches, the number of IVs is also allowed to be fixed. Assumption 4.i. and Assumption 4.iv. show that there is a tradeoff between the degree of dependence in the data and the dimensionality of the IVs allowed. The

[6]For $Z_1\varepsilon_1, \ldots, Z_T\varepsilon_T$ defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and any sub σ-fields $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{A}$, the β-mixing coefficient is given by

$$b_r \equiv \max_{1 \leq l \leq T-r} \beta(\sigma(Z_1\varepsilon_1, \ldots, Z_l\varepsilon_l), \sigma(Z_{l+r}\varepsilon_{l+r}, \ldots, Z_T\varepsilon_T)),$$

where $\sigma(Z_t\varepsilon_t, t \in I)$ is the σ-field generated by $Z_t\varepsilon_t$, $t \in I$, and

$$\beta(\mathcal{A}_1, \mathcal{A}_2) \equiv \frac{1}{2}\sup \left\{ \sum_t \sum_l \mathbb{P}(A_t \cap B_l) - \mathbb{P}(A_t)\mathbb{P}(B_l) \right\},$$

where $\{A_t\}$ and $\{B_l\}$ are any finite partition of Ω in $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively.

larger the small blocks (i.e., the larger the value of $\breve{a}$ and hence r), the further apart the σ-fields that are required to have a decaying β-mixing coefficient are. However, larger values of $\breve{a}$ require larger values of $\tilde{a}$. This decreases $\bar{a}$, and hence the rate at which the number of IVs is allowed to grow. Assumption 4.v. imposes restrictions on the correlation of the product of the IVs with the error term across different points in time. It is satisfied whenever the error terms are independent of the IVs, and have a finite moving-average representation.

No assumption on the first stage (i.e., the relationship between Y and Z) has to be made. This means that the proposed Sup Score test is uniformly valid over all (finite) values of the coefficient on the IVs in the first stage (including arbitrarily weak identification). This also means that no restriction on the factor or sparsity structure of the first stage has to be imposed.

I now state the main theoretical result of this chapter, which ensures that the approach proposed controls the size of the test.[7]

Theorem 2. Under Assumption 4. and the null hypothesis in Equation (3.5),

$$\lim_{T\to\infty} \mathbb{P}(\text{Reject } H_0) \leq \alpha.$$

Proof. See Appendix B.1.1. □

3.4 Simulations

The simulations presented in this section serve a twofold purpose. First, I use the simulations to study how IV selection can lead to problematic post-selection inference. Second, I use the simulations to illustrate the asymptotic validity of the Sup Score test established in the section above, as well as its finite-sample power

[7] It should be noted, however, that–similarly to other sup-based test statistics, such as in Belloni et al. (2012) and Chernozhukov et al. (2018)–the above approach is not efficient. This is to be expected, given the weak assumptions made on the IVs. The (finite-sample) power properties of the above approach will be investigated in the simulation section below. The results show that it has non-trivial power.

properties. Taken together, the simulations hence motivate and further justify applying the Sup Score test proposed in Section 3.2 in practice.[8]

The simulations in this chapter are based on the approach in Mavroeidis et al. (2014). The central difference is that rather than modelling the econometrician as having perfect knowledge of the relevant IVs, and incorrectly employing methods that are not robust to weak identification, I model the econometrician as using exclusively weak-identification robust methods, but not knowing which IVs correspond to the truly relevant ones. Given the extensive literature that pointed out that NKPCs can suffer from weak identification, this setup seems closer to the estimation problem that an econometrician is likely to face.

I embed the NKPC in Equation (3.1) into a dynamic system by specifying that the reduced-form dynamics of the forcing variable and inflation follow a VAR model given by

$$\begin{bmatrix} \pi_t \\ s_t \\ f_t \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} \pi_{t-1} \\ s_{t-1} \\ f_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}, \tag{3.7}$$

where

$$\begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix} \overset{i.i.d.}{\sim} \mathcal{N}\left[0, \begin{bmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & \omega_{23} \\ \omega_{31} & \omega_{32} & \omega_{33} \end{bmatrix}\right],$$

and f_t is a scalar factor variable. All coefficients except for a_{11}, a_{12}, and a_{13} have to be calibrated. The coefficients a_{11}, a_{12}, and a_{13} are backed out of the NKPC

[8]Due to the focus of this chapter on the bias introduced by IV selection, I do not consider the factor-based approaches of Kapetanios et al. (2015) and Mirza and Storjohann (2014). The substantial biases caused by the selection of a small number of IVs can also serve to motivate the use of such factor methods. Unreported simulations show that using as IVs k_s factors from the original set of IVs controls size for all calibrations. The power of this procedure is similar to (and sometimes greater than) the one of the Sup Score test proposed in this chapter when the DGP is calibrated to have a strong factor structure (e.g., large values of a_{23} or ξ), but only has trivial power when the factor structure is weak (e.g., small values of a_{23} or ξ). However, there are no formal guarantees for the validity of this approach. The factor-based GMM approach in Mirza and Storjohann (2014) seems to treat the number of IVs as fixed (and does not provide formal conditions for validity), while the factor AR statistic of Kapetanios et al. (2015) is only applicable in a high-dimensional context if a sufficiently strong factor structure is assumed. In contrast, the Sup Score test proposed in this chapter remains valid and reasonably powerful in high-dimensional contexts regardless of the factor or sparsity structure of the IVs.

based on the Anderson and Moore (1985) algorithm.[9]

High dimensionality of the IVs is introduced by specifying that there exists an $g \times 1$ vector of variables Q_t that follow the process given by

$$Q_t = \xi f_t + u_{4t}, u_{4t} \overset{i.i.d.}{\sim} \mathcal{N}\left[0, I_g\right], \tag{3.8}$$

where ξ is an $g \times 1$ vector of factor loadings. The econometrician does not observe the factor itself, but only observes the forcing variable and inflation, as well as the m variables in Q_t. In this setup, Z_{st} is a subset of the available IVs given by $Z_t = [\pi_{t-1}, s_{t-1}, Q'_{t-1}, 1]'$.

This setup recommends itself for two reasons. First, it constitutes a minimal departure from popular simulations in the existing literature. This ensures that any reported results are not an artefact of a particularly uncharitable setup.[10] Second, it ensures the existence of a sufficiently small set of 'oracle IVs' (given by $\pi_{t-1}, s_{t-1}, f_{t-1}$) without necessarily imposing a sparse setup on the observed first-stage projection (although it can be imposed by setting $a_{13} = a_{23} = \omega_{13} = \omega_{31} = \omega_{23} = \omega_{32} = 0$ or simply $\xi = 0$).[11] The former is desirable because it allows for a comparison of the different ad-hoc inference procedures relative to the most efficient approach (conditional on using the AR statistic). The latter is desirable because it allows for a more general (and perhaps more realistic, see Giannone et al. (2021)) approach to modelling the first stage. This setup is able to achieve both a sparse (unobserved) oracle first stage and an observed first stage that is not necessarily sparse because the elements of Q_{t-1} that have a non-zero corresponding entry in ξ will contain some relevant variation for identification, due to the dependence of the endogenous variables $[\pi_{t+1} - \pi_{t-1}, s_t]'$ on f_{t-1}. Based on the setup above, it is possible to derive

[9] The simulation setup above takes the form of a Gaussian VAR. Hence, the most efficient approach to conducting inference in this setting would be to use Maximum Likelihood Estimation without appealing to IVs. However, due to the focus of this chapter on IV-based limited-information inference on the NKPC, I do not pursue this approach further. I also note that this approach requires making the additional assumption that the econometrician knows that the error terms are Gaussian.

[10] It is, for instance, straightforward to include the variables in Q_t directly in the reduced-form VARs. However, the results from such a DGP are very sensitive to the particular calibration of the parameters chosen.

[11] Ensuring a sufficiently sparse set of 'oracle IVs' further motivates considering only a single lag of a single factor.

two different concentration parameters (μ_O^2 and μ_E^2) that reflect the strength of identification in the sparse unobserved oracle first stage and the observed first stage. The details are given in Appendix B.1.2.

The calibrations are as follows. Throughout, I set $\gamma = 0.8$, $\lambda = 0.05$, $T = 100$, and $\omega_{11} = 0.07$, $\omega_{12} = \omega_{21} = 0.03$, $\omega_{22} = 0.7$, $\omega_{13} = \omega_{31} = \omega_{23} = \omega_{32} = 0$, and $\omega_{33} = 0.4$ (see also Mavroeidis et al. (2014)). The results are not sensitive to this choice of covariance matrix, and this setup makes it possible to create a perfectly sparse observed first stage by setting $a_{23} = 0$. For simplicity, I set $a_{31} = a_{32} = 0$, so that the factor follows an autoregressive process with coefficient given by $a_{33} = 0.7$ (the results do no change appreciably if this is relaxed or a different choice for a_{33} is considered). I set $g = 200$ and $\xi_h = \tau(-1)^h \log\left((h+1)^2/gh\right)$ for $h = 1, \ldots, g$ and $\tau = 0.05$. This is meant to provide a deterministic calibration that balances positive and negative, as well as large and small coefficients. The small value chosen for τ ensures that the information on the factor contained in the observed variables is sufficiently diluted, and that there is some interesting variation in the informational content of the unobserved oracle first stage and the one actually observed.[12] The results are unaffected by different choices of ξ or τ. For all selection procedures, I force the selection of $k_s = 4$ IVs to ensure that the first stage is not overfitted, which again ensures that any distortions in inference are attributable to the selection step itself. For the Sup Score test proposed above, I set $q = 15$, $r = 2$, and the bootstrap replications to 500. I allow a_{21}, a_{22}, and a_{23} to take on different values. The coefficient a_{23} controls how informative the factor is in predicting the endogenous variables, and by extension how informative the variables in Q_{t-1} are.

Table 3.1 shows the size of the AR and Sup Score test following the different selection procedures outlined above for a test with nominal size 10%. The calibrations chosen ensure that a broad range of identification strength and sparsity structures is considered. The first panel for a_{23} corresponds to the perfectly sparse first stage where none of the variables in Q_{t-1} are informative IVs. As a consequence,

[12] I refer to Appendix B.1.2 for more details on this. The derivations also show that choosing small values of τ has a similar effect to choosing a larger term for the variance of the errors in Equation (3.8).

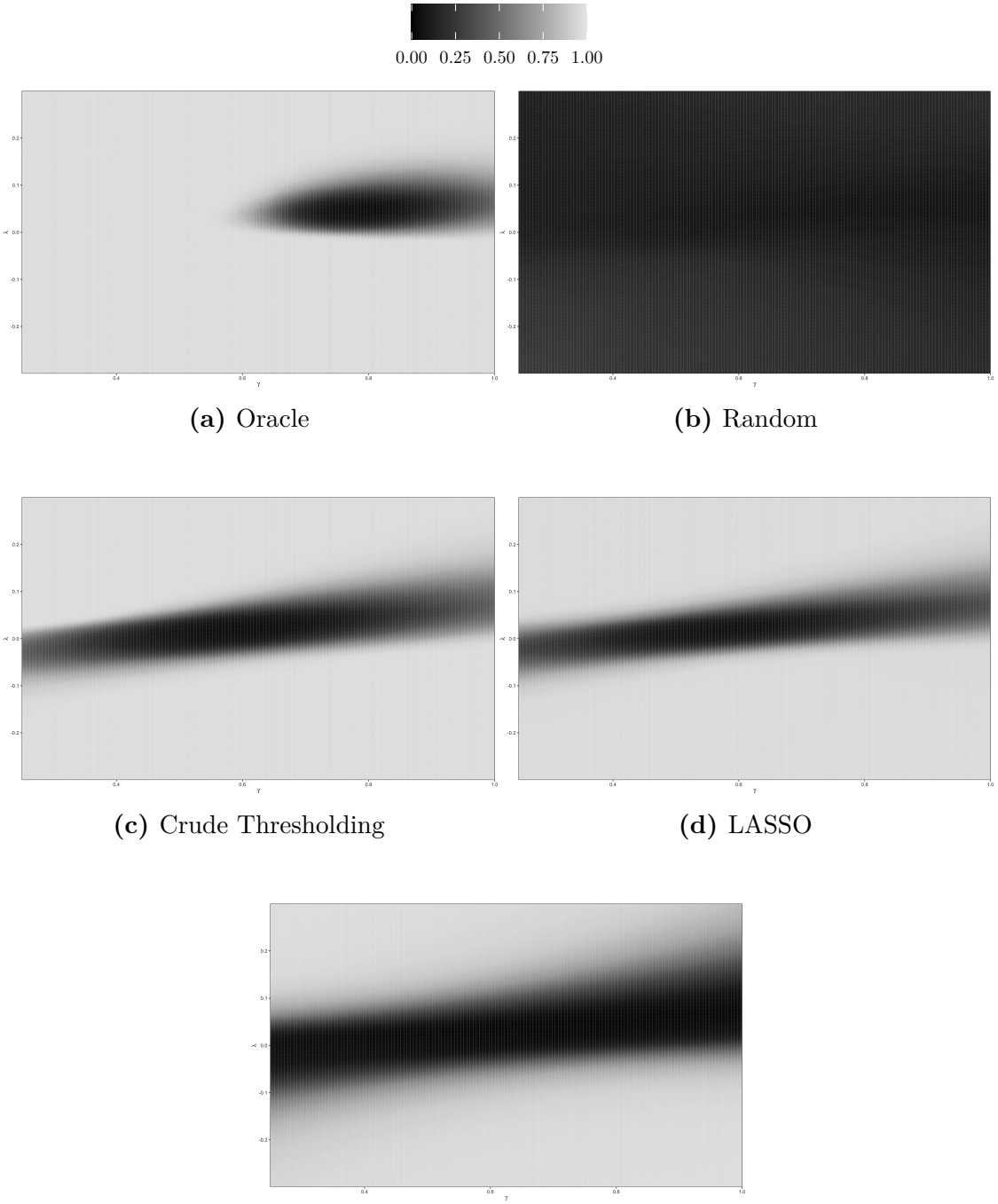
the concentration parameter of the unobserved oracle first stage is the same as the one that is observed. The second and third panel increase the dependence of the two endogenous variables on the unobserved factor. Due to the dense (i.e., not sparse) calibration of ξ, this means that all of the IVs observed by the econometrician are at least somewhat informative. Since the variables in Q_t contain noisy information on the unobserved factor, the concentration parameter in the observed first stage will now be lower than the concentration parameter of the unobserved oracle first stage. As expected, the oracle IVs yield correct, if somewhat conservative, size. Since random selection does not make use of any correlations present in the actual data, the AR statistic with randomly selected IVs also yields roughly correct size, although there is some evidence for finite-sample size distortions. The results for crude thresholding and the LASSO suggest that in all cases size is not controlled, although the distortions appear to be somewhat milder for the LASSO. The results for the Sup Score test proposed in this chapter show that the test controls for size regardless of the DGP considered. In practice, this would translate to (spuriously) small confidence sets.

Figure 3.1 shows the power of the different approaches. I present the results for the case where $a_{21} = a_{22} = a_{23} = 0.450$. The map traced out by the oracle IVs corresponds to the most powerful procedure possible (conditional on exclusively using the AR statistic) that controls for size. The results show that randomly selecting IVs yields no power. This is unsurprising, given that in this setup the first stage is sparse, so that random selection predominantly selects not very informative IVs. The power heatmaps for crude thresholding and the LASSO have a similar shape to the power heatmap of the Sup Score test. Therefore, the Sup Score test compares favourably with these approaches in terms of power, all while controlling for size.

Table 3.1: Simulation results: size.

		$a_{23}=0.000$								
		$a_{21}=0.000$			$a_{21}=0.200$			$a_{21}=0.450$		
	a_{22}	0.000	0.200	0.450	0.000	0.200	0.450	0.000	0.200	0.450
	μ_Q^2	0.000	4.082	24.175	0.000	4.070	23.938	0.000	4.040	23.393
	μ_E^2	0.000	4.082	24.175	0.000	4.070	23.938	0.000	4.040	23.393
Oracle	R.F.	0.049	0.028	0.018	0.055	0.031	0.025	0.047	0.042	0.020
Random	R.F.	0.116	0.122	0.117	0.122	0.112	0.123	0.110	0.116	0.132
C.T.	R.F.	0.197	0.200	0.221	0.205	0.204	0.194	0.192	0.190	0.178
LASSO	R.F.	0.241	0.254	0.239	0.221	0.245	0.219	0.221	0.224	0.224
Sup Score	R.F.	0.024	0.028	0.036	0.028	0.037	0.033	0.042	0.038	0.039
		$a_{23}=0.200$								
		$a_{21}=0.000$			$a_{21}=0.200$			$a_{21}=0.450$		
	a_{22}	0.000	0.200	0.450	0.000	0.200	0.450	0.000	0.200	0.450
	μ_Q^2	11.198	19.501	48.211	12.137	21.030	49.983	13.431	23.283	53.527
	μ_E^2	6.638	14.262	38.910	7.144	15.044	38.117	7.827	16.092	36.214
Oracle	R.F.	0.048	0.027	0.028	0.033	0.025	0.023	0.049	0.028	0.022
Random	R.F.	0.115	0.112	0.129	0.119	0.115	0.121	0.092	0.124	0.112
C.T.	R.F.	0.190	0.207	0.180	0.187	0.180	0.187	0.185	0.191	0.168
LASSO	R.F.	0.235	0.220	0.221	0.228	0.219	0.220	0.259	0.248	0.203
Sup Score	R.F.	0.031	0.038	0.033	0.027	0.033	0.034	0.033	0.031	0.029
		$a_{23}=0.450$								
		$a_{21}=0.000$			$a_{21}=0.200$			$a_{21}=0.450$		
	a_{22}	0.000	0.200	0.450	0.000	0.200	0.450	0.000	0.200	0.450
	μ_Q^2	44.652	65.446	93.830	48.530	69.533	93.493	53.129	72.697	98.668
	μ_E^2	23.965	37.381	51.541	25.116	37.443	47.147	26.298	36.825	42.011
Oracle	R.F.	0.051	0.043	0.030	0.057	0.045	0.037	0.048	0.030	0.037
Random	R.F.	0.116	0.114	0.124	0.105	0.119	0.130	0.110	0.116	0.104
C.T.	R.F.	0.171	0.191	0.182	0.193	0.189	0.195	0.175	0.162	0.201
LASSO	R.F.	0.211	0.221	0.190	0.246	0.226	0.221	0.218	0.221	0.211
Sup Score	R.F.	0.031	0.041	0.024	0.042	0.030	0.037	0.030	0.031	0.040

Notes:
R.F. denotes the rejection frequency.
C.T. denotes crude thresholding.
Nominal test size: 10%.
1,000 Monte Carlo replications.



(a) Oracle **(b)** Random

(c) Crude Thresholding **(d)** LASSO

(e) Sup Score

Figure 3.1: Simulation results: power. $a_{21} = a_{22} = a_{23} = 0.45$. Nominal test size: 10%. 1,000 Monte Carlo replications.

3.5 Empirical Application

Data for the empirical part of this chapter is taken from FRED. I use the non-farm labour share as transformed in Galì and Gertler (1999) as the forcing variable. I use the inflation rate implied by the GDP deflator. I consider the period 1974Q2-2018Q4, and include 90 variables aimed to reflect different parts of the US economy based on the list in McCracken and Ng (2016) with four lags each, transforming them as recommended therein.[13] This yields 179 observations with 361 IVs (inclusive of a constant). Appendix B.2 contains a detailed description of the data.

A natural question to ask is whether considering these 361 IVs is enough to dispel concerns about potential endogeneity biases. Though being more than any number of IVs previously considered in the literature, there are certainly more valid IVs (i.e., additional predetermined variables). However, a substantial endogeneity bias caused by the selection of these 361 IVs would emerge only if the variables were included in the list of McCracken and Ng (2016) based on their correlation with the endogenous variables of this application. This seems very unlikely.

For all ad-hoc selection procedures, I limit the number of IVs selected to four, to ensure that overfitting is not a concern. The results do not change appreciably when other values are chosen. The confidence sets yielded by the traditional IVs and the ad-hoc selection procedures are shown in Figure 3.2, and the corresponding IVs are listed in Table 3.2. Mirroring the results in Section 3.4, the confidence set resulting from random selection is extremely wide, and suggests that the hybrid NKPC is essentially unidentified. The confidence set from applying the LASSO is smaller than the one implied by random selection, but it also does not exclude that the coefficient on expected inflation is in fact equal to zero.

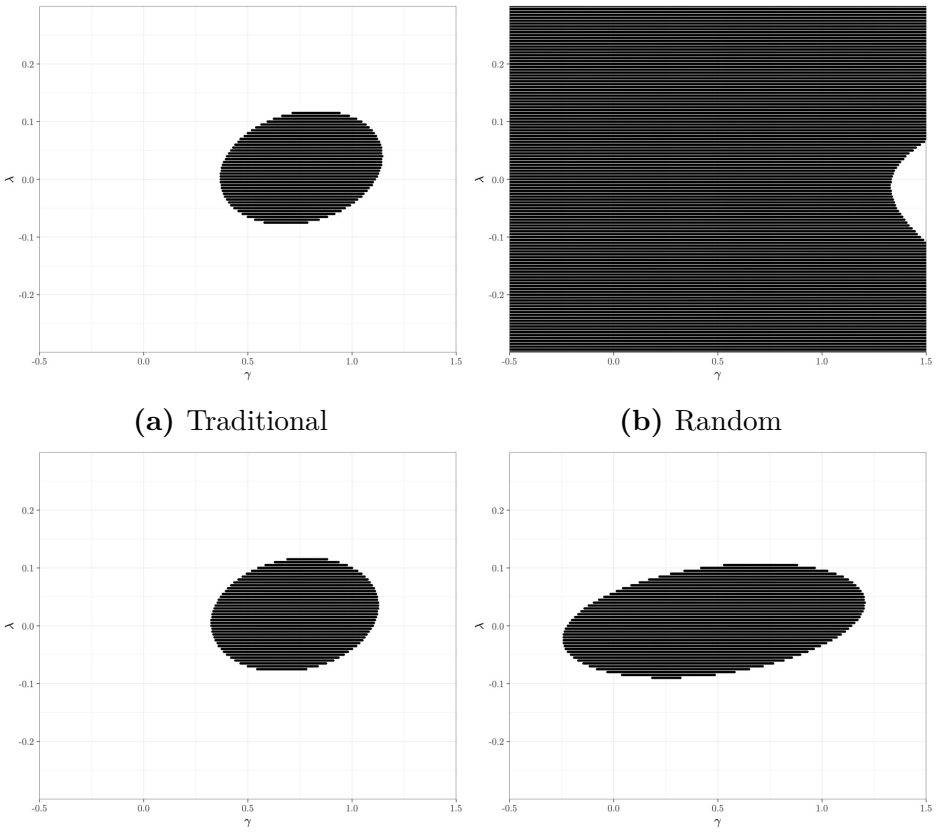
[13] I do not include all the variables listed in McCracken and Ng (2016) since they are not all available over a sufficiently long period of time.

Table 3.2: Identity of the IVs for each of the selection procedures.

Traditional	Random	Crude Thresholding	LASSO
PRS85006173.-1	WILL5000IND.-3	PRS85006173.-1	PRS85006173.-1
PRS85006173.-2	NDMANEMP.-3	PRS85006173.-2	PRS85006173.-2
GDPDEF.-1	EXUSUK.-3	GDPDEF.-1	BAAFFM.-2
GDPDEF.-2	PERMITMW.-4	IPDCONGD.-4	SRVPRD.-3

Notes:

PRS85006173 refers to Nonfarm Business Sector: Labor Share, GDPDEF refers to Gross Domestic Product: Implicit Price Deflator, WILL5000IND refers to Wilshire 5000 Total Market Index, NDMANEMP refers to All Employees, Nondurable Goods, EXUSUK refers to U.S. / U.K. Foreign Exchange Rate, PERMITMW refers to New Private Housing Units Authorized by Building Permits in the Midwest Census Region, IPDCONGD refers to Industrial Production: Durable Consumer Goods, BAAFFM denotes Moody's Seasonsed Aaa Corporate Bond Minus Federal Funds Rate, SRVPRD denotes All Employees, Service-Providing. Negative integers following each variable refer to the lags of the variable in question.



(a) Traditional **(b)** Random

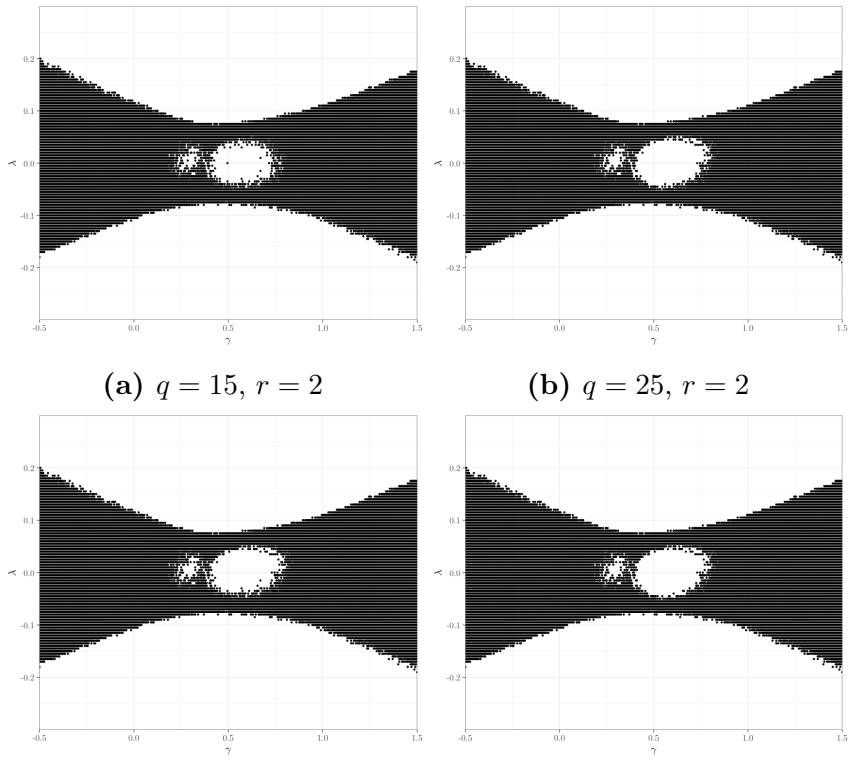
(c) Crude Thresholding **(d)** LASSO

Figure 3.2: 90% confidence sets for the NKPC in Equation (3.1) using the AR statistics with selected IVs by different ad-hoc procedures. The identity of the IVs selected by each of the procedures is shown in Table 3.2.

The confidence sets implied by traditional IVs and crude thresholding are qualitatively very similar. This similarity is explained by the fact that the IVs chosen by crude thresholding are very similar to the traditional IVs, as shown in Table 3.2. This suggests two things. First, it suggests that the ad-hoc selection

procedures used in this chapter may in fact provide a reasonable approximation to the approach taken for selecting IVs in the past literature. Second, given that in the simulation exercise in Section 3.4 crude thresholding yields substantial size distortion, this result suggests that the confidence sets reported in the previous literature are likely to suffer from at least some distortion due to endogeneity bias. In particular, it suggests that the process of trying to find 'strong' IVs may have led to an undercovering of the true parameter values.

The confidence sets of the Sup Score test with 2,500 bootstrap replications for different block lengths are shown in Figure 3.3. The results are not very sensitive to the particular choice of block lengths. For all combinations of block lengths considered, the confidence sets are smaller than the ones yielded by the random selection approach, but wider than for the traditional, crude thresholding, and LASSO approach. This is likely due to a combination of the incorrect size of the latter approaches, and the low power of the Sup Score test documented in Section 3.4. The results suggest that while certain parts of the parameter space considered can be rejected at the 10% level of significance, neither λ nor γ are found to be different from zero for all values of the parameter space considered. This contrasts the results implied by the traditional IVs, where γ (but not λ) is found to be positive and significantly different from zero.



(a) $q = 15$, $r = 2$ **(b)** $q = 25$, $r = 2$

(c) $q = 15$, $r = 5$ **(d)** $q = 25$, $r = 5$

Figure 3.3: 90% confidence sets for the NKPC in Equation (3.1) using the Sup Score test.

The Sup Score test can also help shed some light on what the most relevant IVs are, since there is a (likely) unique IV that maximises the Sup Score statistic in Equation (3.6) for every null hypothesis being tested. I record the identity of the IV maximising the Sup Score statistic for each null hypothesis tested, and report the results in Table 3.3 and Figure 3.4. Table 3.3 shows the identity of the IVs maximising the Sup Score statistic. Figure 3.4 shows in which part of the parameter space the different IVs maximise the Sup Score statistic. Two things stand out. First, the constant maximises the Sup Score statistic for most of the hypotheses considered. One interpretation of this is that the model is misspecified, and that the constant should in fact be included in the NKPC itself (although this can be difficult to motivate from a microfoundational perspective). Second, there are some IVs that maximise the Sup Score statistic that do not feature in Table 3.2, such as the lagged Housing Starts in Northeast Census Region. This relates to the predominant

motivation for wanting to consider very many IVs: the truly relevant IVs can often be 'exotic', in the sense that intuition alone would not point to their relevance.

Table 3.3: Identity of the IVs maximising the Sup Score statistic.

IV	# H_0	Description
Constant	34,548	–
HOUSTNE.-3	5,701	Three-period lag of Housing Starts in Northeast Census Region
HOUSTNE.-4	2,904	Four-period lag of Housing Starts in Northeast Census Region
GDPDEF.-1	2,409	One-period lag of Gross Domestic Product: Implicit Price Deflator
IPDCONGD.-4	1,385	Four-period lag of Industrial Production: Durable Consumer Goods
IPCONGD.-3	705	Three-period lag of Industrial Production: Consumer Goods
TB6SMFFM.-4	539	Four-period lag of 6-Month Treasury Bill Minus Federal Funds Rate
UEMP5TO14.-3	261	Three-period lag of Number Unemployed for 5-14 Weeks
IPDCONGD.-2	28	Two-period lag of Industrial Production: Durable Consumer Goods
CUMFNS.-3	22	Three-period lag of Capacity Utilization: Manufacturing
T1YFFM.-4	16	Four-period lag of 1-Year Treasury Constant Maturity Minus Federal Funds Rate
USCONS.-3	2	Three-period lag of All Employees, Construction

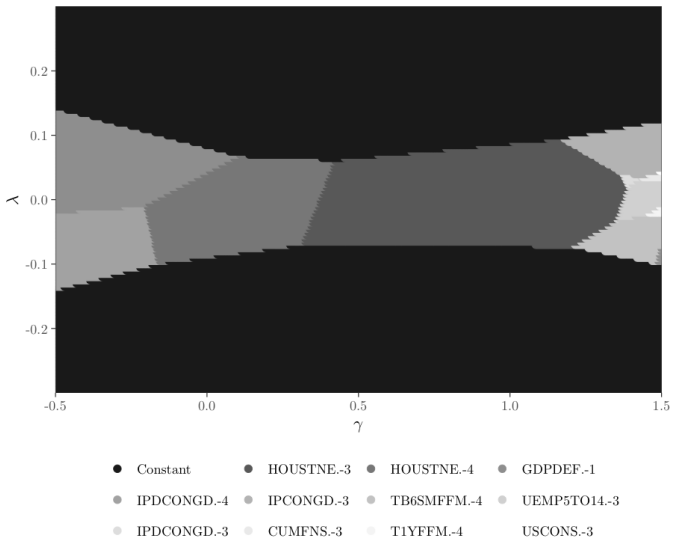


Figure 3.4: Location in the parameter space $\gamma \times \lambda$ where the different IVs in Table 3.3 maximise the Sup Score statistic.

3.6 Conclusion

IV-based limited-information estimation of single equations has become increasingly popular in Macroeconomics over the last 20 years. Using a simulation exercise based on NKPCs, I showed that selecting IVs in ad-hoc ways (random selection, crude

thresholding, and LASSO) can invalidate them, thus yielding invalid inference even if tests with desirable properties (such as robustness to weak identification) are used post-selection. To address this issue, I propose an IV inference procedure (the Sup Score test) that remains valid for high-dimensional IVs and for time series data. In the same simulation exercise that showed that ad-hoc selection procedures can lead to invalid inference, this procedure yielded correct size and reasonable power. Finally, I applied the Sup Score test to conduct inference on the US NKPC with 361 IVs on a sample size of 179 observations. The results showed that the confidence sets implied by the Sup Score test are substantially wider than the ones of all other approaches. In particular, and in contrast to the results in the previous literature, I find no evidence that short-term inflation expectations have a significant impact on inflation today. The simulation results and the empirical application point to the importance of developing further high-dimensional IV methods with good power properties that remain valid under dependence and arbitrarily weak identification.

4

Addressing the Endogeneity of Slack in Phillips Curves

4.1 Introduction

Due to its role in the transmission of monetary policy, the causal effect of economic slack on wage or price inflation is perhaps one of the most widely debated questions in academic and applied Macroeconomics. The vast literature on the topic reports a wide range of estimates for the effect of slack on wages and prices, which remains a central question in Macroeconomics.

One important reason for the wide range of estimates in the literature could be the likely endogeneity of the slack variable (i.e., the non-zero correlation of the slack with the error term), and the steps taken to address it. Indeed, if the slack variable were exogenous and the error term in the Phillips Curve (PC) well-behaved, the coefficient on the slack variable could be consistently estimated by reduced-form OLS, and the results reported in the literature (in the absence of structural breaks) should differ only modestly.

The likely endogeneity of the slack variable could result from omitted variables that are correlated with the slack variable (such as supply shocks, see Gordon (2011)). Monetary policy could also play an important role, since any (successful)

This chapter was published as: ECB working chapter No 2619 / December 2021.

policy action designed to offset demand shocks would lead to a co-movement of slack and the error terms in a reduced-form PC (see, e.g., McLeay and Tenreyro (2019)). Such factors could lead to fundamental differences in estimates from reduced-form approaches and (semi-)structural models (see, e.g., Eser et al. (2020)).

Broadly speaking, the literature on reduced-form estimation of PCs has addressed the likely endogeneity of the slack variable in two ways.[1] The first approach is based on adding to the PC those variables that could affect the transmission of slack to wage growth. In the literature, this has predominantly taken the form of adding variables that ostensibly proxy supply shocks, such as commodity inflation, monetary policy, or an economy's exchange rate (Gordon, 2011; Mazumder, 2014). The second approach involves using instrumental variable (IV) methods, by exploiting some type of exclusion restriction. This approach gained popularity in the wake of the research conducted on structural PCs, but is also used in so-called 'semi-structural' and reduced-form estimation (see, e.g., Bulligan and Viviano (2017), Dufour et al. (2006), Kleibergen and Mavroeidis (2009), and Mazumder (2014)).

Even if these methods could fully address the endogeneity of the slack variable, the high dimensionality of macroeconomic data implies that both approaches are still inherently arbitrary. This is because there is substantial 'specification uncertainty' (Mavroeidis et al., 2014), whereby there is no *a priori* reason for preferring one particular set of available supply-shock proxies or IVs over another.

Against this background, this chapter proposes an alternative approach to address the endogeneity of slack in a reduced-form wage PC for the euro area (EA).

The first contribution of this chapter is to illustrate the problem of specification uncertainty of a standard backward-looking reduced-form wage PC for the EA. We consider a wage PC, as it has received scant attention (relative to price PCs) in the past literature, while noting that the concerns related to supply-shock proxies and questionable IVs raised in the literature on price PCs extend to the case of wage PCs. Considering widely-used supply-shock proxies and IVs, we show that

[1] Due to the focus of this chapter on direct (parametric) estimation of PCs, we do not consider recent 'indirect' approaches to establishing relations between labour market slack and inflation through VARs and DSGE models (see, e.g., Del Negro et al. (2020)).

inference on the slack coefficient depends on which of these are chosen. Since it is not possible to consider all supply-shock proxies or IVs and their transformations, this makes inference on the coefficient of the slack variable dependent on an arbitrary choice of supply-shock proxies or IVs.[2]

As a second contribution, we show how the recently proposed 'Kinky Least Squares' (KLS) estimator of Kiviet (2022) can be used for specification testing in the context of PCs. KLS makes it possible to directly test for the correlation of the error term with any variable by bias-adjusting the OLS estimator for a given degree of endogeneity. Consequently, we can formally test whether a given supply-shock proxy has been incorrectly omitted from a given PC.[3] Equivalently, this test can also be treated as a test of IV validity, so that for the first time in the vast literature on PCs, we are able to directly test the exclusion restrictions of IVs.[4] On either interpretation, this test is especially useful in reduced-form modelling of PCs, since by definition, there is no theoretical or structural reason for considering one set of supply-shock proxies over another or one set of IVs over another.[5]

As a third contribution, we propose to use the KLS estimator also to conduct endogeneity-robust inference on the coefficient of the slack variable without using any supply-shock proxies or IVs. Rather than relying on ad-hoc and not very transparent arguments for preferring one set of supply-shock proxies or IVs over another, we propose to identify the likely sign and magnitude of the coefficient of the slack variable by leveraging the widely-held intuition that the correlation coefficient of the slack variable and the error term is likely *positive* (due, for instance, to omitted supply shocks as argued in Gordon (2011), or monetary policy as argued in Eser et al. (2020)). Indeed, for non-negative as well as moderately negative

[2]It should be noted that recently developed 'big data' econometric methods such as Chernozhukov et al. (2020) are not able to qualitatively solve the issue of specification uncertainty. This is because even these methods are unable to handle an arbitrarily large number of variables.

[3]Previous OLS-based testing of whether a set of supply-shock proxies are significant in a given PC can only provide suggestive evidence, since the validity of such a test hinges on all variables being exogenous, including the slack.

[4]Previously, any test of IV validity was limited to tests of overidentifying restrictions, see Kleibergen and Mavroeidis (2009) and Mavroeidis et al. (2014).

[5]In inference on structural PCs this issue is not as pronounced, since the PC is correctly specified, and any exclusion restriction is valid by microfoundational assumption.

values of the correlation coefficient between the slack variable and the error term, we find a significantly negative coefficient on the slack variable, supporting the view that the wage PC in the EA is alive and well.

Although the focus of this chapter is on conducting inference on the coefficient of the slack variable in wage PCs, the main issues raised and solutions proposed in this chapter are relevant to many other applications in Macroeconomics, such as inference on Taylor rules, and Euler equations. This chapter is hence closely related to recent work by Carvalho et al. (2021), who use a structural model to argue that Taylor rules should be estimated by OLS, since any bias caused by endogeneity is likely to be small. Our chapter extends their line of argument in two important ways. First, if, as in Carvalho et al. (2021), a structural model is at hand to guide the researcher, KLS makes it possible to exploit both the sign and magnitude of the correlation coefficients between the endogenous variables and the error term for inference. In contrast, the OLS approach advocated by Carvalho et al. (2021) makes no formal use of the sign of the correlation between the endogenous variables and the error term. Second, our approach remains valid even in the absence of a structural model. All that is required is an informal (non-parametric) 'prior' on the degree of endogeneity. The remaining assumptions are all testable. This recommends our approach in reduced-form (policy) settings, where practitioners may have intuition grounded in qualitative economic theory, without being committed to a fully-specified macroeconomic structure.

Organisation of the chapter. Section 4.2 specifies the PC used throughout this chapter. Section 4.3 introduces the methodology and the data. Section 4.4 presents the results. Section 4.5 concludes.

4.2 Model

In this chapter we consider the following reduced-form PC:

$$\pi_t^w = \lambda u_t + \sum_{j=1}^{q_w} \gamma_{wj} \pi_{t-j}^w + \sum_{j=1}^{q_p} \gamma_{pj} \pi_{t-j}^p + \epsilon_t, \tag{4.1}$$

where u_t is the labour market slack variable with coefficient λ, π_t^w and π_t^p are the wage and price inflation, γ_{wj} and γ_{pj} are the coefficients on the j^{th} lag of the wage and price inflation, respectively, and ϵ_t is a mean-zero error term (to simplify notation, we treat the variables as demeaned throughout). This reduced-form wage PC with backward-looking inflation expectations is also the workhorse model, e.g., in Nickel et al. (2019), for the analysis of wage developments in the euro area.

We focus on reduced-form PCs out of a desire for generality. Indeed, inference on PCs embedded in a structural model can be seen as a special case, where primitive assumptions imply that the PC is correctly specified, the error terms are well-behaved, the IVs are valid, and the degree of endogeneity is determined by the model's 'deep' parameters. We omit forward-looking variables out of simplicity, and without compromising our ability to conduct inference on the coefficient of the slack variable. Indeed, since under certain (testable) assumptions the KLS-based approach we propose addresses the likely endogeneity of the slack variable, the results will be robust to potentially omitted variables. These could include, for example, forward-looking inflation expectations.[6]

4.3 Methodology and Data

For all approaches, it is useful to cast the PC specified in Equation (4.1) in matrix form:

$$\begin{aligned} \pi &= \lambda u + X\xi_1 + \epsilon \\ &= X^+\beta + \epsilon, \end{aligned} \tag{4.2}$$

where π is the $T \times 1$ stacked vector with element π_t^w, u is the $T \times 1$ stacked vector with element u_t, X is the $T \times (q_w + q_p)$ matrix with columns corresponding to the $T \times 1$ stacked vectors of $\pi_{t-j_w}^w$, $\pi_{t-j_p}^p$ for $j_w = 1, \ldots, q_w$, and $j_p = 1, \ldots, q_p$, respectively, ξ_1 is the $(q_w + q_p) \times 1$ corresponding vector of coefficients, ϵ is the $T \times 1$ stacked vector with element ϵ_t, $X^+ = [u\ X]$, and $\beta = [\lambda, \xi_1']'$. Throughout, we assume that ϵ_t is independent across t. Under this assumption, we can treat

[6] Nickel et al. (2019) find that backward-looking expectations seem to be more relevant to explain wage growth in the euro area through the lens of the Philips Curve than forward-looking expectations.

the matrix of lagged (predetermined) variables, X, as exogenous. However, u is allowed to be correlated with ϵ.

4.3.1 Single-Equation Estimation

The first approach considered in this chapter is to directly proxy for supply shocks in the PC, i.e., estimate the model given by

$$\begin{aligned} \pi &= X^{+}\beta + S\delta + \tilde{\epsilon} \\ &= W\zeta + \tilde{\epsilon}, \end{aligned} \tag{4.3}$$

where S is a $T \times (q_s + 1)g$ matrix consisting of the $g \times 1$ vector of contemporaneous supply-shock proxies, q_s the number of lags, δ the $(q_s + 1) \times 1$ corresponding coefficient vector, $W = [X^{+}\ S]$, and $\zeta = [\beta', \delta']'$. The crucial identifying assumption is that by including these additional supply-shock proxies, all the variables on the RHS are uncorrelated with the new error term, $\tilde{\epsilon}_t$, i.e.

$$\mathbb{E}\left[W'\tilde{\epsilon}\right] = 0.$$

Assuming that other standard regularity conditions hold and that $q_s g$ is fixed (i.e., $q_s g << T$), this moment condition makes it possible to conduct inference on Equation (4.3) by OLS. Such OLS-based approaches first gained popularity with Gordon (1981), and are sometimes termed 'triangle models', since they include demand variables (the slack variable), inertia (lagged inflation variables) and supply variables (the supply-shock variables). They remain popular both in academic settings (Ball and Mazumder, 2019; Gordon, 2011; Mazumder, 2014) and policy settings (Cunningham et al., 2019).

4.3.2 Instrumental Variable Estimation

The second popular approach to address the endogeneity of the slack variable is to use IVs. This involves exploiting moment conditions given by

$$\mathbb{E}[Z'\epsilon] = 0,$$

where Z is a $T \times k$ matrix with columns corresponding to (any) predetermined variable not included in Equation (4.2).

A well-known problem in the context of PCs is that the IVs are likely to be weak, i.e., they are able to explain only a small portion of the exogenous variation in the endogenous variable (Ma, 2002; Mavroeidis, 2005). Therefore, to avoid biases caused by pre-testing, we consider the Anderson and Rubin (1949) (AR) statistic, which is completely robust to (arbitrarily) weak identification, and has other attractive properties vis-à-vis other more sophisticated weak-identification robust methods (see Dufour (2003) for a discussion).

Confidence sets are constructed by inverting the AR statistic under a given null hypothesis. In our application, we have a single endogenous variable, so that the relevant null hypothesis is given by

$$H_0 : \lambda = \lambda_0 \text{ vs. } H_1 : \lambda \neq \lambda_0.$$

Constructing $(1-\alpha)$ confidence sets by inverting the test statistics involves specifying a (sufficiently fine) grid of values of λ_0, and evaluating the test statistics at each of these values. The hypotheses that cannot be rejected at the α level are included in the $(1-\alpha)$ confidence set, whereas the hypotheses rejected at the α level are not.

The AR statistic is given by

$$AR = \frac{\frac{1}{k}\epsilon_0^{*\prime} P_{Z^*} \epsilon_0^*}{\frac{1}{T-k}\epsilon_0^{*\prime} M_{Z^*} \epsilon_0^*},$$

where $\pi^* = M_X \pi$, $u^* = M_X u$, $Z^* = M_X Z$, $\epsilon_0^* = \pi^* - \lambda_0 u^*$, $M_X = I - X\left(X'X\right)^{-1} X'$ $P_{Z^*} = Z^*(Z^{*\prime}Z^*)^{-1}Z^*$, and $M_{Z^*} = I - P_{Z^*}$. Assuming that k is fixed (i.e., $k << T$), and that other regularity conditions hold (Dufour et al., 2006, Appendix B), the AR statistic converges in distribution to $1/k$ times a χ^2_k distributed random variable.

4.3.3 Kinky Least Squares Estimation

The third approach considered in this chapter is KLS. The crucial quantity in KLS is the vector of correlations between the columns in X^+, and the error term ϵ. This vector of correlations will be denoted by the $(1 + q_w + q_p) \times 1$ vector $\rho_{x^+\epsilon}$. Letting r_j denote the assumed (potentially non-zero) correlation coefficient between

X_j^+ and the error term ϵ, the resulting correlation coefficients vector will be given by the $(1 + q_w + q_p) \times 1$ vector $r = [r_1, 0'_{q_w+q_p}]'$ where $0_{q_w+q_p}$ is a $(q_w + q_p) \times 1$ vector of zeros (since we assume that the error terms, ϵ_t, in the PC specified in Equation (4.2) are independent).

KLS works by considering different reasonable values of r_1, and computing estimates of β that correct for the bias that each value of r_1 entails for standard OLS estimates. In particular, the KLS estimator for β is given by

$$\hat{\beta}_{KLS} = \hat{\beta}_{OLS} - \hat{\sigma}_\epsilon S^{-1}_{x^+x^+} S_{x^+} r,$$

where $\hat{\beta}_{OLS} = (X^{+\prime}X^+)^{-1}X^{+\prime}\pi$ is the OLS estimator of the PC in Equation (4.2), $\hat{\sigma}^2_\epsilon = \hat{\sigma}^2_{\epsilon,OLS}/(1-r'S_{x^+}S^{-1}_{x^+x^+}S_{x^+}r)$, $\hat{\sigma}^2_{\epsilon,OLS} = \frac{1}{T}(y-X^+\hat{\beta}_{OLS})'(y-X^+\hat{\beta}_{OLS})$, $S_{x^+x^+} = \frac{1}{T}X^{+\prime}X^+$, and $S^2_{x^+}$ is the matrix containing just the main diagonal of $S_{x^+x^+}$ so that S_{x^+} is the positive diagonal matrix $S_{x^+}S_{x^+} = S^2_{x^+}$.

Under the relatively mild regularity conditions in Kiviet (2022) (the strongest one being homoscedasticity), the limiting distribution of the KLS estimator is given by

$$\sqrt{T}(\hat{\beta}_{KLS} - \beta) \xrightarrow{d} \mathcal{N}[0, \sigma^2_\epsilon V],$$

where V can be consistently estimated by $S^{-1}_{x^+x^+}\hat{\Theta}S^{-1}_{x^+x^+}$, and $\hat{\Theta}$ is the sample analogue of the expression for Θ in Kiviet (2022).

To test whether supply-shock proxies have erroneously been excluded or to test whether a candidate IV satisfies the exclusion restriction, we wish to test the hypothesis

$$H_0 : \mathbb{E}[z_{wt}\epsilon_t] = 0 \text{ vs. } H_1 : \mathbb{E}[z_{wt}\epsilon_t] \neq 0, \tag{4.4}$$

for some $k_w \times 1$ vector of variables z_{wt} (which contains the supply-shock proxies and/or the candidate IVs). Consider the auxiliary regression

$$\pi = X^+\phi_{x^+} + Z_w\phi_{z_w} + \nu,$$

where ϕ_{x^+} is the new $(1 + q_w + q_p) \times 1$ coefficient vector of X^+, Z_w is the $T \times k_w$ stacked matrix of z_{wt}, ϕ_{z_w} is the corresponding $k_w \times 1$ vector of coefficients, and

ν is the $T \times 1$ vector containing the new error terms. Since we are interested in testing whether or not the variables are correlated with the error term, we set $\rho_{[x^+ z_w]\nu} = [\rho'_{x^+\epsilon}, 0'_{k_w}]'$, where 0_{k_w} is a $k_w \times 1$ vector of zeros, run KLS for different values of r_1, and test the exclusion restriction that $\phi_{z_w} = 0$ through a standard Wald test (or a t-test, if $k_w = 1$). As usual, high Wald statistics (lower p-values) imply stronger evidence that the variables in Z_w are in fact correlated with the error term.

Finally, insofar as KLS is simply an endogeneity-corrected OLS estimator, one can use the usual tests developed for OLS to assess whether the assumptions of independent and homoscedastic error terms are fulfilled (see also Kiviet and Kripfganz (2020)). We thus test for heteroskedasticity by regressing the squared residuals at every value of r_1 on the variables included in the PCs, as well as the supply-shock proxies and IVs considered below, and test for autocorrelation by regressing the residuals on different lag lengths of past residuals.

4.3.4 Data

4.3.4.1 Main Data Used in the Estimation of the Wage Phillips Curve

The sample we consider runs from 1999Q1 to 2019Q4. Wage inflation is measured as the year-on-year growth rate of total compensation per employee, which is the standard measure used for the analysis of the EA wage PC.[7] For simplicity, we consider the differenced unemployment rate as our measure of slack.[8] Other measures of the slack variable (e.g., the recent Hamilton (2018) filter) are also possible, and do not change the results obtained in this chapter substantially. The unemployment rate is the number of unemployed persons as a percentage of the labour force. Price inflation is measured as the year-on-year growth rate of the harmonised consumer price index.

[7] As pointed out in Bonam et al. (2021, p. 12), negotiated wages rather than compensation per employee are the conceptually better measure, since the latter includes, amongst other things, one-offs, fiscal shocks, and compositional effects. However, contrary to Bonam et al. (2021), an EA wage PC (as opposed to country-specific wage PCs) is considered, which makes aggregation of negotiated wages across EA countries more difficult, due to differences in methodology and data quality.

[8] We cannot consider the even simpler measure given by the unemployment rate itself, since for the EA it exhibits non-stationary behaviour, see also Galì (2015).

4.3.4.2 Data on Supply-Shock Proxies and Instrumental Variables

Wage growth is likely to be not only influenced by the development of slack (reflecting mainly changes in demand), but also by supply shocks and monetary policy.

To assess the role of these factors, we include standard variables considered in the literature including oil price inflation (covering global price shocks via energy, see Gordon (2011)), labour productivity growth (reflecting changes in labour supply and its composition, see Nickel et al. (2019)), interest rate spreads (covering changes in the monetary policy stance, see Galì and Gambetti (2019), and exchange rate developments (covering international monetary shocks, see Mazumder (2014)).

In the literature that uses IV methods to test PCs, IVs commonly used are lagged values of the endogenous variable and of the dependent variable, as well as variables such as commodity inflation and the interest rate spread (Dufour et al., 2006; Galì and Gertler, 1999; Kleibergen and Mavroeidis, 2009; Mavroeidis et al., 2014). In essence, the IV-based approaches in the literature use as IVs precisely those variables that are directly included in the PC by the approaches in the literature that account for the endogeneity of the slack by including supply-shock proxies.

We measure oil price inflation as the year-on-year growth rate of the oil price. Productivity growth is measured as the year-on-year growth rate of productivity per employee, measured as the gross value added in relation to the total number of employees. The interest rate spread is calculated as the difference between the long-term rate and the short-term rate. The exchange rate considered is the USD-euro exchange rate, aggregated to the quarterly frequency by simple averaging. We construct technical controls by interacting these variables as well as their lags amongst each other. We consider two lags of each supply-shock proxy, so that we are left with 78 candidate supply-shock proxies. We construct IVs in a similar way, but following the IV literature on inference on PCs, also include the interaction with the exogenous variables specified in Equation (4.2) (i.e., the lags of wage inflation and price inflation) and lags of the endogenous slack variable. Throughout, we set $q_w = q_p = 2$. This leaves us with 101 IVs.

These variables are not meant to exhaustively cover all possible supply-shock proxies and IVs. It certainly could be argued that there are other relevant supply-shock proxies and IVs that should be additionally included. However, this illustrates precisely the issue we aim to exemplify.

Indeed, we use this data to show that even with this small and incomplete set of supply-shock proxies, there is substantial specification uncertainty. Arguing that more variables are relevant only worsens the specification uncertainty, and hence strengthens our argument.

Appendix C gives a detailed description of these data, along with their sources.

4.4 Results

4.4.1 Single-Equation Estimation Results

Table 4.1 shows the single-equation estimation results using OLS. Since OLS treats the number of variables as fixed, not all the supply-shock proxies can be considered simultaneously. The only possible approach is to consider the sensitivity of the results to different sets of supply-shock proxies. Column (1) shows the results when no additional supply-shock proxy is added. Columns (2)-(5) show the results when the contemporaneous value, as well as two lags of each respective supply-shock proxy considered are included. Column (6) shows the results when the contemporaneous value of each of the three supply-shock proxies is included (but no lags are included). Column (7) shows the results when five randomly selected supply-shock proxies are chosen from the available 78 technical supply-shock proxies. To keep the results readable, we do not report the point estimates and confidence intervals for the different supply-shock proxies. However, we test the significance of each of the supply-shock proxies included, and report the smallest p-value of these tests in the penultimate row. In the last row we also report the p-value of a Wald test of the hypothesis that the included supply-shock proxies all have a coefficient equal to zero.

The results show that the point estimate of the coefficient of the slack variable is negative across the different specifications considered. However, the 90% confidence sets vary greatly. In column (6) and column (7) we are unable to reject the

hypothesis that the true coefficient of the slack variable is zero. Furthermore, for all cases except for column (4) and column (5), we find that at least one of the included supply-shock proxies has a coefficient significantly different from zero, as can be seen from the the p-values reported in the second row. This suggests two things. First, it suggests that omitted-variable bias caused by omitted supply-shock proxies is a valid concern (although this is not a formal test of the correlation of the supply-shock proxies with the error term of the model without them; see below for a formal direct test of this hypothesis). Second, it suggests that there is no obvious way to select which supply-shock proxies to use. Indeed, there seems to be no principled reason for preferring one set of supply-shock proxies over another, since across the different specifications considered we often find 'significant' supply-shock proxies.[9]

Table 4.1: OLS single-equation estimation results for the coefficient of the slack variable in Equation (4.3).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
u	-0.611 [-1.061, -0.162]	-0.619 [-1.045, -0.193]	-0.727 [-1.329, -0.125]	-0.534 [-0.904, -0.165]	-0.863 [-1.256, -0.470]	-0.174 [-0.784, 0.436]	-0.418 [-0.748, 0.524]
t test p-value	–	0.002	0.001	0.395	0.269	0.002	0.000
χ^2 test p-value	–	0.206	0.085	0.761	0.125	0.552	0.392

Notes:
90% confidence interval based on robust standard errors in brackets.
t test p-value refers to the smallest p-value corresponding to individual significance tests on the coefficients of the supply-shock proxies.
χ^2 test p-value refers to the p-value of a Wald test that tests whether the coefficients of the included supply-shock proxies are jointly equal to zero.
Column (1) does not include any supply-shock proxies. Column (2) uses contemporaneous oil inflation, as well as two lags of oil price inflation, as supply-shock proxies. Column (3) uses contemporaneous labour productivity growth, as well as two lags of labour productivity growth, as supply-shock proxies. Column (4) uses the contemporaneous interest rate spread, as well as two lags of the interest rate spread, as supply-shock proxies. Column (5) uses the contemporaneous USD-euro exchange rate, as well as two lags of the USD-euro exchange rate, as supply-shock proxies. Column (6) uses contemporaneous oil inflation, productivity growth, interest rate spread, and USD-euro exchange rate as supply-shock proxies. Column (7) uses five randomly chosen technical supply-shock proxies. These happen to be the interaction between the one-period and two-period lagged USD-euro exchange rate, the interaction between contemporaneous oil price inflation and the two-period lagged USD-euro exchange rate, the interaction between the one-period lagged interest rate spread and the contemporaneous USD-euro exchange rate, the interaction between two-period lagged oil price inflation and the one-period lagged interest-rate spread, and the interaction between one-period lagged oil price inflation and the two-period lagged interest rate spread.

4.4.2 Instrumental Variable Estimation Results

Figure 4.1 shows the plot of one minus the p-value of the AR statistic for different choices of IVs. To avoid overfitting the relationship between the endogenous variable and the IVs, we limit the number of IVs to five.

[9] Attempting to 'average' these results across different specifications does not seem feasible. For example, a Bonferonni-type correction would be powerless, since even limiting ourselves to the (few) supply-shock proxies we consider in this chapter and to (arbitrarily) setting the number of supply-shock proxies to five, we are left with over 20 million possible specifications.

We present the confidence sets as in Kleibergen (2002). First, we compute the p-value for each hypothesised value of the coefficient of the slack variable in the grid of values considered as outlined in Section 4.3.2. Second, we report in black one minus the p-value for each of the hypotheses tested. Hence, the 90% confidence set for the coefficient of the slack variable is given by the values of λ_0 for which the black curve is below 0.9, which is shown as the light grey horizontal line. Panel (a) shows the results when the one-period lagged slack variable and the one-period lags of each of the four supply-shock proxies are used as IVs. Panels (b)-(d) show the results when five randomly selected IVs are used.[10]

None of the sets of IVs considered indicate that the slack variable is significantly negative. This suggests a problem of weak identification, which in turn stems from the difficulty of predicting a differenced macroeconomic variable such as the measure of slack we consider. Even so, the results point to substantial heterogeneity in the inferential results. Similarly to the case of single-equation estimation above, this raises the question of how to aggregate the results across the very many combinations possible. Furthermore, although the variables considered above are often used as IVs when conducting inference on PCs, the exclusion restriction is questionable, especially in a reduced-form context.

4.4.3 Using Kinky Least Squares to Test Phillips Curves Specifications

Figures 4.2–4.7 test the (joint) hypothesis of uncorrelatedness for the four supply-shock proxies considered in this chapter, as well as their lags. The horizontal axes

[10]These happen to be as follows. Panel (b): interaction between one-period lagged wage inflation and the one-period lagged interest rate spread, interaction between one-period lagged wage inflation and one-period lagged slack, interaction between one-period lagged slack and one-period lagged productivity growth, interaction between one-period lagged inflation and two-period lagged oil price inflation. Panel (c): interaction between one-period lagged oil price inflation and two-period lagged productivity growth, interaction between the one-period lagged spread and two-period lagged slack, interaction between one-period lagged slack and one-period lagged oil price inflation, interaction between one- and two-period lagged inflation. Panel (d): interaction between one-period lagged slack and one-period lagged oil price inflation, interaction between two-period lagged slack and the two-period lagged USD-euro exchange rate, interaction between two-period lagged wage inflation and two-period lagged inflation, interaction between the one-period lagged USD-euro exchange rate and one-period lagged productivity growth, interaction between one-period lagged oil price inflation and two-period lagged slack.

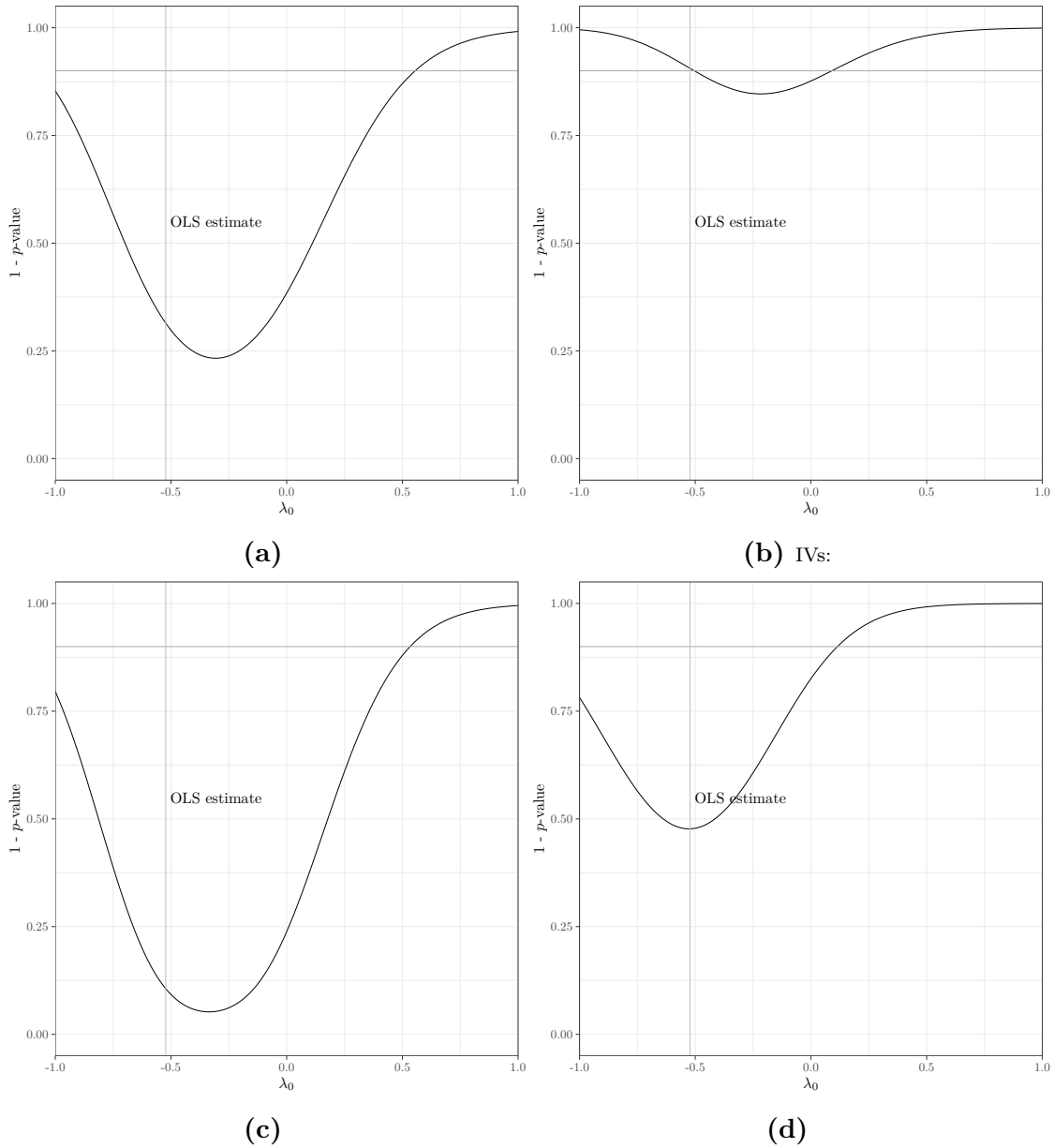


Figure 4.1: AR confidence sets. See the main text for choice of IVs for each panel.

show the posited correlation coefficient between the slack and the error term, r_1 (see Section 4.3.3), while the vertical axes show the p-value of the test of variable exogeneity for the posited degree of endogeneity (see Equation (4.4) in Section 4.3.3). Low p-values suggest that the variable considered is correlated with the error term, and that it hence should be included in the wage PC. The results suggest that for large parts of the parameter space of the correlation coefficient between the slack and the error term, the hypothesis that the supply-shock proxies are uncorrelated

with the error term can overwhelmingly be rejected. There is no set of supply-shock proxies for which we can conclusively claim that they are not missing from the PC.

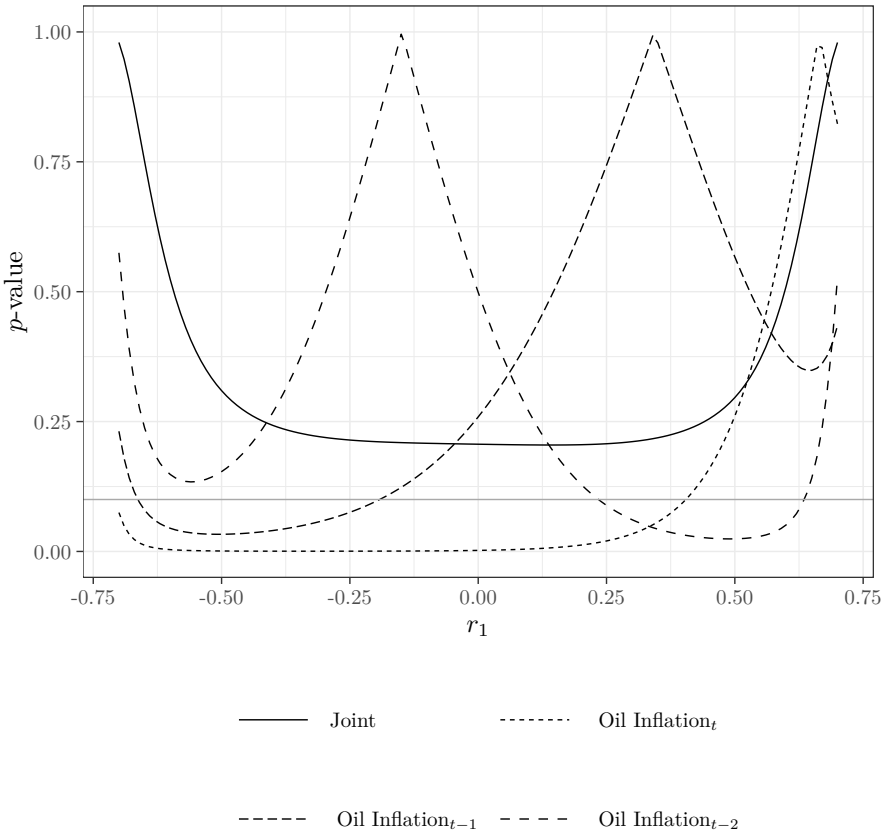


Figure 4.2: Tests of relevance of supply-shock proxies with KLS: oil inflation and its lags.

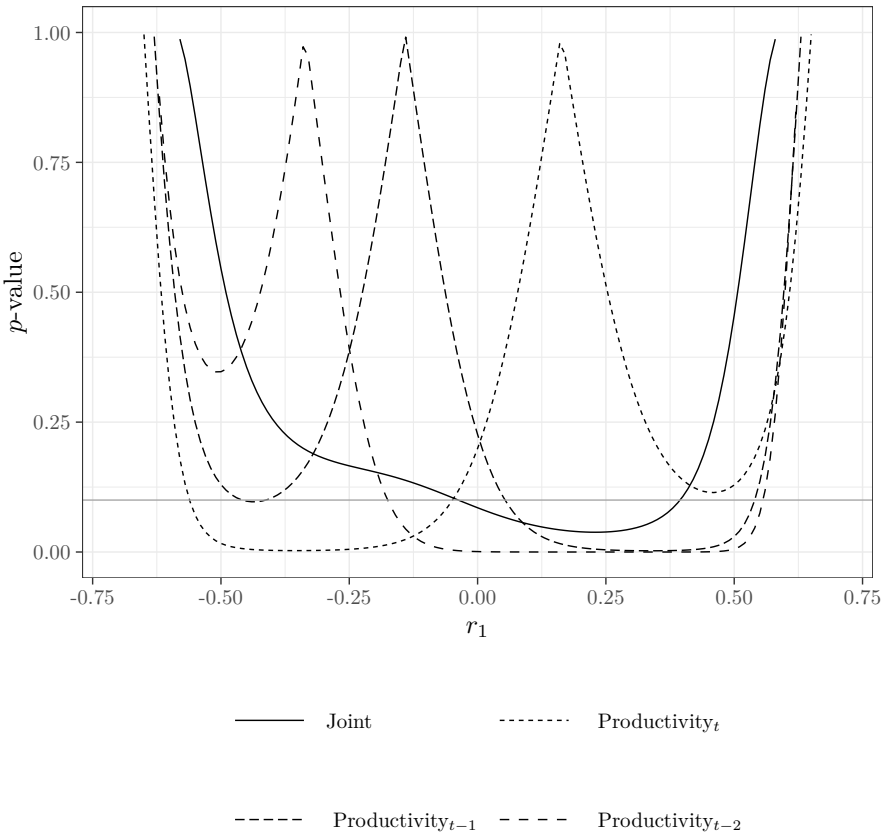


Figure 4.3: Tests of relevance of supply-shock proxies with KLS: labour productivity growth and its lags.

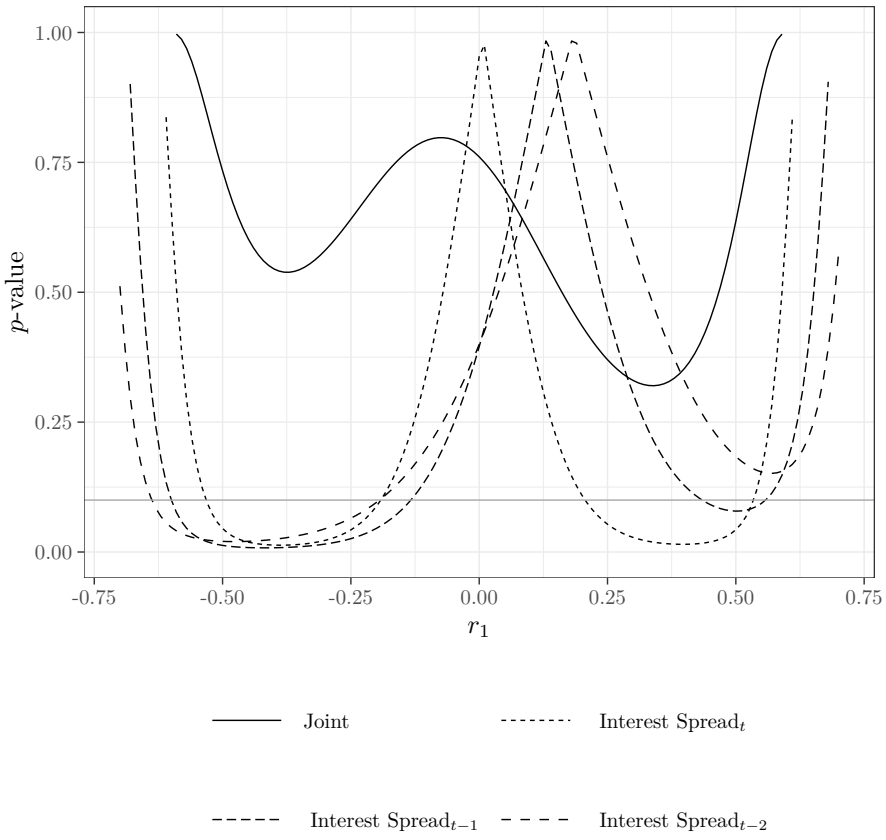


Figure 4.4: Tests of relevance of supply-shock proxies with KLS: interest rate spread and its lags.

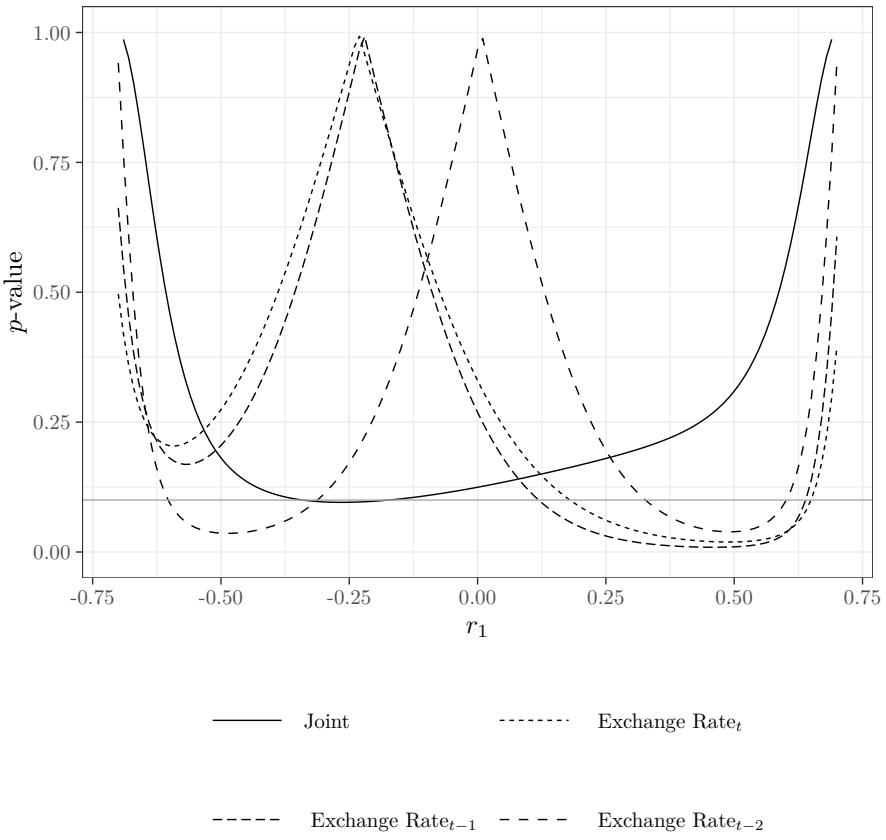


Figure 4.5: Tests of relevance of supply-shock proxies with KLS: USD-Euro exchange rate and its lags.

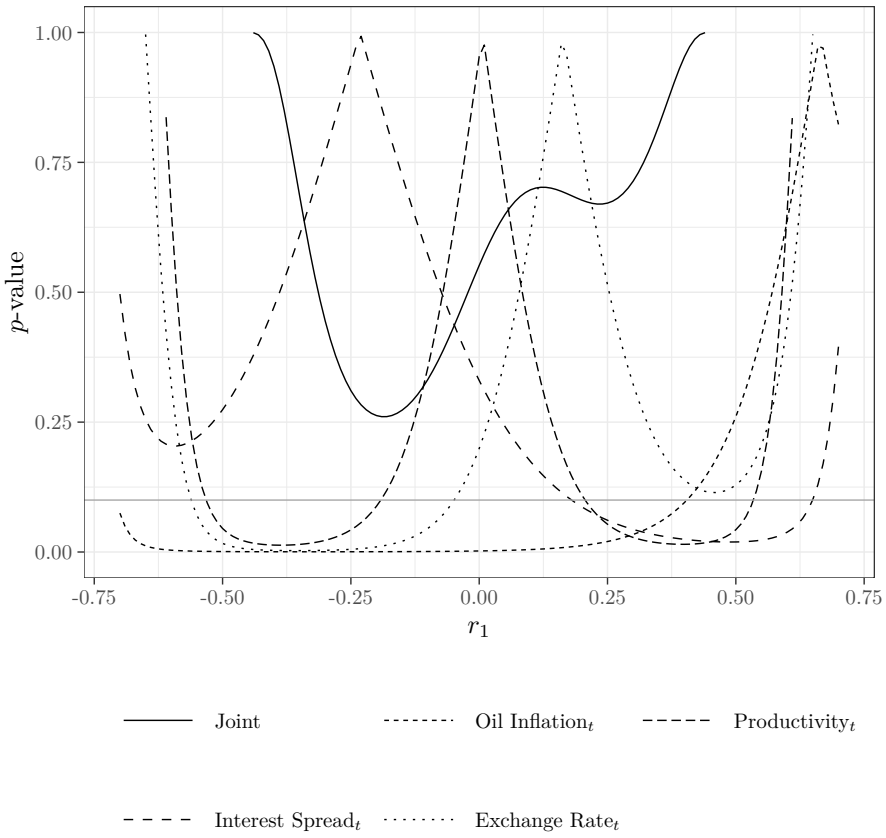


Figure 4.6: Tests of relevance of supply-shock proxies with KLS: contemporaneous oil inflation, labour productivity growth, interest rate spread, USD-Euro exchange rate.

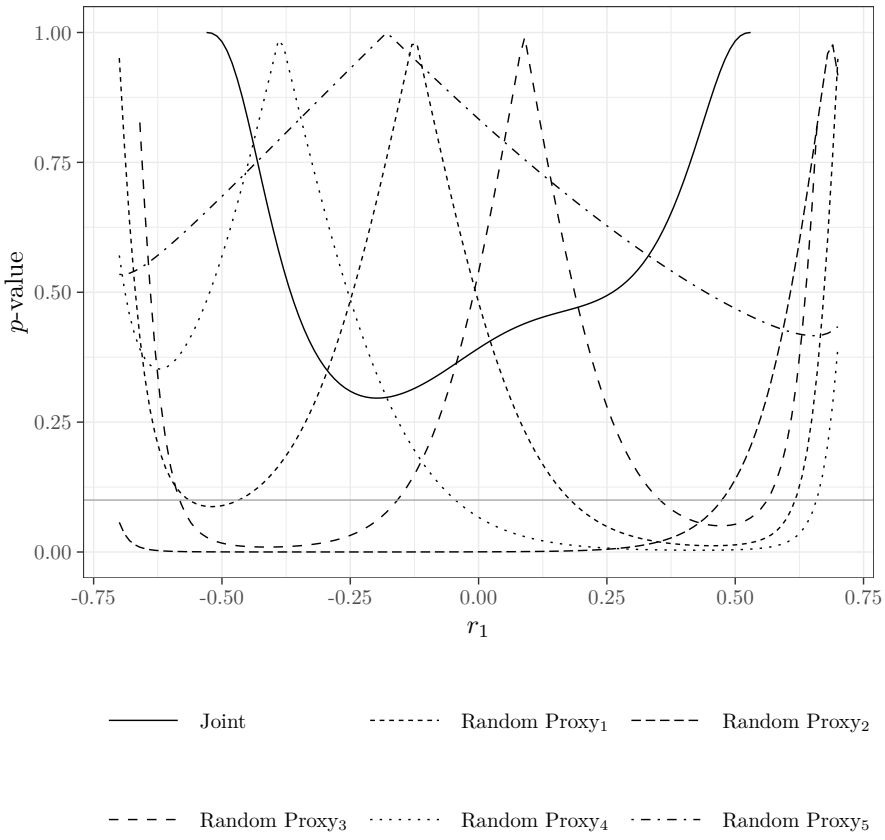


Figure 4.7: Tests of relevance of supply-shock proxies with KLS: random technical supply-shock proxies.

Equivalently, this exercise can also be interpreted as a direct test of the validity of the IVs. Figure 4.8 shows the tests of the individual exclusion restrictions for all of the 101 IVs constructed. The results show that for all degrees of endogeneity, the identifying exclusion restriction of the IVs is unlikely to hold. Therefore, it is not possible to argue for the validity of the IVs by appealing to the intuition that the correlation between the slack and the error term is likely to be positive, since even in this part of the parameter space, the exclusion restriction of these IVs can be rejected.

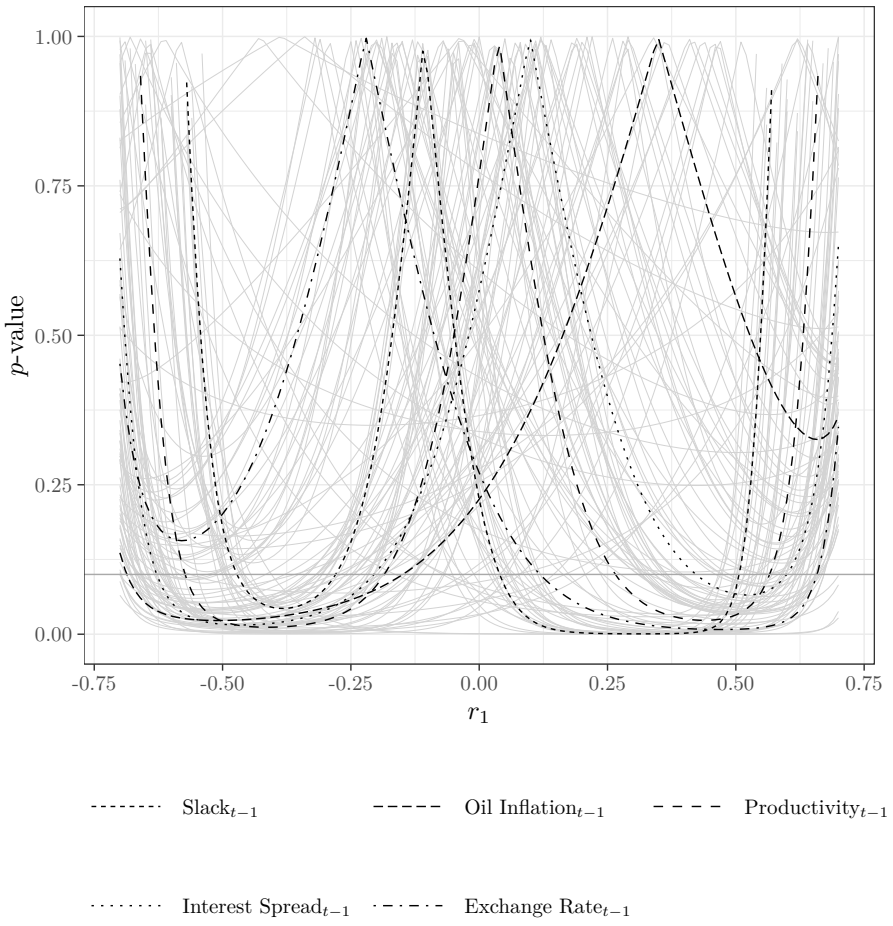


Figure 4.8: Tests of the exclusion restrictions for the 101 possible IVs. The IVs used for the AR statistic in Figure 4.1 Panel (a) are shown in black. Other IVs are shown in grey.

Therefore, in line with Gordon (2011) and references therein, we find evidence that supply-shock proxies play an important role in PCs. Furthermore, by directly testing the exclusion restrictions of the IVs, we find evidence that many of the commonly used IVs are in fact not valid. Both these results motivate the KLS estimation in the next section.

4.4.4 Kinky Least Squares Estimation Results

Figure 4.9 shows the 90% confidence sets implied by KLS.[11,12] The horizontal axis shows the posited correlation coefficient between the slack variable and the error term, while the vertical axis shows the estimated coefficient of the slack variable (black line) and the 90% confidence interval (area shaded in grey) for the posited degree of endogeneity. Under a 'uniform prior' over the correlation coefficient of the slack variable and the error term, nothing interesting can be said on the likely sign and magnitude of the coefficient on the slack variable. However, if we consider values of r_1 above -0.125, we find that the coefficient of the slack variable is significantly negative.

Indeed, the slack variable is usually thought to be *positively* correlated with the error term. This can be motivated structurally as in Eser et al. (2020), who show how the standard three-equation New Keynesian model leads to slack being positively correlated with the error term. It can also be motivated by omitted cost-push shocks (Gordon, 2011), a handful of which were found to be missing from our specification in the section above. Under this assumption (or the weaker assumption that the slack is not 'too negatively' correlated with the error term), the results can be interpreted as saying that the coefficient on the slack variable is significantly negative. One obvious point of criticism is that although this assumption seems plausible (and we are unaware of any theoretical or empirical literature claiming that the slack variable should be highly negatively correlated with the error term), in the absence of a valid IV it does remain an unverifiable assumption. However, we nevertheless believe that this approach is more transparent than ones that use a 'preferred specification'. Indeed, while such specifications often rely on ad-hoc arguments that do not have a clear economic rationale (e.g., preferring import prices over exchange rates as supply-shock proxies), our 'identifying assumption' corresponds to a strong prior often held by macroeconomists and policy makers.

[11] Considering a broader range does not qualitatively change the results (values of r_1 below or above ± 0.5 lead to point estimates that are distinctly above and below zero, respectively), all while making them less readable.

[12] We note that $r_1 = 0$ corresponds to the OLS estimate and confidence interval.

Furthermore, inference on the coefficient of the slack variable using KLS offers three advantages over the other inference procedures considered in this chapter. First, it addresses all types of endogeneity, and is not limited to addressing the endogeneity caused by omitting the handful of supply-shock proxies considered in this chapter. Indeed, although focusing on $r_1 > 0$ is motivated by the important role supply shocks are likely to play, we take no stance on the nature of these shocks, or even require a proxy for them to exist. Second, identification does not rely on questionable exclusion restrictions of the IVs. Third, insofar as no additional variables are used (either as supply-shock proxies or IVs), it also does not suffer from the type of specification uncertainty illustrated above.[13]

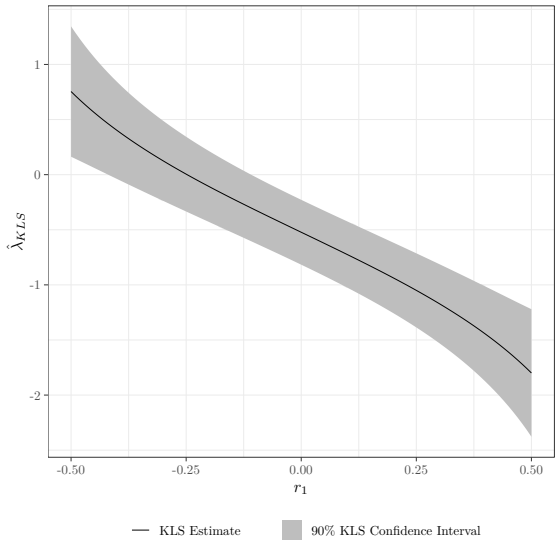


Figure 4.9: 90% KLS confidence sets for the coefficient of the slack variable in Equation (4.2).

The validity of KLS depends on the data being homoscedastic and the error-terms being serially uncorrelated. These are both testable assumptions. Figure 4.10 shows the Breusch-Pagan test for heteroskedasticity as adapted to KLS. The black line shows the p-values that test whether the coefficients of the variables contained in the non-augmented PC (i.e., the variables in Equation (4.2)) are

[13]Specification uncertainty stemming from how to measure slack and (wage) inflation of course remains. However, especially in the reduced-form context considered in this chapter, this type of specification uncertainty seems unavoidable.

significantly different from zero when the squared KLS residuals are regressed on them for different values of r_1. The additional variables in transparent grey show the p-values corresponding to the same test, but where one of the 101 variables considered as IVs or the four contemporaneous supply-shock proxies are added (105 variables in total). The p-values are large across all the parameter space considered for all these different combinations, so we conclude that there is no evidence against the assumption of homoscedasticity.

Figure 4.11 shows the p-values corresponding to the test of whether lagged residuals are significant in a regression of the KLS residuals on lagged values of the KLS residual of increasing order. We find no evidence for autocorrelation for values of r_1 below approximately 0.3, and some evidence for autocorrelation for values between 0.3 and 0.5, and substantial evidence for values above that. Based on recent structural analytic and simulation results in Berriel et al. (2016) and Carvalho et al. (2021), our informal prior is that values of r_1 above 0.5 can reasonably be discounted. Indeed, assuming that the correlation coefficient between the slack and the error term is above 0.5 almost obviates the need for an empirical analysis of the effect of the slack variable on inflation dynamics: under such a dogmatic prior, standard results on OLS bias in the context of endogeneity would likely be sufficient to reach the same qualitative conclusion that slack negatively affects inflation prior to any empirical analysis.

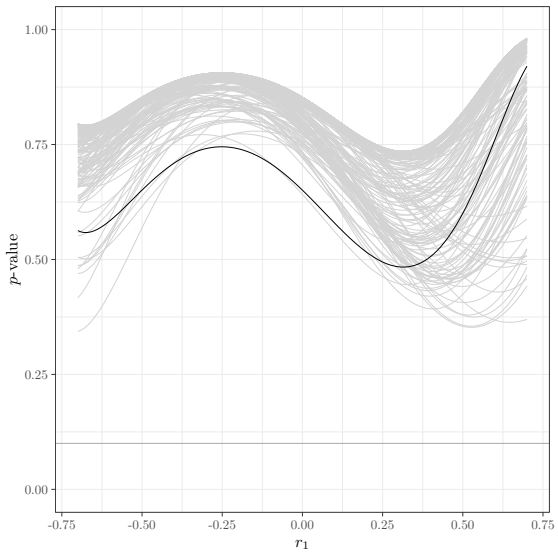


Figure 4.10: Breusch-Pagan test for heteroskedasticity.

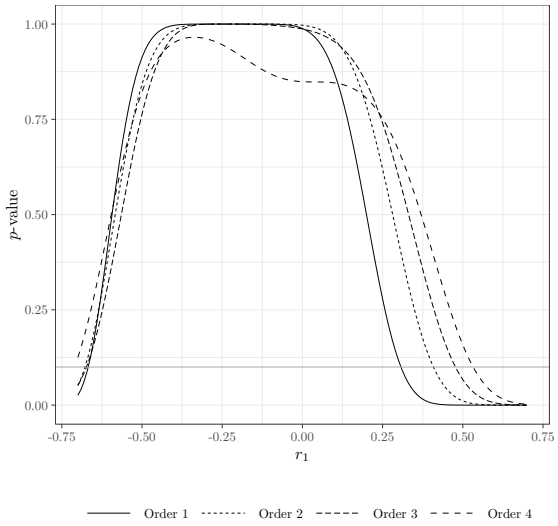


Figure 4.11: Test for autocorrelation.

4.5 Conclusion

This chapter aims to provide endogeneity-robust inference on the coefficient of the slack variable in (reduced-form) PCs. We first identified two broad approaches in the existing literature that seek to account for the endogeneity of the slack variable. The first approach uses proxies of supply-shock variables to address

the endogeneity caused by omitting relevant supply shocks, while the second approach seeks to use IVs.

As the first contribution, we showed that conducting inference using such methods is difficult, if not impossible. Approaches that augment standard PCs with supply shocks require the unverifiable assumption that all the endogeneity is due to the variation in the omitted supply-shock proxies, while IV-based methods require making questionable assumptions on the uncorrelatedness of the chosen IVs with the error term. We also showed that both approaches suffer from specification uncertainty, so that it is unclear how to aggregate the inferential results from the very many different specifications available.

The main methodological contribution of this chapter to the estimation of PCs consists in advocating the use of KLS. First, we showed how KLS makes it possible to formally test whether certain supply-shock proxies should be included in a PC, and (equivalently) assess whether a variable satisfies the exclusion restriction needed for it to be a valid IV. Second, applying KLS we show that the coefficient of the slack variable in the EA wage PC we analyse is indeed significantly negative, as long as the correlation between the slack variable and the error term is not negative with a large absolute value.

5

Robust Inference with Unbalanced Proxy SVARs and the Mundellian Trilemma

5.1 Introduction

'Proxy Structural Vector Autoregressions' (proxy SVARs, sometimes called instrumental variable (IV) SVARs) have become increasingly popular in Macroeconomics (see, e.g., Gertler and Karadi (2015), Mertens and Ravn (2013), and Rey (2016) as well as citing articles). Proxy SVARs identify a column of the covariance matrix of the structural error terms through a variable that is correlated only with the structural error term corresponding to that column (and not with the remaining structural error terms).

The arguably most popular approach to exploit these moment conditions has been two-stage least squares (TSLS). One advantage of using TSLS in this context is that it allows for 'unbalanced' proxy SVARs as in Gertler and Karadi (2015), where the sample for which the IV is observed is shorter than the sample for which data for the reduced-form VAR are observed. This is desirable, since it is often the case that data for the reduced-form VARs are observed over a longer horizon than for the IVs. Indeed, 'standard' variables in reduced-form VARs (such as the policy rate, the consumer price index, and the industrial production index) are often available starting in the 1970s. The IVs, on the other hand, being constructed from more

granular and high-frequency data, are often observed over a relatively shorter horizon. The most prominent example of such an IV is perhaps the aggregation of changes in the three-month Fed Fund futures around monetary policy announcements as introduced by Gürkaynak et al. (2005) and first used in the context of proxy SVARs by Gertler and Karadi (2015), which is only available from 1991 onwards.

One substantial drawback of using TSLS is its lack of robustness against weak identification. This is salient in Macroeconomics, since there is often little relevant exogenous variation in time series data. Pre-testing for IV strength is not a satisfying solution, since selecting IVs based on the rejection decision of a pre-test leads to a test that does not control for size. It also seems unnecessary, since the main drawback of weak-identification robust methods in the wider Economics literature (i.e., only providing confidence sets and no consistent estimator) is not as relevant for inference in proxy SVARs, where the sampling uncertainty of the impulse response functions (IRFs) themselves is the object of inferential interest.

Montiel Olea et al. (2021) recently proposed a weak-identification robust method for inference on impulse response functions (IRFs). This approach, however, assumes that the IVs are observed over the same horizon as the data used for estimation of the reduced-form VARs. Provided that the sample is not excessively unbalanced to preclude estimation of the reduced-form VAR using data only from the sample for which the IV is available, one approach to conduct weak-identification robust inference in unbalanced proxy SVARs would be to shorten the sample (and potentially the lag structure) of the reduced-form VARs to match the sample of the IVs (e.g., Paul (2020)). The drawback of doing so is a potential loss of power, which I verify in the applications I consider in Section 5.3.

The first contribution of this chapter consists in proposing a modification of a weak-identification robust method in Montiel Olea et al. (2021) to allow for unbalanced proxy SVARs. To the best of my knowledge, there is currently no 'ready-to-use' approach (other than shortening the sample) to conduct fully weak-identification robust inference on IRFs identified with an IV that is able to deal with unbalanced proxy SVARs. This method can hence be of use to

practitioners who want to use robust methods of inference on proxy SVARs, but have an unbalanced sample.

As a second contribution, I revisit the applications in Gertler and Karadi (2015) and Rey (2016) who study the effects of US monetary shocks on real and financial variables in the US and other countries, respectively. Applying the approach proposed in this chapter, I find that while the results for the US remain qualitatively unaltered, there is substantially weaker evidence of US monetary spillovers in Canada, Sweden, and the UK. This suggests that these countries' floating exchange rates may in fact insulate them from monetary shocks from abroad, in line with the theoretical predictions of the Mundellian Trilemma. Substantively, this means that the IV-based evidence in favour of a Mundellian *Di*lemma between free capital flows and independent monetary policy is more muted.

Organisation of the chapter. Section 5.2 reviews proxy SVARs and details the proposed inferential procedure. Section 5.3 presents the empirical results. Section 5.4 concludes.

5.2 Methodology

I consider the reduced-form VAR given by

$$Y_t = \sum_{j=1}^{p} A_j Y_{t-j} + u_t, \tag{5.1}$$

where Y_t is an $n \times 1$ vector of variables, A_j is an $n \times n$ matrix of coefficients, p is the number of lags considered, and u_t is the reduced-form zero-mean error term (all variables are hence demeaned). Throughout, and without loss of generality, I assume that the variable whose effects on the other variables is to be studied (US monetary policy in the application below) is ordered first.

The reduced-form error terms are related to an $n \times 1$ vector of structural errors ε_t by the relation $u_t = S\varepsilon_t$, where S is an $n \times n$ matrix, and s_{ab} indicates the entry in the a^{th} row and b^{th} column of S. Identification occurs by exploiting the exclusion restriction given in Assumption 5.

Assumption 5.

$$\begin{aligned} \mathbb{E}\left[Z_t \varepsilon_{1t}\right] &= \Phi \\ \mathbb{E}\left[Z_t \varepsilon_{it}\right] &= 0 \text{ for } i \neq 1, \end{aligned} \tag{5.2}$$

for some variable Z_t.

If $\Phi \neq 0$, it is possible to identify the first column of S (up to scale) by TSLS. Indeed, for $i = 2, \ldots, n$, any element u_{it} of u_t can be written as

$$u_{it} = \frac{s_{i1}}{s_{11}} u_{1t} + \nu_{it}, \tag{5.3}$$

where $\nu_{it} = \sum_{j=2}^{n} (s_{ij} - \frac{s_{i1}s_{1j}}{s_{11}})\varepsilon_{jt}$. Since for $j = 2, \ldots, n$, ν_{it} is simply a linear combination of the structural shocks ε_{jt}, $\mathbb{E}\left[Z_t \nu_{it}\right] = 0$ by the identifying assumption in Equation (5.2). Hence, assuming $\Phi \neq 0$, TSLS is applicable, and provides a consistent estimate of the coefficient s_{i1}/s_{11}. Since the error term is unobserved, it is replaced with the OLS residuals. Use of the reduced-form OLS residuals makes it possible to consider an unbalanced proxy SVAR. This is because one can make full use of the information available to estimate the reduced-form residuals u_t using data for Y_t available in periods $t = 1, \ldots, T$, and subsequently estimate s_{i1}/s_{11} in Equation (5.3) using data that are only available for $t = t_Z, \ldots, T$ for $t_Z \geq 1$. The standard asymptotic arguments apply if $T_Z \to \infty$ (and hence $T \to \infty$), where $T_Z \equiv T + 1 - t_z$ (see, e.g., Gertler and Karadi (2015)). This assumption is relatively weak insofar as it is equivalent to the assumption the researcher would have to make if they conducted inference by balancing the proxy SVAR (i.e., shortening the sample on which the reduced-form VAR is estimated). Indeed, balancing the sample implies that the sample size is T_Z, so that for standard asymptotic arguments to apply the researcher would also require $T_Z \to \infty$.

The crucial assumption of the TSLS approach above is that Φ is sufficiently distant from 0. Montiel Olea et al. (2021) recently proposed an approach that allows Φ to be arbitrarily close to 0. However, they assume that the data are balanced, in the sense that the IV has to be observed over the same horizon as the variables included in the reduced-form VAR. One approach would be to discard any data that falls outside of the sample for which the IV is observed. However,

this would amount to not making use of all the information available to identify the reduced-form coefficients A_j in the VARs described in Equation (5.1), which are often already very densely parameterised. To make full use of the available information in the reduced-form VARs, I modify the bootstrap adaptation of the AR statistic of Montiel Olea et al. (2021). I propose the following procedure to compute whether a given value of the impulse response of variable i at horizon h to a shock to the variable ordered first (denoted λ_{hi}) is in the relevant $(1-\alpha)100\%$ confidence set:

1. Estimate $A = [A_1, \ldots, A_p]$ from Equation (5.1) by OLS using the sample $t = 1, \ldots, T$. Call the estimate $\hat{A}_T$, and save the resulting residuals $\hat{u}_{T,t}$. Also estimate $A = [A_1, \ldots, A_p]$ from Equation (5.1) by OLS using the sample $t = t_Z, \ldots, T$. Call this estimate $\hat{A}_{T_Z}$, and save the resulting residuals $\hat{u}_{T_Z,t}$.

2. Estimate the reduced-form covariance between the residuals and the IV, Γ say, as in Montiel Olea et al. (2021) using the smaller sample $t = t_Z, \ldots, T$, but using the corresponding reduced-form residuals $\hat{u}_{T,t}$ estimated in Step 1. Call this estimate $\hat{\Gamma}_T$. Calculate the analogous quantity using $\hat{u}_{T_Z,t}$ and call this $\hat{\Gamma}_{T_Z}$.

3. Estimate the reduced-form covariance matrix between $\hat{A}_{T_Z}$ and $\hat{\Gamma}_{T_Z}$ by a HAC estimator as in Montiel Olea et al. (2021) using the smaller sample $t = t_Z, \ldots, T$. Call this estimate $\hat{W}_{T_Z}$.

4. Create a grid for each impulse response λ_{hi} of variable i at horizon h, i.e., $\Lambda \equiv \{\lambda_1, \ldots, \lambda_G\}$.

5. Letting m index the bootstrap replications, generate M i.i.d. draws from the model
$$\left[\mathrm{Vec}(A)'_m, \Gamma'_m\right]' \sim \mathcal{N}_{n^2p+n}\left[\left[\mathrm{Vec}(\hat{A}_T)', \hat{\Gamma}'_T\right]', \frac{1}{T_Z}\hat{W}_{T_Z}\right].$$

6. For each $\lambda_g \in \Lambda$ and $m = 1, \ldots, M$ compute and save
$$\sqrt{T_Z}[\left(e_i' C_h\left(A_m\right) - \lambda_g e_1'\right)]\Gamma_m - \sqrt{T_Z}[(e_i' C_h(\hat{A}_{T_Z}) - \lambda_g e_1')]\hat{\Gamma}_{T_Z}, \tag{5.4}$$

where e_j is the j^{th} column of the identity matrix of dimension n, $C_h(A) = \sum_{r=1}^{h} C_{h-r}(A)A_r$ with $C_0(A) = I$, and $A_r = 0$ for $r > p$. Let $\hat{q}_{g,\alpha/2}$ and $\hat{q}_{g,1-\alpha/2}$ denote the $\alpha/2$ and $1-\alpha/2$ quantiles of the computed values in Equation (5.4).

7. Construct the confidence set given by

$$CS_{1-\alpha} = \left\{ \lambda_g \in \Lambda \,\middle|\, \hat{q}_{g,\alpha/2} \leq \sqrt{T_Z}[(e_i' C_h(\hat{A}_{T_Z}) - \lambda_g e_1')]\hat{\Gamma}_{T_Z} \leq \hat{q}_{g,1-\alpha/2} \right\}.$$

To show the validity of the above approach, I make the following high-level assumption on the reduced-form coefficient estimators as in Montiel Olea et al. (2021).

Assumption 6.

$$\sqrt{T_Z}[\text{Vec}(\hat{A}_{T_Z})', \hat{\Gamma}_{T_Z}']' \xrightarrow{d} \mathcal{N}_{n^2p+n}\left[\left[\text{Vec}(A)', \Gamma'\right]', W\right], \tag{5.5}$$

where W is the relevant reduced-form covariance matrix.

I make two additional assumptions.

Assumption 7. $\hat{W}_{T_Z} \xrightarrow{a.s.} W$.

Assumption 8. $\hat{A}_T \xrightarrow{a.s.} A$, $\hat{\Gamma}_T \xrightarrow{a.s.} \Gamma$.

Assumption 7 assumes that the HAC variance estimator is consistent. It is a weak technical assumption identical to the one made in Montiel Olea et al. (2021). Assumption 8 ensures that the estimators of the reduced-form VAR and the covariance estimator based on the entire sample available for the reduced form have the same probability limit as the one evaluated using only the sample for which the IV is available. It is (implicitly) made in all previous TSLS approaches that conduct inference in the context of unbalanced proxy SVARs. While this is a natural assumption if it is assumed that the underlying DGP is relatively stable, it does exclude, for instance, breaks or certain types of time-varying coefficients. I note that only Assumption 8 imposes conditions that do not have to be made by the researcher who conducts inference on the proxy SVAR by balancing the

sample (i.e., shortening the sample on which the reduced-form VAR is estimated), and apply the original method in Montiel Olea et al. (2021).

The proposition below shows the validity of the approach above under these assumptions.

Proposition 2. Under Assumptions 5-8,

$$\lim_{T\to\infty} \mathbb{P}\left(\lambda_{hi} \in CS_{1-\alpha}\right) = 1 - \alpha.$$

Proof. See Appendix D.1.1. □

5.3 The Effect of US Monetary Policy Shocks

I apply the method proposed above to study the effect of US monetary policy shocks on real and financial variables in the US and non-US countries as in Gertler and Karadi (2015) and Rey (2016), respectively. Data are taken from Rey (2016). The US monetary policy variable is given by the one-year U.S. interest rate, and is ordered first in the vector Y_t. The other variables in Y_t correspond to financial and non-financial variables of the different countries considered. For the case of the US, the additional variables are the (log) US consumer price index (CPI), the (log) US industrial production index (IP), the US mortgage spread, the (log) VIX, and the three-month commercial bond spread. For the case of Canada, Sweden, and the UK the additional variables are the respective (log) domestic CPI, the (log) domestic IP, the domestic mortgage spread, the (log) VIX, and the domestic policy rate. The number of lags is set to 12. Z_t is the monthly aggregate of the changes in prices of three-month Fed Fund futures in 30-minute windows around monetary policy announcements.[1] I use dashed lines to report the confidence intervals for the IRFs computed by the wild bootstrap as in Gertler and Karadi

[1]Rey (2016) uses changes in the *current*-month Fed Fund futures whenever using this variable as an IV leads to a higher first-stage F-statistic than using the three-month Fed Fund futures as IVs. I do not apply this approach to avoid the biases that can arise from endogenous IV selection in the context of weakly-identified (time-series) data, see Mikusheva (2021).

(2015) and Rey (2016).[2] The number of bootstrap replication is set to 2,500. Throughout, I report 90% confidence sets. I also report the robust F statistic of Montiel Olea and Pflueger (2013), and the 5% critical value for rejecting a bias caused by weak identification of more than 10%.

Figure 5.1 shows the confidence sets for the US. The confidence sets implied by the robust procedure above are qualitatively similar to the ones implied by the wild bootstrap (although they are substantially wider). The robust F statistic suggests that there may be mild distortions due to weak identification. Nevertheless, the central qualitative conclusions reported in Gertler and Karadi (2015) continue to hold for the US: a monetary shock adversely affects industrial production, increases financial volatility, the commercial- chapter spread, as well as the mortgage spread, and leaves the price level roughly unchanged (i.e., there is no 'price puzzle').

Figures 5.2-5.4 show the IRFs for Canada, Sweden, and the UK. The confidence sets implied by the robust procedure above are substantially different to the ones implied by the wild bootstrap. This is unsurprising, given that the robust F statistic is unable to reject substantial biases due to weak identification. Substantively, the lack of evidence for significant spillovers of US monetary policy to these countries suggests that their floating exchange rates may in fact insulate them from monetary shocks from abroad. This suggests that there is little evidence against the Mundellian Trilemma.

Finally, Figures 5.5-5.7 show the confidence sets when the sample is balanced, and the original approach of Montiel Olea et al. (2021) is used.[3] The results exemplify the benefits in terms of power the approach proposed in this chapter is able to afford vis-à-vis balancing the sample. Indeed, the confidence intervals of the IRFs yielded by balancing the sample are mostly wider for each application considered. Since a relatively small number of observations is dropped for the

[2] As has been noted by Jentsch and Lunsford (2019) this type of wild bootstrap is likely to be invalid even in contexts of strong identification. I nevertheless choose to compute these confidence intervals to ease comparison with the studies from which these applications are taken from.

[3] The results for the UK are not reported because they are identical to the ones reported in Figure 5.4, since data for the reduced-form VAR are observed over a shorter horizon than data for the IV.

case of Canada and Sweden, the difference between Figures 5.2 and 5.3 on one hand and Figures 5.6 and 5.7 on the other are not as pronounced. However, for the case of the US (where balancing the sample leads to dropping a substantial amount of observations), the difference is stark. Indeed, conducting the empirical analysis by balancing the sample suggests no significant effects of US monetary policy shocks on other US economic and financial variables.

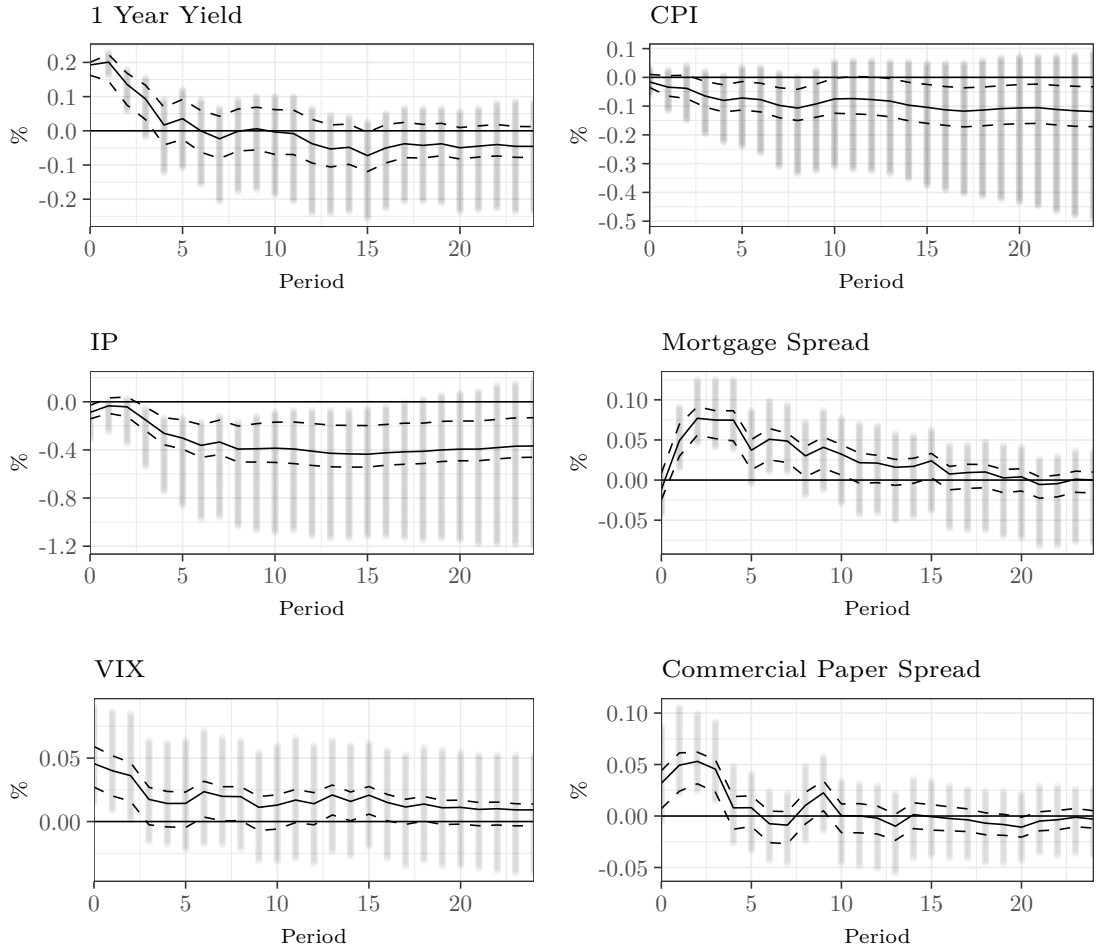


Figure 5.1: Monthly data over the period 1979:07 to 2012:06 for the US. IV available over the period 1991:01 to 2012:06. Robust F statistic: 20.86; critical value: 23.11.

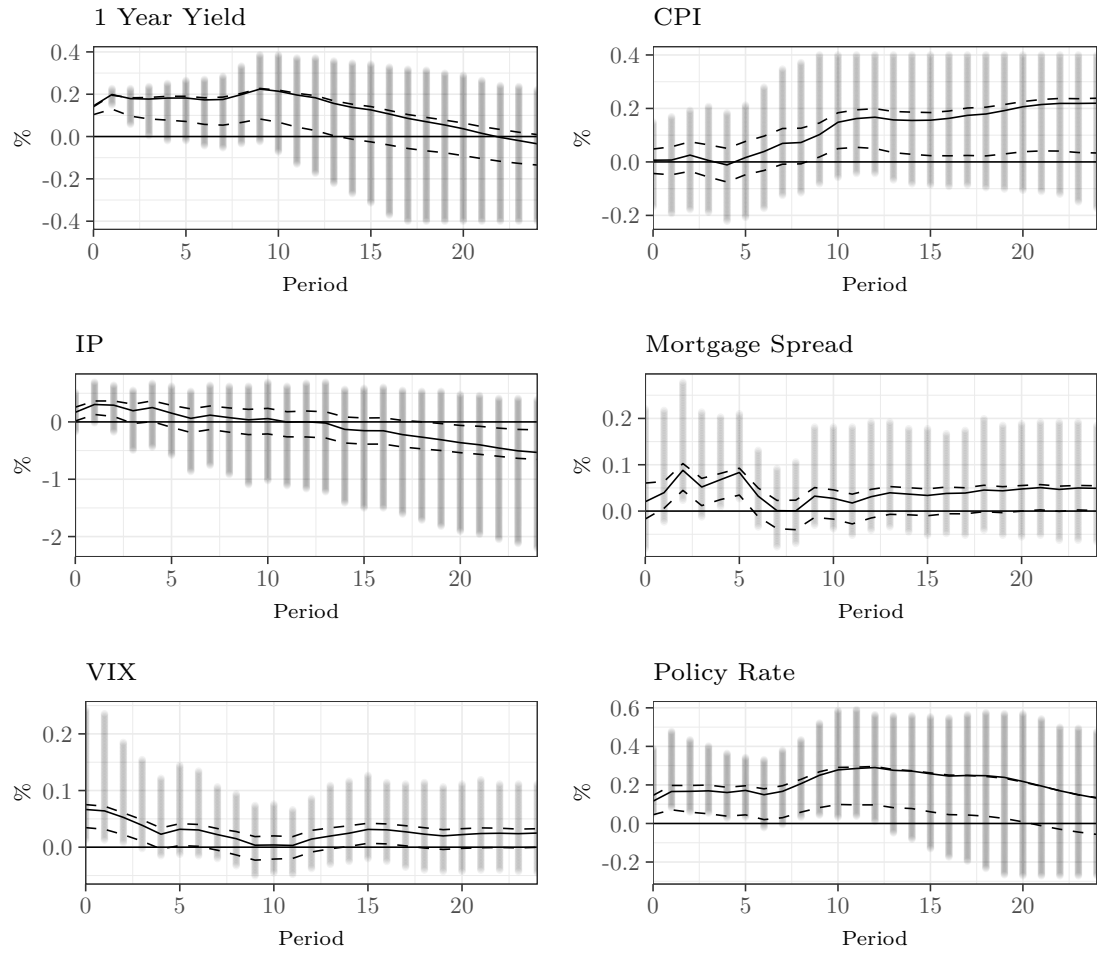


Figure 5.2: Monthly data over the period 1985:07 to 2012:06 for Canada. IV available over the period 1991:01 to 2012:06. Robust F statistic: 9.35; critical value: 23.11.

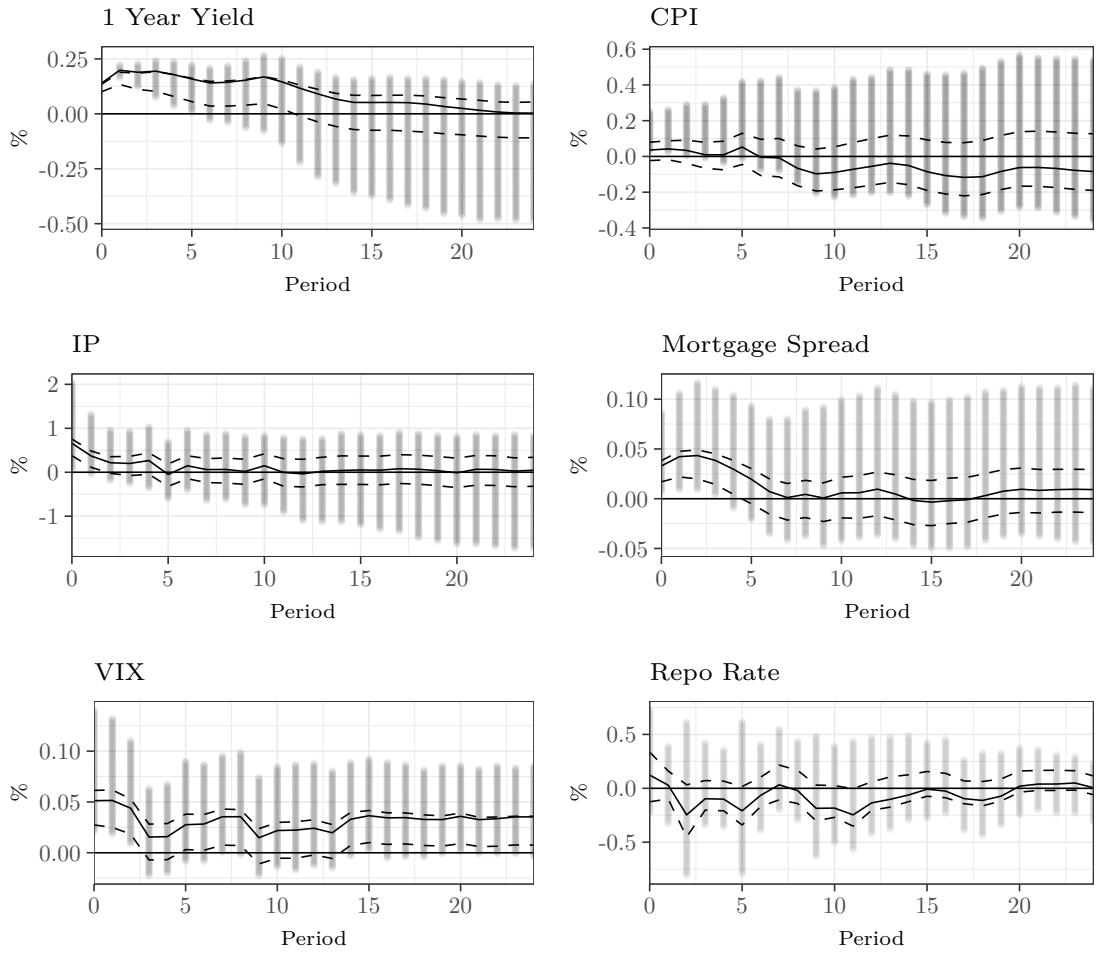


Figure 5.3: Monthly data over the period 1987:01 to 2012:06 for Sweden. IV available over the period 1991:01 to 2012:06. Robust F statistic: 9.40; critical value: 23.11.

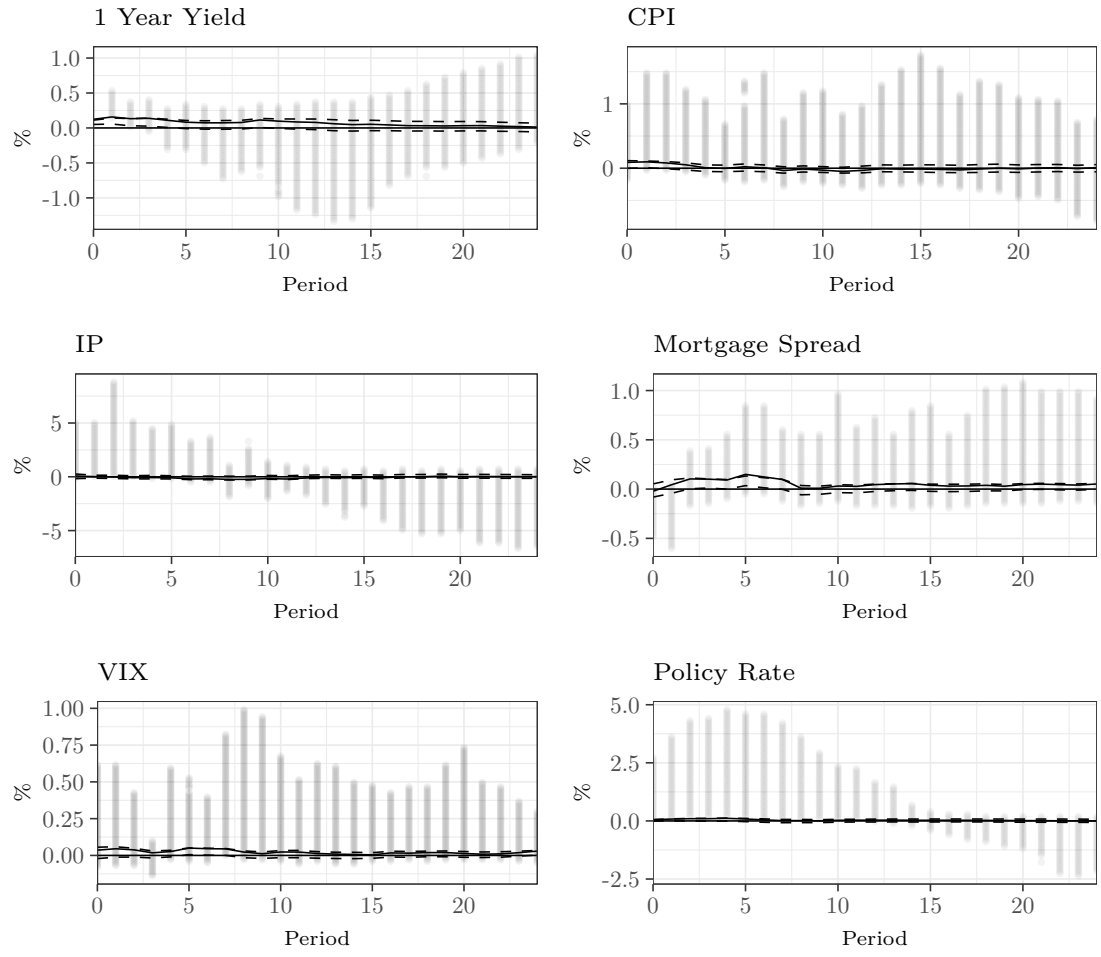


Figure 5.4: Monthly data over the period 1995:01 to 2012:06 for the UK. IV available over the period 1995:01 to 2012:06. Robust F statistic: 2.75; critical value: 23.11.

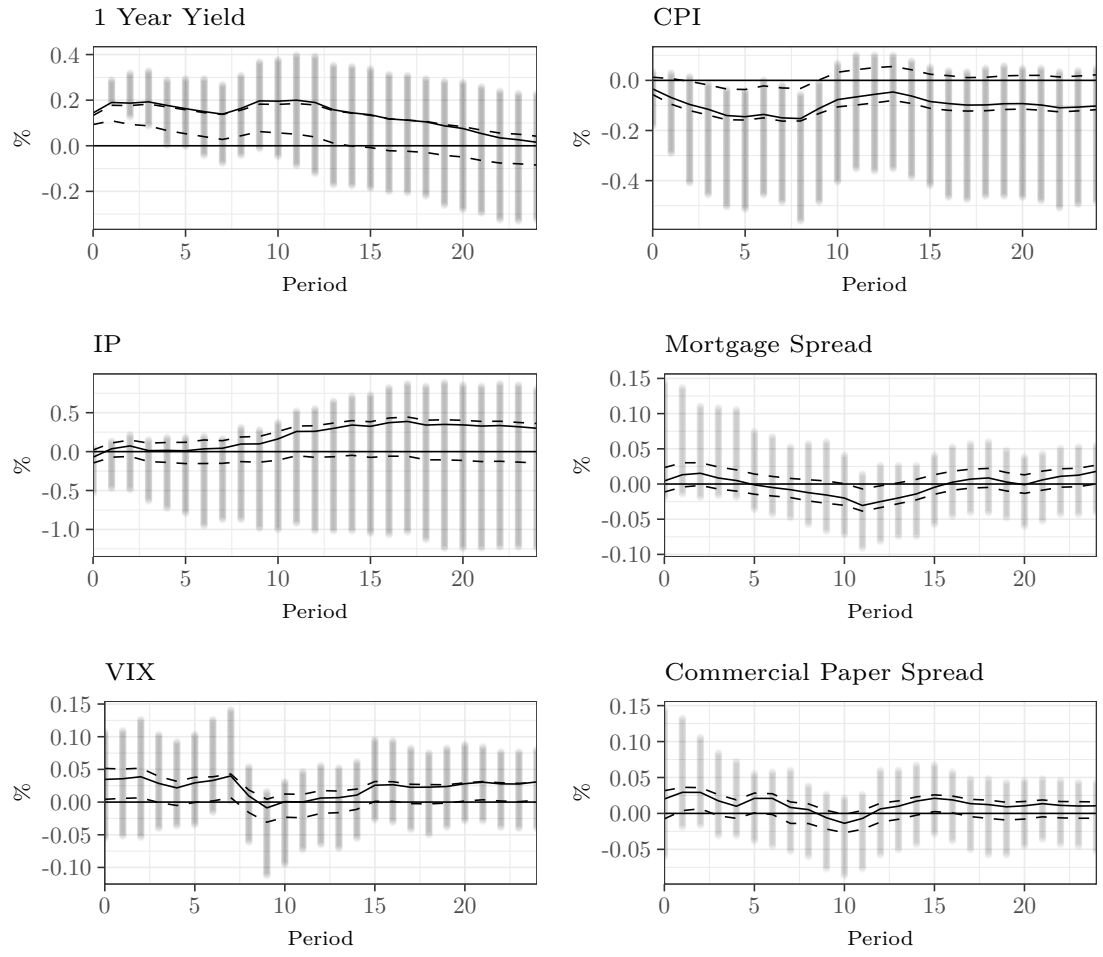


Figure 5.5: Monthly data over the period 1991:01 to 2012:06 for the US. IV available over the period 1991:01 to 2012:06. Robust F statistic: 9.80; critical value: 23.11.

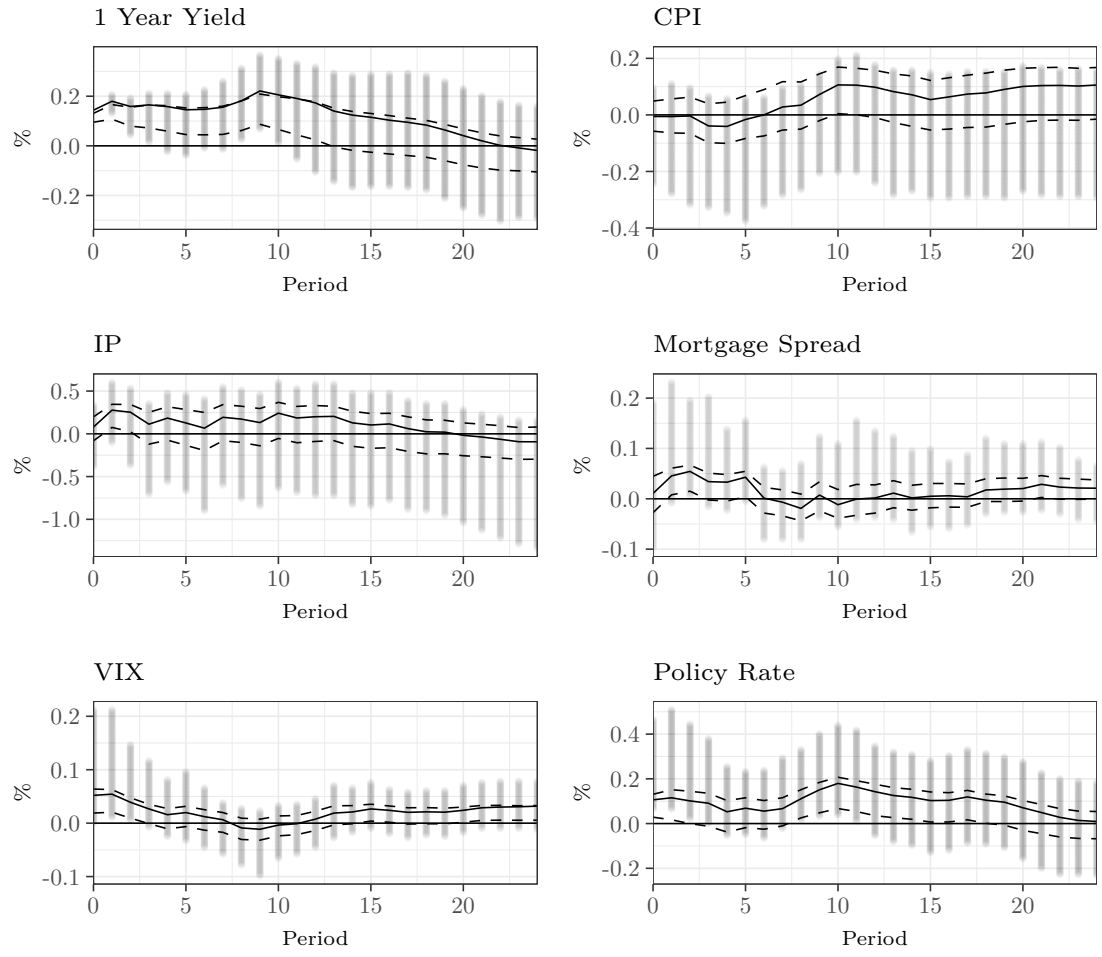


Figure 5.6: Monthly data over the period 1991:01 to 2012:06 for Canada. IV available over the period 1991:01 to 2012:06. Robust F statistic: 10.59; critical value: 23.11.

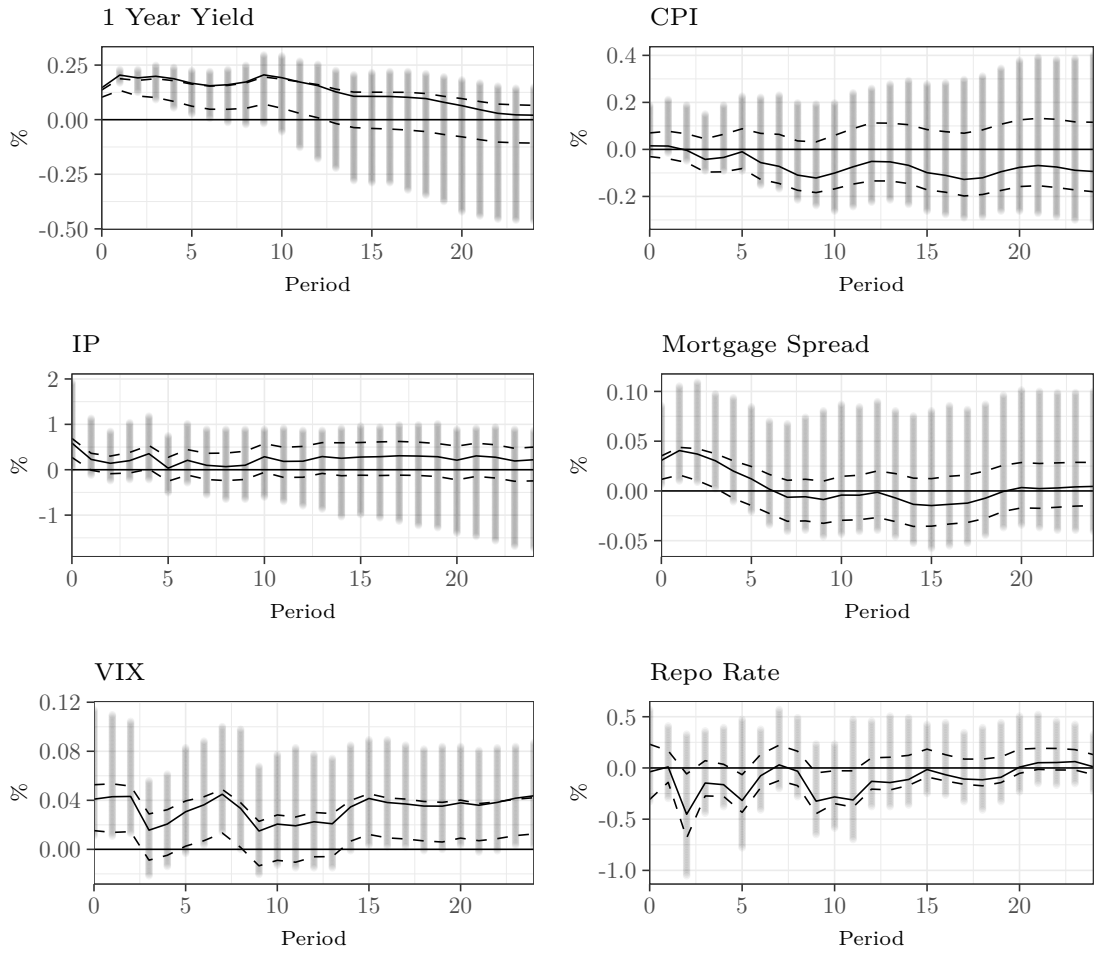


Figure 5.7: Monthly data over the period 1991:01 to 2012:06 for Sweden. IV available over the period 1991:01 to 2012:06. Robust F statistic: 10.89; critical value: 23.11.

5.4 Conclusion

I modified the weak-identification robust proxy-SVAR approach of Montiel Olea et al. (2021) to allow for the case where data for the reduced-form VAR are available over a longer horizon than for the IV. This constitutes the main methodological contribution of this chapter, as it shows that weak-identification robust inference is possible on proxy SVARs without having to balance the sample, and hence not make full use of the available data. I applied this method to analyse the effect of US monetary policy shocks on real and financial variables in the US, Canada, Sweden, and the UK. While the effects on US real and financial variables are qualitatively

similar to those reported in Gertler and Karadi (2015), there is substantially more muted evidence for US monetary spillovers in countries other than the US. This constitutes the substantive contribution of this chapter, as it shows that IV-based evidence against the Mundellian Trilemma may in fact be an artefact of weak IVs.

6
Conclusion

This thesis developed new econometric techniques to (jointly) address the problem of weak identification and the problem of (very) many IVs. This thesis then used these (and other existing) techniques to contribute to different strands of the empirical Macroeconomics literature.

Chapter 2 introduced the ridge-regularised jackknifed Anderson and Rubin (1949) (RJAR) test. The RJAR test provides an improvement over similar tests currently available in the literature, by remaining valid under weaker conditions, and allowing for potentially more IVs than observations. Chapter 3 proposed a Sup Score test that remains valid under dependent data, while continuing to be robust to arbitrarily weak identification and very many IVs. This Sup Score is one of the first weak-identification robust inference procedures that remains valid under very many IVs and time-series data. Chapter 3 further showed the danger of ad-hoc IV selection based on a Monte-Carlo simulation exercise. If IVs are selected based on in-sample correlations in the data, they may not be valid conditional on selection (even if they were valid to begin with). Chapter 3 also applied the Sup Score to conduct inference on a US New Keynesian Phillips Curve (NKPC) using 361 IVs on a sample of 179 observations. Substantially wider confidence sets were found than those commonly reported in the literature. In particular, no substantial role of short-term inflation expectations was found in explaining

inflation dynamics today. Chapter 4 used the recently proposed estimator of Kiviet (2022) to provide a new approach to conduct inference on the coefficient of the slack variable, by restricting the sign and magnitude of the correlation coefficient between the slack variable and the error term of a Euro Area wage Phillips Curve based on macroeconomic intuition. In particular, by restricting attention to positive values of the parameter space of the correlation coefficient between the slack variable and the error term, the coefficient on the slack variable was found to be significantly negative. Chapter 5 provided a weak-identification robust IV method for identifying the structural covariance matrix in 'unbalanced' structural vector autoregressions (SVARs), where data for the reduced-form VAR are available over a longer horizon than for the IV. Chapter 5 subsequently used this method to test for spillovers of US monetary policy in countries that have a floating exchange rate. In line with the predictions of the Mundellian Trilemma, and contrarily to recent approaches that did not use weak-identification robust methods, no significant spillovers of US monetary policy were found.

It is hoped that the methods developed in this thesis will prove useful for economists and other social scientists when they are confronted with any combination of weak, (very) many, and time-series IVs. Additionally, the substantive results and outcomes of this thesis will hopefully contribute to our understanding of important issues in Macroeconomics.

Appendices

A

Appendix for Chapter 2

A.1 Proofs

Throughout this appendix, $C > 0$ denotes a universal constant that can change across lines. CSHNW refers to Chao et al. (2012) from where we also borrow our summation conventions. Since all the proofs in this appendix are under the null hypothesis, we write ε_i and $\hat{\Phi}_{\gamma_n}$ for $e_i(\beta_0)$ and $\hat{\Phi}_{\gamma_n}$, respectively. We also define $\Phi_{\gamma_n} := \frac{2}{r_n}\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^2\mathbb{E}[\varepsilon_i^2]\mathbb{E}[\varepsilon_j^2]$,

The following singular value decomposition of the $n \times k_n$ matrix of instruments Z of rank r_n will be used frequently (see, e.g., Lütkepohl (1996, p. 60)):

$$Z = USQ',$$

where U is an $n \times n$ matrix such that $U'U = UU' = I_n$, Q is a $k_n \times k_n$ matrix such that $Q'Q = QQ' = I_{k_n}$, and S is the $n \times k_n$ matrix given by

$$S = \begin{bmatrix} D & 0_{r_n\times(k_n-r_n)} \\ 0_{(n-r_n)\times r_n} & 0_{(n-r_n)\times(k-r_n)} \end{bmatrix},$$

where D is the diagonal $r_n \times r_n$ matrix containing the singular values of Z. Hence, one can write

$$P^{\gamma_n} = USQ'(QS'SQ' + Q\gamma_n I_{k_n}Q')^{-1}QS'U' = US(S'S + \gamma_n I_{k_n})^{-1}S'U' = U\tilde{D}U', \tag{A.1}$$

where $\tilde{D} = S(S'S + \gamma_n I_{k_n})^{-1}S'$ is the diagonal $n \times n$ matrix with diagonal entries given by given by $\tilde{D}_{ll} = \frac{D_{ll}^2}{D_{ll}^2+\gamma_n} \leq 1$ for $l = 1, \ldots, r_n$, and zero otherwise. Note that the diagonal entries of $\tilde{D}$ are also the the eigenvalues of P^{γ_n}.

A.1.1 Lemmas

The following lemma collects some properties of P^{γ_n}. Recall that Z has rank r_n.

Lemma 1. Fix $n \geq 3$. For all $h, m = 1, \ldots, n$ and $\gamma_n \geq 0$ if $r_n = k_n$ and $\gamma_n > 0$ if $r_n < k_n$ one has

(i) $0 \leq (P^{\gamma_n})_{hh}^j \leq P_{hh}^{\gamma_n} \leq 1$ for all positive integers j,

(ii) $\sum_{i=1}^{n}(P_{hi}^{\gamma_n})^2 = (P^{\gamma_n})_{hh}^2 \leq P_{hh}^{\gamma_n}$,

(iii) $\sum_{i=1}^{n} P_{ii}^{\gamma_n} = \sum_{l=1}^{r_n} \frac{D_{ll}^2}{D_{ll}^2+\gamma_n} \leq r_n$,

(iv) $\sum_{i=1}^{n} (P^{\gamma_n})_{ii}^2 = \sum_{l=1}^{r_n} \frac{D_{ll}^4}{\left(D_{ll}^2+\gamma_n\right)^2} \leq r_n$,

(v) $|P_{hm}^{\gamma_n}| \leq 1$,

(vi) For any $\mathcal{I}_2 \subseteq \{1,\ldots,n\}^2$ and any $\mathcal{I}_3 \subseteq \{1,\ldots,n\}^3$,

(a) $\sum_{\mathcal{I}_2} (P_{ij}^{\gamma_n})^4 \leq r_n$,

(b) $\sum_{\mathcal{I}_3} (P_{ij}^{\gamma_n})^2 (P_{jl}^{\gamma_n})^2 \leq r_n$.

Proof. By the arguments prior to this lemma, the non-zero eigenvalues of P^{γ_n} are $\tilde{D}_{ll} = \frac{D_{ll}^2}{D_{ll}^2+\gamma_n}$ for $l = 1,\ldots,r_n$. (iii) now follows from the trace of P^{γ_n} equaling the sum of its non-zero eigenvalues. Furthermore, the eigenvalues of $(P^{\gamma_n})^2$ are given by $\tilde{D}_{ll}^2 = \frac{D_{ll}^4}{(D_{ll}^2+\gamma_n)^2}$ for $l = 1,\ldots,r_n$, and zero otherwise. (iv) now follows in the same way as (iii).

By Equation (A.1) one has that

$$(P^{\gamma_n})_{hh}^j = \sum_{l=1}^{n} \tilde{D}_{ll}^j U_{hl}^2 \leq \sum_{l=1}^{n} \tilde{D}_{ll} U_{hl}^2 = P_{hh}^{\gamma_n} \leq 1, \tag{A.2}$$

which verifies (i) since $0 \leq \tilde{D}_{ll} \leq 1$; the last inequality following from the largest diagonal entry of P^{γ_n} being bounded from above by its largest eigenvalue, which is no greater than one. Next, by (i),

$$\sum_{i=1}^{n} (P_{hi}^{\gamma_n})^2 = (P^{\gamma_n})_{hh}^2 \leq P_{hh}^{\gamma_n} \leq 1,$$

such that (ii) and (v) follow. Furthermore, by (v) and (iv),

$$\sum_{\mathcal{I}_2} (P_{ij}^{\gamma_n})^4 \leq \sum_{i=1}^{n}\sum_{j=1}^{n} (P_{ij}^{\gamma_n})^2 = \sum_{i=1}^{n} (P^{\gamma_n})_{ii}^2 \leq r_n,$$

and (vi) follows since

$$\sum_{\mathcal{I}_3} (P_{ij}^{\gamma_n})^2 (P_{jl}^{\gamma_n})^2 \leq \sum_{j=1}^{n}\sum_{i=1}^{n} (P_{ij}^{\gamma_n})^2 \sum_{l=1}^{n} (P_{jl}^{\gamma_n})^2 = \sum_{j=1}^{n} \left((P^{\gamma_n})_{jj}^2\right)^2 \leq \sum_{j=1}^{n} (P^{\gamma_n})_{jj}^2 \leq r_n,$$

the penultimate inequality being a consequence of (i) and the last of (iv).

□

Lemma 2. Fix $n \geq 4$. For all $\gamma_n \geq 0$ if $r_n = k_n$ and $\gamma_n > 0$ if $r_n < k_n$ one has

$$\left| \sum_{i<j<l<m} P_{il}^{\gamma_n} P_{jl}^{\gamma_n} P_{im}^{\gamma_n} P_{jm}^{\gamma_n} \right| \leq C r_n.$$

Proof. The proof follows Lemma B2 in CSHNW closely. For $G := \mathrm{diag}(P_{11}^{\gamma_n}, \ldots, P_{nn}^{\gamma_n})$ observe that $(P^{\gamma_n} - G)^4$ equals

$$\begin{aligned}
&(P^{\gamma_n})^4 - (P^{\gamma_n})^3 G - (P^{\gamma_n})^2 G P^{\gamma_n} + (P^{\gamma_n})^2 G^2 - P^{\gamma_n} G (P^{\gamma_n})^2 + P^{\gamma_n} G P^{\gamma_n} G \\
&+ P^{\gamma_n} G^2 P^{\gamma_n} - P^{\gamma_n} G^3 - G(P^{\gamma_n})^3 + G(P^{\gamma_n})^2 G + G P^{\gamma_n} G P^{\gamma_n} \\
&- G P^{\gamma_n} G^2 + G^2 (P^{\gamma_n})^2 - G^2 P^{\gamma_n} G - G^3 P^{\gamma_n} + G^4.
\end{aligned}$$

Since $\mathrm{tr}(A') = \mathrm{tr}(A)$ and $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ for square matrices A and B,

$$\begin{aligned}
\mathrm{tr}((P^{\gamma_n} - G)^4) =& \mathrm{tr}((P^{\gamma_n})^4) - 4\mathrm{tr}((P^{\gamma_n})^3 G) + 4\mathrm{tr}((P^{\gamma_n})^2 G^2) + 2\mathrm{tr}(P^{\gamma_n} G P^{\gamma_n} G) \\
&- 4\mathrm{tr}(P^{\gamma_n} G^3) + \mathrm{tr}(G^4) \\
\leq& \mathrm{tr}((P^{\gamma_n})^4) + 4|\mathrm{tr}((P^{\gamma_n})^3 G)| + 4\mathrm{tr}((P^{\gamma_n})^2 G^2) + 2\mathrm{tr}(P^{\gamma_n} G P^{\gamma_n} G) \\
&+ 4|\mathrm{tr}(P^{\gamma_n} G^3)| + \mathrm{tr}(G^4).
\end{aligned}$$

Next, because $0 \leq (P^{\gamma_n})_{ii}^j \leq P_{ii}^{\gamma_n}$ for all positive integers j by Lemma 1 (i), and $G_{ii} \geq 0$, one gets

$$\begin{aligned}
\mathrm{tr}((P^{\gamma_n} - G)^4) \leq& \mathrm{tr}(P^{\gamma_n}) + 4\mathrm{tr}(P^{\gamma_n} G) + 4\mathrm{tr}(P^{\gamma_n} G^2) + 2\mathrm{tr}(P^{\gamma_n} G P^{\gamma_n} G) \\
&+ 4\mathrm{tr}(P^{\gamma_n} G^3) + \mathrm{tr}(G^4).
\end{aligned}$$

Since $0 \leq P_{ii}^{\gamma_n} \leq 1$ and $G^j \leq I_n$ (elementwise) for all positive integers j, $\mathrm{tr}(P^{\gamma_n} G^j) \leq \mathrm{tr}(P^{\gamma_n}) \leq r_n$, $\mathrm{tr}(P^{\gamma_n} G P^{\gamma_n} G) \leq \mathrm{tr}((P^{\gamma_n})^2) \leq r_n$. Furthermore, $\mathrm{tr}(G^4) \leq \mathrm{tr}(G) = \mathrm{tr}(P^{\gamma_n}) \leq r_n$. Hence,

$$\mathrm{tr}((P^{\gamma_n} - G)^4) \leq 16 r_n. \tag{A.3}$$

As in the proof of Lemma B2 in CSHNW, define the lower-triangular matrix L with entries $L_{ij} = P_{ij}^{\gamma_n} \mathbf{1}_{i>j}$, so that $P^{\gamma_n} = L + L' + G$. Then

$$\begin{aligned}
(P^{\gamma_n} - G)^4 =& (L + L')^4 \\
=& L^4 + L^2 LL' + L^2 L' L + L^2 L'^2 + LL' L^2 + LL' LL'
\end{aligned}$$

$$+ LL'L'L + LL'^3 + L'LL^2 + L'LLL' + L'LL'L + L'LL'^2$$
$$+ L'^2L^2 + L'^2LL' + L'^2L'L + L'^4.$$

Note for all positive integers j, $[(L')^j]' = L^j$. Since $\text{tr}(A') = \text{tr}(A)$ and $\text{tr}(AB) = \text{tr}(BA)$ for any square matrices A and B,

$$\text{tr}((P^{\gamma_n} - G)^4) = 2\text{tr}(L^4) + 8\text{tr}(L^3L') + 4\text{tr}(L^2L'^2) + 2\text{tr}(L'LL'L). \tag{A.4}$$

We consider each of the terms on the right-hand side of Equation (A.4) separately. Before proceeding, note that:

$$\begin{aligned}
(L)^2_{ab} &= \sum_{l=1}^{n} L_{al}L_{lb}, \\
(L')^2_{ab} &= \sum_{l=1}^{n} L_{la}L_{bl} = \sum_{l=1}^{n} L_{bl}L_{la} = (L)^2_{ba}, \\
(L'L)_{ab} &= \sum_{l=1}^{n} L_{la}L_{lb}, \\
(L)^3_{ab} &= \sum_{l=1}^{n} (L)^2_{al}L_{lb} = \sum_{l=1}^{n}\sum_{m=1}^{n} L_{am}L_{ml}L_{lb}, \\
(L)^4_{ab} &= \sum_{l=1}^{n} (L)^2_{al}(L)^2_{lb} = \sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{j=1}^{n} L_{am}L_{ml}L_{lj}L_{jb}, \\
(L^3L')_{ab} &= \sum_{l=1}^{n} (L)^3_{al}(L')_{lb} = \sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{j=1}^{n} L_{am}L_{mj}L_{jl}L_{bl}, \\
(L^2L'^2)_{ab} &= \sum_{l=1}^{n} (L)^2_{al}(L')^2_{lb} = \sum_{l=1}^{n} (L)^2_{al}(L)^2_{bl} = \sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{j=1}^{n} L_{am}L_{ml}L_{bj}L_{jl}, \\
(L'LL'L)_{ab} &= \sum_{l=1}^{n} (L'L)_{al}(L'L)_{lb} = \sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{j=1}^{n} L_{ma}L_{ml}L_{jl}L_{jb}.
\end{aligned} \tag{A.5}$$

Using the expression for $(L)^4_{ab}$ in Equation (A.5),

$$\begin{aligned}
\text{tr}(L^4) &= \sum_{i=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{j=1}^{n} L_{im}L_{ml}L_{lj}L_{ji} = \sum_{i,j,l,m} P^{\gamma_n}_{ji}\mathbf{1}_{j>i}P^{\gamma_n}_{im}\mathbf{1}_{i>m}P^{\gamma_n}_{ml}\mathbf{1}_{m>l}P^{\gamma_n}_{lj}\mathbf{1}_{l>j} \\
&= 0,
\end{aligned} \tag{A.6}$$

since there is no combination of indices i, j, l, m that jointly satisfy each of the indicator functions, as this would require $m < i < j < l < m$.

Using the expression for $(L^3L')_{ab}$ in Equation (A.5),

$$\begin{aligned}\operatorname{tr}(L^3L') &= \sum_{i=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{j=1}^{n} L_{im}L_{mj}L_{jl}L_{il} = \sum_{i,j,l,m} P_{im}^{\gamma_n}\mathbf{1}_{i>m}P_{mj}^{\gamma_n}\mathbf{1}_{m>j}P_{jl}^{\gamma_n}\mathbf{1}_{j>l}P_{il}^{\gamma_n}\mathbf{1}_{i>l} \\ &= \sum_{l<j<m<i} P_{im}^{\gamma_n}P_{mj}^{\gamma_n}P_{jl}^{\gamma_n}P_{il}^{\gamma_n} = \sum_{i<j<l<m} P_{ml}^{\gamma_n}P_{lj}^{\gamma_n}P_{ji}^{\gamma_n}P_{mi}^{\gamma_n} \\ &= \sum_{i<j<l<m} P_{ij}^{\gamma_n}P_{jl}^{\gamma_n}P_{im}^{\gamma_n}P_{lm}^{\gamma_n}.\end{aligned} \tag{A.7}$$

The third equality follows since the product within the summations is non-zero only for $l < j < m < i$. The fourth equality follows by replacing the indices according to the dictionary $\{i : m, j : j, l : i, m : l\}$. The last equality follows from the symmetry of P^{γ_n}.

Using the expression for $(L^2L'^2)_{ab}$ in Equation (A.5),

$$\begin{aligned}\operatorname{tr}(L^2L'^2) &= \sum_{i=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{j=1}^{n} L_{im}L_{ml}L_{ij}L_{jl} \\ &= \sum_{i,j,l,m} P_{im}^{\gamma_n}\mathbf{1}_{i>m}P_{ml}^{\gamma_n}\mathbf{1}_{m>l}P_{ij}^{\gamma_n}\mathbf{1}_{i>j}P_{jl}^{\gamma_n}\mathbf{1}_{j>l} \\ &= \sum_{i>j>l,i>m>l} P_{im}^{\gamma_n}P_{ml}^{\gamma_n}P_{ij}^{\gamma_n}P_{jl}^{\gamma_n} = \sum_{i>j>l,i>m>l} P_{ij}^{\gamma_n}P_{jl}^{\gamma_n}P_{lm}^{\gamma_n}P_{mi}^{\gamma_n} \\ &= \sum_{i>j=m>l} P_{ij}^{\gamma_n}P_{jl}^{\gamma_n}P_{lm}^{\gamma_n}P_{mi}^{\gamma_n} + \sum_{i>j>m>l} P_{ij}^{\gamma_n}P_{jl}^{\gamma_n}P_{lm}^{\gamma_n}P_{mi}^{\gamma_n} \\ &\quad + \sum_{i>m>j>l} P_{ij}^{\gamma_n}P_{jl}^{\gamma_n}P_{lm}^{\gamma_n}P_{mi}^{\gamma_n} \\ &= \sum_{i>j>l} P_{ij}^{\gamma_n}P_{jl}^{\gamma_n}P_{lj}^{\gamma_n}P_{ji}^{\gamma_n} + \sum_{i<j<l<m} P_{ml}^{\gamma_n}P_{li}^{\gamma_n}P_{ij}^{\gamma_n}P_{jm}^{\gamma_n} \\ &\quad + \sum_{i>m>j>l} P_{ij}^{\gamma_n}P_{jl}^{\gamma_n}P_{lm}^{\gamma_n}P_{mi}^{\gamma_n} \\ &= \sum_{i>j>l} P_{ij}^{\gamma_n}P_{jl}^{\gamma_n}P_{lj}^{\gamma_n}P_{ji}^{\gamma_n} + \sum_{i<j<l<m} P_{ml}^{\gamma_n}P_{li}^{\gamma_n}P_{ij}^{\gamma_n}P_{jm}^{\gamma_n} \\ &\quad + \sum_{i<j<l<m} P_{mj}^{\gamma_n}P_{ji}^{\gamma_n}P_{il}^{\gamma_n}P_{lm}^{\gamma_n} \\ &= \sum_{i<j<l} (P_{ij}^{\gamma_n})^2(P_{jl}^{\gamma_n})^2 + 2\sum_{i<j<l<m} P_{ij}^{\gamma_n}P_{il}^{\gamma_n}P_{jm}^{\gamma_n}P_{lm}^{\gamma_n}.\end{aligned} \tag{A.8}$$

The third equality follows since the product within the summations is non-zero only when both $i > j > l$ and $i > m > l$. The fourth equality follows from the symmetry of P^{γ_n}. The sixth equality follows by replacing the indices in the second summation according to the dictionary $\{i : m, j : l, l : i, m : j\}$. The seventh equality follows by replacing the indices in the third summation according to the

dictionary $\{i:m, j:j, l:i, m:l\}$. The last equality follows from the symmetry of P^{γ_n}.

Using the expression for $(L'LL'L)_{ab}$ in Equation (A.5),

$$
\begin{aligned}
\operatorname{tr}(L'LL'L) =& \sum_{i=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{j=1}^{n} L_{mi}L_{ml}L_{jl}L_{ji} \\
=& \sum_{i,j,l,m} P^{\gamma_n}_{mi}\mathbf{1}_{m>i}P^{\gamma_n}_{ml}\mathbf{1}_{m>l}P^{\gamma_n}_{jl}\mathbf{1}_{j>l}P^{\gamma_n}_{ji}\mathbf{1}_{j>i} \\
=& \sum_{i,j,l,m} P^{\gamma_n}_{ij}\mathbf{1}_{i>j}P^{\gamma_n}_{l>j}\mathbf{1}_{l>j}P^{\gamma_n}_{lm}\mathbf{1}_{l>m}P^{\gamma_n}_{im}\mathbf{1}_{i>m} \\
=& \sum_{m=j<i=l} P^{\gamma_n}_{ij}P^{\gamma_n}_{lj}P^{\gamma_n}_{lm}P^{\gamma_n}_{im} + \sum_{m=j<l<i} P^{\gamma_n}_{ij}P^{\gamma_n}_{lj}P^{\gamma_n}_{lm}P^{\gamma_n}_{im} \\
&+ \sum_{m=j<i<l} P^{\gamma_n}_{ij}P^{\gamma_n}_{lj}P^{\gamma_n}_{lm}P^{\gamma_n}_{im} + \sum_{j<m<i=l} P^{\gamma_n}_{ij}P^{\gamma_n}_{lj}P^{\gamma_n}_{lm}P^{\gamma_n}_{im} \\
&+ \sum_{m<j<i=l} P^{\gamma_n}_{ij}P^{\gamma_n}_{lj}P^{\gamma_n}_{lm}P^{\gamma_n}_{im} + \sum_{m<j<l<i} P^{\gamma_n}_{ij}P^{\gamma_n}_{lj}P^{\gamma_n}_{lm}P^{\gamma_n}_{im} \\
&+ \sum_{m<j<i<l} P^{\gamma_n}_{ij}P^{\gamma_n}_{lj}P^{\gamma_n}_{lm}P^{\gamma_n}_{im} + \sum_{j<m<l<i} P^{\gamma_n}_{ij}P^{\gamma_n}_{lj}P^{\gamma_n}_{lm}P^{\gamma_n}_{im} \\
&+ \sum_{j<m<i<l} P^{\gamma_n}_{ij}P^{\gamma_n}_{lj}P^{\gamma_n}_{lm}P^{\gamma_n}_{im} \qquad \text{(A.9)} \\
=& \sum_{j<i}(P^{\gamma_n}_{ij})^4 + \sum_{j<l<i}(P^{\gamma_n}_{ij})^2(P^{\gamma_n}_{lj})^2 + \sum_{j<i<l}(P^{\gamma_n}_{ij})^2(P^{\gamma_n}_{lj})^2 \\
&+ \sum_{j<m<i}(P^{\gamma_n}_{ij})^2(P^{\gamma_n}_{im})^2 + \sum_{m<j<i}(P^{\gamma_n}_{ij})^2(P^{\gamma_n}_{im})^2 \\
&+ 4\sum_{i<j<l<m} P^{\gamma_n}_{li}P^{\gamma_n}_{mi}P^{\gamma_n}_{lj}P^{\gamma_n}_{mj} \\
=& \sum_{j<i}(P^{\gamma_n}_{ij})^4 + 2\sum_{i<j<l}(P^{\gamma_n}_{li})^2(P^{\gamma_n}_{ji})^2 \\
&+ \sum_{i<j<l}(P^{\gamma_n}_{li})^2(P^{\gamma_n}_{lj})^2 + 4\sum_{i<j<l<m} P^{\gamma_n}_{li}P^{\gamma_n}_{mi}P^{\gamma_n}_{lj}P^{\gamma_n}_{mj} \\
=& \sum_{i<j}(P^{\gamma_n}_{ij})^4 + 2\sum_{i<j<l}\left((P^{\gamma_n}_{ij})^2(P^{\gamma_n}_{il})^2 + (P^{\gamma_n}_{il})^2(P^{\gamma_n}_{jl})^2\right) \\
&+ 4\sum_{i<j<l<m} P^{\gamma_n}_{il}P^{\gamma_n}_{jl}P^{\gamma_n}_{im}P^{\gamma_n}_{jm}.
\end{aligned}
$$

The third equality follows by replacing the indices in the summation according to the dictionary $\{i:j, j:i, l:m, m:l\}$. The fourth equality follows since the product within the summation is non-zero only when both $i > j, m$ and $l > j, m$. The fifth equality follows from replacing the indices in the last four summations in the previous line according to the dictionaries $\{i:m, j:j, l:l, m:i\}$, $\{i:l, j:j, l:m, m:i\}$, $\{i:m, j:i, l:l, m:j\}$, $\{i:l, j:i, l:m, m:j\}$, respectively. The sixth equality

follows from replacing the indices in the second, third, fourth and fifth summations in the previous line according to the dictionaries $\{i:l, j:i, l:j\}$, $\{i:j, j:i, l:l\}$, $\{i:l, j:i, m:j\}$, $\{i:l, j:j, m:i\}$, respectively. The last equality follows by the symmetry of P^{γ_n}.

Let $\mathfrak{S} \coloneqq \sum_{i<j<l<m} P^{\gamma_n}_{il}P^{\gamma_n}_{jl}P^{\gamma_n}_{im}P^{\gamma_n}_{jm} + P^{\gamma_n}_{ij}P^{\gamma_n}_{jl}P^{\gamma_n}_{im}P^{\gamma_n}_{lm} + P^{\gamma_n}_{ij}P^{\gamma_n}_{il}P^{\gamma_n}_{jm}P^{\gamma_n}_{lm}$. Substituting the expressions in Equations (A.6)–(A.9) into Equation (A.4) yields

$$\begin{aligned}
\operatorname{tr}((P^{\gamma_n}-G)^4) =& 8\sum_{i<j<l<m} P^{\gamma_n}_{ij}P^{\gamma_n}_{jl}P^{\gamma_n}_{im}P^{\gamma_n}_{lm} \\
&+ 4\left(\sum_{i<j<l}(P^{\gamma_n}_{ij})^2(P^{\gamma_n}_{jl})^2 + 2\sum_{i<j<l<m} P^{\gamma_n}_{ij}P^{\gamma_n}_{il}P^{\gamma_n}_{jm}P^{\gamma_n}_{lm}\right) \\
&+ 2\left(\sum_{i<j}(P^{\gamma_n}_{ij})^4 + 2\sum_{i<j<l}\left((P^{\gamma_n}_{ij})^2(P^{\gamma_n}_{il})^2 + (P^{\gamma_n}_{il})^2(P^{\gamma_n}_{jl})^2\right)\right. \\
&\qquad \left. + 4\sum_{i<j<l<m} P^{\gamma_n}_{il}P^{\gamma_n}_{jl}P^{\gamma_n}_{im}P^{\gamma_n}_{jm}\right) \\
=& 2\sum_{i<j}(P^{\gamma_n}_{ij})^4 \\
&+ 4\sum_{i<j<l}\left((P^{\gamma_n}_{ij})^2(P^{\gamma_n}_{jl})^2 + (P^{\gamma_n}_{il})^2(P^{\gamma_n}_{jl})^2 + (P^{\gamma_n}_{ij})^2(P^{\gamma_n}_{il})^2\right) \\
&+ 8\mathfrak{S}.
\end{aligned}$$

Next, by the triangle inequality and Lemma 1 (vi),

$$\begin{aligned}
|\mathfrak{S}| \le& \frac{1}{4}\sum_{i<j}(P^{\gamma_n}_{ij})^4 \\
&+ \frac{1}{2}\sum_{i<j<l}\left((P^{\gamma_n}_{ij})^2(P^{\gamma_n}_{jl})^2 + (P^{\gamma_n}_{il})^2(P^{\gamma_n}_{jl})^2 + (P^{\gamma_n}_{ij})^2(P^{\gamma_n}_{il})^2\right) \\
&+ \frac{1}{8}\operatorname{tr}((P^{\gamma_n}-G)^4) \\
\le& Cr_n.
\end{aligned} \tag{A.10}$$

Take $\{u_i\}$ to be a sequence of i.i.d. mean-zero and unit variance random variables. Define

$$\begin{aligned}
\Delta_1 &\coloneqq \sum_{i<j<l}\left(P^{\gamma_n}_{ij}P^{\gamma_n}_{il}u_ju_l + P^{\gamma_n}_{ij}P^{\gamma_n}_{jl}u_iu_l + P^{\gamma_n}_{il}P^{\gamma_n}_{jl}u_iu_j\right), \\
\Delta_2 &\coloneqq \sum_{i<j<l}\left(P^{\gamma_n}_{ij}P^{\gamma_n}_{il}u_ju_l + P^{\gamma_n}_{ij}P^{\gamma_n}_{jl}u_iu_l\right), \Delta_3 \coloneqq \sum_{i<j<l}\left(P^{\gamma_n}_{il}P^{\gamma_n}_{jl}u_iu_j\right).
\end{aligned}$$

Then by Lemma 1 (vi),

$$\begin{aligned}\mathbb{E}[\Delta_3^2] &= \sum_{i<j<l} (P_{il}^{\gamma_n})^2 (P_{jl}^{\gamma_n})^2 + 2 \sum_{i<j<l<m} P_{il}^{\gamma_n} P_{jl}^{\gamma_n} P_{im}^{\gamma_n} P_{jm}^{\gamma_n} \\ &\leq r_n + 2 \sum_{i<j<l<m} P_{il}^{\gamma_n} P_{jl}^{\gamma_n} P_{im}^{\gamma_n} P_{jm}^{\gamma_n}.\end{aligned}$$

Furthermore,

$$\mathbb{E}[\Delta_2 \Delta_3] = \sum_{i<j<l<m} P_{ij}^{\gamma_n} P_{il}^{\gamma_n} P_{jm}^{\gamma_n} P_{lm}^{\gamma_n} + \sum_{i<j<l<m} P_{ij}^{\gamma_n} P_{jl}^{\gamma_n} P_{im}^{\gamma_n} P_{lm}^{\gamma_n},$$

and

$$\begin{aligned}\mathbb{E}[\Delta_2^2] = &\sum_{\{i,l\}<j<k} P_{ij}^{\gamma_n} P_{il}^{\gamma_n} P_{mj}^{\gamma_n} P_{ml}^{\gamma_n} + \sum_{i<\{j,l\}<m} P_{ij}^{\gamma_n} P_{jm}^{\gamma_n} P_{il}^{\gamma_n} P_{lm}^{\gamma_n} \\ &+ \sum_{i<j<l<m} P_{ij}^{\gamma_n} P_{im}^{\gamma_n} P_{jl}^{\gamma_n} P_{lm}^{\gamma_n} + \sum_{l<i<j<m} P_{ij}^{\gamma_n} P_{jm}^{\gamma_n} P_{li}^{\gamma_n} P_{lm}^{\gamma_n} \\ = &\sum_{i<j<l} (P_{ij}^{\gamma_n})^2 (P_{il}^{\gamma_n})^2 + \sum_{i<j<l} (P_{ij}^{\gamma_n})^2 (P_{jl}^{\gamma_n})^2 + 2 \sum_{i<m<j<l} P_{ij}^{\gamma_n} P_{il}^{\gamma_n} P_{mj}^{\gamma_n} P_{ml}^{\gamma_n} \\ &+ 2 \sum_{i<j<l<m} P_{ij}^{\gamma_n} P_{jm}^{\gamma_n} P_{il}^{\gamma_n} P_{lm}^{\gamma_n} \\ &+ \sum_{i<j<l<m} P_{ij}^{\gamma_n} P_{im}^{\gamma_n} P_{jl}^{\gamma_n} P_{lm}^{\gamma_n} + \sum_{i<j<l<m} P_{jl}^{\gamma_n} P_{lm}^{\gamma_n} P_{ij}^{\gamma_n} P_{im}^{\gamma_n} \\ = &\sum_{i<j<l} (P_{ij}^{\gamma_n})^2 (P_{il}^{\gamma_n})^2 + \sum_{i<j<l} (P_{ij}^{\gamma_n})^2 (P_{jl}^{\gamma_n})^2 + 2\mathfrak{S} \\ &\leq C r_n,\end{aligned}$$

where the last inequality follows from Lemma 1 (vi) and Equation (A.10). Since $\Delta_1 = \Delta_2 + \Delta_3$, $\mathbb{E}[\Delta_1^2] = \mathbb{E}[\Delta_2^2] + \mathbb{E}[\Delta_3^2] + 2\mathbb{E}[\Delta_2 \Delta_3] \leq C r_n + 2\mathfrak{S} \leq C r_n$. Hence by the triangle inequality and the expression for $\mathbb{E}[\Delta_3^2]$,

$$\begin{aligned}\left| \sum_{i<j<l<m} P_{il}^{\gamma_n} P_{jl}^{\gamma_n} P_{im}^{\gamma_n} P_{jm}^{\gamma_n} \right| &\leq C \left(\mathbb{E}[\Delta_3^2] + r_n \right) \leq C \left(\mathbb{E}[(\Delta_1 - \Delta_2)^2] + r_n \right) \\ &\leq C(\mathbb{E}[\Delta_1^2] + \mathbb{E}[\Delta_2^2] + r_n) \\ &\leq C r_n.\end{aligned}$$

□

Lemma 3. Let $\gamma_- > 0$ and $\Gamma_n = \Gamma(\gamma_-) \coloneqq \{\gamma_n \in \Re : \gamma_n \geq 0 \text{ if } r_n = k_n, \text{ and } \gamma_n \geq \gamma_- > 0 \text{ if } r_n < k_n\}$. Then if P^{γ_n} is a non-diagonal matrix,[1]

$$\gamma_n^* \coloneqq \max \arg\max_{\gamma_n \in \Gamma_n} \sum_{i=1}^{n} \sum_{j \neq i} (P_{ij}^{\gamma_n})^2 \in [0, \infty)$$

exists.

[1]We note that Assumption 3 excludes the case where P^{γ_n} is diagonal, so that γ_n^* exists under the assumptions made in this chapter.

Proof. Notice that

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n}(P_{ij}^{\gamma_n})^2 - \sum_{i=1}^{n}(P_{ii}^{\gamma_n})^2 \\ &= \sum_{i=1}^{n}(P^{\gamma_n})_{ii}^2 - \sum_{i=1}^{n}(P_{ii}^{\gamma_n})^2 \\ &= \sum_{l=1}^{r_n}\left(\frac{D_{ll}^2}{D_{ll}^2+\gamma_n}\right)^2 - \sum_{i=1}^{n}\left(\sum_{l=1}^{r_n}\frac{D_{ll}^2}{D_{ll}^2+\gamma_n}U_{il}^2\right)^2,\end{aligned}$$

where the third equality follows from Lemma 1 (iv) and Equation (A.2). Since

$$\lim_{\gamma_n\to\infty}\frac{D_{ll}^2}{D_{ll}^2+\gamma_n} = 0,$$

and $U_{il}^2 \leq 1$ for $l = 1,\dots,r_n$ and $i = 1,\dots,n$, it follows that

$$\lim_{\gamma_n\to\infty}\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^2 = 0.$$

Hence, the maximum is not attained for arbitrarily large γ_n. Since P^{γ_n} is a non-diagonal matrix by assumption, $\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^2$ is strictly positive. This leaves a compact set over which the non-zero $\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^2$ is maximised, such that a maximiser exists.

□

Lemma 4. Under Assumptions 1, 2 and the null hypothesis in Equation (2.2), $|\hat{\Phi}_{\gamma_n} - \Phi_{\gamma_n}| \xrightarrow{p} 0$.

Proof. Defining $\eta_i := \varepsilon_i^2 - \mathbb{E}[\varepsilon_i^2]$, one can write

$$\hat{\Phi}_{\gamma_n} - \Phi_{\gamma_n} = \frac{2}{r_n}\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^2\left(\eta_i\eta_j + \mathbb{E}[\varepsilon_j^2]\eta_i + \mathbb{E}[\varepsilon_i^2]\eta_j\right),$$

and it follows that

$$\begin{aligned}\left|\hat{\Phi}_{\gamma_n} - \Phi_{\gamma_n}\right| \leq& \frac{2}{r_n}\left|\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^2\eta_i\eta_j\right| + \frac{2}{r_n}\left|\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^2\mathbb{E}[\varepsilon_j^2]\eta_i\right| \\ &+ \frac{2}{r_n}\left|\sum_{i=1}^{n}\sum_{j\neq i}^{n}(P_{ij}^{\gamma_n})^2\mathbb{E}[\varepsilon_i^2]\eta_j\right| \\ \equiv& A_1 + A_2 + A_3.\end{aligned} \tag{A.11}$$

Consider each of $\mathbb{E}[A_1^2]$, $\mathbb{E}[A_2^2]$, and $\mathbb{E}[A_3^2]$ in turn.

$$\begin{aligned}
\mathbb{E}[A_1^2] =& \frac{4}{r_n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{h=1}^{n}\sum_{g\neq h}(P_{ij}^{\gamma_n})^2(P_{hg}^{\gamma_n})^2\mathbb{E}[\eta_i\eta_j\eta_h\eta_g] \\
=& \frac{4}{r_n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{h\neq i}\sum_{g\neq h}(P_{ij}^{\gamma_n})^2(P_{hg}^{\gamma_n})^2\mathbb{E}[\eta_i\eta_j\eta_h\eta_g] \\
&+ \frac{4}{r_n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{g\neq i}(P_{ij}^{\gamma_n})^2(P_{ig}^{\gamma_n})^2\mathbb{E}[\eta_i^2\eta_j\eta_g] \\
=& \frac{4}{r_n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{h\neq i}\sum_{g\notin\{h,i\}}(P_{ij}^{\gamma_n})^2(P_{hg}^{\gamma_n})^2\mathbb{E}[\eta_i]\mathbb{E}[\eta_j\eta_h\eta_g] \\
&+ \frac{4}{r_n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{h\neq i}(P_{ij}^{\gamma_n})^2(P_{ih}^{\gamma_n})^2\mathbb{E}[\eta_i^2\eta_j\eta_h] \\
&+ \frac{4}{r_n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{g\neq i}(P_{ij}^{\gamma_n})^2(P_{ig}^{\gamma_n})^2\mathbb{E}[\eta_i^2\eta_j\eta_g] \\
=& \frac{8}{r_n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{h\neq i}(P_{ij}^{\gamma_n})^2(P_{ih}^{\gamma_n})^2\mathbb{E}[\eta_i^2\eta_j\eta_h] \\
=& \frac{8}{r_n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{h\notin\{i,j\}}(P_{ij}^{\gamma_n})^2(P_{ih}^{\gamma_n})^2\mathbb{E}[\eta_i^2]\mathbb{E}[\eta_j]\mathbb{E}[\eta_h] \\
&+ \frac{8}{r_n^2}\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^4\mathbb{E}[\eta_i^2]\mathbb{E}[\eta_j^2] \\
=& \frac{8}{r_n^2}\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^4\mathbb{E}[\eta_i^2]\mathbb{E}[\eta_j^2] \\
\leq& \frac{C}{r_n^2}\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^4 \\
\leq& \frac{C}{r_n},
\end{aligned}$$

The fourth equality follows from $\mathbb{E}[\eta_i] = 0$ and the symmetry of P^{γ_n}. The sixth equality follows from $\mathbb{E}[\eta_j] = 0$. The first inequality follows from Assumption 1, which implies $\sup_{i\in\mathbb{N}}\mathbb{E}[\eta_i^2] < \infty$. The second inequality follows from Lemma 1 (vi).

Next, concerning $\mathbb{E}[A_2^2]$ one has

$$\begin{aligned}
\mathbb{E}[A_2^2] =& \frac{4}{r_n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{h=1}^{n}\sum_{g\neq h}(P_{ij}^{\gamma_n})^2(P_{hg}^{\gamma_n})^2\mathbb{E}[\varepsilon_j^2]\mathbb{E}[\varepsilon_g^2]\mathbb{E}[\eta_i\eta_h] \\
=& \frac{4}{r_n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{h\neq i}\sum_{g\neq h}(P_{ij}^{\gamma_n})^2(P_{hg}^{\gamma_n})^2\mathbb{E}[\varepsilon_j^2]\mathbb{E}[\varepsilon_g^2]\mathbb{E}[\eta_i]\mathbb{E}[\eta_h] \\
&+\frac{4}{r_n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{g\neq i}(P_{ij}^{\gamma_n})^2(P_{ig}^{\gamma_n})^2\mathbb{E}[\varepsilon_j^2]\mathbb{E}[\varepsilon_g^2]\mathbb{E}[\eta_i^2] \\
=& \frac{4}{r_n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{g\neq i}(P_{ij}^{\gamma_n})^2(P_{ig}^{\gamma_n})^2\mathbb{E}[\varepsilon_j^2]\mathbb{E}[\varepsilon_g^2]\mathbb{E}[\eta_i^2] \\
\leq& \frac{C}{r_n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{g\neq i}(P_{ij}^{\gamma_n})^2(P_{ig}^{\gamma_n})^2 \\
\leq& \frac{C}{r_n},
\end{aligned}$$

by the same reasoning that led to the penultimate display. Finally, similar arguments imply that $\mathbb{E}[A_3^2] \leq \frac{C}{r_n}$ such that by Markov's inequality and Equation (A.11) it follows that $\hat{\Phi}_{\gamma_n} - \Phi_{\gamma_n} = O_p(r_n^{-1/2})$. Thus, by Assumption 2, $|\hat{\Phi}_{\gamma_n} - \Phi_{\gamma_n}| \xrightarrow{p} 0$. □

A.1.2 Proof of Proposition 1

By the assumptions in the statement, P has rank k_n and $P_{ii} \leq 1 - \delta$, which implies that $k_n < n$. Notice that

$$\begin{aligned}
\frac{1}{k_n}\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij})^2 &= \frac{1}{k_n}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}(P_{ij})^2 - \sum_{i=1}^{n}(P_{ii})^2\right) = \frac{1}{k_n}\left(\sum_{i=1}^{n}(P^2)_{ii} - \sum_{i=1}^{n}(P_{ii})^2\right) \\
&= \frac{1}{k_n}\left(\sum_{i=1}^{n}P_{ii} - \sum_{i=1}^{n}(P_{ii})^2\right) \geq \frac{1}{k_n}\left(\sum_{i=1}^{n}P_{ii} - (1-\delta)\sum_{i=1}^{n}P_{ii}\right) = \delta,
\end{aligned}$$

where the third equality follows from the idempotency of P, the first inequality follows since $\max_i P_{ii} \leq 1 - \delta$ by assumption of the statement, and the last equality follows since $\mathrm{tr}(P) = k_n$.

Next, observe that

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^2 =& \sum_{i=1}^{n}\sum_{j\neq i}(P_{ij})^2 + \sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^2 - \sum_{i=1}^{n}\sum_{j\neq i}(P_{ij})^2 \\
=& \sum_{i=1}^{n}\sum_{j\neq i}(P_{ij})^2 + \sum_{i=1}^{n}\sum_{j=1}^{n}(P_{ij}^{\gamma_n})^2 - \sum_{i=1}^{n}\sum_{j=1}^{n}(P_{ij})^2
\end{aligned}$$

$$+\sum_{i=1}^{n}(P_{ii})^2-\sum_{i=1}^{n}(P_{ii}^{\gamma_n})^2.$$

Furthermore, by Equation (A.1) $P_{ii}\geq P_{ii}^{\gamma_n}\geq 0$ for all $\gamma_n\geq 0$ (see also the last equality in Equation (A.2)) such that $\sum_{i=1}^{n}(P_{ii})^2-\sum_{i=1}^{n}(P_{ii}^{\gamma_n})^2\geq 0$. It follows that

$$\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^2\geq\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij})^2-(||P||_F^2-||P^{\gamma_n}||_F^2). \tag{A.12}$$

In addition,

$$\begin{aligned}||P||_F^2-||P^{\gamma_n}||_F^2&=(||P||_F+||P^{\gamma_n}||_F)(||P||_F-||P^{\gamma_n}||_F)\\&\leq 2||P||_F(||P||_F-||P^{\gamma_n}||_F)\\&\leq 2\sqrt{k_n}||P-P^{\gamma_n}||_F,\end{aligned} \tag{A.13}$$

the inequalities following from $||P^{\gamma_n}||_F\leq||P||_F=\sqrt{k_n}$ (using Equation (A.1) compare $||P^{\gamma_n}||_F$ and $||P||_F$). Using Equation (A.1) once more yields that

$$P-P^{\gamma_n}=U(\check{D}-\tilde{D})U',$$

where $\check{D}$ is the $n\times n$ diagonal matrix with upper left block being the $k_n\times k_n$ identity matrix. Therefore,

$$\begin{aligned}||P-P^{\gamma_n}||_F^2&=\mathrm{tr}([P-P^{\gamma_n}]'[P-P^{\gamma_n}])=\mathrm{tr}(U(\check{D}-\tilde{D})^2U')\\&=\sum_{i=1}^{k_n}\left(1-\frac{D_{ii}^2}{D_{ii}^2+\gamma_n}\right)^2.\end{aligned} \tag{A.14}$$

Combining Equations (A.12)–(A.14) implies that

$$\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^2\geq\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij})^2-2\sqrt{k_n}\sqrt{\sum_{i=1}^{k_n}\left(1-\frac{D_{ii}^2}{D_{ii}^2+\gamma_n}\right)^2},$$

such that

$$\frac{1}{k_n}\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^2\geq\frac{1}{k_n}\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij})^2-2\sqrt{\frac{1}{k_n}\sum_{i=1}^{k_n}\left(1-\frac{D_{ii}^2}{D_{ii}^2+\gamma_n}\right)^2}.$$

Since

$$\frac{1}{k_n}\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij})^2\leq\frac{1}{k_n}\sum_{i=1}^{n}\sum_{j=1}^{n}(P_{ij})^2=\frac{1}{k_n}\mathrm{tr}(P^2)=\frac{1}{k_n}\mathrm{tr}(P)=1,$$

and $\gamma_n \mapsto \frac{1}{k_n}\sum_{i=1}^{k_n}\left(1-\frac{D_{ii}^2}{D_{ii}^2+\gamma_n}\right)^2$ is continuous, 0 at 0, and tends to 1 as $\gamma_n \to \infty$, there exists a $\gamma_n^0 > 0$ such that

$$2\sqrt{\frac{1}{k_n}\sum_{i=1}^{k_n}\left(1-\frac{D_{ii}^2}{D_{ii}^2+\gamma_n^0}\right)^2} = \frac{1}{2k_n}\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij})^2$$

ensuring that

$$\frac{1}{k_n}\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n^0})^2 \geq \frac{1}{2k_n}\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij})^2 \geq \delta/2. \tag{A.15}$$

Finally, since γ_n^* is the penalty parameter that maximises $\gamma_n \mapsto \sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^2$,

$$\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n^*})^2 \geq \sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n^0})^2 \geq \frac{k\delta}{2},$$

so that Assumption 3 is satisfied for $\omega = 1$. □

A.1.3 Proof of Theorem 1

We begin by showing that

$$RJAR_{\gamma_n}^* := \frac{1}{\sqrt{r_n}\sqrt{\Phi_{\gamma_n}}}\sum_{i=1}^{n}\sum_{j\neq i}P_{ij}^{\gamma_n}\varepsilon_i\varepsilon_j \tag{A.16}$$

converges in distribution to a standard normal for all sequences of γ_n that satisfy Assumption 3.[2] To this end, define $\mathcal{U}_n := 2\sum_{i=2}^{n}\sum_{j=1}^{i-1}P_{ij}^{\gamma_n}\varepsilon_i\varepsilon_j$, $s_n^2 := \mathbb{E}[\mathcal{U}_n^2]$ and note that by the symmetry of P^{γ_n}, $RJAR_{\gamma_n}^*$ can be written as

$$RJAR_{\gamma_n}^* = s_n^{-1}\mathcal{U}_n.$$

We proceed by establishing that $s_n^{-1}\mathcal{U}_n \xrightarrow{d} \mathcal{N}[0,1]$ as $n \to \infty$. This, in turn, follows from Hall and Heyde (1980, Corollary 3.1)) upon verifying that i) for all $\epsilon > 0$

$$s_n^{-2}\sum_{i=2}^{n}\mathbb{E}[Y_{ni}^2 I(|Y_{ni}| > \epsilon s_n)] \to 0 \qquad \text{as } n \to \infty, \tag{A.17}$$

where $Y_{ni} = 2\sum_{j=1}^{i-1}P_{ij}^{\gamma_n}\varepsilon_i\varepsilon_j$, and ii)

$$s_n^{-2}\mathcal{V}_n^2 \xrightarrow{p} 1 \qquad \text{as } n \to \infty, \tag{A.18}$$

[2]We do not directly invoke Lemma 2 in Hansen and Kozbur (2014) because it would require strengthening the assumptions used in this chapter. In particular, we would have to replace the condition in Assumption 3 that $\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^2 \geq cr_n^{\omega}$ with the stronger condition that $\sum_{i=1}^{n}\sum_{j\neq i}(P_{ij}^{\gamma_n})^2 \geq ck_n$.

where $\mathcal{V}_n^2 = \sum_{i=2}^n \mathbb{E}[Y_{ni}^2|\varepsilon_1,\ldots,\varepsilon_{i-1}]$. Consider first the condition in Equation (A.17) and write

$$\mathbb{E}[Y_{ni}^2] = 4\sum_{j=1}^{i-1}\sum_{h=1}^{i-1} P_{ij}^{\gamma_n} P_{ih}^{\gamma_n} \mathbb{E}[\varepsilon_i^2\varepsilon_j\varepsilon_h] = 4\sum_{j=1}^{i-1}(P_{ij}^{\gamma_n})^2\mathbb{E}[\varepsilon_i^2]\mathbb{E}[\varepsilon_j^2],$$

so that, upon using that $\mathcal{U}_n = \sum_{i=2}^n Y_{ni}$ and $\mathbb{E}[Y_{ni}Y_{nj}] = 0$ for $i \neq j$, one gets

$$s_n^2 = \sum_{i=2}^n \mathbb{E}[Y_{ni}^2] = 4\sum_{i=2}^n\sum_{j=1}^{i-1}(P_{ij}^{\gamma_n})^2\mathbb{E}[\varepsilon_i^2]\mathbb{E}[\varepsilon_j^2]. \tag{A.19}$$

Furthermore,

$$\begin{aligned}\mathbb{E}[Y_{ni}^4] =& 16\sum_{j=1}^{i-1}\sum_{h=1}^{i-1}\sum_{m=1}^{i-1}\sum_{l=1}^{i-1} P_{ij}^{\gamma_n}P_{ih}^{\gamma_n}P_{im}^{\gamma_n}P_{il}^{\gamma_n}\mathbb{E}[\varepsilon_i^4\varepsilon_j\varepsilon_h\varepsilon_m\varepsilon_l] \\ =& 16\sum_{j=1}^{i-1}(P_{ij}^{\gamma_n})^4\mathbb{E}[\varepsilon_i^4]\mathbb{E}[\varepsilon_j^4] + 48\sum_{j=1}^{i-1}\sum_{h\neq j}(P_{ij}^{\gamma_n})^2(P_{ih}^{\gamma_n})^2\mathbb{E}[\varepsilon_i^4]\mathbb{E}[\varepsilon_j^2]\mathbb{E}[\varepsilon_h^2],\end{aligned}$$

so that

$$\begin{aligned}\sum_{i=2}^n \mathbb{E}[Y_{ni}^4] =& 16\sum_{i=2}^n\sum_{j=1}^{i-1}(P_{ij}^{\gamma_n})^4\mathbb{E}[\varepsilon_i^4]\mathbb{E}[\varepsilon_j^4] \\ &+ 48\sum_{i=2}^n\sum_{j=1}^{i-1}\sum_{h\neq j}(P_{ij}^{\gamma_n})^2(P_{ih}^{\gamma_n})^2\mathbb{E}[\varepsilon_i^4]\mathbb{E}[\varepsilon_j^2]\mathbb{E}[\varepsilon_h^2].\end{aligned}$$

It follows that

$$\begin{aligned}\frac{\sum_{i=2}^n \mathbb{E}[Y_{ni}^4]}{s_n^4} &\leq C\frac{\sum_{i=2}^n\sum_{j=1}^{i-1}(P_{ij}^{\gamma_n})^4 + \sum_{i=2}^n\sum_{j=1}^{i-1}\sum_{h\neq j}(P_{ij}^{\gamma_n})^2(P_{ih}^{\gamma_n})^2}{\left(\sum_{i=1}^n\sum_{j\neq i}(P_{ij}^{\gamma_n})^2\right)^2} \\ &\leq C\frac{\sum_{i=1}^n\sum_{j=1}^n(P_{ij}^{\gamma_n})^2 + \sum_{i=1}^n\sum_{j=1}^n(P_{ij}^{\gamma_n})^2\sum_{h=1}^n(P_{ih}^{\gamma_n})^2}{\left(\sum_{i=1}^n\sum_{j\neq i}(P_{ij}^{\gamma_n})^2\right)^2} \\ &= C\frac{\sum_{i=1}^n(P^{\gamma_n})_{ii}^2 + \sum_{i=1}^n\left((P^{\gamma_n})_{ii}^2\right)^2}{\left(\sum_{i=1}^n\sum_{j\neq i}(P_{ij}^{\gamma_n})^2\right)^2} \\ &\leq C\frac{\sum_{i=1}^n(P^{\gamma_n})_{ii}^2 + \sum_{i=1}^n(P^{\gamma_n})_{ii}^2}{\left(\sum_{i=1}^n\sum_{j\neq i}(P_{ij}^{\gamma_n})^2\right)^2} \\ &\leq C\frac{\sum_{i=1}^n(P^{\gamma_n})_{ii}^2}{r_n^{2\omega}} \\ &\leq Cr_n^{1-2\omega} \\ &\to 0,\end{aligned} \tag{A.20}$$

where the first inequality follows from Assumption 1. The second inequality follows from Lemma 1 (v). The third inequality follows from Lemma 1 (i). The fourth

inequality follows from Assumption 3. The fifth inequality follows from Lemma 1 (iii). The limit holds by Assumption 2 and Assumption 3. The condition in Equation (A.17) follows from Equation (A.20).

In order to verify the convergence in Equation (A.18) it suffices to show that

$$s_n^{-4}\mathbb{E}\left[(\mathcal{V}_n^2 - s_n^2)^2\right] = \frac{\mathbb{E}[\mathcal{V}_n^4] + s_n^4 - 2s_n^2\mathbb{E}[\mathcal{V}_n^2]}{s_n^4} = \frac{\mathbb{E}[\mathcal{V}_n^4]}{s_n^4} + 1 - 2\frac{\mathbb{E}[\mathcal{V}_n^2]}{s_n^2} \to 0.$$

Since $\mathcal{V}_n^2 = \sum_{i=2}^n \mathbb{E}[Y_{ni}^2|\varepsilon_1, \ldots, \varepsilon_{i-1}]$ it follows that

$$\mathbb{E}[\mathcal{V}_n^2] = \sum_{i=2}^n \mathbb{E}[Y_{ni}^2] = s_n^2,$$

the last equality following from Equation (A.19). It remains to be verified that

$$\frac{\mathbb{E}[\mathcal{V}_n^4]}{s_n^4} \to 1. \tag{A.21}$$

To this end, observe that

$$\mathcal{V}_n^2 = \sum_{i=2}^n \mathbb{E}[Y_{ni}^2|\varepsilon_1, \ldots, \varepsilon_{i-1}] = 4\sum_{i=2}^n \sum_{j=1}^{i-1} \sum_{h=1}^{i-1} P_{ij}^{\gamma_n} P_{ih}^{\gamma_n} \mathbb{E}[\varepsilon_i^2]\varepsilon_j\varepsilon_h,$$

such that

$$\mathcal{V}_n^4 = 16\sum_{i=2}^n \sum_{j=2}^n \sum_{h=1}^{i-1} \sum_{l=1}^{i-1} \sum_{m=1}^{j-1} \sum_{w=1}^{j-1} P_{ih}^{\gamma_n} P_{il}^{\gamma_n} P_{jm}^{\gamma_n} P_{jw}^{\gamma_n} \mathbb{E}[\varepsilon_i^2]\mathbb{E}[\varepsilon_j^2]\varepsilon_h\varepsilon_l\varepsilon_m\varepsilon_w.$$

For $i \leq j$, tedious but straightforward calculations included for completeness yield

$$\begin{aligned}
&\mathbb{E}\left[\sum_{h=1}^{i-1}\sum_{l=1}^{i-1}\sum_{m=1}^{j-1}\sum_{w=1}^{j-1} P_{ih}^{\gamma_n} P_{il}^{\gamma_n} P_{jm}^{\gamma_n} P_{jw}^{\gamma_n} \varepsilon_h\varepsilon_l\varepsilon_m\varepsilon_w\right] \\
=&\sum_{l=1}^{i-1}\sum_{h=1}^{i-1}\sum_{m=1}^{j-1}\sum_{w\neq m}^{j-1} P_{il}^{\gamma_n} P_{ih}^{\gamma_n} P_{jm}^{\gamma_n} P_{jw}^{\gamma_n} \mathbb{E}[\varepsilon_l\varepsilon_h\varepsilon_m\varepsilon_w] \\
&+\sum_{l=1}^{i-1}\sum_{h\neq l}^{i-1}\sum_{m\neq l}^{j-1} P_{il}^{\gamma_n} P_{ih}^{\gamma_n} (P_{jm}^{\gamma_n})^2 \mathbb{E}[\varepsilon_l]\mathbb{E}[\varepsilon_h\varepsilon_m^2] \\
&+\sum_{l=1}^{i-1}\sum_{h\neq l}^{i-1} P_{il}^{\gamma_n} P_{ih}^{\gamma_n} (P_{jl}^{\gamma_n})^2 \mathbb{E}[\varepsilon_l^3]\mathbb{E}[\varepsilon_h] + \sum_{l=1}^{i-1}\sum_{m=1}^{j-1} (P_{il}^{\gamma_n})^2 (P_{jm}^{\gamma_n})^2 \mathbb{E}[\varepsilon_l^2\varepsilon_m^2] \\
=&\sum_{l=1}^{i-1}\sum_{h=1}^{i-1}\sum_{m=1}^{j-1}\sum_{w\neq m}^{j-1} P_{il}^{\gamma_n} P_{ih}^{\gamma_n} P_{jm}^{\gamma_n} P_{jw}^{\gamma_n} \mathbb{E}[\varepsilon_l\varepsilon_h\varepsilon_m\varepsilon_w] + \sum_{l=1}^{i-1}\sum_{m=1}^{j-1} (P_{il}^{\gamma_n})^2 (P_{jm}^{\gamma_n})^2 \mathbb{E}[\varepsilon_l^2\varepsilon_m^2] \\
=&\sum_{l=1}^{i-1}\sum_{h\neq l}^{i-1}\sum_{m=1}^{j-1}\sum_{w\neq m}^{j-1} P_{il}^{\gamma_n} P_{ih}^{\gamma_n} P_{jm}^{\gamma_n} P_{jw}^{\gamma_n} \mathbb{E}[\varepsilon_l\varepsilon_h\varepsilon_m\varepsilon_w]
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{l=1}^{i-1}\sum_{m\neq l}^{j-1}\sum_{w\neq m}^{j-1}(P_{il}^{\gamma_n})^2P_{jm}^{\gamma_n}P_{jw}^{\gamma_n}\mathbb{E}[\varepsilon_l^2\varepsilon_w]\mathbb{E}[\varepsilon_m]\\
&+\sum_{l=1}^{i-1}\sum_{w\neq l}^{j-1}(P_{il}^{\gamma_n})^2P_{jl}^{\gamma_n}P_{jw}^{\gamma_n}\mathbb{E}[\varepsilon_l^3]\mathbb{E}[\varepsilon_w]+\sum_{l=1}^{i-1}\sum_{m=1}^{j-1}(P_{il}^{\gamma_n})^2(P_{jm}^{\gamma_n})^2\mathbb{E}[\varepsilon_l^2\varepsilon_m^2]\\
=&\sum_{l=1}^{i-1}\sum_{h\neq l}^{i-1}\sum_{m=1}^{j-1}\sum_{w\neq m}^{j-1}P_{il}^{\gamma_n}P_{ih}^{\gamma_n}P_{jm}^{\gamma_n}P_{jw}^{\gamma_n}\mathbb{E}[\varepsilon_l\varepsilon_h\varepsilon_m\varepsilon_w]+\sum_{l=1}^{i-1}\sum_{m=1}^{j-1}(P_{il}^{\gamma_n})^2(P_{jm}^{\gamma_n})^2\mathbb{E}[\varepsilon_l^2\varepsilon_m^2]\\
=&\sum_{l=1}^{i-1}\sum_{h\neq l}^{i-1}\sum_{m\neq h}^{j-1}\sum_{w\neq m}^{j-1}P_{il}^{\gamma_n}P_{ih}^{\gamma_n}P_{jm}^{\gamma_n}P_{jw}^{\gamma_n}\mathbb{E}[\varepsilon_l\varepsilon_h\varepsilon_m\varepsilon_w]\\
&+\sum_{l=1}^{i-1}\sum_{h\neq l}^{i-1}\sum_{w\notin\{h,l\}}^{j-1}P_{il}^{\gamma_n}P_{ih}^{\gamma_n}P_{jh}^{\gamma_n}P_{jw}^{\gamma_n}\mathbb{E}[\varepsilon_l]\mathbb{E}[\varepsilon_w]\mathbb{E}[\varepsilon_h^2]\\
&+\sum_{l=1}^{i-1}\sum_{h\neq l}^{i-1}P_{il}^{\gamma_n}P_{ih}^{\gamma_n}P_{jh}^{\gamma_n}P_{jl}^{\gamma_n}\mathbb{E}[\varepsilon_l^2]\mathbb{E}[\varepsilon_h^2]+\sum_{l=1}^{i-1}\sum_{m=1}^{j-1}(P_{il}^{\gamma_n})^2(P_{jm}^{\gamma_n})^2\mathbb{E}[\varepsilon_l^2\varepsilon_m^2]\\
=&\sum_{l=1}^{i-1}\sum_{h\neq l}^{i-1}\sum_{m\neq h}^{j-1}\sum_{w\neq m}^{j-1}P_{il}^{\gamma_n}P_{ih}^{\gamma_n}P_{jm}^{\gamma_n}P_{jw}^{\gamma_n}\mathbb{E}[\varepsilon_l\varepsilon_h\varepsilon_m\varepsilon_w]+\sum_{l=1}^{i-1}\sum_{h\neq l}^{i-1}P_{il}^{\gamma_n}P_{ih}^{\gamma_n}P_{jh}^{\gamma_n}P_{jl}^{\gamma_n}\mathbb{E}[\varepsilon_l^2]\mathbb{E}[\varepsilon_h^2]\\
&+\sum_{l=1}^{i-1}\sum_{m=1}^{j-1}(P_{il}^{\gamma_n})^2(P_{jm}^{\gamma_n})^2\mathbb{E}[\varepsilon_l^2\varepsilon_m^2]\\
=&\sum_{l=1}^{i-1}\sum_{h\neq l}^{i-1}\sum_{m\neq h}^{j-1}\sum_{w\notin\{m,h\}}^{j-1}P_{il}^{\gamma_n}P_{ih}^{\gamma_n}P_{jm}^{\gamma_n}P_{jw}^{\gamma_n}\mathbb{E}[\varepsilon_l\varepsilon_h\varepsilon_m\varepsilon_w]\\
&+\sum_{l=1}^{i-1}\sum_{h\neq l}^{i-1}\sum_{m\notin\{h,l\}}^{j-1}P_{il}^{\gamma_n}P_{ih}^{\gamma_n}P_{jm}^{\gamma_n}P_{jh}^{\gamma_n}\mathbb{E}[\varepsilon_l]\mathbb{E}[\varepsilon_m]\mathbb{E}[\varepsilon_h^2]\\
&+2\sum_{l=1}^{i-1}\sum_{h\neq l}^{i-1}P_{il}^{\gamma_n}P_{ih}^{\gamma_n}P_{jl}^{\gamma_n}P_{jh}^{\gamma_n}\mathbb{E}[\varepsilon_l^2]\mathbb{E}[\varepsilon_h^2]+\sum_{l=1}^{i-1}\sum_{m=1}^{j-1}(P_{il}^{\gamma_n})^2(P_{jm}^{\gamma_n})^2\mathbb{E}[\varepsilon_l^2\varepsilon_m^2]\\
=&\sum_{l=1}^{i-1}\sum_{h\neq l}^{i-1}\sum_{m\neq h}^{j-1}\sum_{w\notin\{m,h\}}^{j-1}P_{il}^{\gamma_n}P_{ih}^{\gamma_n}P_{jm}^{\gamma_n}P_{jw}^{\gamma_n}\mathbb{E}[\varepsilon_l\varepsilon_h\varepsilon_m\varepsilon_w]\\
&+2\sum_{l=1}^{i-1}\sum_{h\neq l}^{i-1}P_{il}^{\gamma_n}P_{ih}^{\gamma_n}P_{jl}^{\gamma_n}P_{jh}^{\gamma_n}\mathbb{E}[\varepsilon_l^2]\mathbb{E}[\varepsilon_h^2]+\sum_{l=1}^{i-1}\sum_{m=1}^{j-1}(P_{il}^{\gamma_n})^2(P_{jm}^{\gamma_n})^2\mathbb{E}[\varepsilon_l^2\varepsilon_m^2]\\
=&\sum_{l=1}^{i-1}\sum_{h\neq l}^{i-1}\sum_{m\neq h}^{j-1}\sum_{w\notin\{m,h,l\}}^{j-1}P_{il}^{\gamma_n}P_{ih}^{\gamma_n}P_{jm}^{\gamma_n}P_{jw}^{\gamma_n}\mathbb{E}[\varepsilon_l\varepsilon_h\varepsilon_m]\mathbb{E}[\varepsilon_w]\\
&+\sum_{l=1}^{i-1}\sum_{h\neq l}^{i-1}\sum_{m\notin\{h,l\}}^{j-1}P_{il}^{\gamma_n}P_{ih}^{\gamma_n}P_{jm}^{\gamma_n}P_{jl}^{\gamma_n}\mathbb{E}[\varepsilon_l^2]\mathbb{E}[\varepsilon_h]\mathbb{E}[\varepsilon_m]\\
&+\sum_{l=1}^{i-1}\sum_{h\neq l}^{i-1}P_{il}^{\gamma_n}P_{ih}^{\gamma_n}(P_{jl}^{\gamma_n})^2\mathbb{E}[\varepsilon_l^3]\mathbb{E}[\varepsilon_h]+2\sum_{l=1}^{i-1}\sum_{h\neq l}^{i-1}P_{il}^{\gamma_n}P_{ih}^{\gamma_n}P_{jl}^{\gamma_n}P_{jh}^{\gamma_n}\mathbb{E}[\varepsilon_l^2]\mathbb{E}[\varepsilon_h^2]\\
&+\sum_{l=1}^{i-1}\sum_{m=1}^{j-1}(P_{il}^{\gamma_n})^2(P_{jm}^{\gamma_n})^2\mathbb{E}[\varepsilon_l^2\varepsilon_m^2]
\end{aligned}
$$

$$
\begin{aligned}
=&2\sum_{l=1}^{i-1}\sum_{h\neq l}^{i-1} P_{il}^{\gamma_n} P_{ih}^{\gamma_n} P_{jl}^{\gamma_n} P_{jh}^{\gamma_n} \mathbb{E}[\varepsilon_l^2]\mathbb{E}[\varepsilon_h^2] + \sum_{l=1}^{i-1}\sum_{m=1}^{j-1} (P_{il}^{\gamma_n})^2 (P_{jm}^{\gamma_n})^2 \mathbb{E}[\varepsilon_l^2\varepsilon_m^2] \\
=&2\sum_{l=1}^{i-1}\sum_{h\neq l}^{i-1} P_{il}^{\gamma_n} P_{ih}^{\gamma_n} P_{jl}^{\gamma_n} P_{jh}^{\gamma_n} \mathbb{E}[\varepsilon_l^2]\mathbb{E}[\varepsilon_h^2] + \sum_{l=1}^{i-1}\sum_{m\neq l}^{j-1} (P_{il}^{\gamma_n})^2 (P_{jm}^{\gamma_n})^2 \mathbb{E}[\varepsilon_l^2]\mathbb{E}[\varepsilon_m^2] \\
&+ \sum_{l=1}^{i-1} (P_{il}^{\gamma_n})^2 (P_{jl}^{\gamma_n})^2 \mathbb{E}[\varepsilon_l^4] \\
=&2\sum_{h=1}^{i-1}\sum_{l\neq h}^{i-1} P_{ih}^{\gamma_n} P_{il}^{\gamma_n} P_{jh}^{\gamma_n} P_{jl}^{\gamma_n} \mathbb{E}[\varepsilon_h^2]\mathbb{E}[\varepsilon_l^2] \\
&+ \sum_{h=1}^{i-1} (P_{ih}^{\gamma_n})^2 (P_{jh}^{\gamma_n})^2 \mathrm{Var}[\varepsilon_h^2] + \sum_{h=1}^{i-1}\sum_{l=1}^{j-1} (P_{ih}^{\gamma_n})^2 (P_{jl}^{\gamma_n})^2 \mathbb{E}[\varepsilon_h^2]\mathbb{E}[\varepsilon_l^2],
\end{aligned}
$$

where $\mathrm{Var}[\varepsilon_h^2] \coloneqq \mathbb{E}[\varepsilon_h^4] - \mathbb{E}[\varepsilon_h^2]\mathbb{E}[\varepsilon_h^2]$.

Defining $q \coloneqq \min\{i, j\}$, the above display in turn implies

$$
\begin{aligned}
\mathbb{E}[\mathcal{V}_n^4] =&32\sum_{i=2}^{n}\mathbb{E}[\varepsilon_i^2]\sum_{j=2}^{n}\mathbb{E}[\varepsilon_j^2]\sum_{h=1}^{q-1}\sum_{l\neq h} P_{ih}^{\gamma_n} P_{il}^{\gamma_n} P_{jh}^{\gamma_n} P_{jl}^{\gamma_n} \mathbb{E}[\varepsilon_h^2]\mathbb{E}[\varepsilon_l^2] \\
&+ 16\sum_{i=2}^{n}\mathbb{E}[\varepsilon_i^2]\sum_{j=2}^{n}\mathbb{E}[\varepsilon_j^2]\sum_{h=1}^{q-1}(P_{ih}^{\gamma_n})^2(P_{jh}^{\gamma_n})^2\mathrm{Var}[\varepsilon_h^2] \\
&+ 16\sum_{i=2}^{n}\mathbb{E}[\varepsilon_i^2]\sum_{j=2}^{n}\mathbb{E}[\varepsilon_j^2]\sum_{h=1}^{i-1}\sum_{l=1}^{j-1}(P_{ih}^{\gamma_n})^2(P_{jl}^{\gamma_n})^2\mathbb{E}[\varepsilon_h^2]\mathbb{E}[\varepsilon_l^2] \\
\equiv&\mathfrak{A}_n + \mathfrak{B}_n + s_n^4.
\end{aligned}
$$

In order to establish the convergence in Equation (A.21), it is sufficient to show that $\frac{\mathfrak{A}_n+\mathfrak{B}_n}{s_n^4} \to 0$. To this end, note that

$$
\begin{aligned}
\mathfrak{A}_n &\leq C\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{h=1}^{q-1}\sum_{l\neq h} P_{ih}^{\gamma_n} P_{il}^{\gamma_n} P_{jh}^{\gamma_n} P_{jl}^{\gamma_n} \\
&= C\left(\sum_{i=1}^{n}\sum_{j\neq i}\sum_{h=1}^{q-1}\sum_{l\neq h} P_{ih}^{\gamma_n} P_{il}^{\gamma_n} P_{jh}^{\gamma_n} P_{jl}^{\gamma_n} + \sum_{i=1}^{n}\sum_{h=1}^{q-1}\sum_{l\neq h}(P_{ih}^{\gamma_n})^2(P_{il}^{\gamma_n})^2\right) \\
&\leq C\left(\sum_{i=1}^{n}\sum_{j<i}\sum_{h<j}\sum_{l\neq h} P_{ih}^{\gamma_n} P_{il}^{\gamma_n} P_{jh}^{\gamma_n} P_{jl}^{\gamma_n} + r_n\right) \\
&\leq C\left(\sum_{i=1}^{n}\sum_{j<i}\sum_{h<j}\sum_{l<h} P_{ih}^{\gamma_n} P_{il}^{\gamma_n} P_{jh}^{\gamma_n} P_{jl}^{\gamma_n} + r_n\right) \\
&\leq C\left(\left|\sum_{i=1}^{n}\sum_{j<i}\sum_{h<j}\sum_{l<h} P_{ih}^{\gamma_n} P_{il}^{\gamma_n} P_{jh}^{\gamma_n} P_{jl}^{\gamma_n}\right| + r_n\right) \\
&\leq Cr_n
\end{aligned}
\tag{A.22}
$$

The first inequality follows from Assumption 1. The second inequality follows from Lemma 1 (vi). The third inequality follows from the symmetry of P^{γ_n}. The fifth inequality follows from Lemma 2. Next,

$$\mathfrak{B}_n \leq C \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{h=1}^{q-1} (P_{ih}^{\gamma_n})^2 (P_{jh}^{\gamma_n})^2 \leq C r_n, \tag{A.23}$$

the first inequality following from Assumption 1 and the second from Lemma 1 (vi). Thus,

$$\frac{\mathfrak{A}_n + \mathfrak{B}_n}{s_n^4} \leq C \frac{r_n}{\left(\sum_{i=1}^{n} \sum_{j \neq i} (P_{ij}^{\gamma_n})^2\right)^2} \leq C r_n^{1-2\omega} \to 0. \tag{A.24}$$

where the second inequality follows from Assumption 3 and the convergence from Assumption 2 and Assumption 3. Equation (A.24) verifies the condition in Equation (A.21), which in turn implies that the condition condition in Equation (A.18) holds.

Having verified the conditions in Equation (A.17) and Equation (A.18) we conclude that $RJAR^*_{\gamma_n} \xrightarrow{d} \mathcal{N}[0,1]$. By Lemma 4 $|\hat{\Phi}_{\gamma_n} - \Phi_{\gamma_n}| \xrightarrow{p} 0$ and hence also $\hat{\Phi}_{\gamma_n}/\Phi_{\gamma_n} \xrightarrow{p} 1$ under Assumptions 1 and 3. Hence, under the null hypothesis, the continuous mapping theorem implies that $RJAR_{\gamma_n}(\beta_0) \xrightarrow{d} \mathcal{N}[0,1]$.

□

B

Appendix for Chapter 3

B.1 Proofs

B.1.1 Proof of Theorem 2

Throughout, it is assumed that the null hypothesis in Equation (3.5) holds, so that ε_{0t} is replaced by ε_t. The proof is a straightforward application of the multiplier bootstrap for dependent data in Appendix B of Chernozhukov et al. (2018) (CCK in the sequel).

Notice that

$$\begin{aligned}\mathcal{R} &= \max_{1\leq j\leq k}\left|\frac{1}{\sqrt{T}}Z_j'\varepsilon_0\right| \\ &= \max_{1\leq j\leq 2k}\sqrt{T}\tilde{\mu}_j,\end{aligned}$$

which is equivalent to the test statistic considered by CCK in Appendix B.1. I hence proceed to check the conditions of Theorem B.1 in CCK.

The assumptions of non-degeneracy and bounded variance ($c_1 \leq \underline{\sigma}^2(q) \leq \bar{\sigma}^2(r)\vee\bar{\sigma}^2(q) \leq C_1$ in their notation) holds by Assumption 4.ii. and Assumption 4.v..

By Assumption 4.i., Assumption 4.iii., and 4.iv., the first dimensionality assumption of CCK ($\max\{mb_r, (r/q)\log^2 p\} \leq C_1 n^{-c_2}$ in their notation) holds.

Since $\frac{1}{2} - a - \tilde{a} > 0$ by Assumption 4.iii., Assumption 4.iv. implies that $\log k \lesssim T^{\frac{2}{5}(\frac{1}{2}-a-\tilde{a})} \lesssim T^{\frac{2}{5}(\frac{1}{2}-a-\tilde{a})} - \log 2T$. Furthermore, $Z_{tj}\varepsilon_t - \mathbb{E}[Z_{tj}\varepsilon_t]$ is bounded almost surely by Assumption 4.ii.. Taken together, this verifies the second dimensionality assumption of CCK ($qD_n\log^{5/2}(pn) \leq C_1 n^{1/2-c_2}$ in their notation).

Since $\mathbb{E}[Z_t\varepsilon_t] = 0$, it is possible to invoke the second consequence of Theorem B.1 in CCK, which yields

$$\left|\mathbb{P}\left(\mathcal{R} > c(\alpha)\right) - \alpha\right| \lesssim T^{-c_1},$$

for some $c_1 > 0$. Letting $T \to \infty$ yields the required result.

B.1.2 Derivation of Concentration Parameters for Simulations

I first derive the concentration parameter for the unobserved oracle first stage. The derivations for this concentration parameter are very similar to those in the online appendix of Mavroeidis et al. (2014).

The endogenous variables in the model the econometrician estimates (Equation (3.2)) can be written in terms of the excluded IVs as[1]

$$\begin{bmatrix} \pi_{t+1} - \pi_{t-1} \\ s_t \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{E_1} \begin{bmatrix} \pi_{t+1} \\ s_{t+1} \\ f_{t+1} \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}}_{E_2} \begin{bmatrix} \pi_t \\ s_t \\ f_t \end{bmatrix} + \underbrace{\begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{E_3} \begin{bmatrix} \pi_{t-1} \\ s_{t-1} \\ f_{t-1} \end{bmatrix}. \quad \text{(B.1)}$$

Define $p_t = [\pi_{t+1} - \pi_{t-1}, s_t]'$, $R_t = [\pi_t, s_t, f_t]'$, $u_t = [u_{1t}, u_{2t}, u_{3t}]'$,

$$\Psi = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \text{ and } \Omega = \begin{bmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & \omega_{23} \\ \omega_{31} & \omega_{32} & \omega_{33} \end{bmatrix}.$$

Equation (B.1) can now be written as

$$\begin{aligned} p_t &= E_1 R_{t+1} + E_2 R_t + E_3 R_{t-1} \\ &= (E_1\Psi^2 + E_2\Psi + E_3)R_{t-1} + (E_1\Psi + E_2)u_t + E_1 u_{t+1} \\ &= DR_{t-1} + w_t, \end{aligned}$$

for $D = E_1\Psi^2 + E_2\Psi + E_3$, $w_t = (E_1\Psi + E_2)u_t + E_1 u_{t+1}$.

Assuming that R_t is stationary, and letting $\Gamma = \mathbb{V}[R_t]$,

$$\text{vec}(\Gamma) = (I_9 - \Psi \otimes \Psi)^{-1} \text{vec}(\Omega).$$

The population projection of p_t on R_{t-1} has coefficient matrix given by

$$\begin{aligned} M &= \mathbb{E}[p_t R_{t-1}']\Gamma^{-1} \\ &= \mathbb{E}\left[(DR_{t-1} + w_t)R_{t-1}'\right]\Gamma^{-1} \\ &= D\mathbb{E}\left[R_{t-1}R_{t-1}'\right]\Gamma^{-1} \\ &= D, \end{aligned}$$

since $\mathbb{E}[w_t R_{t-1}'] = \mathbb{E}\left[((E_1\Psi + E_2)u_t + E_1 u_{t+1})\, R_{t-1}'\right] = 0$.

The projection error of the unobserved oracle first stage is given by

$$\begin{aligned} e_t &= p_t - MR_{t-1} \\ &= DR_{t-1} + w_t - DR_{t-1} \\ &= w_t. \end{aligned}$$

[1] As in Mavroeidis et al. (2014, Online Appendix), the constant can be omitted for the purposes of deriving the concentration matrix since it is not an informative IV.

The variance of the population projection error of the unobserved oracle first stage $\Sigma = \mathbb{V}[e_t]$ is hence given by

$$\begin{aligned}\Sigma &= \mathbb{V}[w_t] \\ &= (E_1\Psi + E_2)\Omega(E_1A + E_2)' + E_1\Omega E_1'.\end{aligned}$$

The concentration matrix of the unobserved oracle first stage is then given by

$$C = T\Sigma^{-1/2}D\Gamma D'\Sigma^{-1/2'},$$

where $\Sigma^{-1/2}\Sigma^{-1/2,'} = \Sigma^{-1}$. The minimum eigenvalue μ_O^2 of matrix C gives the concentration parameter of the unobserved oracle first stage.

The steps to derive the concentration matrix corresponding to the first stage observed by the econometrician are similar. Define $\tilde{R}_t = [\pi_t, s_t, Q_t']'$, and

$$\Xi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0_{m\times 1} & 0_{m\times 1} & \xi \end{bmatrix}, F = \begin{bmatrix} 0_{2\times 2} & 0_{2\times m} \\ 0_{m\times 2} & I_m \end{bmatrix},$$

so that $\tilde{R} = \Xi R_t + \tilde{u}_t$, where $\tilde{u}_t = [0, 0, u_{4t}']'$, $\mathbb{V}[\tilde{u}_t] = F$.

Assuming that R_t is stationary, $\tilde{R}_t$ is also stationary. Letting $\tilde{\Gamma} = \mathbb{V}[\tilde{R}_t]$,

$$\begin{aligned}\tilde{\Gamma} &= \mathbb{V}\left[\Xi R_t + \tilde{u}_t\right] \\ &= \Xi\mathbb{V}\left[R_t\right]\Xi' + F \\ &= \Xi\Gamma\Xi' + F.\end{aligned}$$

The population projection of p_t on $\tilde{R}_{t-1}$ has coefficient matrix given by

$$\begin{aligned}\tilde{M} &= \mathbb{E}[p_t\tilde{R}_{t-1}']\tilde{\Gamma}^{-1} \\ &= \mathbb{E}[p_t(R_{t-1}'\Xi' + \tilde{u}_t')]\tilde{\Gamma}^{-1} \\ &= \mathbb{E}[p_tR_{t-1}']\Xi'\tilde{\Gamma}^{-1} + \mathbb{E}[p_t\tilde{u}_t']\tilde{\Gamma}^{-1} \\ &= \mathbb{E}[(DR_{t-1} + w_t)R_{t-1}']\Xi'\tilde{\Gamma}^{-1} \\ &= D\mathbb{E}[R_{t-1}R_{t-1}']\Xi'\tilde{\Gamma}^{-1} + \mathbb{E}[w_tR_{t-1}']\Xi'\tilde{\Gamma}^{-1} \\ &= D\Gamma\Xi'\tilde{\Gamma}^{-1}.\end{aligned}$$

The projection error for the observed first stage is given by

$$\begin{aligned}\tilde{e}_t &= p_t - \tilde{M}\tilde{R}_{t-1} \\ &= DR_{t-1} + w_t - \tilde{M}\tilde{R}_{t-1} \\ &= DR_{t-1} + w_t - \tilde{M}(\Xi R_{t-1} + \tilde{u}_{t-1}) \\ &= (D - \tilde{M}\Xi)R_{t-1} + w_t - M\tilde{u}_{t-1}.\end{aligned}$$

The variance of the population projection error of the observed first stage $\tilde{\Sigma} = \mathbb{V}[\tilde{e}_t]$ is hence given by

$$\begin{aligned}\tilde{\Sigma} &= \mathbb{V}[(D - \tilde{M}\Xi)R_{t-1}] + \mathbb{V}[w_t] + \mathbb{V}[\tilde{M}\tilde{u}_{t-1}] \\ &= (D - \tilde{M}\Xi)\Gamma(D - \tilde{M}\Xi)' + \Sigma + \tilde{M}F\tilde{M}'.\end{aligned}$$

The concentration matrix of the unobserved oracle first stage is then given by

$$\tilde{C} = T\tilde{\Sigma}^{-1/2}\tilde{M}\tilde{\Gamma}\tilde{M}'\tilde{\Sigma}^{-1/2'},$$

where $\tilde{\Sigma}^{-1/2}\tilde{\Sigma}^{-1/2,'} = \tilde{\Sigma}^{-1}$. The minimum eigenvalue μ_E^2 of matrix $\tilde{C}$ gives the concentration parameter of the observed first stage.

B.2 Data

Code	Description	Type
RPI	Real Personal Income	G
PERMIT	New Private Housing Units Authorized by Building Permits	G
INDPRO	Industrial Production Index	G
PERMITNE	New Private Housing Units Authorized by Building Permits in the Northeast Census Region	G
CUMFNS	Capacity Utilization: Manufacturing	D
PERMITMW	New Private Housing Units Authorized by Building Permits in the Midwest Census Region	G
IPFINAL	Industrial Production: Final Products (Market Group)	G
PERMITS	New Private Housing Units Authorized by Building Permits in the South Census Region	G
IPCONGD	Industrial Production: Consumer Goods	G
PERMITW	New Private Housing Units Authorized by Building Permits in the West Census Region	G
IPDCONGD	Industrial Production: Durable Consumer Goods	G
DPCERA3M086SBEA	Real personal consumption expenditures (chain-type quantity index)	G
IPNCONGD	Industrial Production: Nondurable Consumer Goods	G
CMRMTSPL	Real Manufacturing and Trade Industries Sales	G
IPBUSEQ	Industrial Production: Business Equipment	G
UMCSENT	University of Michigan: Consumer Sentiment	G
IPMAT	Industrial Production: Materials	G
M1SL	M1 Money Stock	G
IPMANSICS	Industrial Production: Manufacturing (SIC)	G
M2SL	M2 Money Stock	G
IPB51222s	Industrial Production: Residential utilities	G
TOTRESNS	Total Reserves of Depository Institutions	G
IPFUELS	Industrial Production: Fuels	G
BUSLOANS	Commercial and Industrial Loans, All Commercial Banks	G
CLF16OV	Civilian Labor Force Level	G
REALLN	Real Estate Loans, All Commercial Banks	G
UNRATE	Unemployment Rate	N
NONREVSL	Total Nonrevolving Credit Owned and Securitized, Outstanding	G
UEMPMEAN	Average Weeks Unemployed	D
DTCOLNVHFNM	Consumer Motor Vehicle Loans Owned by Finance Companies, Outstanding	G
UEMPLT5	Number Unemployed for Less Than 5 Weeks	G
DTCTHFNM	Total Consumer Loans and Leases Owned and Securitized by Finance Companies, Outstanding	G

UEMP5TO14	Number Unemployed for 5-14 Weeks	G
INVEST	Securities in Bank Credit, All Commercial Banks	G
UEMP15OV	Number Unemployed for 15 Weeks & Over	G
FEDFUNDS	Effective Federal Funds Rate	G
UEMP15T26	Number Unemployed for 15-26 Weeks	G
TB3SMFFM	3-Month Treasury Bill Minus Federal Funds Rate	N
UEMP27OV	Number Unemployed for 27 Weeks & Over	G
TB6SMFFM	6-Month Treasury Bill Minus Federal Funds Rate	N
PAYEMS	All Employees, Total Nonfarm	G
T1YFFM	1-Year Treasury Constant Maturity Minus Federal Funds Rate	N
USGOOD	All Employees, Goods-Producing	G
T5YFFM	5-Year Treasury Constant Maturity Minus Federal Funds Rate	N
CES1021000001	All Employees, Mining	G
T10YFFM	10-Year Treasury Constant Maturity Minus Federal Funds Rate	N
USCONS	All Employees, Construction	G
AAAFFM	Moody's Seasoned Aaa Corporate Bond Minus Federal Funds Rate	N
MANEMP	All Employees, Manufacturing	G
BAAFFM	Moody's Seasoned Baa Corporate Bond Minus Federal Funds Rate	N
PRS85006173	Nonfarm Business Sector: Labor Share	GG
TWEXMMTH	Trade Weighted U.S. Dollar Index: Major Currencies, Goods	G
DMANEMP	All Employees, Durable Goods	G
EXSZUS	Switzerland / U.S. Foreign Exchange Rate	G
NDMANEMP	All Employees, Nondurable Goods	G
EXJPUS	Japan / U.S. Foreign Exchange Rate	G
SRVPRD	All Employees, Service-Providing	G
EXUSUK	U.S. / U.K. Foreign Exchange Rate	G
USTPU	All Employees, Trade, Transportation, and Utilities	G
EXCAUS	Canada / U.S. Foreign Exchange Rate	G
USWTRADE	All Employees, Wholesale Trade	G
WPSFD49502	Producer Price Index by Commodity for Final Demand: Personal Consumption Goods	G
USTRADE	All Employees, Retail Trade	G
WPSID61	Producer Price Index by Commodity for Intermediate Demand by Commodity Type: Processed Goods for Intermediate Demand	G
USFIRE	All Employees, Financial Activities	G
WTISPLC	Spot Crude Oil Price: West Texas Inter mediate (WTI)	G
USGOVT	All Employees, Government	G
PPICMM	Producer Price Index by Commodity Metals and metal products: Primary nonferrous metals	G
CES0600000007	Average Weekly Hours of Production and Nonsupervisory Employees, Goods-Producing	G
CPIAUCSL	Consumer Price Index for All Urban Consumers: All Items in U.S. City Average	G
AWOTMAN	Average Weekly Overtime Hours of Production and Nonsupervisory Employees, Manufacturing	D
CUSR0000SAC	Consumer Price Index for All Urban Consumers: Commodities in U.S. City Average	G
AWHMAN	Average Weekly Hours of Production and Nonsupervisory Employees, Manufacturing	D
CUSR0000SAD	Consumer Price Index for All Urban Consumers: Durables in U.S. City Average	G
CES0600000008	Average Hourly Earnings of Production and Nonsupervisory Employees, Goods-Producing	G
CUSR0000SAS	Consumer Price Index for All Urban Consumers: Services in U.S. City Average	G
CES2000000008	Average Hourly Earnings of Production and Nonsupervisory Employees, Construction	G
PCEPI	Personal Consumption Expenditures: Chain-type Price Index	G
CES3000000008	Average Hourly Earnings of Production and Nonsupervisory Employees, Manufacturing	G

DDURRG3M086SBEA	Personal consumption expenditures: Durable goods (chain-type price index)	G
HOUST	Housing Starts: Total: New Privately Owned Housing Units Started	G
DNDGRG3M086SBEA	Personal consumption expenditures: Nondurable goods (chain-type price index)	G
HOUSTNE	Housing Starts in Northeast Census Region	G
DSERRG3M086SBEA	Personal consumption expenditures: Services (chain-type price index)	G
HOUSTMW	Housing Starts in Midwest Census Region	G
GDPDEF	Gross Domestic Product: Implicit Price Deflator	G
HOUSTS	Housing Starts in South Census Region	G
WILL5000IND	Wilshire 5000 Total Market Index	G
HOUSTW	Housing Starts in West Census Region	G
GDPC1	Real Gross Domestic Product	G

Notes:
N refers to no transformation of the data.
G refers transforming variable w_t w_t by computing $100\left(\log(w_t) - \log(w_{t-1})\right)$.
D refers transforming variable w_t by computing $w_t - w_{t-1}$.
GG refers transforming variable w_t by computing $0.1226 \times 100 \log(w_t/100)$ (as in Galì and Gertler (1999) and Kleibergen and Mavroeidis (2009)).

C Appendix for Chapter 4

C.1 Data

Table C.1: Data description.

Variable	Description	Data Source
Compensation per employee	euro area 19 total compensation of employees	ECB `MNA.Q.N.I8.W2.S1.S1._Z.COM_PS._Z._T._Z.EUR.V.N`
Unemployment rate	euro area 19 unemployment rate (as a % of labour force)	ECB `STS.Q.I8.S.UNEH.RTTT000.4.000`
Consumer price index	euro area HICP overall index	ECB `ICP.M.U2.N.000000.4.INX`
Oil Price	Crude Oil Prices: Brent - Europe (USD)	FRED `DCOILBRENTEU`
Productivity per employee	Gross value added in relation to the total number of employees	ECB `MNA.Q.Y.I8.W2.S1.S1._Z.SAL._Z._T._Z.PS._Z.N`
Short-term rate	euro area Eonia rate	ECB `FM.Q.U2.EUR.4F.MM.EONIA.HSTA`
Long-term rate	euro area 10-year government benchmark bond yield	ECB `FM.Q.U2.EUR.4F.BB.U2_10Y.YLD`
USD-euro exchange rate	U.S. Dollars to one Euro	FRED `DEXUSEU`

Notes:
Interest rate spread calculated as the difference between the long-term rate and the short-term rate.
Inflation rate for variable w_t calculated as $100 \times \big(\log(w_t) - \log(w_{t-1})\big)$.
ECB refers to ECB Warehouse Code. FRED refers to FRED Economic Data.

D

Appendix for Chapter 5

D.1 Proofs

D.1.1 Proof of Proposition 2

Proof. The procedure constructs confidence sets by inverting the test

$$H_0 : \lambda_{hi} = \lambda_g \text{ vs. } H_1 : \lambda_{hi} \neq \lambda_g, \tag{D.1}$$

so that the remainder of the proof follows under the null hypothesis (i.e., $\lambda_{hi} = \lambda_g$). Define the function $g(A, \Gamma) := e_i' C_h(A)\Gamma - \lambda_g e_1 \Gamma$. Notice that when this function is evaluated at the true values of A, Γ, it is equal to zero, since

$$\begin{aligned} e_i C_h(A)\Gamma - \lambda_g e_1'\Gamma &= \frac{e_i C_h(A)\Gamma}{e_1'\Gamma} e_1'\Gamma + \lambda_g e_1'\Gamma \\ &= \lambda_{hi} e_1'\Gamma - \lambda_g e_1'\Gamma = 0, \end{aligned}$$

where the second equality follows from the definition of the IRF, see, e.g., Montiel Olea et al. (2021, Equation (2.9)), and the third equality from the null hypothesis in Equation (D.1). For simplicity, replace $\hat{W}_{T_Z}$ with W (i.e., its population counterpart). The statement in the proposition follows if the quantiles of

$$\sqrt{T_Z}\left(g(A_m, \Gamma_m) - g(\hat{A}_T, \hat{\Gamma}_T)\right) \Big| \hat{A}_T, \hat{\Gamma}_T$$

approximate the quantiles of

$$\sqrt{T_Z}\left(g(\hat{A}_{T_Z}, \hat{\Gamma}_{T_Z}) - g(A, \Gamma)\right) = \sqrt{T_Z}[(e_i' C_h(\hat{A}_{T_Z}) - \lambda_g e_1')]\hat{\Gamma}_{T_Z}.$$

As stated in Montiel Olea et al. (2021, Appendix A.4), in the context of balanced data, this can be shown under Assumptions 5-6 and through an appeal to the Delta argument for bootstrap regimes. The proof for the bootstrap procedure proposed in

this chapter is very similar, but is given in full, since to the best of my knowledge it does not follow directly from existing results. The rest of the proof is virtually the same as the proof of Theorem 23.5 in van der Vaart (2007). The only difference is that the bootstrap is conditioned on a larger data set. As much as possible without clashing with the notation in the main text, the notation follows the one used in the proof of Theorem 23.5 in van der Vaart (2007).

Define $\hat{\theta}_T := [\text{vec}(\hat{A}_T)', \hat{\Gamma}_T']'$, $\hat{\theta}_{T_Z} := [\text{vec}(\hat{A}_{T_Z})', \hat{\Gamma}_{T_Z}']'$, and $\hat{\theta}^*_{T_Z} := [\text{vec}(A_m)', \Gamma_m']'$. For ease of notation, further define $Q := \mathcal{N}[0, W]$ and $\tilde{g}(\theta) := g(A, \Gamma)$. Then by Assumption 5 and 6,

$$\sqrt{T_Z}(\hat{\theta}_{T_Z} - \theta) \xrightarrow{d} Q. \tag{D.2}$$

By the sampling procedure in Step 5 above, it is also the case that

$$\sqrt{T_Z}(\hat{\theta}^*_{T_Z} - \hat{\theta}_T) \xrightarrow{d} Q. \tag{D.3}$$

Since $g(A, \Gamma)$ is a differentiable function, the standard Delta argument and Equation (D.2) imply that

$$\sqrt{T_Z}\left(\tilde{g}(\hat{\theta}_{T_Z}) - \tilde{g}(\theta)\right) \xrightarrow{d} \tilde{g}^\dagger_\theta(Q),$$

where $\tilde{g}^\dagger_\theta$ is the gradient of $\tilde{g}$ evaluated at θ.

I now show that a similar Delta argument applies to Equation (D.3). By the mean value theorem, since $\tilde{g}$ is continuously differentiable everywhere, $\tilde{g}(\hat{\theta}^*_{T_Z}) - \tilde{g}(\hat{\theta}_T)$ can be written as $\tilde{g}^\dagger_{\tilde{\theta}}(\hat{\theta}^*_{T_Z} - \hat{\theta}_T)$ for a point $\tilde{\theta}$ between $\hat{\theta}^*_{T_Z}$ and $\hat{\theta}_T$ if the latter two are in a ball around θ.

By the continuity of the derivative, there exists for every $\eta > 0$ a constant $\delta > 0$ such that $||\tilde{g}^\dagger_{\theta^*} h - \tilde{g}^\dagger_\theta h|| < \eta||h||$ for every h and every $||\theta^* - \theta|| \leq \delta$. If T_Z is sufficiently large, δ sufficiently small, $\sqrt{T_Z}||\hat{\theta}^*_{T_Z} - \hat{\theta}_T|| \leq L$ for some (large) number L, and $||\hat{\theta}_T - \theta|| \leq \delta$, then

$$\begin{aligned}
R_{T_Z} &:= \left|\left|\sqrt{T_Z}\left(\tilde{g}(\hat{\theta}^*_{T_Z}) - \tilde{g}(\hat{\theta}_T)\right) - \sqrt{T_Z}\left(\tilde{g}^\dagger_\theta(\hat{\theta}^*_{T_Z} - \hat{\theta}_T)\right)\right|\right| \\
&= \left|\left|\sqrt{T_Z}\left(\tilde{g}^\dagger_{\tilde{\theta}}(\hat{\theta}^*_{T_Z} - \hat{\theta}_T)\right) - \sqrt{T_Z}\left(\tilde{g}^\dagger_\theta(\hat{\theta}^*_{T_Z} - \hat{\theta}_T)\right)\right|\right| \\
&= \left|\left|\left(\tilde{g}^\dagger_{\tilde{\theta}} - \tilde{g}^\dagger_\theta\right)\sqrt{T_Z}(\hat{\theta}^*_{T_Z} - \hat{\theta}_T)\right|\right| \\
&\leq \eta L,
\end{aligned}$$

where the last inequality follows by setting $h = ||\sqrt{T_Z}(\hat{\theta}^*_{T_Z} - \hat{\theta}_T)||$.

Fix a number $\epsilon > 0$, and fix L. For η sufficiently small to ensure $\eta L < \epsilon$,

$$\begin{aligned}
\mathbb{P}\left(R_{T_Z} > \epsilon|\hat{\theta}_T\right) &\leq \mathbb{P}\left(R_{T_Z} > \eta L|\hat{\theta}_T\right) \\
&\leq \mathbb{P}\left(\sqrt{T_Z}||\hat{\theta}^*_{T_Z} - \hat{\theta}_T|| > L \text{ or } ||\hat{\theta}_T - \theta|| > \delta|\hat{\theta}_T\right) \\
&\leq \mathbb{P}\left(\sqrt{T_Z}||\hat{\theta}^*_{T_Z} - \hat{\theta}_T|| > L|\hat{\theta}_T\right) + \mathbb{P}\left(||\hat{\theta}_T - \theta|| > \delta|\hat{\theta}_T\right) \\
&\xrightarrow{a.s.} 0,
\end{aligned}$$

where the limiting result follows from Assumption 8 and the data-generating process for the bootstrap draws $\hat{\theta}^*_{T_Z}$ (cf. Equation (D.3)), and the fact that $\mathbb{P}\left(\sqrt{T_Z}||\hat{\theta}^*_{T_Z} - \hat{\theta}_T|| > L\right)$ can be made arbitrarily small by choice of L. Then an application of Slutsky's lemma implies

$$\sqrt{T_Z}\left(\tilde{g}(\hat{\theta}^*_{T_Z}) - \tilde{g}(\hat{\theta}_T)\right)\Big|\hat{\theta}_T \xrightarrow{d} \tilde{g}^\dagger_\theta(Q),$$

and therefore the quantiles of $\sqrt{T_Z}(g(A_m, \Gamma_m) - g(\hat{A}_T, \hat{\Gamma}_T))|\hat{A}_T, \hat{\Gamma}_T$ do indeed approximate the ones of $\sqrt{T_Z}(g(\hat{A}_{T_Z}, \hat{\Gamma}_{T_Z}) - g(A, \Gamma))$ as required. Finally, notice that by Slutsky's lemma, W can be replaced with a consistent estimator given by $\hat{W}_{T_Z}$. This completes the proof.

□

References

Adamek, Robert, Stephan Smeekes, and Ines Wilms (2020). "Lasso Inference for High-Dimensional Time Series ". In: *arXiv* 2007.10952, pp. 1–52.

Anatolyev, Stanislav (2012). "Inference in regression models with many regressors". In: *Journal of Econometrics* 170.2, pp. 368–382.

Anatolyev, Stanislav and Nikolay Gospodinov (2011). "Specification Testing in Models with Many Instruments". In: *Econometric Theory* 27.2, pp. 427–441.

Anatolyev, Stanislav and Pavel Yaskov (2017). "Asymptotics of Diagonal Elements of Projection Matrices Under Many Instruments/Regressors". In: *Econometric Theory* 33, pp. 717–738.

Anderson, Gary and George Moore (1985). "A Linear Algebraic Procedure for Solving Linear Perfect Foresight Models". In: *Economics Letters* 17.3, pp. 247–252.

Anderson, Theodore and Herman Rubin (1949). "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations". In: *The Annals of Mathematical Statistics* 20.1, pp. 46–63.

Andrews, Donald and James Stock (2007). "Testing with many weak instruments". In: *Journal of Econometrics* 138.1, pp. 24–46.

Ascari, Guido, Leandro Magnusson, and Sophocles Mavroeidis (2021). "Empirical evidence on the Euler equation for consumption in the US". In: *Journal of Monetary Economics* 117, pp. 129–152.

Bai, Jushan and Serena Ng (2010). "Instrumental Variable Estimation in a Data Rich Environment". In: *Econometric Theory* 26.6, pp. 1577–1606.

Ball, Laurence and Sandeep Mazumder (2019). "A Phillips Curve for the Euro Area". In: *NBER Working Paper* 26450, pp. 1–26.

Bayar, Omer (2018). "Weak instruments and estimated monetary policy rules". In: *Journal of Macroeconomics* 58, pp. 308–317.

Bekker, Paul and Federico Crudu (2015). "Jackknife instrumental variable estimation with heteroskedasticity". In: *Journal of Econometrics* 185.2, pp. 332–342.

Belloni, Alexandre, Daniel Chen, Victor Chernozhukov, and Christian Hansen (2012). "Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain". In: *Econometrica* 80.6, pp. 2369–2429.

Berriel, Tiago, Marcelo Medeiros, and Marcelo Sena (2016). "Instrument selection for estimation of a forward-looking Phillips Curve". In: *Economics Letters* 145, pp. 123–125.

— (2019). "Regularization and Identification of the New Keynesian Phillips Curve". In: *European Conferences of the Econometrics Community Conference Paper*, pp. 1–27.

Blandhol, Christine, John Bonney, Magne Mogstad, and Alexander Torgovitsky (2022). "When is TSLS *Actually* LATE?" In: *Becker Friedman Institute Working Papers* 2022-16, pp. 1–68.

Bonam, Dennis, Jakob de Haan, and Duncan van Limbergen (2021). "Time-varying wage Phillips curves in the euro area with a new measure for labor market slack". In: *Economic Modelling* 96, pp. 157–171.

Bulligan, Guido and Eliana Viviano (2017). "Has the wage Phillips curve changed in the euro area?" In: *IZA Journal of Labor Policy* 6.9, pp. 1–22.

Bun, Maurice, Helmut Farbmacher, and Rutger Poldermans (2020). "Finite sample properties of the GMM Anderson-Rubin test". In: *Econometric Reviews* 39.10, pp. 1042–1056.

Card, David (2009). "Immigration and Inequality". In: *American Economic Review Papers and Proceedings* 99.2, pp. 1–21.

Carrasco, Marine and Guy Tchuente (2016). "Regularization Based Anderson Rubin Tests for Many Instruments". In: *University of Kent, School of Economics Discussion Papers* 1608, pp. 1–34.

Carvalho, Carlos, Fernanda Nechio, and Tiago Tristao (2021). "Taylor Rule Estimation by OLS". In: *Journal of Monetary Economics* 124, pp. 140–154.

Chao, John, Norman Swanson, Jerry Hausmann, Whitney Newey, and Tiemen Woutersen (2012). "Asymptotic Distribution of JIVE in a Heteroskedastic IV Regression With Many Instruments". In: *Econometric Theory* 28, pp. 42–86.

Chernozhukov, Victor, Denis Chetverikov, and Kengo Kato (2018). "Inference on Causal and Structural Parameters using Many Moment Inequalities". In: *The Review of Economic Studies* 86.5, pp. 1867–1900.

Chernozhukov, Victor, Christian Hansen, and Martin Spindler (2015). "Post-Selection and Post-Regularization Inference in Linear Models with Many Controls and Instruments". In: *American Economic Review* 105.5, pp. 486–490.

Chernozhukov, Victor, Wolfgang Härdle, Chen Huang, and Weining Wang (2020). "LASSO-Driven Inference in Time and Space". In: *The Annals of Statistics (forthcoming)*, pp. 1–76.

Crudu, Federico, Giovanni Mellace, and Zsolt Sándor (2021). "Inference in Instrumental Variable Models with Heteroskedasticity and Many Instruments". In: *Econometric Theory* 37.2, pp. 281–310.

Cunningham, Rose, Vikram Rai, and Kristina Hess (2019). "Exploring Wage Phillips Curves in Advanced Economies". In: *Bank of Canada Staff Discussion Paper*, pp. 1–27.

Del Negro, Marco, Giorgio Primiceri, Michele Lenza, and Andrea Tambalotti (2020). "Whats Up with the Phillips Curve?" In: *Brookings Papers on Economic Activity*, pp. 301–373.

Dufour, Jean-Marie (2003). "Identification, Weak Instruments, and Statistical Inference in Econometrics". In: *The Canadian Journal of Economics* 36.3, pp. 767–808.

Dufour, Jean-Marie, Lynda Khalaf, and Maral Kichian (2006). "Inflation dynamics and the New Keynesian Phillips Curve: An identification robust econometric analysis". In: *Journal of Economic Dynamics and Control* 30.9-10, pp. 1707–1727.

Eser, Fabian, Peter Karadi, Philip Lane, Laura Moretti, and Chiara Osbat (2020). "The Phillips Curve at the ECB". In: *The Manchester School* 88.S1, pp. 50–85.

Gabaix, Xavier and Ralph Koijen (2020). "Granular Instrumental Variables". In: *NBER Working Paper Series* 28204, pp. 1–98.

Galì, Jordi (2015). "Hysteresis and the European Unemployment Problem Revisited". In: *NBER Working Paper Series* 21430, pp. 1–51.

Galì, Jordi and Luca Gambetti (2019). "Has the U.S. Wage Phillips Curve Flattened? A Semi-Structural Exploration". In: *NBER Working Paper Series*, pp. 1–34.
Galì, Jordi and Mark Gertler (1999). "Inflation dynamics: A structural econometric analysis". In: *Journal of Monetary Economics* 44.2, pp. 195–222.
Gertler, Mark and Peter Karadi (2015). "Monetary Policy Surprises, Credit Costs, and Economic Activity". In: *American Economic Journal: Macroeconomics* 7.1, pp. 44–76.
Giannone, Domenico, Michele Lenza, and Giorgio Primiceri (2021). "Economic Predictions with Big Data: The Illusion of Sparsity". In: *Econometrica* 89.5, pp. 2409–2437.
Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift (2020). "Bartik Instruments: What, When, Why, and How". In: *American Economic Review* 110.8, pp. 2586–2624.
Gordon, Robert (1981). "Inflation, Flexible Exchange Rates, and the Natural Rate of Unemployment". In: *NBER Working Paper Series* 708, pp. 1–99.
— (2011). "The History of the Phillips Curve: Consensus and Bifurcation". In: *Economica* 78.309, pp. 10–50.
Gürkaynak, Refet, Brian Sack, and Eric Swanson (2005). "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements". In: *International Journal of Central Banking* 1.1, pp. 55–93.
Hall, Peter and Chris Heyde (1980). *Martingale Limit Theory and Its Application*. Academic Press.
Hamilton, James (2018). "Why You Should Never Use the Hodrick-Prescott Filter". In: *Review of Economics and Statistics* 100.5, pp. 831–843.
Hansen, Christian and Damian Kozbur (2014). "Instrumental variables estimation with many weak instruments using regularized JIVE". In: *Journal of Econometrics* 182.2, pp. 290–308.
Horowitz, Joel (2018). "Non-Asymptotic Inference in Instrumental Variables Estimation". In: *arXiv* 1809.03600, pp. 1–33.
Jentsch, Carsten and Kurt Lunsford (2019). "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States: Comment". In: 109.7, pp. 2655–2678.
Kaffo, Maximilien and Wenjie Wang (2017). "On bootstrap validity for specification testing with many weak instruments". In: *Economics Letters* 157, pp. 107–111.
Kapetanios, George, Lynda Khalaf, and Massimiliano Marcellino (2015). "Factor-Based Identification-Robust Interference in IV Regressions". In: *Journal of Applied Econometrics* 31.5, pp. 821–842.
Kiviet, Jan (2022). "Instrument-free inference under confined regressor endogeneity and mild regularity". In: *Econometrics and Statistics*.
Kiviet, Jan and Sebastian Kripfganz (2020). "Reassessment of classic case studies in labor economics with new instrument-free methods". Unpublished manuscript.
Kleibergen, Frank (2002). "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression". In: *Econometrica* 70.5, pp. 1781–1803.
— (2005). "Testing Parameters in GMM Without Assuming that They Are Identified". In: *Econometrica* 73.4, pp. 1103–1123.
Kleibergen, Frank and Sophocles Mavroeidis (2009). "Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve". In: *Journal of Business & Economic Statistics* 27.3, pp. 293–311.
Lütkepohl, Helmut (1996). *Handbook of Matrices*. Wiley.

Ma, Adrian (2002). "GMM estimation of the new Phillips curve". In: *Economics Letters* 76.3, pp. 411–417.

Mavroeidis, Sophocles (2005). "Identification Issues in Forward-Looking Models Estimated by GMM, with an Application to the Phillips Curve". In: *Journal of Money, Credit and Banking* 37.3, pp. 421–448.

— (2010). "Monetary Policy Rules and Macroeconomic Stability: Some New Evidence". In: *American Economic Review* 100.1, pp. 491–503.

Mavroeidis, Sophocles, Mikkel Plagborg-Møller, and James Stock (2014). "Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve'". In: *Journal of Economic Literature* 52.1, pp. 124–188.

Mazumder, Sandeep (2014). "Inflation in China: Old Versus New Phillips Curves". In: *Europe-Asia Studies* 66.5, pp. 659–709.

McCracken, Michael and Serena Ng (2016). "FRED-MD: A Monthly Database for Macroeconomic Research". In: *Journal of Business & Economic Statistics* 34.4, pp. 574–589.

McLeay, Michael and Silvana Tenreyro (2019). "Optimal Inflation and the Identification of the Phillips Curve". In: *NBER Working Paper Series* 25892, pp. 1–56.

Mertens, Karel and Morten O Ravn (2013). "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States". In: *American Economic Review* 103.4, pp. 1212–1247.

Mikusheva, Anna (2021). "Many Weak Instruments in Time Series Econometrics". Unpublished manuscript.

Mikusheva, Anna and Liyang Sun (2022). "Inference with Many Weak Instruments". In: *Review of Economic Studies (forthcoming)*, pp. 1–37.

Mirza, Harun and Lidia Storjohann (2014). "Making Weak Instrument Sets Stronger: Factor-Based Estimation of Inflation Dynamics and a Monetary Policy Rule". In: *Journal of Money, Credit and Banking* 46.4, pp. 643–664.

Montiel Olea, José Luis and Carolin Pflueger (2013). "A Robust Test for Weak Instruments". In: *Journal of Business & Economic Statistics* 31.3, pp. 358–369.

Montiel Olea, José Luis, James Stock, and Mark Watson (2021). "Inference in Structural Vector Autoregressions Identified With an External Instrument". In: *Journal of Econometrics* 225.1, pp. 74–87.

Newey, Whitney and Frank Windmeijer (2009). "Generalized Method of Moments With Many Weak Moment Conditions". In: *Econometrica* 77.3, pp. 687–719.

Ng, Serena and Jushan Bai (2009). "Selecting Instrumental Variables in a Data Rich Environment". In: *Journal of Time Series Econometrics* 1.1, pp. 1–34.

Nickel, Christiane, Elena Bobeica, Gerrit Koester, Eliza Lis, and Mario Porqueddu (2019). "Understanding low wage growth in the euro area and European countries". In: *European Central Bank Occasional Paper Series*, pp. 1–88.

Paul, Pascal (2020). "The Time-Varying Effect of Monetary Policy on Asset Prices". In: *The Review of Economics and Statistics* 102.4, pp. 690–704.

Phillips, Peter C B and Wayne Yuan Gao (2017). "Structural inference from reduced forms with many instruments". In: *Journal of Econometrics* 199.2, pp. 96–116.

Rey, Hélène (2016). "International Channels of Transmission of Monetary Policy and the Mundellian Trilemma". In: *IMF Economic Review* 64.1, pp. 6–35.

Stock, James and Jonathan Wright (2000). "GMM with Weak Identification". In: *Econometrica* 68.5, pp. 1055–1096.

van der Vaart, Adrianus Willem (2007). *Asymptotic Statistics*. Cambridge University Press.

Wong, Kam Chung, Zifan Li, and Ambuj Tewari (2020). "Lasso Guarantees for β-Mixing Heavy Tailed Time Series ". In: *Annals of Statistics* 48.2, pp. 1124–1142.

Yogo, Motohiro (2004). "Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak". In: *The Review of Economics and Statistics* 86.3, pp. 797–810.