Essays in Empirical Macroeconomics and Finance

Martin McCarthy
Merton College

Word count: 83,030 words

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2023

Acknowledgements

I'd like to thank my DPhil supervisors, Jennifer Castle, Stephen Bond and Michael Devereux, and my MPhil supervisor Sarah Clifford. Their advice has greatly improved my thesis and made me a better researcher. I am indebted to my co-authors, Stephen Snudden, Prakash Loungani and Umberto Collodel, from whom I have learned a great deal. I am also grateful to the Reserve Bank of Australia for funding my post-graduate studies. Finally, I'd like to thank my parents, who raised me to value education, and my brothers, who have always been there to support me.

Contents

Introduction

This thesis comprises three self-contained chapters. All relate in some way to expectations about macroeconomic or financial variables.

The first chapter provides empirical evidence on how one class of agent, international organisations, forms expectations about a key variable, real Gross Domestic Product growth. It compares the expectations of four different organisations, finding few statistically significant differences in accuracy. It then tests the rationality (or efficiency) of their forecasts. For all four organisations, the null of rationality is rejected for most countries. Finally, this chapter explores three possible explanations for the rejection of these rationality tests: the use of conditioning assumptions; the practice of making modal forecasts; and incentives faced by international organisations to make biased forecasts.

The second chapter provides evidence on how one should form expectations for bilateral exchange rates or effective exchange rates. International organisations, central banks and other practitioners often need to form expectations for month-average or quarter-average exchange rates. We show that the extensive literature on forecasting these exchange rates suffers from major limitations. First, the literature tests if these exchange rates are predictable by comparing forecasting models to a benchmark forecast, but use the wrong benchmark. Specifically, they compare the models to the assumption that the period-average will remain constant. However, gains relative to this benchmark are expected *by construction*. Instead, one should compare the models to the assumption that the period-average in the future will equal the current end-of-period level. We address this limitation by performing the first test of the predictability of period-average exchange rates against the correct benchmark. Second, the literature only uses models that are estimated on period-average data, rather than end-of-period or daily data. We show that both recursive and direct forecasting models perform much better with end-of-period or daily data.

The final chapter takes it as given that share prices reflect the forward-looking expectations of share market participants. Under this assumption, if share prices will be affected by a policy, and the announcement of the policy makes it likely that the policy will be implemented, then share prices will respond as soon as the policy is announced. This chapter tests if Australian share prices responded to the announcement of a change to Australia's system for taxing dividends, finding no evidence of a response. This policy change would have increased tax paid by Australian shareholders, but not foreign shareholders. The absence of a share price response is consistent with the hypothesis that share prices in a small economy like Australia's are only determined by foreign investors. This has important implications for the effect of dividend taxes on investment and for the economic incidence of dividend taxes.

1

Real GDP Forecasts by International Organisations

Martin McCarthy, Umberto Collodel[1], *Prakash Loungani*[2,3]

Abstract

We study forecasts for real Gross Domestic Product growth by international organisations. These forecasts influence policy decisions at international organisations and national governments, and receive significant attention from the private-sector. We make three contributions. First, we compare the accuracy of forecasters and test the rationality of forecasts using a large dataset of forecasts by four international organisations (the International Monetary Fund, the World Bank, the European Commission and the Organisation for Economic Cooperation and Development). Second, we derive a new decomposition of differences in forecast accuracy into contributions from disagreement and the accuracy of the average forecast. Finally, we present theoretical results and empirical evidence on three explanations for the rejection of rationality tests: the use of conditioning assumptions, the practice of making modal forecasts, and incentives to make biased forecasts for countries receiving assistance from the organisation.

[1]Central Bank of Malta

[2]Johns Hopkins University

[3]We thank Zidong An for providing the initial dataset and helpful discussions. We also thank conference participants at the 42nd International Symposium on Forecasting for their comments and suggestions.

1.1 Introduction

International organisations put substantial resources into forecasting real Gross Domestic Product (GDP) growth. The forecasting processes at these organisations take multiple weeks of work by dozens of economists. This high amount of effort reflects the importance of the forecasts for the decisions of international organisations and others, and hence their importance for economic outcomes more broadly.

International organisations use their forecasts to inform key policy decisions. The International Monetary Fund (IMF) and World Bank draw on the published forecasts for a country when deciding whether to make a loan or grant to that country. An overly optimistic forecast could, for instance, cause the IMF to provide too little assistance, resulting in a country experiencing more severe drops in output, employment and living standards. The European Commission (EC) uses its growth forecasts when setting its annual budget. If the EC is unable to accurately forecast growth, it will have difficulty using its budget to offset business cycle fluctuations, and may even amplify them. The IMF, World Bank, EC and Organisation for Economic Cooperation and Development (OECD) all rely on their growth forecasts when giving macroeconomic policy advice to national governments. Poor quality forecasts could therefore lead to poor advice, which may lead to poor policies. This can occur even if national governments are sceptical of the advice. The IMF often requires a number of specific policy actions as a condition of receiving loans (Independent Evaluation Office 2008).

International organisation forecasts are also important due to their effects on external users. This includes national governments, whose own forecasts are informed by those made by the organisations. A survey of national governments shows that a substantial majority agree that IMF forecasts 'are valuable inputs to the economic policy process in my country' (Genberg and Martinez 2014b). Other external users include financial market participants, with JP Morgan and Citigroup both reporting that clients pay attention to IMF forecasts (Genberg, Martinez, and Salemi 2014). Since international organisation forecasts influence the expectations of both governments and the private-sector, they affect investment and other decisions.

Empirical work finds they have causal effects on economic outcomes. Afonso (2010) shows that EC growth forecasts influence bond yields, while Beaudry and Willems (2022) find that if the IMF makes overly optimistic growth forecasts, this causes an economic contraction later, which they attribute to governments and the private-sector accumulating excessive debt in response to the optimistic forecast.

Both academic researchers and international organisations have investigated the accuracy and rationality of the forecasts. Comparisons of accuracy hold forecasters accountable, and assist users in choosing how much attention to pay to different forecasters. Tests of rationality, together with explanations for the failure of rationality tests, help forecasters assess how they can improve their forecasts. This chapter makes three contributions to that literature.

Firstly, we compare forecast accuracy and test rationality using a large dataset of forecasts. The dataset comprises vintages of forecasts by four international organisations (the IMF, the World Bank, the EC and the OECD) and for reference, the average forecast of respondents to the Consensus survey. One strength of our dataset is that it covers as many countries as possible, allowing us to identify features of the forecasts that are general, rather than specific to particular groups of countries. Another strength is that it covers forecasts as recently as 2023, which makes the results relevant to current practice at international organisations. For all pairs of forecasters analysed, we find statistically significant differences in accuracy for only a few countries. In contrast, for each forecaster, we find statistically significant departures from forecast rationality for the majority of countries. The nature of this irrationality varies by country, but often takes the form of optimistic bias, revisions that are too large, or forecasts that are too extreme.

Secondly, we extend the literature on comparing accuracy by deriving a new decomposition. We show that the difference in squared errors of a pair of forecasts can be decomposed into contributions from disagreement between the forecasters and the accuracy of the average forecast. Forecast accuracy tends to differ by more at longer horizons. Using our decomposition, we show that this occurs because of lower accuracy at longer horizons, and to a lesser extent, more disagreement at longer horizons.

Finally, we investigate three explanations for the failure of rationality tests that may be especially relevant to international organisations. One explanation is that international organisations often make forecasts that are conditional on future outcomes of a set of explanatory variables. We derive conditions under which this results in the failure of rationality tests. Another explanation is that organisations may make 'modal' forecasts, which means they report the most likely outcome rather than the expected outcome. We present empirical evidence on how much bias this could introduce into the forecasts. Finally, international organisations may face incentives to make deliberately biased forecasts for some countries. Previous literature found that the IMF tends to make optimistic forecasts for countries to which it provides loans and grants, and has interpreted this as evidence of deliberate bias. We find that IMF forecasts are optimistic for program countries, and that the optimism is greater for larger programs. However, we find that the same is true for Consensus, even though private-sector forecasters do not have the same incentives to make biased forecasts. This raises doubt about deliberate bias, but cannot rule it out.

The remainder of this chapter is organised as follows. Section 1.2 reviews the literature, while section 1.3 describes our dataset. Section 1.4 compares the accuracy of the forecasters, including presenting our new decomposition of differences in forecast accuracy. The results of rationality tests are presented in section 1.5. Section 1.6 investigates three explanations for the rejection of rationality tests. Finally, section 1.7 concludes. Supplementary material is provided in appendix A.

1.2 Literature Review

1.2.1 Comparing Accuracy across Forecasters

A number of papers have compared the accuracy of real GDP growth forecasts by different international organisations. Some papers compare measures of forecast performance, but do not formally test for differences in accuracy. For example, Hong and Tan (2014) find that the United Nations growth forecasts performed slightly better than those by the IMF and World Bank over 2000 to 2012. Fioramanti et al. (2016) find that EC growth forecasts performed similarly to the OECD and better than the European Central Bank, IMF or Consensus over 2000 to 2014. Other papers formally test for accuracy. Timmermann (2007) finds that there a few countries where IMF and Consensus had statistically significant differences in accuracy. Melander et al. (2007) also find few countries where the IMF, OECD, Consensus or EC had significant differences in performance, using a sample of European countries from 1990 to 2005. Most recently, Abreu (2011) use a sample of 9 advanced economies over 1991 to 2009, and find few statistically significant differences in accuracy when comparing international organisations (IMF, EC, OECD) to private-sector forecasters (Consensus or The Economist).

The lack of statistically significant differences in accuracy in the previous literature may reflect the small sample sizes available. This argument was made, for instance, by Timmermann (2007). Our chapter includes more recent forecasts than earlier studies, increasing our sample size. Another explanation for the lack of significant differences is that the forecasts themselves tend to be quite similar. This similarity could arise because the forecasters independently make similar forecasts to one another due to having similar information sets. It could also arise if forecasters deliberately adjust their forecasts to be closer to others. The empirical evidence on whether forecasters deliberately 'herd' by making their forecasts more similar to one another is mixed (Clements 2018).

1.2.2 Forecast Rationality

The literature describes a number of related notions of rationality. We follow Elliott, Komunjer, et al. (2008) in defining a forecaster to be 'rational' if their forecasts minimise the conditional expectation of the forecaster's loss function given their information set. To formalise this, define:

- y_{t+h} is realised real GDP growth in period $t + h$
- $y_{t+h|t}$ is a forecast made in period t for real GDP growth in period $t + h$
- $v_{t+h|t} \equiv y_{t+h} - y_{t+h|t}$ is the forecast error, which is the actual less the forecast.
- $L : \mathbb{R} \to \mathbb{R}_{\geq 0}$ is the forecaster's loss function
- $\mathcal{I}_t$ is the forecaster's information set at time t

The forecaster is rational if:

$$y_{t+h|t} = \arg\min_{z} E\left[L(y_{t+h} - z)|\mathcal{I}_t\right] \quad \forall t$$

Here, the conditional expectation is computed using the conditional probabilities implied by the true data generating process (DGP). This is quite a stringent definition, as it assumes that the forecaster has a correctly specified model. Any mis-specification, such as the forecaster viewing growth as stationary when it is instead subject to breaks in the mean, would violate forecast rationality.

The literature often focusses on the special case where the forecaster's loss function is the squared error, $L(v_{t+h}) = v_{t+h}^2$. Jointly assuming forecast rationality and a squared error loss function is equivalent to assuming that the forecast equals the conditional expectation of real GDP growth given the forecaster's information set. This is called 'strong efficiency' by Nordhaus (1987).

$$y_{t+h|t} = E\left[y_{t+h}|\mathcal{I}_t\right] \quad \forall t$$

If the forecaster's forecasts equal the conditional expectation of growth given their information set, then the forecast errors must be orthogonal to any vector

$\boldsymbol{x}_t$ of variables known at time t. This orthogonality condition is called 'strong rationality' by Stekler (2004) and is the basis for the regression-based rationality tests common in the literature.[4] If the forecast error satisfies the linear regression model (1.1), where u_{t+h} is a mean-zero error, then orthogonality states that the parameter vector $\boldsymbol{\theta}$ is the zero vector.

$$v_{t+h|t} = \boldsymbol{x}_t\boldsymbol{\theta} + u_{t+h} \quad \forall t \tag{1.1}$$

The literature on growth forecasts contains various regressions of forecast errors on information available at the time of the forecast. Much of this literature tests the rationality of the average of private-sector forecasters (Coibion and Gorodnichenko 2012, 2015a) or individual private-sector forecasters (Bordalo, Gennaioli, Ma, et al. 2020; Broer and Kohlhas 2022). However, a number of papers test the rationality of international organisations forecasts for multiple countries, as we do.[5] These papers include rationality regressions with a variety of explanatory variables, including:

- Regressions of forecast errors on an intercept to test for bias. The results are mixed. For the IMF, some papers have found an optimistic bias for many countries (Artis 1996; Timmermann 2007), while others have found optimistic biases for some countries and pessimistic biases for others (Beach 1999; Takagi and Kucur 2006). For the OECD, Vogel (2007) finds that current-year forecasts are unbiased, while year-ahead forecasts have an optimistic bias. Tsuchiya (2023) finds that the same is true of World Bank forecasts since the global financial crisis. Finally, for the EC, Melander et al. (2007) finds no evidence of bias for all but a few countries.

- Regressions of forecast errors on past revisions to test if the forecaster tends to revise forecasts by too much or too little. Ashiya (2006) performs this test on

[4]The orthogonality condition rules out linear relationships between forecast errors and variables in the information set. This is weaker than jointly assuming forecast rationality and squared error loss, which rules out *any* relationship between forecast errors and variables in the information set.

[5]Some papers study international organisation forecasts for a single country in detail, including Jerić et al. (2020) (for Croatia), Ziembińska (2021) (for Poland), Baumgartner (2002) (for Austria) and Tsuchiya (2016) (for Japan).

IMF and OECD forecasts for Group of 7 countries, rejecting rationality in 40% of cases, typically because the revisions are too small.

- Regressions of forecast errors on an intercept and forecast. This can be used to test if the forecasts are 'too extreme' in the sense that they should be scaled up or scaled down in magnitude. Equivalently, one can regress actual outcomes on an intercept and forecast, as in Mincer and Zarnowitz (1969).[6] Tsuchiya (2023) finds that this test is rejected for one-year-ahead forecasts by the World Bank, while Vogel (2007) finds the same for the OECD.

- Regressions of forecast errors on errors known at the time. If forecast errors are positively correlated with past forecast errors known to the forecaster, then the forecaster is failing to learn from their mistakes. Timmermann (2007) finds that IMF forecast errors have positive sample autocorrelations, but do not formally test the null of no autocorrelation. Melander et al. (2007) does not reject the null of rationality for EC forecasts for most countries.

We test the rationality of multiple international organisations for as many countries as possible using forecasts made as recently as 2023. In contrast, much of the literature tends to test the rationality of only one organisation. A notable exception is Abreu (2011), which tested the rationality of forecasts for 9 advanced economies over 1991 to 2009. A key takeaway from Abreu (2011) is that the departures from rationality tend to be similar across forecasters.[7] This is useful, but leaves open the possibility that their findings are specific to these economies during the pre-financial-crisis period. By covering more countries over a longer time period, we can have more confidence that we have identified general features of international organisation forecasts.

[6]Mincer and Zarnowitz (1969) regress the actual outcome on an intercept and the forecast, and test if the intercept is 0 and the coefficient on the forecast is 1. This is equivalent to regressing the forecast error on an intercept and forecast and then testing if both the intercept and coefficient on the slope are 0.

[7]For example, for Spring next-year forecasts, Abreu (2011) finds that for all three of the IMF, EC and OECD, they can reject unbiasedness for Germany, France and Italy at a 10% level, but not for any of the other six countries. Similarly, they find that for the IMF and EC, they can reject the null of no autocorrelation in the errors for Italy and Spain at a 10% level, but not for other countries.

The literature also studies the consequences of departures of rationality for the forecasts. For example, Clements (2022), finds evidence that inefficiency (i.e. irrationality) in forecasts contributes to persistent differences in accuracy across forecasters. Some literature studies the implications of irrationality for the behaviour of the economy. Bordalo, Gennaioli, and Shleifer (2018) show that a model of irrationality called 'diagnostic expectations' can explain credit cycles, while Bordalo, Gennaioli, Kwon, et al. (2021) shows that this model can explain asset price bubbles.

1.2.3 Explaining Rejection of Rationality Tests

The literature provides a wide variety of explanations for the rejection of the null of rationality in standard forecast rationality tests. These explanations are in three broad categories: alternative loss functions; information rigidities; or forecasters not minimising a loss function.

1.2.3.1 Loss Functions other than Squared Errors

A forecaster may be rational, but be minimising a loss function other than squared errors. Asymmetry in the loss function could explain optimistic or pessimistic bias in the forecasts (Elliott, Komunjer, et al. 2008). For example, if the forecaster is especially averse to under-predicting growth, then they will tend to make forecasts that are too optimistic from the perspective of a symmetric loss function. While asymmetry in the loss function could explain bias in the forecasts, it is less obvious how a loss function could explain other reasons for the failure of the rationality tests. Specifically, it is unclear if there is a sensible loss function such that a forecaster with rational expectations would make forecast errors that are correlated with their own forecasts, their own forecast revisions, or with their past forecast errors. One response would be to view the forecaster as having some highly unusual loss function. However, for the rational expectations model to be meaningful, it must impose some restriction on forecaster' behaviour. If we claim that forecasters act as if they make forecasts to minimise a loss function, we have to restrict our attention to certain kinds of loss functions (e.g. losses that depend only on forecast errors) in order for this theory to have any empirical content.[8] For this reason, if the forecasts have strange properties such as being correlated with past errors, it is useful to describe the forecaster's behaviour using a model other than rational expectations.

It is difficult to provide empirical evidence on the loss functions of international organisations.[9] While it is not possible to test the loss function in isolation, it is possible to jointly test the assumption of forecast rationality and a particular loss function. Elliott, Komunjer, et al. (2008) shows that one can do this by regressing 'generalised forecast errors' on the vector of variables known at the time $\boldsymbol{x}_t$. However, if one departs from the standard assumption of squared error loss, it is not obvious what specific alternative loss function one would choose. Elliott, Komunjer, et al. (2008) propose assuming the forecaster has a loss function in a flexible family indexed by an asymmetry parameter α and an exponent p. One can then test the joint hypothesis that the forecaster has rational expectations and a loss function in this family. This weaker assumption about the loss function comes at a price. Since the loss function has additional parameters to be estimated, the power of the test to detect departures from rational expectations is reduced relative to the standard rationality tests, which is concerning given the modest sample sizes of macroeconomic forecasts available.

In this chapter we investigate three explanations for the failure of rationality tests by international organisations. One of those explanations is that the forecaster aims to report the conditional mode of future growth, rather than the conditional mean. Reporting the conditional mode is rational if one's loss function applies no penalty to being exactly correct (to the published precision of the forecasts, usually 1 decimal place) and applies the same penalty to being incorrect by any amount. If the conditional distribution if symmetric and unimodal, then the mean and mode

[8] For example, if we observed that a forecaster always set its growth forecast equal to the temperature in Washington DC on the forecast date, we could rationalise those forecasts by assuming their loss function is the deviation of the forecast from the temperature in DC on the forecast date.

[9] One piece of suggestive evidence is that international organisations often use root mean square forecast error (RMSFE) to evaluate their own forecasts, as in Genberg and Martinez (2014a). This suggests that the staff conducting the evaluations have a squared error loss function, though it is possible that the staff producing the forecasts have a different loss function.

will coincide, but otherwise, the mean and mode may differ. Two papers have tried to explain optimistic bias in growth forecasts in a similar way. Firstly, Kangur et al. (2019) finds that the IMF's World Economic Outlook (WEO) forecasts for the euro area have an optimistic bias. They suggest that WEO may be making modal forecasts, which will tend to be higher than mean forecasts because the distribution of growth outcomes is negatively skewed. Secondly, Burgess et al. (2021) finds that the distribution of real GDP per capita growth is negatively skewed in most countries, and that this could in principle explain the bias.[10] They then show that, empirically, mean growth and median growth tend to be similar, and argue on this basis that the bias in WEO forecasts cannot be fully explained by the skew of growth outcomes.[11]

This chapter advances on these previous efforts in two ways.

- We are the first to quantify the effect of making modal rather than mean forecasts on bias. Burgess et al. (2021) did a similar exercise, but they compared the mean with the *median* instead. However, this provides little information about the effect of making modal forecasts, because even when the mean and median are similar, the mode could be quite different. For example, if growth follows a skew-normal distribution with location 2, scale 1 and shape -5, then the mean will be 1.2, the median 1.3 and the mode 1.6.

- We argue that the practice of making modal forecasts could lead to bias (as noted by Kangur et al. (2019)), but we additionally argue this practice could cause the failure of rationality tests by causing the forecast errors to be correlated with variables known on the forecast date.

[10] Bekaert and Popov (2019) also find negative skewness in most countries, but find positive skewness in many others. They argue that low income countries may have positive skewness because they experience occasional 'growth spurts', such as an increase in growth due to the discovery of oil.

[11] Modal forecasts have also been explored in the context of nominal interest rates forecasting, where the presence of the zero lower bound implies that the distribution of future interest rates is right-skewed, so the mean forecast will exceed the modal forecast (Bauer and Rudebusch 2016; Dison and Elliott 2015).

1.2.3.2 Information Rigidities

Suppose that a rationality test has found that forecast errors are correlated with variables $\boldsymbol{x}_t$ that are in-principle knowable at the time the forecast was made. One possibility is that the forecast had rational expectations, but those variables were not in their information set. Many macroeconomic models assume 'full information rational expectations' (FIRE), where each forecaster's information set contains all information about the past and present. However, many models assume that the forecaster has rational expectations but is affected by information rigidities. There are a wide variety of such models. In 'sticky information' models, forecasters update their information sets and hence forecasts infrequently (Mankiw and Reis 2002; Mankiw, Reis, and Wolfers 2004). In 'noisy information' models, forecasters update their information sets in every period, but only receive noisy signals about the world (Sims 2003; Woodford 2003). Both sticky and noisy information models can explain the empirical observation that the average forecast of a set of forecasters tends to be revised too little. Recently, the noisy information models have been extended so that they can also explain the tendency of individual forecasters to revise their forecasts too much. For example, Kohlhas and Robertson (2022) assume forecasters don't know the accuracy of the signals, so must estimate the weight to place on them. Kohlhas and Walther (2021) assume that the forecasters receive signals about different components of output, and the accuracy of these signals can differ by component.

1.2.3.3 Forecasts do not minimise loss given information

A forecaster may fail rationality tests because their forecasts do not minimise a loss function given their information set. This could occur because the forecaster intends to minimise the expected loss given their information set, but fails to do so because they suffer from psychological biases. For example, they may be systematically overconfident in their private information (Broer and Kohlhas 2022).

Alternatively, they may fail rationality tests because, even if the forecasters are capable of forming rational forecasts, they choose their public forecasts to pursue an objective that cannot be described as minimising the expected loss from forecast errors. Two of the explanation we give for the failure of rationality tests are in this category.

One such explanation is that forecasters often make 'conditional forecasts', defined as forecasts that are conditional on assumed future paths for a set of explanatory variables. We characterise the circumstances under which making conditional forecasts will cause standard rationality tests to be rejected. We show that these tests may be rejected because the assumptions about the explanatory variable may be irrational, or because the 'response function' that relates growth to the explanatory variables has certain properties, such as strict convexity or strict concavity.

A related paper, Faust and Wright (2008), develops a method that tests if a set of observed growth forecasts are rational given a set of observed interest rate assumptions. Their approach is useful, as forecasters who make conditional forecasts will be interested to know if they are making rational conditional forecasts. However, their method can only be applied where the assumptions about the explanatory variable are observed, which is not true of some forecasters, and not true of assumptions about qualitative developments such as a war. Moreover, their method can only be used when the relationship between growth and the explanatory variable is linear, which may not be the case for some variables, such as the climate. For this reason, researchers are likely to continue running standard rationality tests on conditional forecasts. In such cases, the theoretical results in this chapter provide guidance on whether the conditional nature of the forecasts can explain the failure of these standard tests.[12]

The third explanation for the failure of rationality tests that we investigate is that international organisations may have incentives to make biased forecasts for certain countries. We study the specific case of whether the IMF makes optimistic forecasts for countries to which it provides loans or grants. The IMF may have a few reasons to do this. First, since the Fund becomes a 'de facto' creditor when intervening in a country with a program, a higher forecast increases the likelihood the country will

[12]Engelke et al. (2019), Glas and Heinisch (2023), Page and Lambrias (2019) and Lewis and Pain (2014) all show that the accuracy of the assumptions affects the accuracy of the growth forecasts, but do not consider how the practice of making conditioning assumptions affects rationality.

attract new foreign capital, thus increasing its chances of repayment. Second, unlike the IMF's WEO forecasts, the program forecasts are the result of discussions and a subsequent agreement between Fund officials and country authorities: it follows that "if the authorities are leaning towards very optimistic projections, the final result of the process will be biased in that direction [too]" (Luna 2014).

The literature suggests mixed evidence on the matter. The database of IMF forecasts for countries receiving loans or grants is divided into an early sample (up to 2003) and a later sample, owing to classification and other changes. Using the early sample, Musso and Phillips (2002) finds IMF forecasts have a statistically significant optimistic bias, while Baqir et al. (2005) does not. Using the later sample, Luna (2014) finds statistically significant optimistic bias, but Eicher et al. (2019) does not, though their point estimates are suggestive of optimistic bias. We use the later sample too, but find statistically significant optimistic bias. One reason for this difference is that Eicher et al. (2019) use a sample of 156 programs while we use a larger sample of 214 programs. Our key contribution on this topic is to show that the Consensus forecasts for countries receiving large IMF loans and grants are just as optimistic as the IMF ones. This is evidence against deliberate bias, since the private-sector lacks the incentives to make biased forecasts discussed above.

1.3 Data

In this chapter, we examine annual real GDP growth forecasts at horizons of up to 2 years ahead. To this end, we assemble a large dataset incorporating the forecasts of the most prominent international organisations and the average of private-sector forecasters. In this section, we describe the different databases we combine in our analysis, clarify the terminology we use in the rest of the chapter and discuss the issues that can arise when comparing forecasts from different sources.

A strength of our dataset is that it covers as many countries as possible, and contains forecast vintages as recently as June 2023. The large set of countries allows us to draw conclusions about international organisation forecasts in general, rather than worrying that our conclusions only apply to forecasts for a specific unusual group of countries, such as Group of 7 countries. The inclusion of recent forecast vintages that were unavailable to previous researchers increases our sample sizes, which increases the relevance of our results to current practice at international organisations, and increases the power of our tests.

1.3.1 Forecast Vintages

We use short-term forecasts for annual real GDP growth from four international organisations:

- The IMF publishes forecasts in the WEO. A full WEO is released in each April and each October, which we call the 'Spring' and 'Fall' editions respectively, and provides forecasts for all 192 countries that make up the IMF membership. A WEO update is published in each January and June, which provides forecasts for selected countries and for country groups. We retrieve forecasts from the IMF's internal WEO database, as the public version of the database only contains forecasts from full WEOs.

- The World Bank publishes forecasts in the *Global Economic Prospects* (GEP) publication each January and July. The GEP provides forecasts for emerging and developing economies, and for selected advanced economies. These forecasts are not publicly available in an electronic format.

- The EC publishes forecasts for European Union member states and for selected other economies each Spring and Autumn, and sometimes at other times as well. We use the publicly available AMECO database.

- The OECD publishes forecasts for OECD members and major non-members in the *Economic Outlook* publication each December and June, and occassionally at other times. We use the publicly available OECD Economic Outlook database.

We also use the average of the forecasts made by respondents to the Consensus survey. The forecasters asked for a given country's forecasts will typically be a mix of global institutions (e.g. global banks) and country-specific institutions (e.g. country-specific banks or think tanks). Around 20-30 forecasters are surveyed for each country.

Table 1.1 summarises the forecasts available. This rich variety of data ensures a well-rounded view on the forecast performance of all major institutional and private economic actors allowing us to draw general conclusions about the economic forecasting profession in the last 30 years. Both WEO and Consensus increased the number of countries for which forecasts are available over time, which means our evaluation is able to study more countries than similar evaluations conducted earlier. The World Bank and EC produced forecasts for a mostly constant set of countries over time.

Table 1.1: Forecasts availability

Forecaster	**Typical release months**	**Years with vintages in our dataset**
IMF	April & October	1990-2022
	Update: January & June	Update: 2010-2022
World Bank	January & July	2010-2021
EC	May & November	2011-2023
OECD	December & June	1985-2023
Consensus - Mean	Every month	1990-2022

1.3.2 Computing Forecast Horizons

We define the horizon of a forecast as the number of days from the publication date of the forecast to the end date of the target year.

- The publication date of a forecast is assumed to be the 15th day of the publication month.[13]

- International organisations and Consensus forecast real GDP growth in calendar years for most countries, but forecast growth in fiscal years in others (see Appendix A.1 for a list). For example, WEO forecasts US real GDP growth in calendar years, so in the October 2019 WEO, the end date of the first target year will be 31 December 2019. However, WEO forecasts Indian real GDP growth in fiscal years, so in the October 2019 WEO, the end date of the first target year will be 31 March 2020.

When comparing forecasters or testing the rationality of forecasts, we will report results for a set of forecasts with a specific horizon. In the terminology of Clements (1997), we perform 'rolling event' regressions rather than 'fixed event' regressions, because our regressions are estimated on samples of forecasts with a fixed horizon and a varying forecast date. However, a subtlety is that we round the forecast horizons to ensure we have adequate sample sizes. For example, when we test the rationality of 2-year-ahead forecasts by EC, we estimate our regression on all EC forecasts whose horizon in days is between 548 days and 913 days. An alternative to pooling together forecasts with similar horizons is to construct a set of artificial fixed-event forecasts, an approach that has been critiqued by Yetman (2018).

[13] The schedules on which international organisations have published their forecasts has changed over time. We attempted to collect the publication date of each individual vintage for each forecaster, but the information was often unreliable. E.g. The OECD provided us with a spreadsheet of publication months. The OECD iLibrary lists precise publication dates, but they often contradict the publication months.

1.3.3 Separating Forecasts and Actuals

The WEO, OECD, World Bank and EC databases provide vintages comprising actual outcomes and forecasts. To distinguish between actuals and forecasts, we need to make an assumption about how long it takes for statistical agencies to publish actual outcomes after the end of the target year. Our assumption is that this takes 90 days.[14]

1.3.4 Computing Forecast Errors

We compute forecast errors as actual outcomes less forecasts. For WEO, OECD and EC, we compute forecast errors by comparing the institution's forecast with their own actuals. An advantage of this approach is that the forecasts and actuals for a country will definitely be comparable, as they will both be recorded over the same calendar or fiscal years.

Unfortunately, Consensus lacks data on actual outcomes, and GEP only provides a small amount of data on actual outcomes. Hence, we compute errors in Consensus forecasts and GEP forecasts by comparing them with WEO actuals. One might be concerned that the forecasts for a country may be recorded over a different year as the actuals, since they are taken from different sources. Fortunately, Consensus uses the same years as WEO for all countries where it makes forecasts. Unfortunately, there are some countries for which the World Bank forecasts are recorded over different years to the WEO actuals, and we are unable to compute forecast errors in these countries (see Appendix A.1 for the specific countries).

For all forecasters, we compute errors using the latest vintage of the actual at the time of writing. An alternative approach would have been to take the first vintage of the actual. Revisions to real GDP data can be substantial (Genberg and Martinez 2014a), so this would lead to different results.Our choice of approach rests on two claims:

[14]To illustrate, consider WEO vintages of real GDP growth in India. WEO reports Indian real GDP growth over fiscal years that end on 31 March of each year. Growth for fiscal year on 2016-17 will be published on 29 June 2017, which is 90 days after the end of the year. As discussed above, we assume the publication date of a forecast is the 15th day of the publication month. Hence, the June 2017 WEO vintage is assumed to have been published on 15 June 2017, so we assume that the value reported by WEO for 2016-17 is a forecast rather than an actual outcome.

1. **International organisations should aim to forecast the best estimate of true real GDP growth.** True growth determines other economic variables and welfare, so a forecast for the best estimate of true growth is the best guide to policy.

2. **Later vintages of actuals are better estimates of 'true' growth than earlier vintages.** A major source of revisions to real GDP data is that the statistical agency receives additional data over time. For example, a statistical agency may produce the first vintage of a component of GDP based solely on survey data, but produce more accurate later vintages using tax data as well. This makes later vintages better estimates of true growth. Another source of revisions is changes to international and national standards for the measurement of GDP. Whether such revisions also make the data a better estimate of true growth is less clear. One view is that the changed standards better approximate some ideal notion of real GDP. Another view is that the later vintages are estimates of a slightly different variable to that which the forecaster had in mind.[15] In our view, even if one is concerned about changing standards, this second claim would still hold because the later vintages of GDP benefit so much from additional data.

   A major source of revisions to real GDP data is that the statistical agency receives additional data over time. For example, a statistical agency may produce the first vintage of a component of GDP based solely on survey data, but produce more accurate later vintages using tax data as well. This makes later vintages better estimates of true growth. Another source of revisions is changes to the standards that guide the measurement of GDP. This includes the United Nations of System of National Accounts, which were changed in 1993 and 2008, and the standards of individual countries. Whether such revisions

[15] For example, suppose that a country's standards were revised to say that its real GDP measures should include estimates of production of illegal drugs. This may lead to revisions to the history of real GDP growth for that country. A forecaster may have made an accurate prediction for growth excluding illicit drugs, but be judged by us to have made a forecast error as its forecast differed from growth including illicit drugs.

also make the data a better estimate of true growth is less clear. One view is that the changed standards better approximate some ideal notion of real GDP. Another view is that the later vintages are estimates of a slightly different variable to that which the forecaster had in mind.[16] In our view, even if one is concerned about changing standards, this second claim would still hold because the later vintages of GDP benefit so much from additional data.

Taken together, these claims imply that international organisations should aim to forecast the latest vintage of the actual, which makes this latest vintage the proper standard by which the accuracy and rationality of their forecasts should be assessed.

A limitation of our approach is that it reduces the comparability of earlier and later errors, since forecasts from earlier vintages will be compared to actuals that have been revised many times, while forecasts from later vintages will be compared to actuals that have been revised only a few times. On balance, however, we judge that it is appropriate to compute errors usign the latest vintage of the actuals.

[16]For example, suppose that a country's standards were revised to say that its real GDP measures would include estimates of production of illegal drugs. This may lead to revisions to the history of real GDP growth for that country. A forecaster may have made an accurate prediction for growth excluding illicit drugs, but be judged afterwards to have made a forecast error because its forecast differed from growth including illicit drugs.

1.4 Comparing Accuracy across Forecasters

1.4.1 Method

We compare the accuracy of different forecasters using Diebold-Mariano (DM) tests (Diebold 2015). We use WEO forecasts as our benchmark, and compare each other forecaster to the WEO. We perform this comparison separately for each country and forecast horizon. WEO forecasts are available for a comprehensive set of countries, so using them as our benchmark maximises the number of countries for which we can perform comparisons.

In a DM test, the null hypothesis is that the pair of forecasters have equal forecast accuracy (i.e. equal expected loss), while the alternative is unequal forecast accuracy.

$$H_0 : E\left[L\left(v_{t+h|t}^{\text{other}}\right)\right] = E\left[L\left(v_{t+h|t}^{\text{WEO}}\right)\right]$$
$$H_1 : E\left[L\left(v_{t+h|t}^{\text{other}}\right)\right] \neq \left[L\left(v_{t+h|t}^{\text{WEO}}\right)\right]$$

A ‘loss differential’ is the difference between the loss of the two forecasters, $d_{t+h|t} \equiv L\left(v_{t+h|t}^{\text{other}}\right) - L\left(v_{t+h|t}^{\text{WEO}}\right)$. The null hypothesis of the DM test is equivalent to stating that, in a regression of the loss differential on an intercept, the intercept is zero.

$$d_{t+h|t} = \mu + e_{t+h|t} \quad \forall t$$

Under the assumption that the loss differential is covariance-stationary, the OLS estimator of the intercept is asymptotically normal. We perform the DM test by performing a two-sided t-test of the null that $\mu = 0$. We use heteroscedasticity and autocorrelation consistent (HAC) standard errors, as we want to allow for the possibility that the loss differentials are serially correlated.[17]

There are two further econometric complications, which are similar to those that arose when comparing accuracy. Firstly, our approach is only asymptotically valid, but the sample sizes in macroeconomic forecast evaluation are often modest. As Kiefer et al. (2000) explains, conducting t-tests with HAC standard errors in finite samples leads to rejecting null hypotheses more often than the chosen level of

[17]We use the HAC estimator in the ‘sandwich’ package in R (Zeileis 2004) with the default options.

significance. The importance of this issue depends on the forecasters being compared. As shown in table 1.1, our dataset includes many forecast vintages for the IMF, OECD and Consensus (e.g. 97 vintages from the IMF), but fewer for the EC and World Bank (e.g. 30 vintages from the EC).

Secondly, the forecasts by international organisations and the private-sector are often informed by models whose parameters are estimated on finite samples of input data. Even where the forecaster relies solely on judgement, they can be viewed as using an implicit mental model that is informed by the finite sample of input data. Diebold (2015) discusses alternative tests that aim to address this issue, and argues that using a DM test rather than the alternative tests is appropriate provided that the forecast errors are *approximately* covariance-stationary.

1.4.1.1 Timing Issues

When computing loss differentials for a country, we must ensure that the two forecasters record that country's growth rates over the same period. Fortunately, WEO, OECD, EC and Consensus all record each country's growth rates over the same years (see Appendix A). There are some countries where WEO and GEP record growth rates over different periods, so we do not compute loss differentials for these countries.[18] As an additional check, whenever both forecasters have data on actual outcomes, we check that the actual outcomes are the same.[19]

We also need to ensure that the loss differentials are computed with forecasts made around the same time. When comparing Consensus with an international organisation, we always compute loss differentials using forecasts made in the same month for the same target year. E.g. If we wanted to compute a sequence of 14-month-ahead loss differentials, we'd compute one differential as 'October 2020 WEO

[18] Recall that GEP errors are computed using GEP forecasts and WEO actuals. We abstain from computing GEP errors for any country for which GEP and WEO record growth rates over different periods. Hence for these countries, we lack GEP errors, so we couldn't compute loss differentials even if we wanted to.

[19] Where we can compare actuals in this way, we remove the risk of accidentally comparing loss differentials whose actuals are recorded over different periods. Moreover, we avoid the risk of comparing loss differentials whose actuals differ for some other reasons, such as differences in rounding conventions or data sources.

forecast for 2021' with 'October 2020 Consensus forecast for 2021', and compute the next differential as 'October 2021 WEO forecast for 2022' with 'October 2021 Consensus forecast for 2022'.

When comparing international organisations with each other, we compute loss differential using forecasts made in the same *or adjacent* months for the same target year. E.g. We could compute one differential as 'October 2020 WEO forecast for 2021' with 'November 2020 EC forecast for 2021', and compute the next differential as 'October 2021 WEO forecast for 2022' with 'November 2021 EC forecast for 2022'. A limitation of this approach is that one forecaster will have an additional month of information relative to the other, giving them additional information, and giving them an unfair advantage in our comparisons of forecast accuracy. In practice, EC always has one more month of information than WEO, GEP has the same or less information than WEO, and the OECD sometimes has more information and sometimes less (see Table 1.2).[20]

Table 1.2: Mismatch in timing of forecasts used to compute loss differentials

Publication Month	EC	OECD	GEP
1 month before WEO	0	18	9
Same month as WEO	0	2	14
1 month after WEO	28	22	0

1.4.2 Results

1.4.2.1 Diebold-Mariano tests for individual countries

We conducted a DM test for each country and horizon. We report the p-value of each DM test. In addition, we report the ratio of the RMSFE of one forecaster

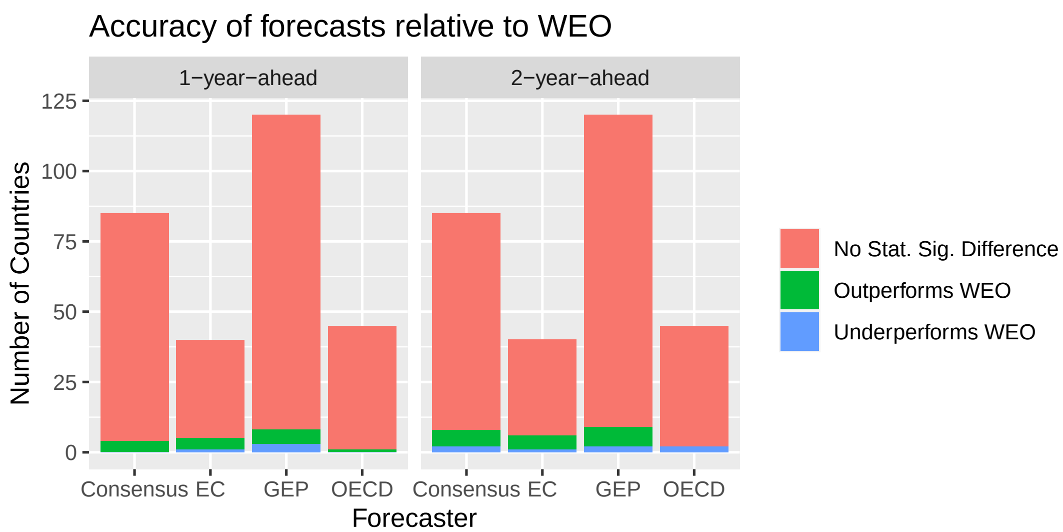
[20] The timing mismatches raise a subtle issue with forecast horizons. Each DM test is performed on a sequence of loss differentials with a specified forecast horizon, such as $h = 1$ year ahead. Where a loss differential compares two forecasts published in the same month for the same target year, their forecast horizons are the same, and that loss differential can be viewed as having that horizon. However, where a loss differential compares two forecasts published in adjacent months for the same target year, their forecast horizons differ by one month. In this case the loss differential is labelled with the average of the horizons of the two forecasts being compared.

to the RMSFE of the WEO.[21]

For the vast majority of countries, we find no statistically significant difference at a 5% level in the performance of the WEO forecasts and the other forecasters (Figures 1.1). For the 1-year-horizon there are a handful of countries where Consensus and EC are found to outperform or underperform WEO, and at the 2-year-horizon there are a few countries where each forecaster outperforms or underperforms.

Since we perform DM tests separately for dozens of countries with a 5% level of significance, even if the null hypothesis is true we would expect around 5% of countries to show a significant result. This would be just 2 or 3 countries for EC, OECD and Consensus, and 0 or 1 countries for GEP. Hence, the handful of countries with statistically significant differences could be spurious. If one of the forecasters was genuinely better or worse in general, we would expect to see that forecaster outperform or underperform for many countries, but this isn't the case. The results for each individual country at a 1-year-ahead horizon are available in Appendix A.2.

Figure 1.1



The lack of significant results could occur because there is genuinely no meaningful difference in performance, or because our sample sizes are too small to detect any

[21]This ratio is not the point estimate of the DM test statistic, which would be the mean square forecast error of one forecaster less than mean square forecast error of WEO. We prefer to report the ratio of RMSFEs as it is easier to compare across countries than differences in sample means, which is larger in magnitude for countries whose forecasts tend to be less accurate.

performance differences. The comparisons of OECD or Consensus with WEO use forecast vintages from 1990 onwards, so our results for these forecasters provide quite strong evidence that there is little difference in the performance of these forecasters. However, the comparisons of GEP or EC with WEO only use forecast vintages from 2010 or 2011 onwards. Since the GEP and EC each make just two forecasts per year, this results in sample sizes of around two dozen forecasts, raising the risk that there is a difference in forecast performance that we were unable to detect.

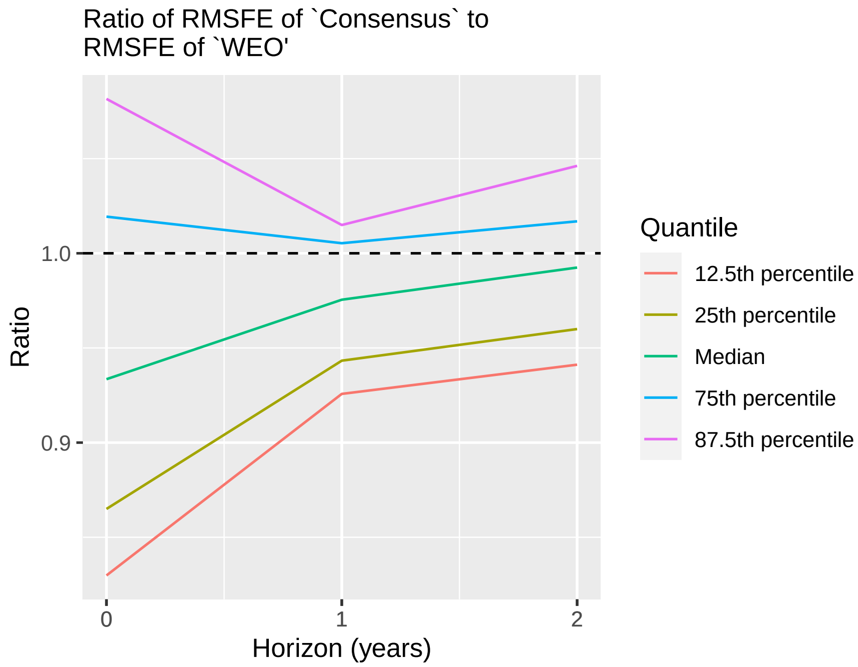
1.4.2.2 Cross-country Evidence

The hypothesis tests considered each country in isolation. A natural question is whether one can better detect differences in forecast performance by drawing on the results for many different countries. Unfortunately, it is not straightforward to use the observed performance for different countries in a formal hypothesis test. The reason is that actual outcomes are correlated across countries (e.g. high growth in France tends to coincide with high growth in Germany), and forecasts tend to be correlated across countries too (e.g. if the IMF expects high growth in the US it likely expects high growth in Mexico too). As such, the forecast errors for different countries do not provide independent sources of information.

Nevertheless, the results for different countries suggest some possibilities that are worth investigating formally once longer samples become available. First we compare Consensus to WEO. For each horizon, we sorted countries from the lowest ratio of Consensus RMSFE to WEO RMSFE to the highest. At 0 or 1 year horizon, roughly 75% of countries had a lower Consensus RMSFE than WEO RMSFE (Figure 1.2). Forecast combination tends to improve forecast accuracy (Clements and Harvey 2009), which could explain why the Consensus mean had lower RMSFEs in our sample than WEO.

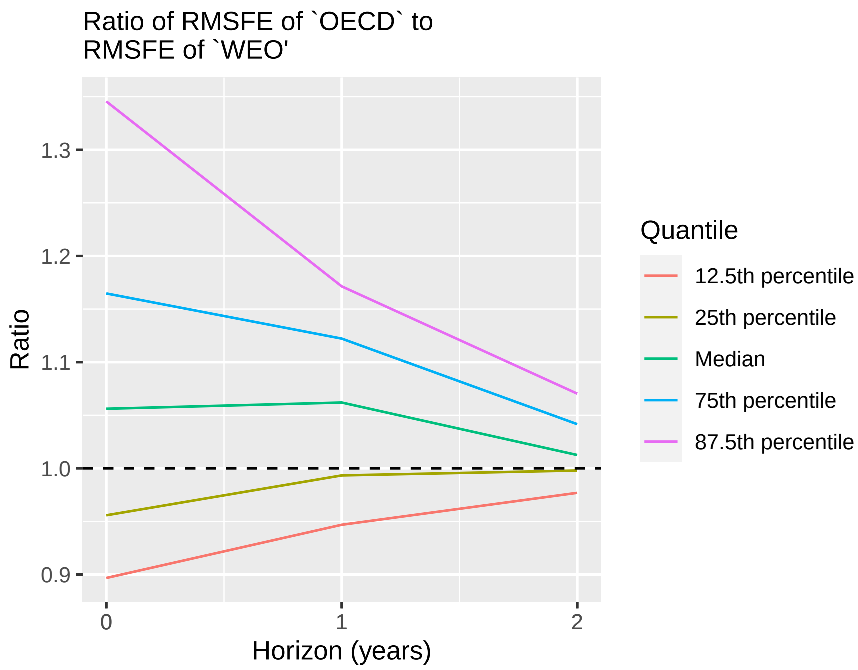
Second, we compared OECD to WEO using the same exercise. At a 1 year horizon, around 75% of countries had a higher OECD RMSFE than WEO RMSFE (Figure 1.3). This cannot be attributed to differences in the timing with which each forecaster publishes their forecasts, because although the OECD has a 1 month

Figure 1.2



timing advantage in some vintages, it has a 1 month timing disadvantage in a similar number of vintages (table 1.2).

Figure 1.3



We also compared EC to WEO and compared GEP to WEO. These did not have a clear pattern in either direction. These point estimates are noisier, as the comparisons are based on smaller set of vintages than OECD or Consensus.

1.4.3 Decomposing differences in Forecast Performance

A loss differential measures the difference in accuracy of two forecasts for the same outcome. We show that in the case of squared error loss, a loss differential can be decomposed into two components: the accuracy of the average forecast and the disagreement between forecasters.

Theorem 1. *Suppose the loss differential between a forecaster c and another forecaster b is computed with squared error loss:*

$$d_{t+h|t} \equiv L\left(v^c_{t+h|t}\right) - L\left(v^b_{t+h|t}\right) = \left(v^c_{t+h|t}\right)^2 - \left(v^b_{t+h|t}\right)^2$$

Then the absolute loss differential can be written:

$$\underbrace{\left|d_{t+h|t}\right|}_{\text{Absolute Loss Differential}} = 2 \times \underbrace{\left|y_{t+h} - \frac{1}{2}\left(y^c_{t+h|t} + y^b_{t+h|t}\right)\right|}_{\text{Absolute Error of Average Forecast}} \times \underbrace{\left|y^c_{t+h|t} - y^b_{t+h|t}\right|}_{\text{Absolute Difference in Forecasts}}$$

Proof. See appendix A.2. □

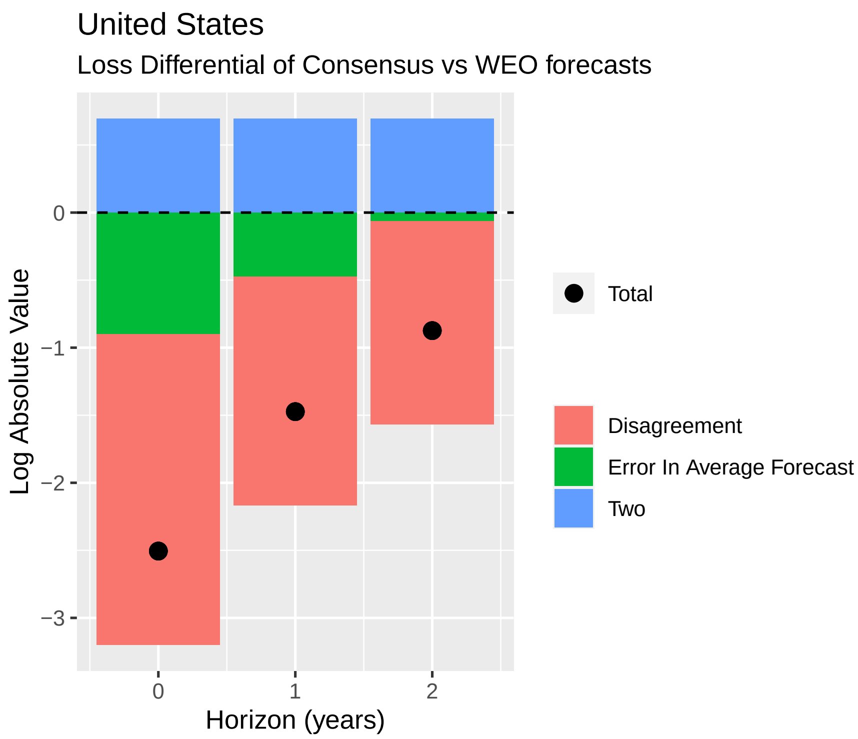
If we assume that the loss differential, the error in the average forecast and the absolute difference in forecasts are all non-zero, we can take natural logs, giving:

$$log\left(\left|d_{t+h|t}\right|\right) = log(2) + log\left(\left|y_{t+h} - \frac{1}{2}\left(y^c_{t+h|t} + y^b_{t+h|t}\right)\right|\right) + log\left(\left|y^c_{t+h|t} - y^b_{t+h|t}\right|\right)$$

1.4.3.1 The Relationship between Performance and Forecast Horizon

We can use this decomposition to understand how the relative performance of a pair of forecasters for a country depends on the forecast horizon. To do this, we compute sample means of the log absolute loss differentials across all forecast dates, and sample means of the components. Figure 1.4 shows this decomposition for the Consensus and WEO forecasts for the US. At longer horizons, the log absolute loss differential is less negative, which means the absolute loss differential is larger. Our decomposition shows that this occurred because longer horizons had both larger amounts of disagreement and larger errors in average forecasts.

Figure 1.4



We also performed the decomposition separately for each country, and then summarised the results by estimating a panel regression (equation (1.2)). The dependent variable is the log absolute loss differential for a particular country i, forecast month t and horizon h. The explanatory variables are country fixed effects and dummies for each horizon. We include a horizon dummy for each horizon $h = 1, ..., H$, which means the base case is a horizon of $h = 0$ years.[22] The country fixed effects $\alpha_i, \beta_i, \gamma_i$ are included because the magnitude of loss differentials is likely to differ across countries.[23]

$$log\left(\left|d^i_{t+h|t}\right|\right) = \alpha_i + \phi_1 \mathbb{I}(h=0) + ... + \beta_H \mathbb{I}(h=H) + \varepsilon_{t+h|t} \quad \forall i,t,h \tag{1.2}$$

[22] For example, the January 2021 WEO includes forecasts for growth in 2020 since actual outcomes are published with a lag. These forecasts would be considered to have a horizon of 0 years, since their horizon in days rounds to 0 years (see section 1.3).

[23] For example, real GDP growth in an emerging market like India is typically higher and more volatile than in an advanced economy like Japan. Hence, the loss differentials are likely to be larger in India than in Japan. We include country fixed effects, but assume the same coefficients on the horizon dummies for all countries. This implies that we allow for absolute loss differentials to be larger in some countries than others on average across all horizons, but we impose that the effect on absolute loss differentials of changing horizon is the same across all countries.

We then perform two more panel regressions with the same explanatory variables, but where the dependent variable is a term from our decomposition (equation (1.3)).

$$\begin{aligned} log\left(\left|y_{t+h}-\frac{1}{2}\left(y^c_{t+h|t}+y^b_{t+h|t}\right)\right|\right) &= \beta_i+\psi_1\mathbb{I}(h=0)+...+\beta_H\mathbb{I}(h=H)+\eta_{t+h|t} \\ log\left(\left|y^c_{t+h|t}-y^b_{t+h|t}\right|\right) &= \gamma_i+\lambda_1\mathbb{I}(h=0)+...+\gamma_H\mathbb{I}(h=H)+\omega_{t+h|t} \\ &\forall i,t,h \end{aligned} \tag{1.3}$$

Figures 1.5 show the estimated coefficients on the horizon dummies for the comparison of Consensus to WEO. Figures 1.6, 1.7 and 1.8 do the same for comparisons involving EC, GEP and OECD respectively. Across all four pairs of forecasters, we find that, relative to the 0-year-ahead base case, the log absolute loss differentials are larger at a 1 year horizon, and larger still at a 2 year horizon. Our results suggest that this occurs because the errors in the average forecast tend to be larger at longer horizons, and to a lesser extent, because disagreement between the forecasters is larger at longer horizons.[24] It makes sense that the forecast errors and disagreement are both larger at longer horizons, since forecasters have less information at these horizons. In particular, the forecasts at a 0 year horizon typically benefit from having data for some quarters of the year, but this is not available at 1 or 2 year horizons. Our results suggests that practitioners should be cautious when interpreting squared errors at different horizons: if the difference in squared errors become larger in magnitude at longer horizons, this does not necessarily indicate changes in the relative performance of the two forecasters, but rather could arise simply because both forecasters are less accurate at longer horizons.

[24] The confidence intervals are computed with 'classical' standard errors. These confidence intervals may be too narrow or too wide if the error terms in the panel regressions, $\varepsilon_{t+h|t}, \eta_{t+h|t}, \omega_{t+h|t}$, are correlated across countries i or over time t. However, a change in the width of the confidence intervals is unlikely to change the qualitative findings, which are common across the the pairs of forecasters analysed.

Figure 1.5

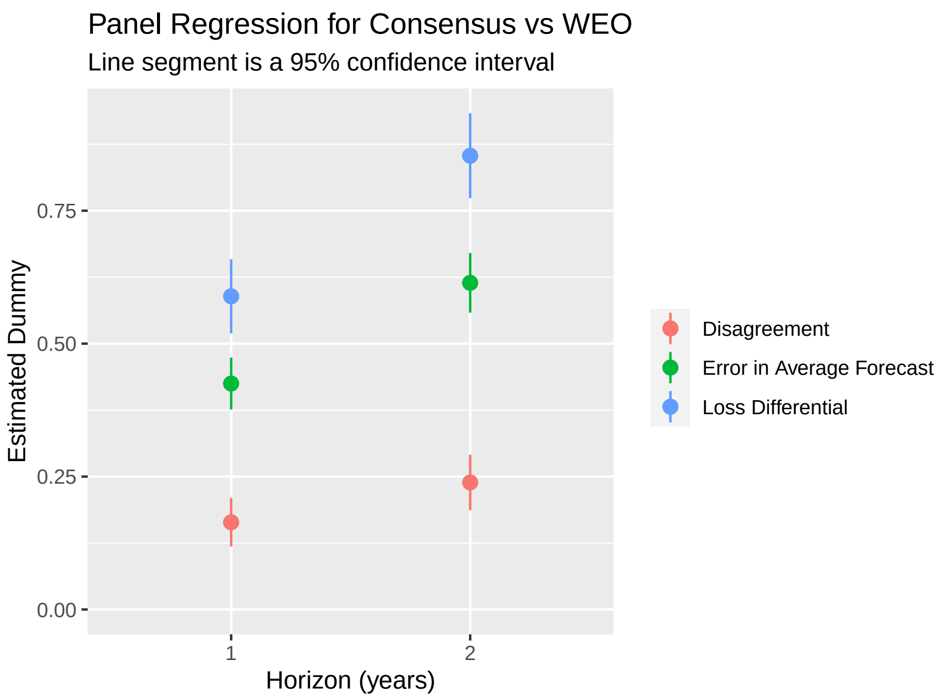


Figure 1.6

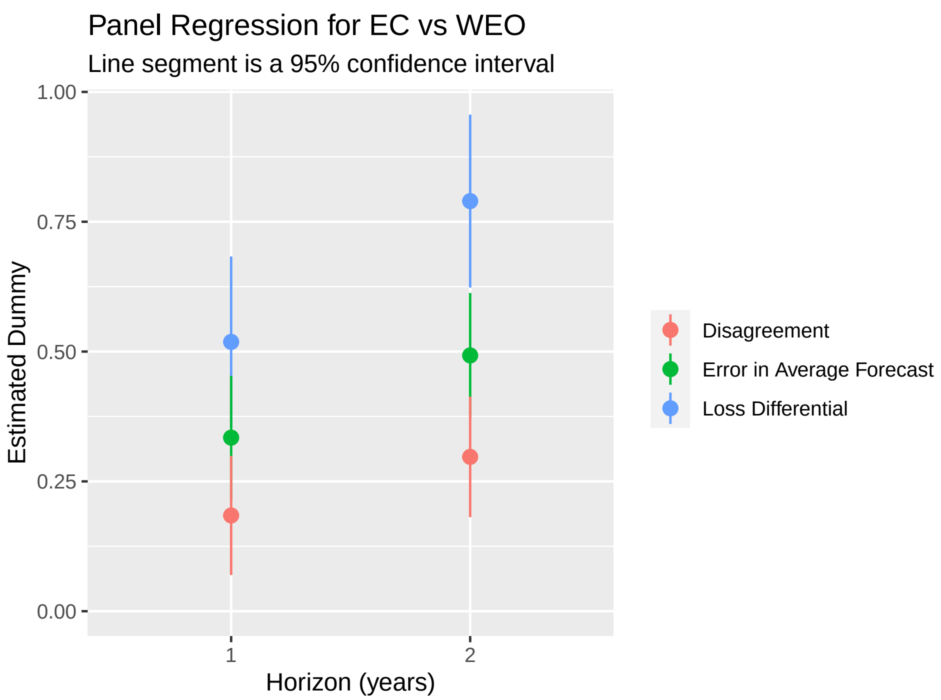


Figure 1.7

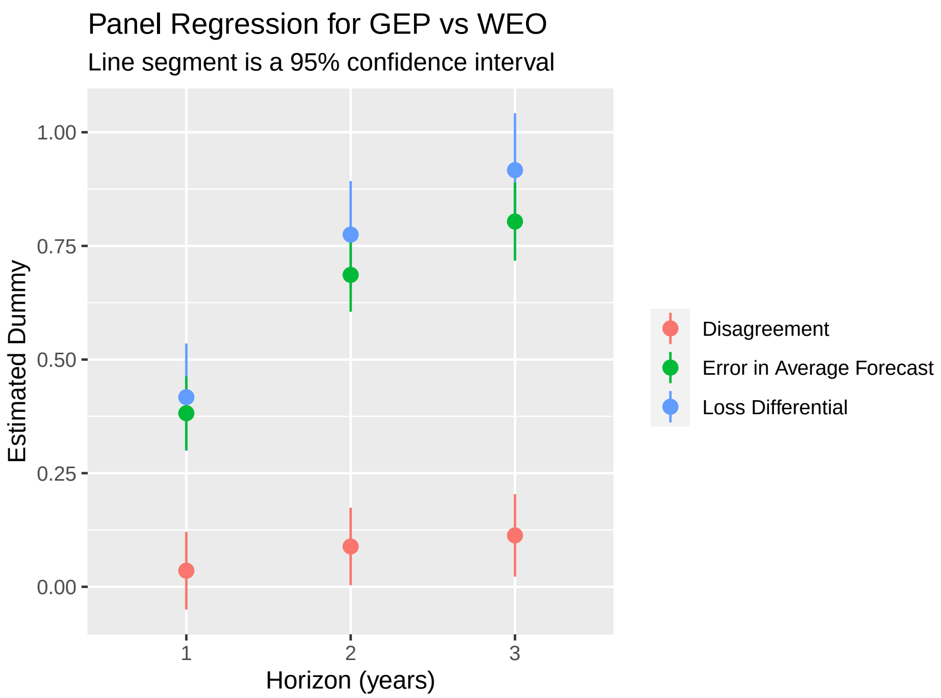
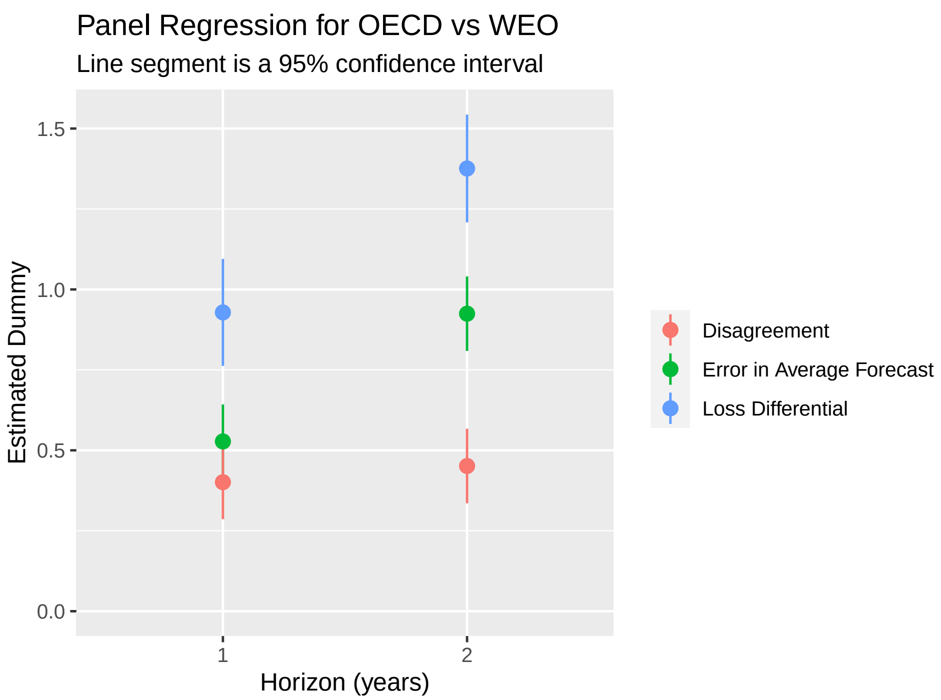


Figure 1.8



1.5 Testing Forecast Rationality

1.5.1 Methods of Testing Forecast Rationality

We perform rationality regressions on fixed-event forecasts to test the joint assumption of rationality and squared error loss. Specifically, we test the null that the parameter vector is the zero vector in equation (1.1).

$$H_0 : \boldsymbol{\theta} = \mathbf{0}$$
$$H_1 : \boldsymbol{\theta} \neq \mathbf{0}$$

The key assumption is that the error in this regression is covariance-stationary.[25] However, we allow for the possibility that the error in this regression is heteroscedastic and auto-correlated.[26] Under our assumptions, the OLS estimator of $\boldsymbol{\theta}$ is asymptotically normal, so we can perform asymptotically valid t-tests and Wald-tests of hypotheses concerning $\boldsymbol{\theta}$ using a HAC estimator of the covariance matrix of the OLS estimator (see the introduction of Kiefer et al. (2000)). We use the HAC estimator in the 'sandwich' package in R with the default options (Zeileis 2004), just as we did in the DM tests.

Our rationality test raises similar econometric issues to the DM test (as discussed in 1.4.1), which is to be expected given that both are tests about the coefficients of a regression involving forecast errors. Firstly, the test has an asymptotic justification, but as with any evaluation of macroeconomic forecasts, we only have modest sample sizes. This problem is less severe in this chapter than some other papers, as our sample

[25] I.e. We assume that the part of the forecast errors not explained by the variables in our vector $\boldsymbol{x}_t$ is covariance-stationary. It is possible, for instance, that the forecast errors trend up over time, as long as this is explained by some variable in $\boldsymbol{x}_t$, so the errors do not trend up.

[26] Autocorrelation in the error u_{t+h} would occur if, for instance, the error in the April 2020 forecast for growth in 2022 is correlated with the error in the April 2021 forecast for growth in 2023. It is plausible that these forecast errors could be serially correlated. For example, in both 2020 and 2021 the forecaster may forecast high growth in 2022 and 2023 respectively, so if an unexpected recession occurs that reduces growth in 2022 and 2023, the forecast errors for these two years will have the same sign. This could arise even if the forecaster is rational, because when the forecaster makes a forecast in April 2021 they have no way of knowing the error in the forecast they made in April 2020. One common response to autocorrelation in the error term is to model the dependence by including lags of the dependent variable to the set of explanatory variable. This solution is not appropriate here, as we want all the explanatory variables of a rationality regression to be known on the forecast date, and the lag of the dependent variable is a forecast error that is not known in the forecast date if the horizon is more than 1 period or the actuals are published with a lag.

includes recent forecast vintages that were unavailable to researchers in previous years. Secondly, the forecasts may be produced using explicit models that are estimated on finite samples, or using judgement that is informed by finite samples. Some authors, such as West and McCracken (1998), have shown that even in the presence of parameter estimation error, one can test hypotheses about the forecast errors that would arise absent that error. However, we are not able to conduct such an exercise as there is no single explicit model for each forecaster. Hence we do not try to account for estimation error in model parameters, following Rossi and Sekhposyan (2016).

In brief, we find that international organisation growth forecasts fail rationality tests for most countries. The source of the failure of the test varies by country. In many cases, it arises because the forecasts are too extreme, the forecast revisions are too large, or the forecasts have an optimistic bias.

1.5.2 Regression Specifications

1.5.2.1 The General Specification

Our general regression specification is:

$$v_{t+h|t} = \alpha + \underbrace{\beta_0 y_{t+h|t} + \beta_1 y_{t+h|t-1}}_{\text{Forecasts for target year } (t+h)} + \underbrace{\gamma_0 v_{t-p|t-p-h} + \gamma_1 v_{t-p-1|t-p-h-1}}_{\text{Errors in } h\text{-ahead forecasts}} + u_{t+h} \quad (1.4)$$

This regression includes the following elements:

- The dependent variable $v_{t+h|t}$, is the error in the forecast for period $(t+h)$ made in period t.
- The forecasts for the target year. This includes the forecast that was used to compute the error, $y_{t+h|t}$, and one earlier forecast, $y_{t+h|t-1}$.
- Errors in h-step-ahead forecasts. We let p denote the number of periods it takes for statistical agencies to publish the actual real GDP growth outcome after the end of a period. This implies that, in period t, the latest actual known to the forecaster is y_{t-p}. Our regression includes the latest forecast error known to the forecaster, $v_{t-p|t-p-h}$, and one earlier error, $v_{t-p-1|t-p-h-1}$.

In the notation of equation (1.1), the vector $\boldsymbol{x}_t$ comprises the intercept, forecasts for the target year and errors in h-ahead forecasts, and the vector of parameters is $\boldsymbol{\theta} = (\alpha, \beta_0, \beta_1, \gamma_0, \gamma_1)$. The rationality test is a Wald test of $H_0 : \boldsymbol{\theta} = \mathbf{0}$ against $H_0 : \boldsymbol{\theta} \neq \mathbf{0}$.

It would be straightforward to add more lags of forecasts or errors to this regression. We choose not to, as it would result in the parameters being estimated very imprecisely given the modest sample sizes of macroeconomic forecasts available.

When we estimate these regressions, we use a dataset constructed with real-time information. For example, in one observation in this dataset, the dependent variable is the error in the April 2020 WEO forecast for 2022. The explanatory variables include two forecasts for that target year, specifically the April 2020 WEO forecast for 2022, the Jan 2020 WEO forecast for 2022. The explanatory variables also include two errors in 2-year-ahead forecasts. As of April 2020, the most recent actual is likely to be 2019, so the two latest 2-year-ahead errors are the April 2018 forecast for 2019 and the April 2017 forecast for 2018.

$$\begin{aligned} v_{2022|Apr2020} =& \alpha + \beta_0 y_{2022|Apr2020} + \beta_1 y_{2022|Jan2020} + \gamma_0 v_{2019|Apr2018} \\ & + \gamma_1 v_{2018|Apr2017} + u_{t+h} \end{aligned}$$

1.5.2.2 Interpreting the General Specification

Interpreting the parameters of the rationality regression is difficult. In particular, the intercept is *not* the forecaster's unconditional bias, but rather is the bias conditional on past forecasts and past errors all being zero. To facilitate interpretation, it is useful to rearrange equation (1.4) as follows.

$$\begin{aligned} v_{t+h|t} = & \underbrace{\left(\alpha + \beta_0 E[y_{t+h|t}] + \beta_1 E[y_{t+h|t-1}] + \gamma_0 E\left[v_{t-p|t-p-h}\right] + \gamma_1 E\left[v_{t-p-1|t-p-h-1}\right]\right)}_{\text{Unconditional Bias}} \\ & + \underbrace{\beta_0 \left(y_{t+h|t} - E[y_{t+h|t}]\right) + \beta_1 \left(y_{t+h|t-1} - E[y_{t+h|t-1}]\right)}_{\text{Contribution of Demeaned Forecasts for Target Year}} \\ & + \underbrace{\gamma_0 \left(v_{t-p|t-p-h} - E\left[v_{t-p|t-p-h}\right]\right) + \gamma_1 \left(v_{t-p-1|t-p-h-1} - E\left[v_{t-p-1|t-p-h-1}\right]\right)}_{\text{Contribution of Demeaned Forecast Errors for Horizon}} \\ & + u_{t+h} \end{aligned} \tag{1.5}$$

The right-hand-side terms each have a nice interpretation.

- The first term is the unconditional bias, which is the expectation of the forecast error. This differs from the intercept, which is the conditional bias given that all the explanatory variables are zero.

- The contribution of demeaned forecasts tells us if the forecasts tend to be 'too extreme' or 'not extreme enough'. For example, suppose the coefficient on the latest forecast is positive $\beta_0 > 0$. If the forecaster predicted higher growth than they usually do, $y_{t+h|t} > E[y_{t+h|t}]$, then the forecast error is likely to be positive $v_{t+h|t} > 0$, so the forecast error would have been smaller if the forecast were even higher. Conversely, if the forecaster had predicted lower growth than usual, the forecast error would likely be negative, so the forecast error should have been even lower. For this reason, we describe $\beta_j > 0$ as indicating that the jth most recent forecast was not extreme enough, and $\beta_j < 0$ as indicating that it was too extreme.

- The contribution of demeaned forecast errors tells us if the forecaster's h-period-ahead errors are related to the latest h-period-ahead-errors known on the forecast date, which would indicate that the forecaster is not making efficient use of the information contained in their past errors. For example, the forecaster may be slow to recognise that a change in the mean of the series has occurred, and hence persistently over-predict growth over a long period.

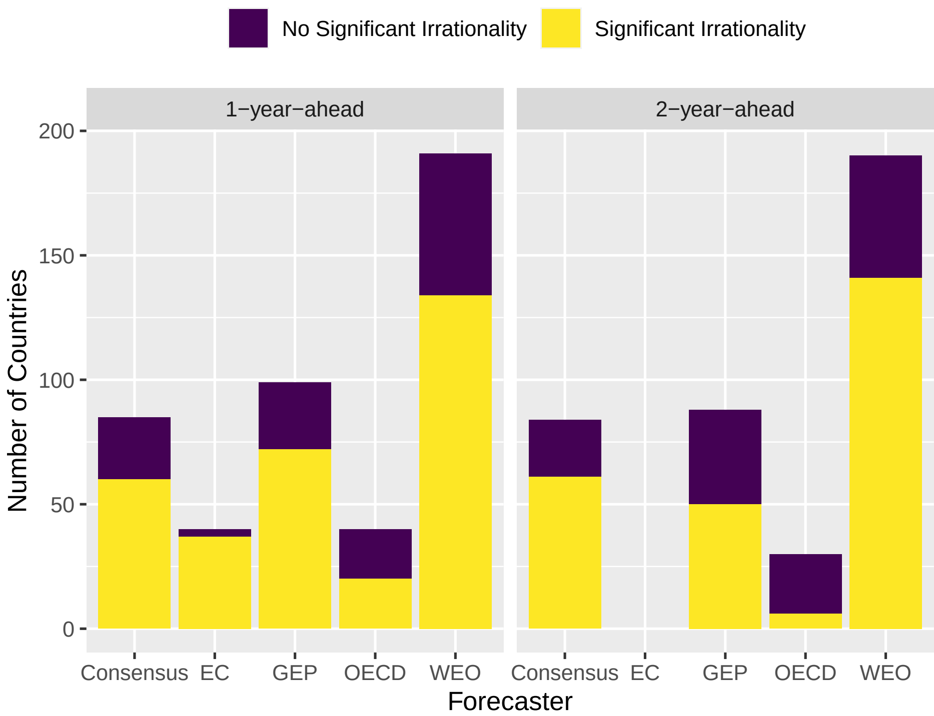
We cannot estimate the rearranged regression directly, because this would require knowledge of the population mean of forecasts and the population mean of errors. Instead, we estimate the original regression (equation (1.4)), but we bear the rearranged regression in mind when interpreting the coefficients. Since we cannot estimate the original regression, we cannot use is to test the hypothesis of zero unconditional bias. Instead, we need to estimate a regression of errors on an intercept alone, as in section 1.5.3.3.

1.5.3 Results

1.5.3.1 General specification

We perform a Wald test of the null hypothesis of forecast rationality (all parameters equal zero) against the alternative of irrationality. For most forecasters, we reject rationality of 1-year-ahead and 2-year-ahead forecasts at a 5% level for most countries (Figure 1.9. The exception is the OECD, where we only reject rationality for a minority of countries. The samples of forecasts available for the OECD are smaller than for WEO and Consensus, so this might partly reflect lower power of tests for the OECD.

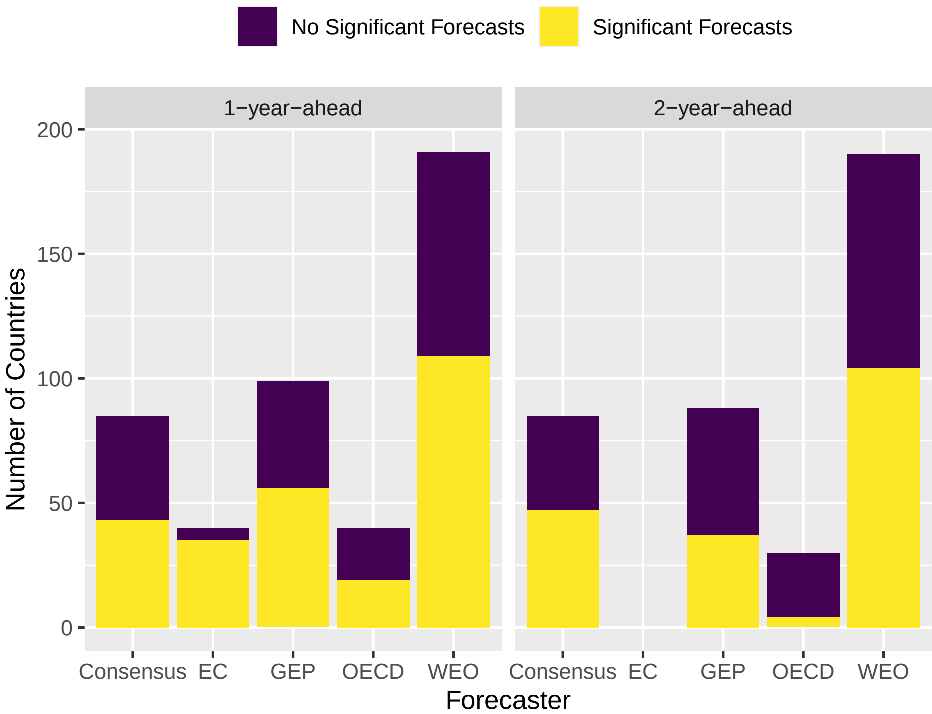
Figure 1.9: Wald Test of All Terms in General Specification



Note: The sample of 2-year-ahead EC forecasts is small, so this column is not shown.

Having found widespread irrationality, we now investigate the source of irrationality. First we perform Wald tests of the null that the coefficients on past forecasts are zero, which we reject for most countries (Figure 1.10). As explained in section 1.5.2.2, this indicates that the forecasters have a tendency to make forecasts that are too extreme or not extreme enough.

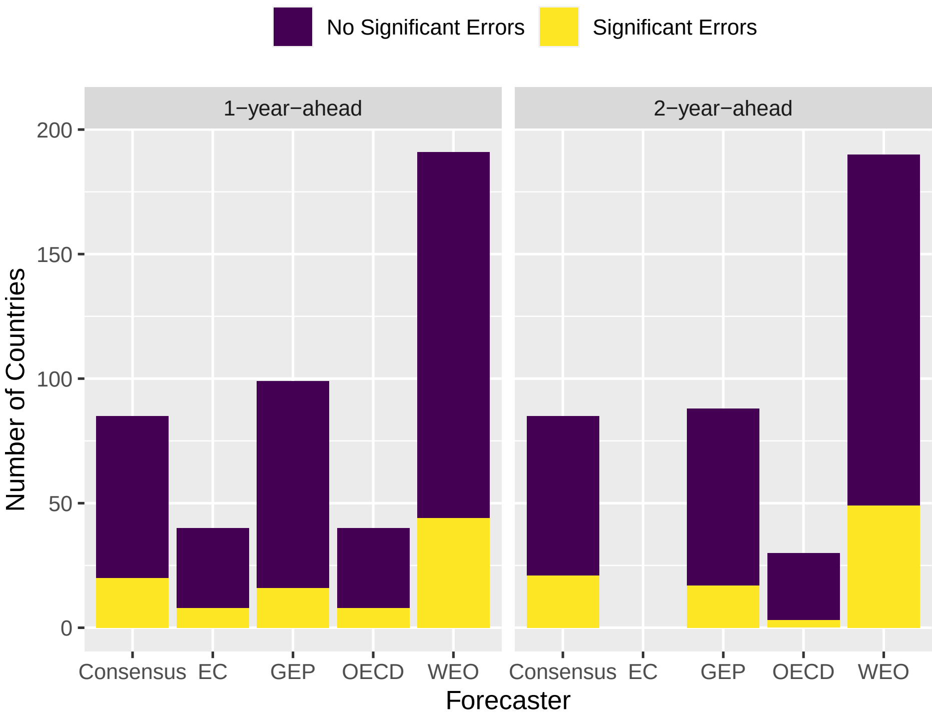
Figure 1.10: Wald Test of Forecasts in General Specification



Note: The sample of 2-year-ahead EC forecasts is small, so this column is not shown.

Second, we perform Wald tests of the null that the coefficients on past errors are zero. Some forecasters, including the IMF, OECD and EC, only review their forecast performance every few years (Genberg, Martinez, and Salemi 2014), so it's plausible that they would fail to make efficient use of any the information contained in their recent errors. We reject the null that the coefficients on past errors are zero for a minority of countries (Figure 1.11). Since we conduct a separate hypothesis test at the 5% level for each country and forecaster, we would expect the Wald test to be rejected for 5% of countries if the null hypothesis were true for all countries. These results provide tentative evidence that a failure to make use of the information in past errors is a source of irrationality for some forecasters for some countries, but suggest this issue is not too widespread.

Figure 1.11: Wald Test of Past Errors in General Specification



Note: The sample of 2-year-ahead EC forecasts is small, so this column is not shown.

1.5.3.2 Intercept and Past Forecasts Specification

The Wald tests of the general specification suggested that failing to account for information in past errors was a relatively unimportant source of irrationality. For this reason, we consider a special case of the general specification where these past errors are removed.

$$v_{t+h|t} = \alpha + \beta_0 y_{t+h|t} + \beta_1 y_{t+h|t-1} + u_{t+h} \tag{1.6}$$

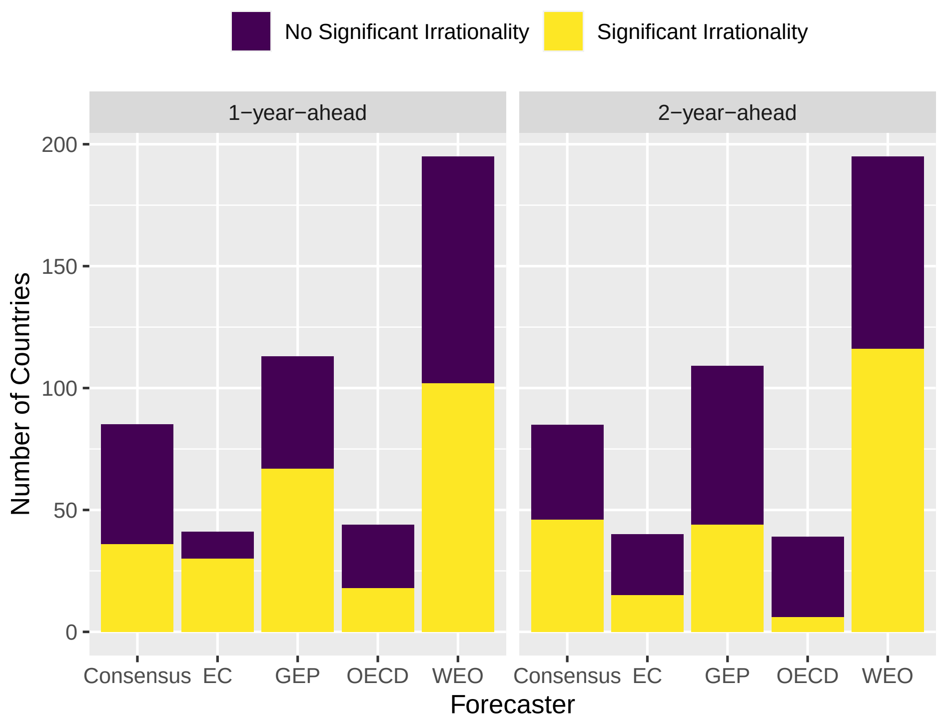
We can view this regression through the lens of forecast revisions, which are are changes in forecasts for a given target year from one vintage to the next. We rearrange the specification with forecasts only to obtain a specification with a forecast revision and a past forecast (equation (1.7)).

$$v_{t+h|t} = \alpha + \psi_0 \underbrace{\left(y_{t+h|t} - y_{t+h|t-1}\right)}_{\text{Latest forecast revision}} + \underbrace{\psi_1 y_{t+h|t-1}}_{\text{Second-latest forecast}} + u_{t+h} \tag{1.7}$$

The coefficient on the latest forecast revision tell us if the forecast revisions tend to be too large or too small. For example, if $\psi_0 > 0$, it implies that if the forecaster revised up their forecast in the last period, $(y_{t+h|t} - y_{t+h|t-1}) > 0$, then the forecast error is likely to be positive, $v_{t+h|t} > 0$, which implies that the forecast error would be smaller if the upward forecast revision had been larger. Hence whenever $\psi_0 > 0$ we say that the latest forecast revision tends to be too small, indicating that the forecaster tends to underreact to the news received between $(t-1)$ and t.[27]

We test rationality in specification (1.7) by performing a Wald test of the null that all coefficients are zero (Figure 1.12). As with the general specification, we reject rationality for most countries for all forecasters other than the OECD.

Figure 1.12: Wald Test of All Terms in Intercept, Revision and Past Forecast Specification



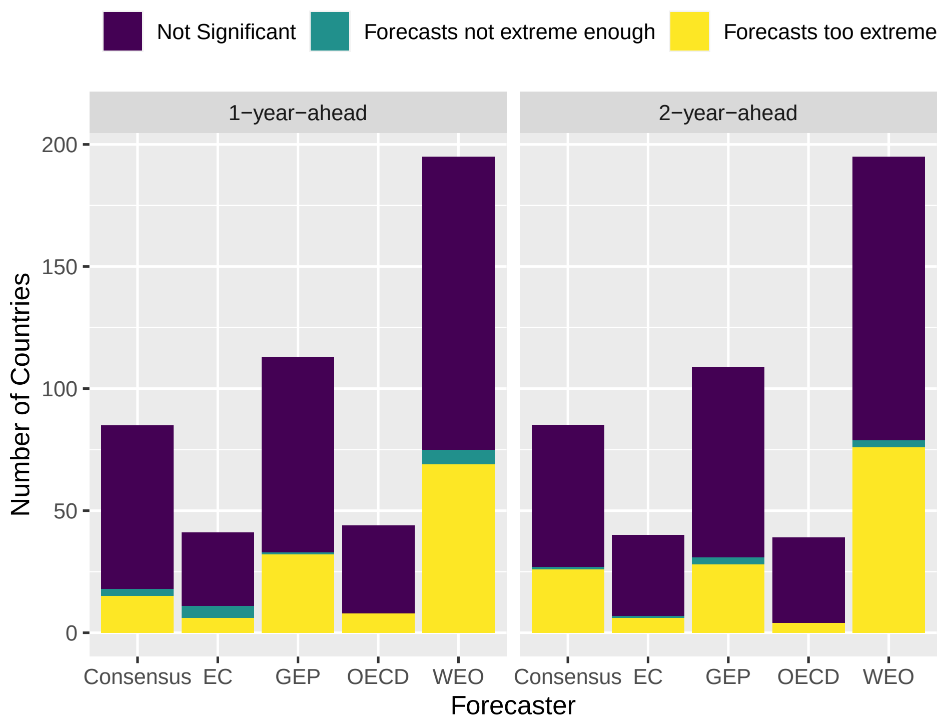
We know that this irrationality is not solely due to bias, because if we do a Wald

[27]If we had regressed the forecast error on the past revision alone, then it would be unclear if the coefficient on the past revision reflects the forecasters reaction to news (as we have claimed) or because the level of the forecast $y_{t+h|t}$ was too high or low (as argued by Fuhrer (2018)). However, our regression should provide separate estimates of the response of whether the forecaster responds efficiently to news (given by the coefficient on the revision, $(y_{t+h|t} - y_{t+h|t-1})$) and whether the forecasts tend to be too extreme or not extreme enough (given by the coefficient of the second-latest forecast, $y_{t+h|t-1}$).

test of the coefficient on the revision and on the past forecast, we reject the null for the majority of countries for all forecasters. To understand the role of revisions versus past forecasts, we do t-tests of each, noting that for some countries neither variable may be individually significant even though the two variables are jointly significant.

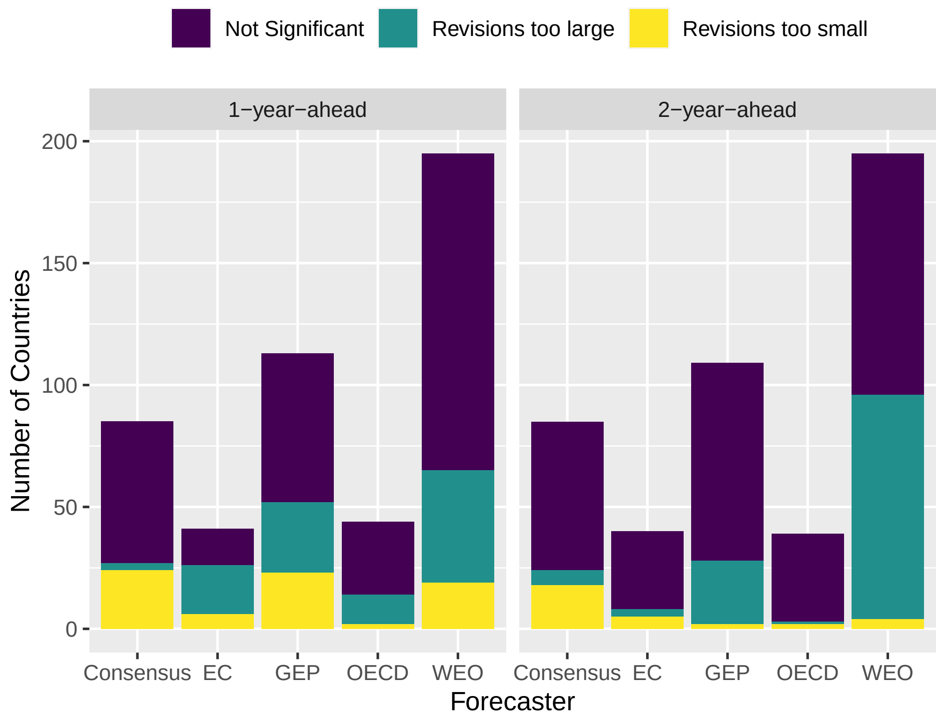
The t-tests of the coefficient on the past forecast show that this coefficient is insignificant in most countries (Figure 1.13). Where it is significant, the forecasters are typically too extreme. When the forecaster expects higher-than-average growth, it could improve accuracy by making a less optimistic forecast, and when the forecaster expects lower-than-average growth, it could improve accuracy by making a less pessimistic forecast.

Figure 1.13: t-test of Past Forecast in Intercept, Revision and Past Forecast Specification



The t-tests show that the coefficient on past revisions is significant in a sizable minority of countries for most forecasters, and a majority of countries for EC (Figure 1.14). However, the direction of the inefficiency differs. In each international organisation, it is more common for revisions to be too large than for them to be too small. For the average of Consensus forecasters, it is more common for revisions to be too small. This is similar to the finding that individual private-sector forecasters make overly large forecast revisions (Bordalo, Gennaioli, Ma, et al. 2020; Broer and Kohlhas 2022). In contrast, Consensus forecasts for some countries have revisions that are too small, consistent with previous evidence that the average of private-sector forecasts has overly small forecast revisions (Coibion and Gorodnichenko 2012, 2015a).

Figure 1.14: t-test of Revision in Intercept, Revision and Past Forecast Specification



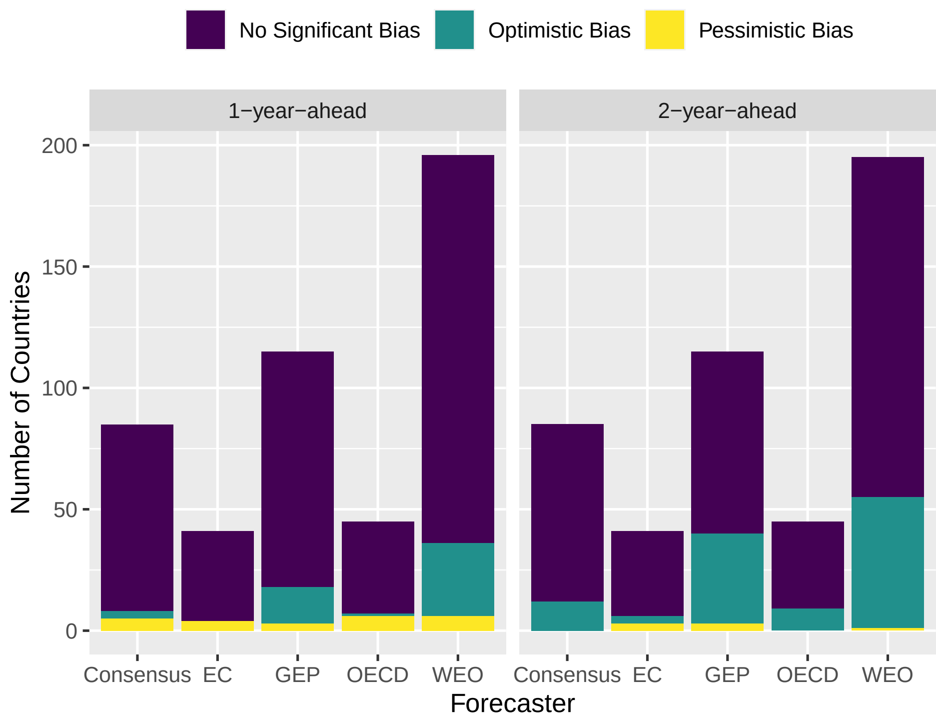
An alternative to regressing errors on forecasts (as we have just done) would be to regress forecast revisions on their own lags. Nordhaus (1987) showed that, if forecasters incorporate all new information each time they update their forecasts, then the sequence of forecasts for a fixed target date will be serially uncorrelated. Such a sequence of forecasts is said to satisfy 'weak efficiency', which is a necessary condition for rationality. We chose to instead regress errors on lagged revisions so that our analysis of revisions is done in the same framework as our analysis of other sources of irrationality.

1.5.3.3 Intercept Only Specification

We also estimate a specification with the intercept alone. By doing this we are not claiming that the revision and past forecast are absent from the true model. As we saw earlier, the forecast revision and past forecast are jointly significant in most countries. Rather, we estimate this specification so that we can test for unconditional bias by performing a t-test of the intercept, as is standard in the literature. Although we could provide point estimates of unconditional bias in each country using the previous specifications, we were not able to formally test if it was equal to zero.

For all forecasters, we find that there is no significant unconditional bias for most countries at at 1-year or 2-year horizon (figure 1.15), though it is more common at a 2-year horizon for GEP and WEO. Where bias is present, it is almost always optimistic rather than pessimistic.

Figure 1.15: t-test of Intercept in Intercept Only Specification



To summarise, we have shown that international organisation growth forecasts fail rationality tests for the majority of countries. The reason for the rejection of this test differs by country, but common reasons are that the forecasts are too extreme (Figure 1.13), revisions being too large (Figure 1.14), and optimistic bias (Figure 1.15).

1.6 Explanations for Rejection in Rationality Tests

In section 1.5, we conducted rationality tests for international organisation and Consensus forecasts by regressing forecast errors on an intercept, past forecasts for the target year and past forecast errors. Where the null hypothesis of the rationality test is false, then at least one of the following must be true:

- The forecaster does not satisfy forecast rationality;
- The forecast must not have a squared error loss function; or
- The vector of explanatory variables, $\boldsymbol{x}_t$ must not be in the forecaster's information set

If the null is false then this is likely because the forecaster does not satisfy forecast rationality or does not have squared error loss, rather than because the explanatory variables are absent from their information set. The reason is that the explanatory variables in our regressions are the forecaster's own forecast or the errors in their own forecasts, which are almost certainly known to the forecaster.

An extensive literature provides explanations for why different types of agents, such as consumers or firms, may make irrational forecasts. However, the explanations for irrationality at international organisations may be quite different, given that they make forecasts using different methods and face different incentives. In this section we consider three explanations that are based on specific features of international organisations, and of similar institutions such as central banks and ministries of finance.

- International organisations make forecasts that are conditional on assumed future values of a variety of explanatory variables. Such forecasts violate forecast rationality under fairly general conditions (Section 1.6.1).
- International organisations may report a 'modal' forecast, rather than a mean forecast (Section 1.6.2).

- International organisations may face incentives to make deliberately biased forecasts for some countries. We study if this is true of IMF forecasts for countries receiving IMF financing (Section 1.6.3).

1.6.1 Conditioning Assumptions

1.6.1.1 The Rational Expectations Approach

Future growth y_{t+h} depends on the future values of a variety of explanatory variables, $\boldsymbol{x_{t+h}}$, such as crude oil prices, exchange rates, nominal interest rates or government spending. As discussed in section 1.5, if a forecaster has rational expectations and a squared error loss function, then their forecast will be a conditional expectation of growth given their information set.

By the law of iterated expectations, this rational forecast equals:

$$E[y_{t+h}|I_t] = E\left[E[y_{t+h}|\boldsymbol{x_{t+h}}]|I_t\right]$$

That is, the forecaster can make a rational forecast by:

1. Computing a conditional expectation of future growth given a value of the future explanatory variables, $E[y_{t+h}|\boldsymbol{x_{t+h}}]$, for each possible value of the explanatory variables; and then

2. Averaging this conditional expectation $E[y_{t+h}|\boldsymbol{x_{t+h}}]$ over the various possible paths of the explanatory variables, weighted by the conditional probabilities of each possible value of the explanatory variables given the information set

For example, one would compute the expected path for growth for each possible scenario for crude oil prices, and then compute a weighted average of those different paths for growth weighted by the likelihood of each crude oil price scenario.

1.6.1.2 The Current Approach at International Organisations

International organisations make forecasts that are conditional on assumed future values of the explanatory variables.

In each WEO, the IMF publicly states its assumptions for exchange rates, oil prices, bond yields, monetary policy and fiscal policy. In addition, the IMF makes assumptions about other issues on an ad-hoc basis. In the April 2021 WEO, for instance, the IMF made assumptions about COVID-19 vaccination and testing (IMF 2021).

The EC and OECD follow a similar practice of making assumptions about future explanatory variables such as exchange rates and then making growth forecasts conditional on those assumptions (Genberg, Martinez, and Salemi 2014). Though central banks don't form part of our dataset, it's worth noting that they follow this practice as well (see, for instance, Kohler (2023)). It's unclear what practices private-sector forecasters follow, as their forecasting processes are rarely as well documented or publicised as those of international organisations.

1.6.1.3 Current Practice Typically Differs from Rational Expectations

We will use the term 'current practice' forecast to refer to the conditional expectation of future growth y_{t+h} conditional on an assumed future value of the explanatory variable $\widetilde{\boldsymbol{x_{t+h}}}$.

$$y_{t+h|t}^{\text{current practice}} = E\left[y_{t+h}|\widetilde{\boldsymbol{x_{t+h}}}, I_t\right]$$

Current practice will typically be inconsistent with rational expectations. This is true whether the assumed values of the explanatory variables are rational forecasts or not. We will consider each case in turn.

Case with Rational Forecasts for Explanatory Variable

First, suppose that the forecaster's assumed future value of the explanatory variable is a rational forecast for that variable, $\widetilde{\boldsymbol{x_{t+h}}} = E\left[\boldsymbol{x_{t+h}}|I_t\right]$. Then the current practice forecast will be:

$$y_{t+h|t}^{\text{current practice}} = E\left[y_{t+h}|E\left[\boldsymbol{x_{t+h}}|I_t\right], I_t\right]$$

Whether the current practice forecast of growth coincides with the rational forecast depends on two factors. Firstly, what is the relationship between future growth and the future explanatory variable? This relationship is described by the 'response function', $r(x_{t+h}) = E\left[y_{t+h}|x_{t+h}, I_t\right]$. Secondly, what is the conditional distribution of the future explanatory variable given the information set? This is described by the conditional probability density function (pdf) $p(x_{t+h}|I_t)$.

The next proposition provides sufficient conditions for the current practice forecasts to be rational. For brevity, the theorems and proofs omit the time subscripts, omit the information set, and denote $\mu \equiv E(x)$. The proofs are in appendix A.3.

Theorem 2. *The rational forecast will equal the current practice forecast if:*

(a) The response function r is affine. I.e. $r(x) = a + bx$ for constants a and b; or

(b) The response function r has 180° rotational symmetry about the point $(\mu, r(\mu))$, and the conditional pdf p is symmetric about its mean, $p(\mu - z) = p(\mu + z)$.

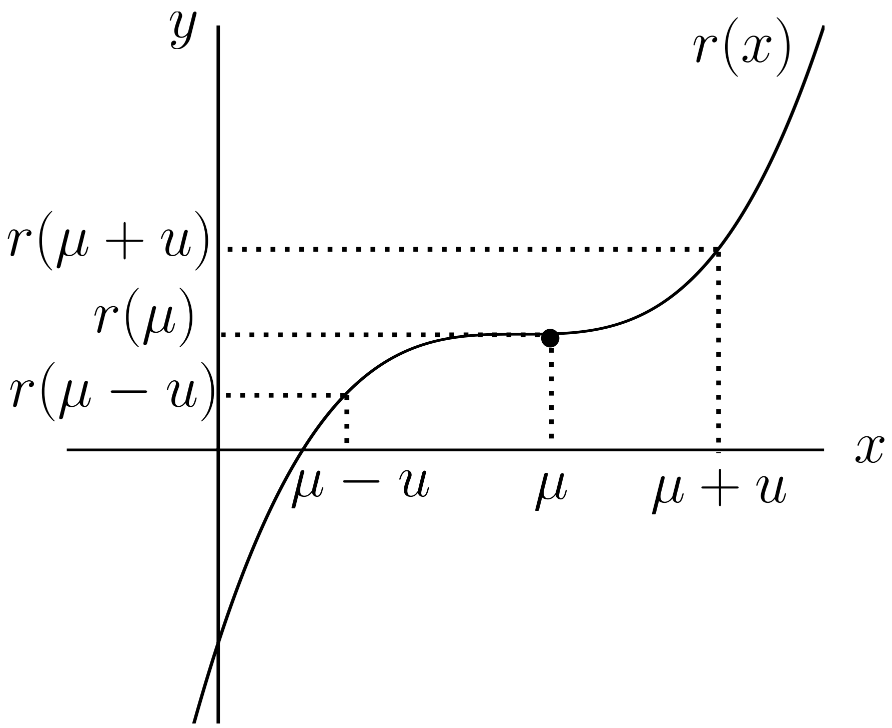
Intuitively, the rational forecast of y considers the possibility that x could be above its mean and that x could be below its mean. Under the assumptions of this theorem, these two considerations have exactly offsetting effects, so the rational forecast coincides with the current practice forecast, which assumes that x will equal its mean. Consider case (b) where r satisfies the rotational symmetry assumption. As illustrated by Figure 1.16, if the explanatory variable is a distance u above its mean, the response function would change by $(r(\mu + u) - r(\mu))$. If the explanatory variable were the same distance below its mean, the response function would change by $(r(\mu - u) - r(\mu))$. The rotational symmetry assumption states that these two changes are the same in size and opposite in sign.

$$r(\mu - u) - r(\mu) = -\left(r(\mu + u) - r(\mu)\right)$$

Since the future explanatory variable has a symmetric distribution about its mean, it is just as likely to be above its mean by u units as below its mean by the same amount. Hence, the rational forecast equals the current practice forecast.[28]

[28]An affine function is a special case of a function satisfying the rotational symmetry assumption. In this case, we do not need to assume that x is symmetric, as our argument no longer proceeds by considering upward and downward deviations of equal size. Instead, we note that since μ is the mean of x, the expected size of upward deviations, $E[x|x > \mu]$ is equal to the expected size of

Figure 1.16: Example of Response Function satisfying Rotational Symmetry Assumption



If either of these symmetry conditions fail, then the current practice forecast may differ from the rational forecast, as we now show.

Theorem 3. *Suppose the future explanatory variable has a non-degenerate distribution.*

- *If the response function r attains a strict global minimum (maximum) at the rational forecast for the explanatory variable, μ, then the rational forecast will be strictly greater than (strictly less than) the current practice forecast.*

- *If the response function r is strictly convex (strictly concave), then the rational forecast will be strictly greater than (strictly less than) the current practice forecast*

downward deviations, $E[x|x < \mu]$. Since r is affine, it is linear in any deviation of the explanatory variable from its mean. Since the expected size of upward and downward deviations is the same, and growth is linear in the deviations, the rational forecast that considers these deviations equals the current practice forecast that assumes the explanatory variable will equal its mean.

To illustrate, consider the dependence of real GDP growth on rainfall, which may be important in low income countries where agriculture is a large share of GDP. Typically, the kinds of agriculture undertaken in a given location are those that provide high yields when rainfall is typical for that location. For example, rice in areas with high rainfall, and wheat in areas with lower rainfall. Hence it may be approximately correct to assume that real GDP growth in a low income country attains a strict global maximum at the average level of rainfall, or that growth is strictly concave in rainfall. If either of these assumptions is true, theorem 3 implies that the rational forecast will be lower than the current practice forecast.

Case with Irrational Forecasts for Explanatory Variables

The preceding discussion assumed that the assumed future values of the explanatory variables are rational forecasts for those variables. This is likely to be approximately true when those assumptions are based on financial markets, as when crude oil price assumptions are taken from crude oil price futures. However, when assumptions are made in other ways they may be far from rational forecasts.

Assumptions about future fiscal policy will often differ greatly from rational forecasts. Where the IMF or EC are providing financing to a country, they assume that the country will fully implement any policies it agreed to as a condition of the financing (Genberg, Martinez, and Salemi 2014). It is unlikely that expecting full compliance is a rational forecast, given that countries often comply only in part. For other countries, the IMF and EC approach: "assumes that policy changes will occur during the forecast period only if they have been legislated in advance or if there is substantial knowledge that suggests they will occur. Otherwise, policy is assumed to remain constant during the forecast period." (Genberg, Martinez, and Salemi 2014). This is also unlikely to be a rational forecast, as it ignores the systematic tendency of governments of advanced economies to vary their fiscal policy to smooth out business cycle fluctuations.

1.6.2 Modal Forecasts

1.6.2.1 Which Forecasters make Modal Forecasts?

For a given conditional distribution of future growth, a forecaster could report the mean, the mode, the median, or some other summary statistic. Surveys of professional forecasters typically do not specify the type of point forecasts respondents should provide. For example, Elliott, Komunjer, et al. (2008) notes that the US Survey of Professional Forecasters does not specify the type of point forecasts that respondents should provide. Similarly, international organisations refer to their forecasts using vague terms such as 'central forecast'.

It's likely that at least some international organisations see themselves as making modal forecasts. One piece of evidence is that some major central banks make modal forecasts.[29] This suggests international organisations may do the same, given that they use similar forecasting models, are staffed by economists with similar training, and follow similar internal processes during forecasting rounds.

Some private-sector forecasters may also see themselves as making modal forecasts. This view was held, for instance, by JP Morgan (Davies 2017).

1.6.2.2 How would Modal Forecasts Affect Rationality Tests?

Consider a forecaster who is able to compute the conditional distribution of growth given the information available on the forecast date.

- The forecaster could report the conditional expectation of growth, $y_{t+h|t} = E\left[y_{t+h}|\mathcal{I}_t\right]$. In this case, their forecasts would satisfy rationality tests. i.e. Their forecast errors $v_{t+h|t}$ would be orthogonal to any vector of known variables $\boldsymbol{x}_t$ (see equation (1.1)).

- If the conditional distribution is unimodal, the forecaster could instead report the conditional mode of growth, $y_{t+h|t} = \text{Mode}\left[y_{t+h}|\mathcal{I}_t\right]$. If the conditional

[29]The Bank of England refers to its forecasts as modal forecasts in its main public communications, such as its monetary policy report (Bank of England 2023). Other central banks view themselves as making modal forecasts, but rarely use this term in public communications. Examples include Federal Open Markets Committee participants (Rudebusch 2008), the Federal Reserve Bank of New York (Alessi et al. 2014) and the Reserve Bank of Australia (Debelle 2017).

> distribution is also symmetric, then the mode would coincide with the mean, and these forecasts would satisfy rationality tests. However, if the conditional distribution is asymmetric, the mode would differ from the mean, and the forecasts may not satisfy rationality tests.

Modal forecasts may fail rationality tests both due to bias (as recognised in Kangur et al. (2019)) but also due to correlations of forecast errors with other explanatory variables (which has not been recognised previously). Biased forecasts will cause the rejection of rationality tests whenever the vector $\boldsymbol{x}_t$ includes a constant, because the bias would lead to the coefficient on that constant being non-zero. As shown in Bekaert and Popov (2019) and Burgess et al. (2021), past growth outcomes have been negatively skewed in most countries. If the conditional distribution of *future* growth given the forecaster's information set is also negatively skewed, then the mode will exceed the mean, and growth forecasts will have an optimistic bias.

Modal forecasts may result in correlations between forecast errors and other explanatory variables. For example, suppose that growth is a realisation from one of two non-overlapping distributions: a 'crisis' distribution centred on -5, or a 'non-crisis' distribution centred on 2. Suppose that the probability of a crisis next year is low. Then the mean will depend on both the crisis and non-crisis distribution, while the mode will only depend on the non-crisis distribution. Suppose some variable in $\boldsymbol{x}_t$, such as debt-to-GDP, is positively correlated with the probability of a crisis next year. Then that conditional expectation of growth will be a decreasing function of that variable, while the conditional mode of growth will be unrelated to that variable. This implies that, if the forecaster makes a modal forecast, their resulting forecast errors will be positively correlated with that variable, causing the failure of rationality tests.

1.6.2.3 Empirical Evidence on whether Making Modal Forecasts introduces Bias

We have just argued that the practice of making modal forecasts could result in biased forecasts. We now show empirically that modal growth outcomes tend to exceed mean growth outcomes, which suggests that this bias might arise in practice, but the effect is small. We also raise a methodological limitation that affects both our method and the similar method of Burgess et al. (2021). Due to this limitation, it remains possible that making modal forecasts contributes greatly to bias or not at all.

Conceptual Overview of Method

Suppose that for any given country and forecast horizon, the conditional mode for a period t equals the sum of a constant μ, the conditional mean $E\left[y_t|I_{t-h}\right]$, and a mean-zero error e_t.

$$\text{Mode}\left[y_t|I_{t-h}\right] = \mu + E\left[y_t|I_{t-h}\right] + e_t \quad \forall t \tag{1.8}$$

The constant μ is the expected difference between conditional mode and a conditional mean. This parameter determines whether a hypothetical forecaster that knew the conditional distribution of future growth would make biased forecasts if it reported the conditional mode rather than the conditional mean. Our aim is estimate this parameter, so our empirical strategy is focussed on growth outcomes (which are relevant to the hypothetical forecaster) rather than on the growth forecasts of any particular organisation.

Ideally, we would compute the mean and mode of the conditional distribution of future growth for a sample of periods. Using this sample, we could estimate regression (1.8), giving us an estimate of the amount of bias that modal forecasting causes for that country, $\hat{\mu}$. We could also test the hypothesis that modal forecast does not introduce bias, $\mu = 0$, against the alternative that it does.

Unfortunately, we don't observe the conditional distribution of growth given particular information sets. Rather, we observe a set of growth outcomes, which provides information about the unconditional distribution of growth. Hence, we need to use our model of conditional distributions (equation (1.8)) to make a prediction about unconditional distributions. We do this by taking expectations of equation (1.8) over possible information sets.[30]

$$E\left[\text{Mode}\left[y_t|I_{t-h}\right]\right] = \mu + E\left[y_t\right] \quad \forall t$$

[30]By the law of iterated expectations: $E\left[E\left[y_t|I_{t-h}\right]\right] = E\left[y_t\right]\ \forall t$. The error is zero on average across information sets by assumption.

We then make the further assumption that the unconditional mode, $\text{Mode}\left[y_t\right]$, equals the expectation of the conditional modes, $E\left[\text{Mode}\left[y_t|I_{t-h}\right]\right]$. This assumption is a limitation of the method, as discussed below. With this assumption:

$$\text{Mode}\left[y_t\right] = \mu + E\left[y_t\right] \quad \forall t \tag{1.9}$$

Given a sample of growth outcomes, we would like to estimate the unconditional mode and the unconditional mean, as the difference between them would provide an estimate of μ. A challenge is that a country's growth may be a non-stationary process, so its mean and mode may vary over time. For example, some East Asian economies such as Japan experienced high 'catch-up' growth followed by more moderate growth. To allow for this, we assume that a country's growth is the sum of a deterministic component b_t and a stationary stochastic component z_t.

$$y_t = b_t + z_t$$

Substituting this into equation (1.9), and recognising that deterministic components can be taken outside of the mode or expectation operator, we have:

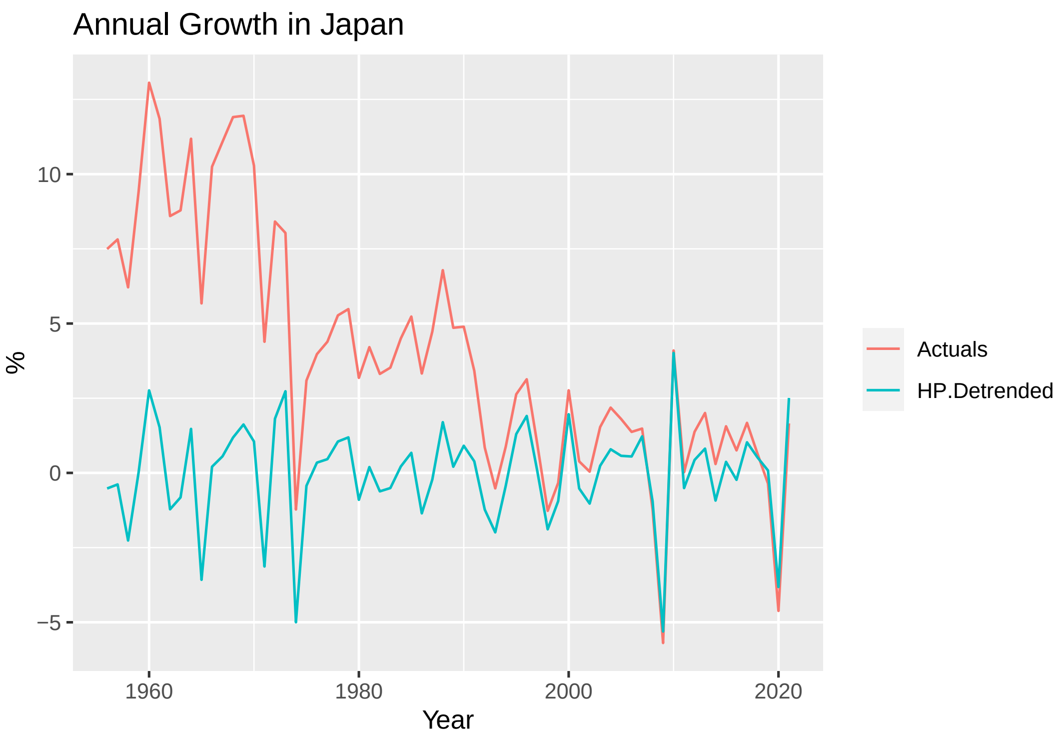
$$\mu = \text{Mode}\left[z_t\right] - E\left[z_t\right]$$

Hence, we estimate the bias introduced by making modal forecasts, μ, as the estimated mode of de-trended growth less the sample mean of de-trended growth. This is somewhat similar to Burgess et al. (2021), who compare the sample mean of de-trended growth to the sample median of de-trended growth.

Implementation and Results

We use annual real GDP growth data from the WEO database for all countries with at least 40 observations. To estimate the stochastic component, we convert the growth rates into log levels, compute detrended log levels, and then compute growth rates of the detrended log levels. We detrend log levels using a two-sided Hodrick-Prescott filter (Hodrick and Prescott 1997), with the parameter value $\lambda = 6.25$, following the recommendation for annual data given by Ravn and Uhlig (2002). Figure 1.17 compares actual growth rates y_t and our two estimates of the stationary stochastic component z_t for Japan.

Figure 1.17



We estimate the mode of de-trended growth using the half-sample mode estimator as implemented by the 'modeest' R package (Poncet 2019). This estimator assumes that the data are drawn from a continuous unimodal distribution, but doesn't require the assumption of any specific parametric distribution, and has the advantage of being insensitive to outliers (Bickel and Frühwirth 2006). We estimate the mean of de-trended growth using the sample mean.

We find that 95 of 148 countries have an estimated mode above the sample mean. These differences are small in magnitude, however. On average across these countries, the estimated $\hat{\mu}$ is 0.2 percentage points.

As a robustness check, we also compute the sample skewness of detrended growth for each country. Skewness is important because, if the distribution of the de-trended growth is unimodal, then its mode will exceed its mean if and only if the distribution is negatively skewed.

$$\widehat{Skew(z_t)} = \frac{\frac{1}{n}(z_t - \bar{z}_t)^3}{\left[\frac{1}{n}(z_t - \bar{z}_t)^2\right]^{\frac{3}{2}}}$$

We find that 103 of 148 countries have negative sample skewness. This supports our finding that the unconditional mode exceeds the unconditional mean in most countries. Without this complementary evidence, one might have wondered if our results were driven by the specifics of our mode estimator.

Limitations

Our method assumes that the unconditional mode, $\text{Mode}\,[y_t]$, equals the expectation of the conditional modes, $E\,[\text{Mode}\,[y_t|I_{t-h}]]$. This assumption holds exactly if the conditional mode does not vary with respect to variables in the information set. This is possible in principle. Suppose that there are two possible information sets: one that makes a crisis very unlikely, and one that makes a crisis somewhat unlikely. For both of these information sets, the crisis will probably not occur, so the conditional mode of growth is the same.

It is more likely, however, that the conditional mode of growth does vary with respect to the information set. In this case, the assumption may or may not hold. We will illustrate this in the extreme case where the information sets fully determine growth. Suppose that there are three possible information sets: 'pre-bust', 'ordinary' and 'pre-boom', which guarantee that growth will be -2, 2 or 6 respectively. First we consider the symmetric case, where 'pre-bust' and 'pre-boom' each occur with probability $\frac{1}{4}$ and 'ordinary' occurs with probability $\frac{1}{2}$. In this case, the assumption holds:

$$\text{Mode}\,[y_t] = 2$$

$$E\,[\text{Mode}\,[y_t|I_{t-h}]] = \frac{1}{4} \times -2 + \frac{1}{2} \times 2 + \frac{1}{4} \times 6 = -\frac{1}{2} + 1 + \frac{3}{2} = 2$$

In practice, business cycles are likely to be asymmetric. Suppose that the probability of 'pre-bust' is $\frac{1}{4}$, of 'ordinary' is $\frac{3}{4}$ and 'pre-bust' is 0. Then:

$$\text{Mode}\,[y_t] = 2$$

$$E\,[\text{Mode}\,[y_t|I_{t-h}]] = \frac{1}{4} \times -2 + \frac{3}{4} \times 2 = -\frac{1}{2} + \frac{3}{2} = 1$$

As this discussion illustrates, unconditional modes may differ from the expected conditional mode in the face of asymmetric business cycles, which is a situation we are interested in. Hence, one should be cautious both about the results of our method (which suggest that modal forecasts may contribute slight optimistic bias), but also about any other effort to explain forecast bias using the unconditional distribution of growth. In particular, comparing the sample median and sample mean of growth as in Burgess et al. (2021) provides an unbiased estimate of the difference between the conditional median and conditional mean only if one assumes that $E\left[\text{Median}\left[y_t|I_{t-h}\right]\right] = \text{Median}\left[y_t\right]$, which is subject to an analagous critique.

1.6.3 Optimistic Bias in Forecasts for Program Countries

The failure of rationality tests could occur because the forecaster makes deliberately biased forecasts for some countries because they are motivated by considerations other than forecast performance. One setting where this could occur is the IMF program forecasts. The literature often claims that the IMF is motivated by concerns other than forecast performance when making these forecasts, such as a desire to justify the provision of financial resources, to encourage investors to provide financing to the country, or to placate governments who would prefer to receive optimistic forecasts (see section 1.2.3.3).

In this section, we show that the IMF forecasts have an optimistic bias, and that the bias is larger for larger programs. In isolation, this could be interpreted as evidence for deliberate bias. However, we also show that Consensus forecasts are also more optimistic for larger programs. Since the private-sector forecasters lack the same incentives to make biased forecasts, this casts some doubt on claims of deliberate bias by the IMF. It suggests instead that the optimistic bias arises from some source that affects both IMF and private-sector forecasters, such as a tendency for forecasters to be too optimistic in the face of severe crises. Our results cannot, however, rule out deliberate bias. It is possible that the IMF forecasts have a deliberate optimistic bias, and that because private-sector forecasters try to make similar forecasts to the IMF, their forecasts are biased as well.

1.6.3.1 Evidence for IMF Program Forecasts

We use some additional data specific to this subsection.

- **Forecasts**: We use IMF real GDP growth forecasts from the Monitoring of Fund Arrangements (MONA) database. The IMF makes MONA forecasts for a country at the inception of each program and revises them at each program review.[31] This differs from the IMF's WEO forecasts used elsewhere in this chapter, which are produced at a regular frequency for all countries.

- **Data on Programs**: We use data on a sample of 214 IMF programs approved in the period 2002-2018. Table 1.3 reports summary statistics for the variable of interest, the amount of the program: there is a large heterogeneity in the size of programs approved, ranging from a minimum of 5% of a country's IMF quota (Rwanda 2002) to a maximum of 3211.8% (Greece 2010) with a standard deviation of roughly 360 percentage points.

Table 1.3: Descriptive Statistics - Amount Programs

Statistic	N	Mean	St. Dev.	Min	Max
Amount (% of quota)	214	228.9	363.8	5.0	3,211.8

Note: Table shows descriptive statistics for IMF programs size (in % of member country quota at Board approval) over the period 2002-2018. Data retrieved from MONA database.

We study this question using the rationality testing framework as in section 1.5. We assume the linear relationship in equation (1.1) between the IMF's forecast error $v_{t+h|t}$ and a vector $\boldsymbol{x}_t$ of explanatory variables known at the time the forecasts were made, and then test hypotheses about individual parameters in the vector $\boldsymbol{\theta}$ using t-tests with HAC standard errors. However, rather than estimating a separate time series regression for each country, we estimate a single pooled regression using the data for all countries in all countries. We take this approach because a typical

[31]The full database is divided into two periods: 1993–2003 and 2002 to present. The reason behind this distinction is the reclassification and restructuring of several economic variables that occurred in the early 2000s (Luna 2014). For this reason, we concentrate only on the second part of the database.

country only has a handful of programs, so the only way to obtain reasonably precise estimates is to pool data across countries.

Table 1.4: Rationality Regressions for IMF Program Forecasts

	Dependent variable:									
	GDP forecast error (current year)					GDP forecast error (year ahead)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Post-GFC		0.340					−0.275			
		(0.225)					(0.314)			
Total amount (% quota)			-0.001^{**}	-0.001^{**}	-0.001^{*}			-0.001^{**}	-0.001^{**}	-0.001^{***}
			(0.0003)	(0.0003)	(0.0003)			(0.0004)	(0.0004)	(0.0005)
Remaining months				0.024					0.034	
				(0.033)					(0.046)	
Non-Concessional					0.211					-0.563^{*}
					(0.250)					(0.340)
Constant	-0.348^{***}	-0.525^{***}	−0.174	−0.303	−0.323	-0.665^{***}	-0.504^{**}	-0.425^{**}	-0.610^{**}	−0.026
	(0.113)	(0.162)	(0.132)	(0.219)	(0.221)	(0.154)	(0.240)	(0.179)	(0.303)	(0.300)
Observations	225	225	225	225	225	216	216	216	216	216
F Statistic		2.282	6.081^{**}	3.305^{**}	3.394^{**}		0.767	6.467^{**}	3.514^{**}	4.635^{**}

Note: Dependent variable winsorized at the 10% level. For columns 1-5 the dependent variable is equal to the current year forecast error, while for column 5-10 is equal to the year-ahead forecast error. Post-GFC is a dummy equal to 1 for programs approved after 2009. Remaining months is a variable that represents the number of months remaining before the end of the year from the date of program approval. Non-concessional is a dummy equal to 1 for non-concessional GRA programs (see note 24). Heteroskedasticity robust standard errors in parentheses.***: significant at 1% level, **: significant at 5% level, *: significant at 10% level.

Table 1.4 shows 5 different regression specifications for current and year-ahead forecasts. Each specification corresponds to a different choice of vector $\boldsymbol{x}_t$. The first two specifications test for bias in the IMF's MONA forecasts over different samples. In specification (1), we test for bias over the full sample by setting $\boldsymbol{x}_t = 1$. We find that the current-year forecasts have an optimistic bias of -0.3% on average while the year-ahead forecasts have an optimistic bias of -0.7% on average. In specification (2), we test if the amount of bias differs between MONA forecasts for programs approved before the GFC versus programs approved after the GFC. To do this, we set $\boldsymbol{x}_t = [1\ (PostGFCdummy)]$, where the dummy equals 1 for all observations after the GFC. We find somewhat limited evidence of this: if anything, optimistic biases are smaller for current-year forecasts, but not for year-ahead.

The remaining three specifications test if the amount of bias is related to the amount of financing provided under the program. Specification (3) tests this by setting $\boldsymbol{x}_t = [1 \;\; (Totalamount)]$. We find that a 1% higher amount corresponds to a forecast error lower by -0.001%, indicating that forecasts for countries receiving larger programs have a more optimistic bias. One possible source of omitted variable bias is that the number of months remaining in the year may directly affect bias (as it makes the forecast horizon longer) and may be correlated with the size of IMF programs. To address this, specification (4) adds the number of remaining months, and still shows that forecasts for countries with larger programs have a more optimistic bias. Finally, specification (5) adds a dummy for whether the program is concessional or non-concessional.[32] If larger programs are concessional and at the same time present larger forecast errors, the coefficient would be negatively biased. The introduction of the dummy has no effect for current year forecasts. However, the dummy has a statistically significant negative coefficient of -0.5 for the year-ahead, indicating that non-concessional programs exhibit more optimistic forecasts. Nevertheless, the coefficient on the amount of financing does not lose magnitude and remains significant: this means that the relationship holds independently from the difference concessional/non-concessional.

The evidence reported in Table 1.4 supports the hypothesis of more optimistic forecasts for bigger programs.

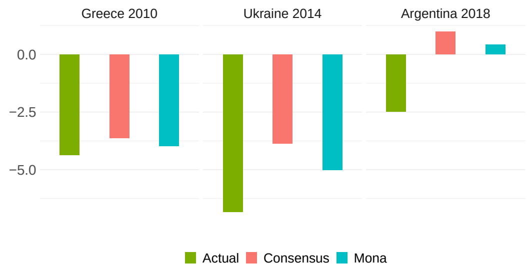
1.6.3.2 Evidence for Consensus Forecasts

Private-sector forecasters lack the same incentives to make optimistic forecasts for program countries. Nonetheless, they have made overly optimistic forecasts for some countries receiving high-profile IMF programs. As shown by figure 1.18, in some instances Consensus was even more optimistic than the IMF (e.g. Argentina 2018).

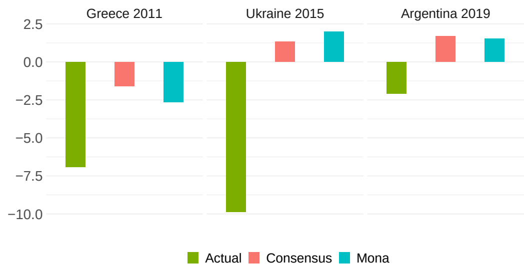
[32] Non-concessional programs come from the General Resources Account and are Stand by Arrangements, Extended Fund Facilities, Precautionary and Liquidity Lines and Precautionary Credit Lines.

Figure 1.18: Comparison with Consensus: Anecdotal Evidence

(a) Current year



(b) Year-ahead



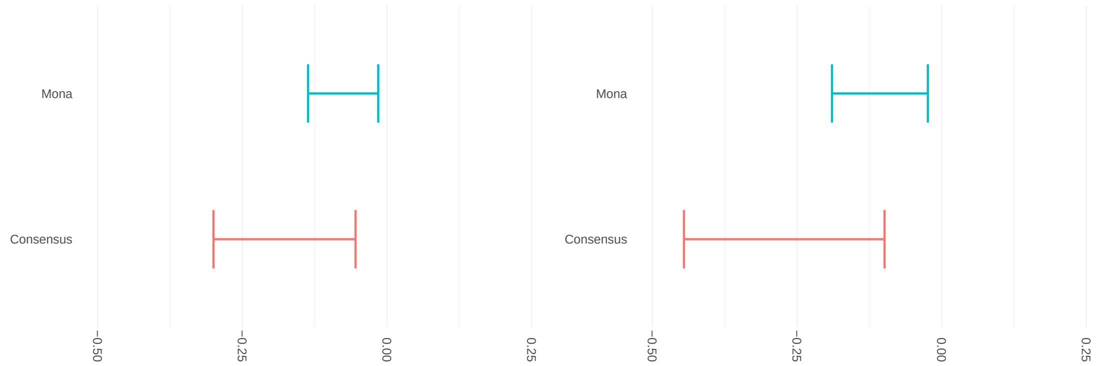
Note: The figure shows the real GDP growth forecasts formulated at program approval for Greece 2010, Ukraine 2014 and Argentina 2018 SBAs from MONA database and Consensus survey (mean value). Actual growth rate is the real GDP growth from the October WEO issue of the following year.

We test formally this equivalence substituting MONA forecasts with Consensus forecasts produced in the same month, calculate the forecast error and estimate the same regressions of Table 1.4 (columns 4-9). Figure 1.19 shows the estimated relationship between forecast errors and the total amount of the program for MONA and Consensus.[33] For both IMF and Consensus, the forecasts are more optimistic for countries receiving larger programs.

[33]Since Consensus forecasts are not available for all programs, we estimate both regressions on a subset of 70 programs.

Figure 1.19: Comparison with Consensus: Statistical Relationship

(a) Current year **(b)** Year-ahead



Note: The figure shows 95% confidence intervals obtained regressing the forecast errors for programs on the amount of the program (in % of country quota) and multiplying by 100. Data for forecasts are, respectively, from MONA and Consensus. The sample of programs included in the regressions corresponds to data availability for Consensus (70 programs).

1.7 Conclusion

This chapter provided new evidence and theory concerning real GDP forecasts by international organisations. The chapter began by comparing the accuracy of four international organisations and the average of Consensus forecasts, finding statistically significant differences for only a few countries. It presented a new decomposition of differences in accuracy into contributions of disagreement and the accuracy of the average forecast. This decomposition was then used to show that, on average across countries, differences in accuracy are larger at longer forecast horizons due to both greater disagreement and lower accuracy of the average forecast. The chapter then turned to forecast rationality. For all forecasters, we rejected the null of the rationality test for most countries. Often the irrationality took the form of optimistic bias, overly large revisions or forecasts that were too extreme. Finally, we contributed the literature that attempts to explain why rationality is often rejected in these tests. We focussed on three explanations: the use of conditioning assumptions; the practice of making modal forecasts; and incentives to make biased forecasts for specific countries.

Our findings have important implications for the practices of international organisations. Firstly, we found that the international organisations are quite similar to one another in terms of accuracy and the ways in which they fail rationality tests. This is unsurprising given the close collaboration between the international organisations, especially the IMF and World Bank. However, it suggests that if international organisations want to obtain alternative views on their forecasts, they should prioritise seeking views from outside the community of international organisations. Secondly, our findings highlight the importance of forecasters being transparent, including about their conditioning assumptions and whether they see themselves as making modal forecasts. This would assist users in interpreting the forecasts. Knowing that a forecaster is making modal forecasts, for instance, would tell users that they should not view the forecasts as incorporating tail risks like financial crisis. Transparency would also make it easier to choose among competing explanations for the failure of rationality tests, which would help organisations determine what actions to take in response.

The lessons of this chapter could also be applied to other types of forecasters. For example, one could ask whether differences in the accuracy of private-sector forecasters are larger at longer horizons, and if so, use our decomposition to assess if this reflects greater disagreement or lower accuracy. Researchers could also investigate whether other types of forecasters make conditioning assumptions or modal forecasts.

2

Temporal Aggregation and the Spurious Predictability of Exchange Rates

Martin McCarthy, Stephen Snudden[1,2]

Abstract

Forecasts of period-average exchange rates require the use of disaggregated data to both construct efficient forecasts and to correctly test the null of no-predictability. However, in a survey of the literature, we show that all existing studies of period-average exchange rates failed to do either. To rectify this, we construct real-time vintages at the daily frequency for all measures of exchange rates and for all countries. We document that the end-of-period no-change forecast, which reflects the null of no-predictability, substantially outperforms the period-average no-change forecast used as a benchmark in existing studies. Moreover, we find that forecasting models are more accurate at short horizons when estimated with end-of-period or daily inputs than with period-average inputs. Finally, we show that models using such high-frequency information almost always outperform the period-average no-change benchmark. In contrast, when these models are compared against the end-of-period no-change benchmark, we find little evidence of predictability at short horizons for nominal bilateral exchange rates, but some evidence of predictability for real and effective exchange rates.

[1]Wilfrid Laurier University

[2]We thank Reinhard Ellwanger and conference participants at the 43rd International Symposium on Forecasting for their discussions, comments, and suggestions

2.1 Introduction

Exchange rate forecasting plays a pivotal role in guiding economic decisions, ranging from policymaking to investment strategies. Of particular interest in macroeconomic analysis are period-average exchange rates, both effective and bilateral, particularly when expressed in real terms. These metrics provide valuable insights into a country's competitiveness, trade dynamics, inflation, and broader economic conditions. In macroeconomics, the average exchange rate over a period of time ('period-average' rates) are often more useful than exchange rates at specific points in time ('point-sampled' rates). This is because period-average rates are more relevant to variables measured as flows over time, such as Gross Domestic Product and net exports. For this reason, many policymakers including international organizations and central banks, make assumptions about period-average exchange rates as part of their routine forecasting processes (see, for instance, Glas and Heinisch 2023; International Monetary Fund 2023; Wieland and Wolters 2013). Unfortunately, both policymakers and the academic literature have given scant attention to temporal aggregation challenges inherent in period-average exchange rates. In this chapter, we explore this critically overlooked facet of exchange rate forecasting, with a focus on rectifying limitations that pervade the existing literature. This chapter makes four contributions, which we now describe in turn.

Surveying Temporal Aggregation Aspects of Exchange Rate Forecasting Literature

The first contribution is to carefully survey the temporal aggregation aspects of the existing literature on forecasting exchange rates. We consider four types: nominal effective exchange rates (NEERs), real effective exchange rates (REERs), bilateral nominal exchange rates (NERs) and bilateral real exchange rates (RERs). We systematically assess the methods used in each paper, including whether it forecasts levels or returns, as well as the temporal dimensions of the data used in estimation and the benchmark against which forecasts are evaluated. Additionally, we consider whether the forecasts were constructed in real-time. In doing so, we complement existing surveys of this extensive literature by extending research from the last decade and summarizing temporal methods (i.e. Engel, Mark, West, et al. 2007; Frankel and Rose 1995; Rogoff 1996; Rossi 2013). We show that the literature suffers from three limitations.

The first limitation identified in the literature survey concerns the choice of benchmark used to test the predictability of exchange rates. A large literature tests the predictability of exchange rates by comparing forecasts from a candidate model against naive no-change forecasts. We show that when this literature tests the predictability of *period-average* exchange rates, they always compare against the period-average no-change benchmark, which assumes that the period-average level in each future period will equal the current period-average level. Unfortunately, gains relative to this benchmark are expected *by construction*, whether one is forecasting levels or returns.

- If the daily levels follow a random walk (and are therefore unpredictable by definition), growth rates of the period-average levels will be serially correlated (Working 1960). Hence, models for forecasting growth rates are theoretically expected to outperform the benchmark that the growth rates will equal zero, which is equivalent to the period-average no-change benchmark for the levels.

- Similarly, if the daily levels follow a random walk (or any other autoregressive integrated moving average process), models for forecasting period-average levels are theoretically expected to outperform the period-average no-change benchmark (Ellwanger and Snudden 2023a; Marcellino 1999; Weiss 1984).

Since this outperformance of the benchmark arises by construction, it has been typically referred to as 'spurious predictability'. To avoid such spurious predictability, forecasts of period-averages need to be compared against the end-of-period no-change forecast, which assumes that the period-average level in each future period will equal the current end-of-period level. This is because only the end-of-period no-change reflects the null hypothesis that all future exchange rates, averaged or not, are conditionally unpredictable (Ellwanger and Snudden 2023a).

The second limitation identified in the literature survey concerns the efficiency of the forecasts. We find that all existing studies evaluating forecasts of period-average exchange rates estimate models on period-average inputs rather than daily or end-of-period inputs. This is concerning, as constructing forecasts with period-average inputs can substantially worsen forecast accuracy (Amemiya and Wu 1972; Ellwanger and Snudden 2021; Kohn 1982; Lütkepohl 1986; Wei 1978).

The final limitation is that, while several studies have examined real-time forecasts of point-sampled bilateral NERs, (e.g. Clarida, Sarno, et al. 2003; Clarida and Taylor 1997; Faust, Rogers, et al. 2003), no study has yet examined real-time forecasts of any EERs, or for point-sampled bilateral RERs. It is therefore unclear if methods advocated for by the existing literature for these exchange rates would be useful if operationalized in practice.

New theory on directional accuracy

The second contribution of the chapter is to explore how to assess the directional accuracy of forecasts of temporal aggregates. We show that only comparisons against the end-of-period no-change benchmark reflect the null of no-directional accuracy for forecasts of temporally aggregated data. We prove that, if one computes a success ratio (SR) using a month-average no-change benchmark, then the SR is expected to be above half even if the daily series is a random walk. We then use simulations to show that if the test of Pesaran and Timmermann 2009 is performed with the period-average no-change benchmark, the type I error rate will be higher than the intended level of significance, and power will be reduced. These results complement previous literature on testing mean-squared accuracy of temporally aggregated series (Ellwanger and Snudden 2023a), and on testing directional accuracy in the absence of temporal aggregation (Pesaran and Timmermann 1992, 2009).

Constructing a real-time exchange rate dataset

The third contribution of the chapter is to construct *real-time* vintages for the four types of daily exchange rate (bilateral NERs, bilateral RERs, NEERs and REERs), for every country. This is the first real-time dataset for EERs, which has thus far only been examined for bilateral rates (i.e. Clarida and Taylor 1997). Moreover, it is first time that daily time series of REERs have been constructed for any country. The daily effective exchange rates (EERs) are computed in a consistent way across countries, and use the official International Monetary Fund (IMF) weights.

New evidence on forecasting period-average exchange rates

The final contribution of the paper is to quantify the real-time information gains from temporal disaggregation using the newly constructed daily measures of exchange rates. This includes an evaluation of the effects of temporal aggregation for model-based and no-change forecasts. The investigation also includes testing, for the first time, real-time out-of-sample forecasts of period average exchange rates against the null of no predictability. We find three empirical results regarding the quantification of the importance of temporal aggregation bias in exchange rate forecasting.

The first empirical finding is that, for all measures of exchange rates and for almost all countries, the month-average no-change benchmark is less accurate than the end-of-month no-change benchmark. The difference in performance is large because daily exchange rates are highly persistent. This evidence is consistent with prior evidence on the information loss from end-of-period versus monthly average no-change forecasts in commodity prices, stock prices, and a bilateral exchange rate (Ellwanger and Snudden 2021, 2023a). This finding suggests that studies that found forecast improvements relative to the period-average no-change are unlikely to find that such gains translate when compared to the end-of-period no-change. The second empirical finding is that both direct and recursive forecasts estimated with month-average data perform substantially worse than forecasts estimated with daily or end-of-month inputs. This is found to be very robust across exchange rate measures and for most countries. Once again, this substantiates theoretical concerns regarding the loss in forecast accuracy when exchange rates are temporally aggregated. These findings are also encouraging, as they show that substantial gains in model-based forecast accuracy of period average exchange rates can be achieved in real-time using information from daily exchange rates, relative to current methods that are employed. Moreover, the results suggest that the point-sampled forecasts of EERs (Ca'Zorzi et al. 2022; Kohlscheen et al. 2017; Zhang, Dufour, et al. 2016) and bilateral RERs (i.e. Chen, Jackson, et al. 2014; Froot and Ramadorai 2005) could be potentially quite informative on the desirable methods to forecast and the general predictability of period-average exchange rates.

While our empirical results relate to month-average exchange rates, the information loss caused by temporal aggregation is likely even greater for quarter-average exchange rates. This issue is especially relevant to countries whose official Consumer Price Index (CPI) data is only available quarterly (such as Australia and New Zealand), as exchange rate forecasts for these countries are often done with quarterly data.[3]

The chapter is organized as follows. Section 2.2 surveys the literature. Section 2.3 presents analytical results and simulations to show that one should assess the directional accuracy of forecasts by comparison to an end-of-period no-change benchmark. Sections 2.4, 2.5 and 2.6 describe the methods, data, and results of our out-of-sample forecast evaluation, respectively.

[3]Information is also likely to be lost by intraday temporal aggregation. This suggests estimating models on end-of-day exchange rate data rather than day-average exchange rate data, though this is only likely to be relevant at very short horizons of up to a few weeks ahead.

2.2 Literature Survey

This literature review offers a comprehensive analysis of research on forecasting EERs and bilateral exchange rates. For each paper, we assess the type of exchange rate targeted, including if real or nominal and if in levels or returns. We also summarize the frequency and sampling of the data of the forecast target, in estimation, and the benchmark against which forecasts are evaluated. We confine our examination to papers published or accepted for publication as of 2022. We also record if forecast analysis was conducted in 'real-time', defined as forecasts made with models estimated only on data available at the time of the forecast (see for example Clarida and Taylor (1997)). Specifically, if the exchange rates are expressed in real terms, then this requires that they are computed using CPI observations available at the time of the forecast (noting that CPI data is published with a lag and is subject to revision). Moreover, for EERs, this requires real-time treatment of the trade weights.

As the main focus of the survey is the temporal methods used for the forecasts, our survey separately documents forecasts of point-sampled and period average exchange rates. We also delineate studies into those that examine EERs, Section 2.2.1, and bilateral exchange rates, Section 2.2.2. In cases where papers forecast multiple types of exchange rates, we include them in each section. The surveyed papers appear in a dedicated reference list at the end of this thesis.

2.2.1 Effective Exchange Rates

Our initial focus is on forecasts of EERs, which are prominent in macroeconomics. REERs are important because they reveal relative price levels between a nation and its trade partners, which influences trade flows. NEERs are useful summaries of a country's nominal exchange rate with its trading partners. Among other things, they can be used to forecast the extent to which nominal exchange rate movements will contribute to domestic inflation (Dornbusch 1987; Forbes et al. 2018; Goldberg and Knetter 1996; Shambaugh 2008).

2.2.1.1 Forecasts for Period-Average Effective Exchange Rates

We found 17 papers that examined forecasts of period-average EERs, as summarized in Table 2.1. Around half of these papers concentrate on forecasts of the level of EERs rather than returns in EERs, with the focus on real versus nominal EERs also approximately split. Most studies forecast month-average EERs, although there's a recent trend towards forecasting quarter-average EERs.

Table 2.1: Papers Forecasting Period-average Effective Exchange Rates

Paper	Level or Return	Frequency	Forecast Target	Benchmark	Model Estimation	Real or Nominal	Real-time
Hooper and Morton (1982)	Level	M, Q	Average	Average	Average	Both	N
Meese and Rogoff (1983a)	Level	M	Average	Average	Average	Nominal	N
Meese and Rogoff (1983b)	Level	M	Average	Average	Average	Real	N
Boughton (1986)	Both	M	Average	Average	Average	Both	N
Throop (1993)	Return	Q	Average	Average	Average	Real	N
MacDonald (1997)	Level	Q	Average	Average	Average	Real	N
Amano and Norden (1998a)	Return	M	Average	Average	Average	Real	N
Amano and Norden (1998b)	Level	M	Average	Average	Average	Real	N
Siddique and Sweeney (1998)	Level	M	Average	Average	Average	Real	N
Sarantis (1999)	Level	M	Average	Average	Average	Real	N
Gourinchas and Rey (2007)	Return	Q	Average	Average	Average	Nominal	N
Adrian et al. (2009)	Return	M	Average	Average	Average	Nominal	N
Chen et al. (2010)	Return	Q	Average	Average	Average	Nominal	N
Chen et al (2014)	Level	A	Average	Average	Average	Nominal	N
Ca'Zorz et al. (2016)	Level	M	Average	Average	Average	Real	N
Ca'Zorz et al. (2017)	Level	Q	Average	Average	Average	Real	N
Hatzinikolaou and Polasek (2019)	Return	Q	Average	Average	Average	Nominal	N

Note: "Benchmark" refers to the no-change forecast that the forecast was compared against. "Model Estimation" refers to the data used in estimation.

The three limitations of the exchange rate forecasting literature identified in the introduction (section 2.1) apply to the period-average EER literature. All studies that test the predictability of period-average EERs by comparing candidate forecasts to a naive forecast have done so using the period-average no-change benchmark. This is problematic, because as explained in section 2.1, it is possible to outperform this benchmark even if the daily series is a random walk and hence unpredictable by definition.

Secondly, the literature on period-average EERs always uses models estimated on period-average data. However, this is expected to compromise forecast accuracy due to information loss from temporal aggregation, as discussed in the introduction. The degree of the information loss is an empirical question, quantified in Section 2.6.

Finally, we find that no study has conducted a real-time forecast evaluation for any period-average EERs. Hence, it remains unclear if the methods proposed in existing studies would be useful in practical applications if adopted by policymakers or other forecasters. The lack of real-time forecast evaluations may reflect the absence of real-time EER data vintages that account for the delay in the publication of trade weights, a gap that we remedy with our dataset (see Section 2.4).

2.2.1.2 Forecasts for Point-sampled Effective Exchange Rates

Only three studies evaluate forecasts for end-of-period EERs, see Table 2.2. As was the case for period-average EERs, none of the studies use real-time methods. Forecasts for end-of-period NEERs were examined by Kohlscheen et al. (2017) and Zhang, Dufour, et al. (2016). Zhang, Dufour, et al. (2016) specifically discuss the information loss from temporal aggregation in their motivation of daily forecasts of NEERs. Additionally, Ca'Zorzi et al. (2022) stand alone in examining forecasts of end-of-period real EERs, which they construct for a basket of eight advanced economies. These studies compared forecasts against end-of-period no-change benchmarks and, hence, unlike the studies examining period average exchange rates, correctly tested against the null of no predictability.

Table 2.2: Papers Forecasting Point-sampled Effective Exchange Rates

Paper	Level or Return	Frequency	Forecast Target	Benchmark	Model Estimation	Real or Nominal	Real-time
Kohlscheen et al. (2016)	Return	D	EoP	EoP	EoP	Nominal	N
Zhang et al. (2016)	Return	D	EoP	EoP	EoP	Nominal	N
Ca'Zorzi et al. (2022)	Level	Q	EoP	EoP	EoP	Real	N

Note: "Benchmark" refers to the no-change forecast that the forecast was compared against. "EoP" refers to end-of-period sampling. "Model Estimation" refers to the data used in estimation.

The valid hypothesis testing in these papers is potentially informative on the predictability of period-average EERs exchange rates. This is because, under certain conditions, a forecast for the end-of-period EER can be an excellent forecast of the period-average at short horizons when the underlying series is persistent, and converges to the efficient forecast at long horizons (Ellwanger and Snudden 2021). However, this question can only be answered quantitatively. Due to the interest in the forecastability of period-average EERs in macroeconomics, we examine the efficiency of point-sampled forecasts for period averages for all countries in Section 2.6.

2.2.2 Bilateral Exchange Rates

2.2.2.1 Forecasts for Period-Average Bilateral Exchange Rates

We now examine the literature on forecasting period-average bilateral exchange rates. Bilateral exchange rates provide insights into relative price levels between a pair of countries and are therefore relevant to flows between them. The body of research on period-average bilateral exchange rates is less extensive than that on EERs, with only twelve papers, see Table 2.3. Moreover, only three papers examine period-average bilateral RERs, and only one of those forecasts the level. In contrast to EERs, a few papers employ real-time methods for period-average bilateral exchange rates in nominal terms (Abbate and Marcellino 2018; Carriero, Kapetanios, et al. 2009; Molodtsova, Nikolsko-Rzhevskyy, et al. 2008; Wright 2008) and one in real terms (Kilian and Taylor 2003).

Table 2.3: Papers Forecasting Period-Average Bilateral Exchange Rates

Paper	Level or Return	Frequency	Forecast Target	Benchmark	Model Estimation	Real or Nominal	Real-time
Amano and Norden (1995)	Return	M	Average	Average	Average	Real	N
Kilian and Taylor (2003)	Level	Q	Average	Average	Average	Real	Y
Islam and Hasan (2006)	Level	Q	Average	Average	Average	Nominal	N
Issa et al. (2008)	Return	Q	Average	Average	Average	Real	N
Molodtsova et al. (2008)	Return	Q	Average	Average	Average	Nominal	Y
Wright (2008)	Return	M, Q	Average	Average	Average	Nominal	Y
Carrieroa et al. (2009)	Level	M	Average	Average	Average	Nominal	Y
Molodtsova and Papell (2009)	Return	M	Average	Average	Average	Nominal	N
Giacomini and Rossi (2010)	Return	M	Average	Average	Average	Nominal	N
Fratzscher et al. (2015)	Return	M	Average	Average	Average	Nominal	N
Abbate and Marcellino (2018)	Level	M	Average	Average	Average	Nominal	Y
Eichenbaum et al. (2021)	Return	Q	Average	Average	Average	Nominal	N

Note: “Benchmark" refers to the no-change forecast that the forecast was compared against. “Model Estimation" refers to the data used in estimation.

Unfortunately, like for EERs, all papers summarized are found to compare forecasts to the period-average no-change benchmark, and never to the end-of-period no-change benchmark. As with the EER literature, forecasts are expected to outperform the period-average no-change benchmark by construction, even if the daily series is a random walk and hence unpredictable by definition. This reveals that for both bilateral and EERs, there is a critical gap in the understanding of the forecastability of period-average exchange rates. Moreover, like EERs, these studies universally use period-average inputs in estimation, potentially jeopardizing forecast accuracy. In essence, our understanding of the predictability of period-average bilateral exchange rates remains limited.

2.2.2.2 Forecasts for Point-Sampled Bilateral Exchange Rates

Lastly, we delve into the literature which has examined point-sampled bilateral exchange rates. Researchers may favor bilateral point-sampled exchange rates over bilateral period-average rates when precision is paramount, such as in asset valuation or trade settlements at specific time intervals. Our survey documents 14 studies examining real rates and 89 studies examining nominal rates. The literature examining point-sampled bilateral RERs is presented in Table 2.4. We also discuss papers that have examined point-sampled bilateral NERs, for which summary tables are reported in appendix B.2.

In all cases, papers are found to construct forecasts using point-sampled data and compare them to point-sampled no-change forecasts. This suggests that conclusions derived from hypothesis testing in these papers are valid, and do not suffer from the concerns of spurious predictability discussed in the last sections. Again, the valid hypothesis testing for RERs is potentially quite informative on the predictability of period-average bilateral exchange rates and will be quantified in Section 2.6.

Finally, no paper has investigated real-time forecasts of point-sampled bilateral RERs. This disparity suggests a knowledge gap regarding real-time forecasts for bilateral RERs. In contrast, since Clarida and Taylor (1997), 16 out of 69 studies of point-sampled bilateral NERs have employed real-time forecasts.

Table 2.4: Papers Forecasting Point-Sampled Bilateral RERs

Paper	Level or Return	Frequency	Forecast Target	Benchmark	Model Estimation	Real or Nominal	Real-time
Boughton (1987)	Both	M	EoP*	EoP*	EoP*	Real	N
Meese and Rogoff (1988)	Level	M	EoP	EoP	EoP	Real	N
Throop (1993)	Return	Q	EoP*	EoP*	EoP*	Real	N
Jorion and Sweeny (1996)	Level	M	EoP	EoP	EoP	Real	N
Taylor et al. (2001)	Level	M	EoP	EoP	EoP	Real	N
Chen and Rogoff (2003)	Level	Q	EoP	EoP	EoP	Real	N
Froot and Ramadorai (2005)	Return	D	MoP	MoP	MoP	Real	N
Rapach and Wohar (2006)	Level	M	EoP*	EoP*	EoP*	Real	N
Engel and West (2006)	Level	M	EoP	EoP	EoP	Real	N
Clements and Fry (2008)	Return	Q	EoP	EoP	EoP	Real	N
Mumtaz et al. (2012)	Level	Q	EoP*	EoP*	EoP*	Real	N
Chen and Chen (2014)	Level	M	EoP	EoP	EoP	Real	N
Ca'Zorzi and Rubaszek (2020)	Return	M	EoP	EoP	EoP	Real	N
Liu and Shaliastovich (2022)	Return	M	EoP	EoP	EoP	Real	N

Note: "*" is used in cases where the paper did not provide information whether exchange rates are average or point sampled, and so point-in-time sampling was assumed. "EoP" and "MoP" refer to end-of-period and middle-of-period sampling, respectively. "Benchmark" refers to the no-change forecast that the forecast was compared against. "Model Estimation" refers to the data used in estimation.

2.3 Assessing Directional Accuracy

The literature on exchange rate forecasting often focusses on mean-square accuracy, often by computing ratios of root mean square forecast errors (RMSFEs) of a candidate model to a benchmark, or conducting Diebold-Mariano tests of the candidate against a benchmark (Diebold 2015). However, the literature also studies the directional accuracy of forecasting models, which also requires specifying a benchmark. In this section, we show that one should assess directional accuracy of model against the period-end no-change benchmark, rather than the period-average no-change benchmark as is common in the literature.

There are two main findings. Firstly, we show that the usual measure of directional accuracy, SRs, should be computed with the period-end no-change benchmark. An SR of 0.5 is interpreted as indicating that a model is no better than a coin flip at guessing the direction in which a series will move. We prove that if one uses a period-average no-change benchmark, it is easy to find models whose expected SR above 0.5 if the daily data is a random walk. In contrast, if we use an end-of-period no-change benchmark, then for any candidate model, the expected SR will be 0.5 when the daily data is a random walk. Secondly, we show that hypothesis tests of directional accuracy should also be done with the period-end no-change benchmark. Our simulations show that, if one tests directional accuracy using the period-average no-change benchmark, then the null will be rejected much more than the intended level of significance.

2.3.1 Theory

The SR is a common measure of directional accuracy, and is the proportion of times that a candidate forecast correctly predicts the direction of change relative to the time of the forecast. If the level follows a random walk (and is therefore unpredictable by definition), the SR of any candidate forecast is expected to be 0.5. In other words, no candidate can perform better than a coin flip. The candidates are said to have 'no directional accuracy'.

Assessing the directional accuracy of forecasts for a period-average requires special care. The usual approach is to compute the SR using changes relative to the latest period-average level. As we show, if the underlying high-frequency series is a random walk and hence unpredictable, it is trivial to find candidate forecasts whose SR is expected to be above 0.5. The apparent predictability of period-averages assessed by this measure is spurious, as it arises by construction. Instead, one should compute the SR using changes relative to the latest end-of-period level. This measure has the desirable property that, when the high-frequency series is a random walk, forecasts for the period-average have an expected SR of 0.5.

To make these claims precise, we introduce some notation. Our results apply to any period-average of a high-frequency series, but for concreteness, we will refer to the high-frequency series as being daily and the period-averages as month-averages. The daily level on day $t = 1, 2, \ldots, T$ is denoted D_t. The number of days in a month is denoted n, and is assumed equal across all months. The average level in month $m = 1, 2, \ldots, M$, where $T = nM$, is $A_m = \frac{1}{n}\sum_{i=1}^{n} D_{(m-1)n+i}$. The end-of-month level in month m is denoted Z_m and equals D_{mn}.

Consider a forecaster in month m making a forecast for the level of the average h-months ahead, A_{m+h}. The directional accuracy of a 'candidate' forecast is defined as:

$$DA_{m,h} \equiv \mathbb{1}\left\{sgn(A_{m+h} - \hat{A}^{bench}_{m+h|m}) = sgn(\hat{A}^{candidate}_{m+h|m} - \hat{A}^{bench}_{m+h|m})\right\}, \tag{2.1}$$

where $\hat{A}^{candidate}_{m+h|m}$ is the candidate forecast, $\hat{A}^{bench}_{m+h|m}$ is the benchmark forecast, and $\mathbb{1}[\cdot]$ is an indicator function with 1 if true and 0 otherwise. The sign function is:

$$sgn(x) \equiv \begin{cases} 1 & x > 0 \\ -1 & x \leq 0. \end{cases}$$

The benchmark forecast is always a no-change forecast, so directional accuracy can be interpreted as whether the candidate forecast correctly guessed the direction in which the series moved. When the benchmark is the month-average no-change, directional accuracy of the candidate forecast equals 1 if the change in the month-average series, $(A_{m+h} - A_m)$ is in the same direction expected by the candidate, $\left(\hat{A}^{candidate}_{m+h|m} - A_m\right)$.

$$DA_{m,h} \equiv \mathbb{1}\left\{sgn(A_{m+h} - A_m) = sgn\left(\hat{A}^{candidate}_{m+h|m} - A_m\right)\right\}, \tag{2.2}$$

An alternative benchmark is the end-of-month no-change, which is the last disaggregated observation in the forecasters information set. In this case, the change in the series is measured from the last day of the forecast month m to the month-average of the future month $(m + h)$.

$$DA_{m,h} \equiv \mathbb{1}\left\{sgn(A_{m+h} - Z_m) = sgn\left(\hat{A}^{candidate}_{m+h|m} - Z_m\right)\right\}, \tag{2.3}$$

The SR is the sample mean of the directional accuracy of a sample of forecasts (Pesaran and Timmermann 1992). Given a full sample of months $m = 1, ..., M$, one selects a subsample of months $m_{min}, (m_{min} + 1), ..., m_{max}$, where $m_{min} \geq 1$ and $m_{max} \leq (M - h)$. For each month in the subsample, a candidate model is used to generate an h-step-ahead forecast using the data available up to the end of that month. The success ratio of this sample of h-step-ahead forecasts is defined as:

$$SR_h \equiv \frac{1}{(m_{max} - m_{min} + 1)} \sum_{m=m_{min}}^{m_{max}} DA_{m,h}. \tag{2.4}$$

Suppose the data generating process of the daily frequency data is a random walk:

$$D_t = D_{t-1} + e_t \quad \forall\ t \tag{2.5}$$

where the initial level D_0 is a constant, each error e_t is independent of the error in other periods and the past levels of all other variables, and the distribution of each error e_t is continuous and symmetric about 0.

We now introduce our main theorem, which shows how the SRs compare to a half in the case when the high-frequency series is a random walk.

Theorem 4. *Suppose the data generating process of the daily frequency data is a random walk, as given by equation (2.5).*

(a) Suppose the candidate is end-of-month no-change (EOM) and the benchmark is month-average no-change (Avg). Then the expected SR is strictly greater than half.

$$E\left[SR_h^{EOM\ vs\ Avg}\right] = E\left[DA_{m,h}\right] > \frac{1}{2}$$

(b) *Suppose the candidate is any function of the levels of variables up to the time the forecast is made, and the benchmark is end-of-month no-change. Then the expected SR is a half.*

$$E\left[SR_h^{Candidate\ vs\ EOM}\right] = E\left[DA_{m,h}\right] = \frac{1}{2}$$

Proof. See Appendix B.1 □

Theorem 4(a) shows that if the SR is computed relative to the latest period-average level, it is trivial to find candidates whose expected SR is above 0.5, with the end-of-period no-change being one such example. This is undesirable, as it implies that the SR will often suggest that a candidate model can guess the direction of changes, but this arises by construction, and hence does not indicate that the model is useful. Theorem 4(b) shows that an SR computed relative to the end-of-period level will equal 0.5 in expectation for any candidate model, which is appropriate given the underlying series is unpredictable.

2.3.2 Simulation Evidence

Simulation experiments now quantify the assessment of directional accuracy for forecasts of temporally aggregated data. Let the daily data, D_t, be represented by an autoregressive model with one lag, AR(1), $D_t = \rho D_{t-1} + e_t$, such that $\rho = 1$ is the random walk model, equation 2.5, and $e_t \sim N(0, 1)$.[4] The simulated data is aggregated, A_m to weekly, monthly, or quarterly frequency, with $n = 5$, 21, or 62, respectively.[5] The baseline simulations burn the first 500 daily observations and use 40 years worth of daily data, consistent with applications where daily data has been available since the early-1980s.

[4] Success ratios and tests of Pesaran and Timmermann 2009 do not depend on the initial level or variance of the daily series.

[5] Temporally aggregated economic data are typically built using daily data such as simple averages of closing values on business days (e.g. exchange rates, commodity prices, and interest rates), which is why we assume there are $n = 5$ days in a week, $n = 21$ in a month and 62 in a quarter.

2.3.2.1 Comparison of No-Change Forecasts

Out-of-sample period-average no-change forecasts, A_m, are compared to the end-of-period no-change forecasts, Z_m, and vice versa, see Table 2.5. The simulations help us quantify theorem 4. First, the expected SR of any model against the end-of-month benchmark is 0.5, as shown by theorem 4(b). Table 1, columns 2 to 4 verify that this is the case for the alternative no-change forecasts of a random walk. When the daily data follows a random walk, $\rho = 1$, only when the directional accuracy of forecast of the aggregate is compared to the end-of-period no-change does the success ratio correctly reflect random chance of directional accuracy at all horizons (i.e. 50-50 or 50%).

Table 2.5: Directional Accuracy is Expected when Compared to the Period Average No-Change

	Versus End of Period			*Versus Period Average*		
Horizon	Weekly	Monthly	Quarterly	Weekly	Monthly	Quarterly
1	0.50	0.50	0.50	0.70	0.74	0.75
	(0.016)	(0.023)	(0.040)	(0.015)	(0.020)	(0.034)
3	0.50	0.50	0.50	0.60	0.61	0.61
	(0.016)	(0.023)	(0.040)	(0.014)	(0.020)	(0.034)
6	0.50	0.50	0.50	0.57	0.58	0.58
	(0.016)	(0.023)	(0.040)	(0.014)	(0.020)	(0.020)
12	0.50	0.50	0.50	0.55	0.55	0.55
	(0.016)	(0.023)	(0.041)	(0.015)	(0.020)	(0.036)

Notes: Success ratios from forecasts of end-of-period and period average no-change forecasts. 5000 simulations, using 40 years of data. Standard deviation of the success ratios in brackets.

In contrast, the expected SR of an end-of-month forecast against the month-average forecast is greater than 0.5, as shown by theorem 4(a). However, the theorem cannot say how much above 0.5. Table 1, columns 5 to 7, quantifies that the success ratios for the end-of-period no-change forecast relative to the period-average no-change forecast in the case where the random walk error, e_t, is normally distributed. Mean directional accuracy is 70, 74, and 75 percent at the one-step-ahead for weekly, monthly, and quarterly forecasts, respectively. Thus, sizable gains in directional accuracy are spuriously expected when forecasts are compared to the period-average no-change, even though the daily data is inherently unpredictable.

2.3.2.2 Bottom-up Forecasts

Suppose one wanted to test the null hypothesis that the expected success ratio is 0.5. That is, the candidate forecast is no better than a coin flip at predicting the direction in which the actual outcome will differ from the benchmark. To achieve this, we test that the categorical random variables $sgn(A_{m+h} - \hat{A}^{bench}_{m+h|m})$ and $sgn(\hat{A}^{candidate}_{m+h|m} - \hat{A}^{bench}_{m+h|m})$ are independent of each other using the Pesaran and Timmermann 2009 test.[6] This test is ideal for our situation, as it is valid even if each of the categorical random variables is serially correlated.

To see how the benchmark itself affects power, we need to keep the candidate forecast the same and vary only the benchmark. A bottom up forecast is a good candidate forecast of the period average forecast of the AR(1) model. Specifically, an AR(1) model is estimated with ordinary least squares on the daily data. The estimate of ρ is used it to construct dynamic forecasts of the daily series, and then these forecasts are ex-post averaged to the desired lower frequency. We reserve the second half of the sample to evaluate out-of-sample forecasts, and reestimate the model with an expanding window in every period.

The simulations in table 2.6 provide evidence on the power of hypothesis tests of directional accuracy using 40, 20 and 10 years of data. When the daily data is unpredictable, $\rho = 1$, testing relative to the end-of-period no-change closely reflects the 5 percent significance level. This is desirable as when ρ is 1, then the daily series is a random walk and by definition, no candidate model should exhibit improvements in directional accuracy.

Another way to implement the test of Pesaran and Timmermann 2009 is to compare relative to the period-average no-change benchmark. When forecasting a month-average at a 1-month horizon, the null is rejected 100 percent of the time when $\rho = 1$. This may be a valid test of whether $sgn(A_{m+h} - \hat{A}^{bench}_{m+h|m})$ is independent of $sgn(\hat{A}^{candidate}_{m+h|m} - \hat{A}^{bench}_{m+h|m})$. However, it is an invalid test of our underlying question,

[6] Newey and West 1987 standard errors use $4(N/100)^{2/9}$ lags, where N is the observations in the forecast evaluation sample, and are evaluated using the standard normal.

Table 2.6: Comparison Relative to the Period Average Results in Type I Error and Power Loss

			Versus End of Period			Versus Period Average		
Years	*DGP*	*Horizon*	Weekly	Monthly	Quarterly	Weekly	Monthly	Quarterly
40	ρ=1	1	0.05	0.05	0.06	1.00	1.00	1.00
		3	0.05	0.05	0.06	1.00	0.98	0.71
		6	0.05	0.05	0.07	1.00	0.80	0.40
		12	0.05	0.05	0.07	0.94	0.48	0.19
	ρ=0.9	1	1.00	1.00	1.00	1.00	1.00	1.00
		3	1.00	1.00	1.00	1.00	1.00	1.00
		6	1.00	1.00	1.00	1.00	1.00	1.00
		12	1.00	1.00	1.00	1.00	1.00	1.00
20	ρ=1	1	0.06	0.06	0.07	1.00	1.00	0.96
		3	0.05	0.06	0.08	1.00	0.84	0.44
		6	0.05	0.06	0.08	0.95	0.50	0.22
		12	0.06	0.06	0.07	0.69	0.27	0.10
	ρ=0.9	1	1.00	1.00	1.00	1.00	1.00	0.98
		3	1.00	1.00	1.00	1.00	1.00	0.98
		6	1.00	1.00	0.99	1.00	1.00	0.96
		12	1.00	1.00	0.99	1.00	1.00	0.91
10	ρ=1	1	0.05	0.07	0.09	1.00	0.99	0.78
		3	0.06	0.07	0.09	0.96	0.57	0.28
		6	0.06	0.07	0.08	0.72	0.31	0.13
		12	0.06	0.06	0.05	0.41	0.13	0.05
	ρ=0.9	1	1.00	1.00	0.95	1.00	1.00	0.86
		3	1.00	1.00	0.94	1.00	1.00	0.83
		6	1.00	1.00	0.88	1.00	0.99	0.74
		12	1.00	1.00	0.70	1.00	0.99	0.54

Notes: Forecast performance of bottom-up forecasts. The daily random walk is given by $\rho = 1$; whereas $\rho = 0.9$ implies a predictable data generating process. 5000 simulations, using 40, 20, and 10 years worth of data. Tested at the 5 percent level.

which is whether there are candidates with directional accuracy, since it rejects more than 5 percent of the time when the daily series is a random walk.

Now consider when the daily data is a predictable stationary AR(1) with $\rho = 0.9$. When comparing relative to the end of period no-change forecast, the test exhibits substantial power at all horizons, even with only tens of years of data (i.e. 5 years of forecasts). In contrast, when comparing relative to the period average no-change, the last three columns, the null is often rejected, but not quite as often. The testing of directional accuracy relative to the monthly average no-change exhibits loss of predictive power relative to comparisons against the end-of-period no-change.

Together, the findings show that care needs to be taken to assess directional accuracy of temporal aggregates. Assessing directional accuracy of the temporal aggregate relative to the period average no-change benchmark results in inflated type I error under the null and loss of power under the alternative. Instead, one should assess directional accuracy relative to the end-of-period no-change benchmark.

2.4 Data

We construct a comprehensive dataset of four types of exchange rate: bilateral NERs, bilateral RERs, NEERs and REERs.

For each type of exchange rate and country, we construct a sequence of real-time vintages of daily exchange rates, as explained in section 2.4.1. There is one vintage per month, which is intended to reflect all information that would have been available to a forecaster at the end of the month. Our decision to construct *real-time* vintages is not motivated by revisions, since NERs, CPIs and trade weights are rarely revised. Rather, our aim is to account for the typical delays in the publication of data the CPI and trade weights data.

For each monthly vintage, we also construct a vintage of month-average exchange rates and a vintage of end-of-month exchange rates. We do this using the corresponding vintage of daily exchange rates, as explained in 2.4.2.

Our dataset fills gaps left by official data sources. Firstly, while daily bilateral exchange rates are widely available, and the Bank for International Settlements publishes daily NEERs, we are the first to construct daily REERs. Secondly, while some authors have constructed real-time datasets of bilateral exchange rates, we are the first to construct a real-time dataset of EERs that accounts for the typical delay in weights becoming available. Of course, such a dataset is necessary but not sufficient for the literature to adopt best practices. Daily bilateral exchange rates have been available for some time, but many papers nevertheless evaluate forecasts for period-average bilateral rates by comparison to the no change of period-average benchmark.

2.4.1 Monthly Vintages of Daily Frequency Exchange Rates

Sections 2.4.1.1 describes the calculation of bilateral NERs and bilateral RERs respectively. Section 2.4.1.2 describes the calculation of NEERs and REERs. Detail on the inputs into these calculations (bilateral NERs, CPIs and trade weights) is provided in Appendix B.3.

2.4.1.1 Bilateral Exchange Rates

Constructing monthly vintages of bilateral NERs is straightforward. Bilateral NERs are available daily and observed in real time. As such, a bilateral NER vintage for a month is simply the daily NER on each day until the end of that month. For example, the March 2023 vintage of Canada's bilateral NER is simply its daily bilateral NER on each day up to 31 March 2023.

To construct monthly vintages of daily bilateral NERs we need data on both daily bilateral RERs and monthly CPI. To describe the calculations precisely, we introduce some notation. Let NER_t^i denote the bilateral NER of country i on day t. This is the value of the currency in terms of US dollars. Let CPI_m^i denote the CPI level in country i in month m. Finally, let RER_t^i denote the bilateral RER of country i on day t. This is the cost of goods and services in country i relative to the cost of goods and services in the United States.

The daily bilateral RER on day t of month m is the daily bilateral NER of that country multiplied by the ratio of country i's CPI level to the US CPI level.

$$RER_t^i \equiv NER_t^i \times \frac{CPI_m^i}{CPI_m^{\text{US}}} \tag{2.6}$$

An alternative approach would have been to combine the daily nominal price with daily CPI levels, where the daily CPI levels have been estimated by interpolating monthly levels. We chose the current approach because it is more transparent than the alternative, since it avoids needing to take a stand on how to perform the interpolation.

A complication is that CPI data is published with a lag that differs by country. For example, as at the end of March 2023, the latest CPI data for the United States or Canada is for February 2023, which is a one month lag. In contrast, some low or middle income countries may only publish their CPI two or three months later. When constructing a monthly vintage, we only use the monthly CPI data likely to have been known at the time. To overcome this, we nowcast the missing monthly CPI levels by assuming that CPI inflation remains constant at the latest rate known at the time.

To illustrate, suppose we are computing the March 2023 vintage of Canada's daily bilateral RER. As at end March 2023, a forecaster would have had access to a daily NER up to 31 March 2023, and the monthly CPI of the US and Canada up to February 2023. With this data:

- To compute the daily RER for any day t up to 28 February 2023, we apply equation (2.6) to the day t daily NER, the month m Canadian monthly CPI, and the month m US monthly CPI, since all of that data would have been known to a forecaster as at end March 2023.

- To compute the daily RER for any day t from 1 March 2023 to 31 March 2023, we would apply equation (2.6) to the day t daily NER (which would have been known at the time), a nowcast of Canada's March 2023 monthly CPI, and a nowcast of the US's March 2023 monthly CPI.

2.4.1.2 Effective Exchange Rates

We also construct monthly vintages of daily EERs. This is more complex, both because a number of EER formulas are available, and because we only want each vintage to be constructed only using CPI and trade weights data available at the time.

We compute daily EERs by adapting the formulas used for monthly EERs by the IMF. We use the IMF's method because we want our method to be consistent with our choice of weights, and we use the IMF weights because they are the most comprehensive in terms of countries and time periods. Other institutions use different formulas for computing EERs.[7]

[7] For REERs, the formulas differ in how they combine NERs and prices.

- The Bank for International Settlements' approach is aggregate the bilateral nominal exchange rates to obtain an NEER, separately aggregate the price levels, and then compute the REER by adjusting the NEER by aggregate price levels (Klau and Fung 2006, Turner and Van 't dack 1993).

- The IMF's previous approach was to directly aggregate the bilateral RERs (Bayoumi et al. 2006).

- The IMF's current approach is to compute the REER as a ratio of products of bilateral NERs and CPIs.

For both NEERs and REERs, the formulas differ in how they aggregate across countries, which affects the properties of the series. For example, Vartia and Vartia 1984 show that the NEER used by the Bank of Finland at the time had an upward bias, unlike alternative index number formulas such as a Fisher index or Tornqvist index.

To describe our method, we must define some terms. We use the term 'reporter' to refer to the country whose effective exchange rate we are computing, and we use the term 'partner' to refer to any other country included in the calculation. The 'weight reference period' is the period of a few years with which a set of weights is associated. For example, there is a set of weights based on the trade flows during the '2010 to 2012' weight reference period (see B.3.3 for details). Let $w^b_{r,j}$ denote the weight that reporter r puts on partner j in weight reference period b.

If we only have data for a single weight reference period, then we can compute the daily EER using a 'fixed weight' formula. Equation (2.7) is used to compute the daily REER of a reporter r on a day t in month m and weight reference period b. The numerator is the reporter's NER in US dollars multiplied by the reporter's price level. To compute the denominator, we multiply each partner's NER in US dollars with that partner's price level, and then aggregate across partners. To compute the NEER, simply set the CPI terms equal to 1.

$$REER^{r,b}_t = \frac{NER^r_t \times CPI^r_m}{exp\left(\sum_{j=1}^{J} w^{r,j}_b ln\left(NER^j_t \times CPI^j_m\right)\right)} \tag{2.7}$$

For each weight reference period, we only use partners whose exchange rates are available on all days in the period.[8] Additionally, if over half of partners by weight have missing exchange rates for a weight reference period, then we don't compute the REER for that period.

Typically, we want to compute the EER over a longer time period that spans multiple weight reference periods. In this case, we compute an EER by 'chaining' the fixed-weight indexes. The chained EER is set equal to 1 on the first day of our sample. For each subsequent day, the growth in the chained EER is set equal to the growth in the relevant fixed-weight EER. Formally:

$$\frac{EER^r_t}{EER^r_{t-1}} = \frac{EER^{r,b}_t}{EER^{r,b}_{t-1}}$$

[8] An exception is that, when computing EERs over the 1990-1995 weight reference period, we compute EERs from 1990 to 1992 using partners whose exchange rates are available from 1990 to 1992, and then compute REERs from 1993 to 1995 using partners whose exchange rates are available from 1993 to 1995. This materially increases the number of partners included in the 1993 to 1995 calculations, because the number of countries with NER data increases materially from the start of the IMF NERs on 1 January 1993.

where b is the weight reference period that contains day t.

This formula ensures that the numerator and denominator both use the same set of weights. For example, if we compute growth in the United Kingdom's chained-EER on 1 Jan 1996, which is the first day of the '1996 to 2003' weight reference period, we would compute:

$$\frac{EER^{\text{UK}}_{\text{1 Jan 1996}}}{EER^{\text{UK}}_{\text{31 Dec 1995}}} = \frac{EER^{\text{UK,1996 to 2003}}_{\text{1 Jan 1996}}}{EER^{\text{UK,1996 to 2003}}_{\text{31 Dec 1995}}}$$

To compute the chained REER on 31 Dec 1995, which the last day of the '1990 to 1995' weight reference period, we would compute:

$$\frac{EER^{\text{UK}}_{\text{31 Dec 1995}}}{EER^{\text{UK}}_{\text{30 Dec 1995}}} = \frac{EER^{\text{UK,1990 to 1995}}_{\text{31 Dec 1995}}}{EER^{\text{UK,1990 to 1995}}_{\text{30 Dec 1995}}}$$

For each monthly vintage, we compute daily EERs by applying the above formulas to the data that a forecaster would have had access to as at the end of each month.

- Daily NERs are published without delay
- Monthly CPI is published with a lag of 1 or more months, which we overcome by nowcasting the missing observations as explained in section 2.4.1.1.
- Trade weights are published with very long lags. Trade weights are assumed to be unavailable for 5 years after the end of the period to which the weights relate. For example, the trade weights based on 2013-2015 trade flows are assumed to become available from the January 2021 vintage onwards. We assume a 5-year lag to emulate current practice at the IMF, since they are our source of trade weights.

2.4.2 Monthly Vintages of Monthly Exchange Rates

We derive month-average and end-of-month series from the daily series. For each vintage of daily exchange rates (of any type), we make a corresponding vintage of month-average exchange rates (by averaging the daily rates over each month) and a vintage of end-of-month exchange rates (by extracting the last daily rate of each month).

For bilateral RERs, an alternative would be to apply the bilateral RER formula to a month-average NER and a monthly CPI. However, this is exactly equivalent to our approach of computing an average of daily bilateral RERs.

$$RER^i_m \equiv \frac{1}{n}\sum_{t=1}^{n} RER^i_t = \frac{1}{n}\sum_{t=1}^{n} NER^i_t + \frac{CPI^i_m}{CPI^{\mathrm{US}}_m}$$

Similarly, one could instead compute NEERs by applying the EER formula to month-average NERs, or compute REERs by applying the formula to month-average NERs and monthly CPI. This alternative approach gives similar growth rates in practice, though differ greatly during periods of hyperinflation.[9]

[9] It is not obvious whether policymakers faced with hyperinflation should aim to forecast a month-average of daily EERs (as we do) or forecast the result of applying the EER formula to monthly data. This depends on how different measures of EERs relate to the underlying economic concepts that the policymaker is interested in, which is out of scope for this paper.

2.5 Method

2.5.1 Out-of-Sample Evaluation

We conduct an out-of-sample evaluation of forecasts of the month-average exchange rate. The results that follow pertain to the same set of 83 countries for consistent comparison. Although exchange rates are observed for all countries, this sample represents countries for which all types of exchange rates (bilateral NER, bilateral RER, NEER, REER) start no later than 1 January 1994. Using a common sample period and set of countries facilitates comparisons between the results for different types of exchange rates.

For each of these countries, we produce real-time forecasts using each monthly vintage. We make a forecast using each monthly vintage from January 2004 to September 2022. By making January 2004 the earliest vintage, we ensure that all of these forecasts are made with models estimated on at least 10 years of data. When computing forecast errors for bilateral RERs, we target the actual outcome computed from the bilateral NERs and CPI data as at end June 2023. For EERs, we target the actual outcome computed with the bilateral NERs and CPI levels as at end June 2023, but with the weights known on the forecast date. In this case, the forecaster needs to predict the combined effect of changes in bilateral NERs and CPI levels, but not the weights. We choose this approach since it best reflects the aims of policymakers, who typically do not try to predict the effect of future changes in weights, in part because new weights will typically not be released until after the end of the forecast horizon. This approach also ensures that later forecast vintages are treated in the same way as earlier forecast vintages, since the forecasts will need to anticipate weight changes.

The sample of 83 countries includes those with various exchange arrangements (floating, fixed, other managed arrangements), including countries whose exchange arrangements changed part-way through the sample period (e.g. Lithuania, whose currency was pegged to the USD, then pegged to the euro, and then replaced by the euro). We include all countries in the forecast exercise and treat them equally. In doing so, we do not attempt to account for structural breaks such as changes in exchange rate regimes. We do this because we aim to quantify the effects of temporal aggregation generally, rather than to take a stand on the best forecast practices for any specific country.

We employ two common real-time forecast evaluation criteria.

The first forecast evaluation criteria is the ratio of the RMSFE of a candidate model relative to the RMSFE of the benchmark. Specifically, the RMSFE ratio at horizon h, $RMSFE_h^{ratio}$, is computed as the quotient of the RMSFE of the model-based forecast and the RMSFE of the alternative forecast:

$$RMSFE_h^{ratio} = \frac{\sqrt{\sum_{m=1}^{M}\left(A_{m+h} - \hat{A}_{m+h|m}^{candidate}\right)^2}}{\sqrt{\sum_{m=1}^{M}\left(A_{m+h} - \hat{A}_{m+h|m}^{bench}\right)^2}} \tag{2.8}$$

where $\hat{A}_{m+h|m}^{candidate}$ represents the real-time candidate forecast for the h step ahead of forecast target A_{m+h}, and $\hat{A}_{m+h|m}^{bench}$ is the alternative benchmark forecast, for all periods of the evaluation sample, denoted as $m = 1, ..., M$. We also perform Diebold-Mariano tests of the null that expected squared error loss is equal. To perform the test for a horizon h, we compute a loss differential for forecasts at that horizon (i.e. difference in squared errors). We then regress the loss differentials on an intercept, and perform a two-sided test of the null that the intercept is zero using standard normal critical values. We compute the statistic using heteroscedasticity and autocorrelation (HAC) standard errors to allow for the possibility that the loss differentials are autocorrelated.[10]

The second forecast criteria is the SR, which measures directional accuracy. We also test the null of no directional accuracy by testing if the categorical random variables $sgn(A_{m+h} - \hat{A}_{m+h|m}^{bench})$ and $sgn(\hat{A}_{m+h|m}^{candidate} - \hat{A}_{m+h|m}^{bench})$ are independent of each other, as described in section 2.3.1.

2.5.2 Description of Forecasting Methods

This subsection describes the forecasting methods. These methods are in three broad categories: no change forecasts; recursive forecasts; and direct forecasts. We

[10]We use the HAC estimator in the 'sandwich' package in R (Zeileis 2004) with the default options including not performing prewhitening.

consider both recursive and direct forecasts for generality as both have advantages and disadvantages, so it is not obvious a priori which will perform better (see section 2.7.7 of Petropoulos, Apiletti, Assimakopoulos, Babai, Barrow, Ben Taieb, et al. 2022b).

While our aim is to forecast the level of the exchange rates, we estimate the models using log levels. We take the natural log of the exchange rate, make a no-change, autoregressive or direct forecast for the log of the period-average exchange rate, and take the exponent to give a forecast for the level of the period-average exchange rate. We do this because log variables are more likely to be closer to satisfying the assumptions of typical forecasting models. For example, if we think it is equally likely that an exchange rate could appreciate by 1% or depreciate by 1%, then we should model the *log* exchange rate using a model with symmetric errors, rather than modelling the exchange rate itself as having symmetric errors.

As in section 2.3.1, we denote daily, month-average and end-of-month exchange rates by D_t, A_m and Z_m respectively. We now denote log levels by lower case letters: d_t, a_m and z_m. We continue to assume that there are n days in each month to simplify our notation. We let M denote the current month, which means the forecaster has access to data from months $m = 1, ..., M$ when making a forecast for a future month $(M + h)$.

2.5.2.1 No Change Forecasts

We consider two types of no-change forecasts:

i) Month-average no-change. Our forecast for the month-average in any future month $(M + h)$ is the last observed month-average:

$$\widehat{a}_{M+h|M} = a_M \quad \forall h$$

ii) End-of-month no-change. Our forecast for the month-average in any future month $(M + h)$ is the current end-of-month level. Equivalently, we assume that the end-of-month level is a random walk, and use the end-of-month value for the forecast of all future monthly average values. Hence, this method is an example of Period End Price Sampling (PEPS) (as in Ellwanger and Snudden 2023a).

$$\widehat{a}_{M+h|m} = z_M \quad \forall h$$

Note that the 'daily no-change' forecast, where our forecast for the month-average level in any future month $M + h$ would be the latest daily level, d_{Mn}, is exactly equivalent to the end-of-month no-change forecast. This is because, in our forecast evaluation, the forecasts are always constructed on the last day of each month.

2.5.2.2 Recursive AR(1) Forecasts

We make recursive forecasts using autoregressive models of order 1 (AR(1)) estimated on exchange rate levels using OLS.[11] We consider three types:

i) Recursive of month-average inputs. We estimate an AR(1) model of month-average exchange rates.

$$a_{m+1} = \alpha + \beta a_m + e_m \quad \forall m = 1, ..., M - 1 \tag{2.9}$$

We then use this model to make recursive forecasts for all future horizons.

ii) Recursive PEPS. We estimate an AR(1) model of end-of-month exchange rates.

$$z_{m+1} = \alpha + \beta z_m + e_m \quad \forall m = 1, ..., M - 1 \tag{2.10}$$

We then use this model to make recursive forecasts for the end-of-month exchange rate for all future horizons consistent with Ellwanger and Snudden 2021. We then treat these forecasts for end-of-month rates as if they were forecasts for month-average rates. This is another example of a PEPS forecast.

iii) Recursive Bottom-up. We estimate an AR(1) on daily exchange rates.

$$d_{t+1} = \alpha + \beta d_t + e_t \quad \forall t = 1, ..., Mn - 1 \tag{2.11}$$

We use this model to make recursive forecasts for the daily exchange rate for all future days. We then average those daily forecasts to obtain month-average forecasts, which makes this an example of a bottom-up forecast implemented in real-time for averaged real data (Benmoussa et al. 2020).

$$\widehat{a}_{M+h|M} = \frac{1}{n} \sum_{t=1}^{n} \widehat{d}_{(M+h-1)n+t} \quad \forall h$$

[11] For a handful of countries, the estimated AR(1) model had a coefficient that was outside $(-1, 1)$, suggesting that exchange rates were non-stationary. Where this occurred, the country was excluded from our results.

As a robustness check, we also estimated recursive AR(1) models on returns, and then used the resulting forecasts for returns to compute forecasts for levels. We found that this had little effect on performance for bilateral NERs or NEERs, but worsened performance for bilateral RERs and REERs. Since estimating on returns results in similar or worse performance, we have shown recursive models estimated on levels in this chapter. As another check, we also produced recursive forecasts using the 'automatic ARIMA' procedure of Hyndman, Athanasopoulos, et al. 2022, which chooses whether to difference the series based on a stationarity tests and then selects the number of autoregressive and moving average terms using information criteria. This procedure gave qualitatively similar results to estimating AR(1) models on levels.

2.5.2.3 Direct Forecasts

We construct direct forecasts using linear regressions estimated on exchange rate levels using OLS. We consider three approaches:

i) Direct of month-average inputs. For each horizon in months, h, we estimate a regression of the month-average exchange rate in $(m + h)$ on the month-average exchange rate in month m.

$$a_{m+h} = \alpha_h + \beta_h a_m \quad \forall m = 1, ..., M - h \tag{2.12}$$

We then use the estimated model to directly forecast the month-average exchange rate in h months.

$$\widehat{a}_{M+h} = \widehat{\alpha}_h + \widehat{\beta}_h a_M$$

ii) Direct PEPS. For each horizon, we estimate a regression of the end-of-month exchange rate in $(m + h)$ on the end-of-month exchange rate in m.

$$z_{m+h} = \alpha_h + \beta_h z_m \quad \forall m = 1, ..., M - h \tag{2.13}$$

We then use this estimated model to produce an h-month-ahead forecast of the end-of-month exchange rate:

$$\widehat{z}_{M+h} = \widehat{\alpha}_h + \widehat{\beta}_h z_M$$

Finally, we treat the forecast of the end-of-month exchange rate as if it were a forecast for the month-average exchange rate, $\hat{a}_{M+h} = \hat{z}_{M+h}$. This is another example of PEPS (Benmoussa et al. 2020; Ellwanger and Snudden 2021).

iii) Direct UMIDAS. For each horizon h we estimate a regression of the month-average exchange rate in $(m + h)$ on the latest daily observation known on the forecast date. We estimate the parameters of the model with OLS without any restrictions. Since the latest daily observation available on the forecast date is the current end-of-month exchange rate, z_m, this model can be written:

$$a_{m+h} = \alpha_h + \beta_h z_m \quad \forall m = 1, ..., M - h \tag{2.14}$$

We then use the estimated model to forecast the month-average exchange rate directly. This is an example of an unrestricted Mixed Data Sampling (UMIDAS) model, as described in Foroni, Marcellino, and Schumacher 2015.[12]

$$\widehat{a}_{M+h} = \widehat{\alpha}_h + \widehat{\beta}_h z_M$$

[12]As a robustness check, we also tested a UMIDAS model where the period-average exchange rate was regressed on the seven latest daily observations, and the results were qualitatively similar.

2.6 Results

This section reports the quantitative results regarding the importance of temporal aggregation bias for exchange rates. All forecasts are constructed in real-time as described in section 2.5 using the data documented in section 2.4.

2.6.1 Comparison of No-Change Benchmarks

We begin by examining the extent to which the end-of-period no-change forecast outperforms the monthly average no-change forecast. As a point of comparison, we also report the results for a simulated random walk at the daily frequency aggregated to monthly data with $n = 21$. We simulate 30 years worth of data in addition to burning the first 500 daily observations. We then apply our out-of-sample evaluation methodology to the simulated data, and iterate 5000 times. Table 2.9 reports the median RMSFE ratios at various forecast horizons for the end-of-month no-change forecast relative to the month-average no-change forecast for the simulated random walk and the four exchange rates.

When the data follows a random walk, the end-of-month no-change forecast substantially outperforms the monthly average no-change forecast. The gains in the RMSFE are largest at the one-month ahead, showing a 17 percent reduction. Moreover, consistent with theory, the differences decline with the forecast horizon (Ellwanger and Snudden 2023a). These patterns are also present for directional accuracy. For 1-month-ahead forecasts, the median SR is 0.74, which means that the end-of-month no-change predicted the direction in which the month-average exchange rate moved 74% of the time. The SRs also decline at longer horizons, but remain above 0.5 up to 12 months ahead.

The pattern of the forecast gains observed for the simulated random walk are also observed for all the exchange rates. In particular, the median RMSFE ratios for NER are nearly identical to those obtained from a random walk in practice. The results suggest that the NER exhibits properties most similar to a random walk followed by the RER, NEER, and REER. That said, even for EERs, the end-of-month no-change forecast out-performs the monthly average no-change forecast at

Table 2.7: Median Performance of End-of-Month No-change Forecasts Versus Monthly Average No-Change Forecasts

Months Ahead	1	3	6	12	24	36
Measure			RMSFE Ratio			
Random-Walk	0.73	0.94	0.97	0.99	0.99	1.00
NER	0.76	0.93	0.97	1.00	1.00	1.00
NEER	0.93	0.97	0.98	0.99	1.00	1.00
RER	0.87	0.96	0.98	1.00	1.00	1.00
REER	0.97	0.99	0.99	1.00	1.00	1.00
			Success Ratio			
Random-Walk	0.74	0.61	0.58	0.55	0.54	0.53
NER	0.71	0.60	0.59	0.53	0.52	0.49
NEER	0.71	0.63	0.59	0.55	0.55	0.52
RER	0.68	0.59	0.56	0.52	0.51	0.49
REER	0.69	0.59	0.56	0.52	0.52	0.50

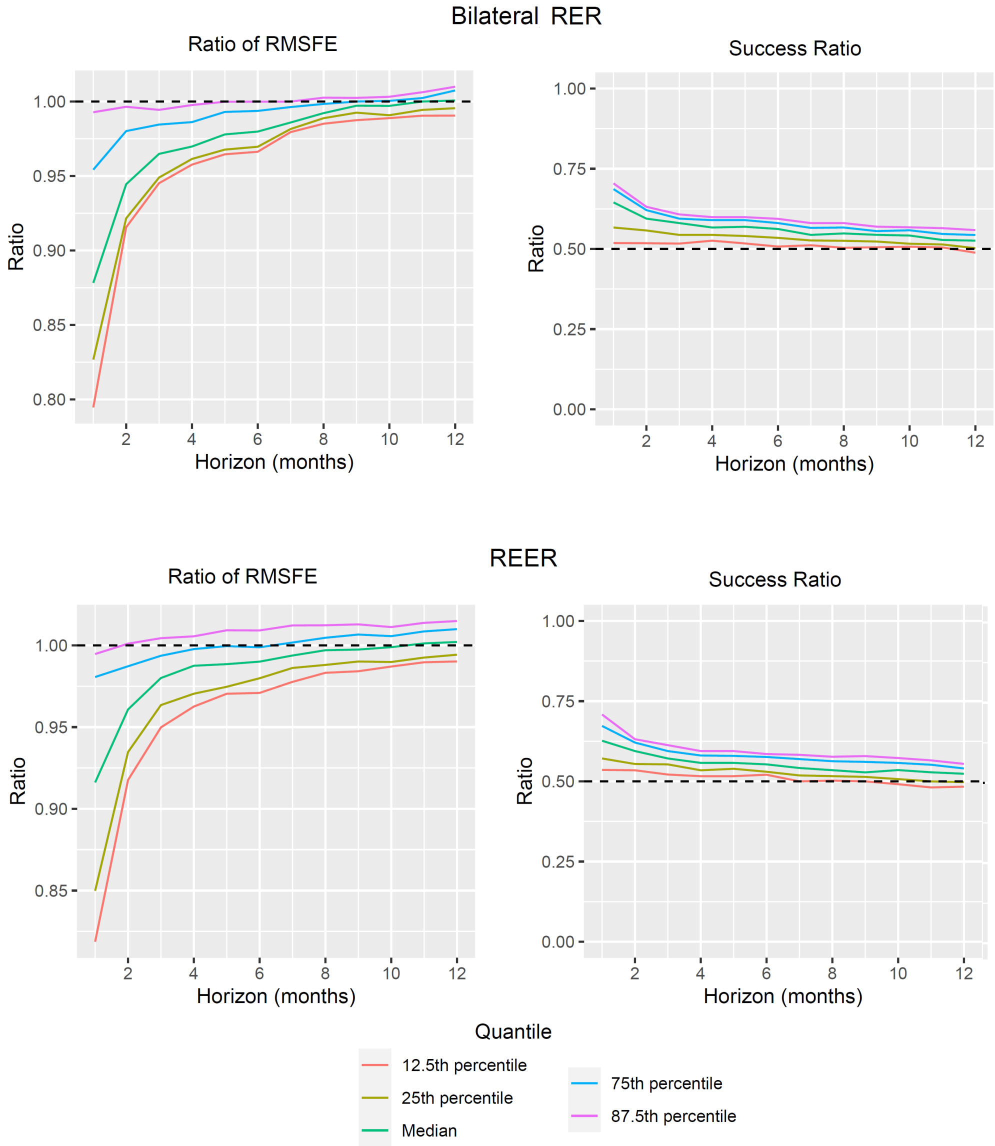
Note: Forecast accuracy of end-of-month no-change forecast versus monthly average no-change forecast. Reports the median across countries. “end-of-month" inputs in model estimation follow PEPS and use the point forecast as the forecast of the average. Recursive daily forecasts constructed using the bottom-up approach. Direct forecasts use UMIDAS restricted to the end-of-month observation. “Random Walk" is simulated using 5000 iterations and 30 years of data.

the one-month ahead by 7 and 3 percent for the NEER and REER, respectively. Moreover, the end-of-month no-change does at least as well as the monthly average up to 12 months ahead for all four exchange rate measures.

Regarding directional accuracy, the gains in the SR are also very substantial but much more consistent across exchange rate measures. For all exchange rates, at one-month ahead, gains of around 40 percent are found relative to a coin flip. Moreover, even at the six-month ahead horizons, gains of 12 to 18 percent are found for all exchange rates. The results clearly indicate that the end-of-month no-change is a more accurate naive forecast than the monthly average no-change.

Now we explore how robust these forecast gains are across countries. Figure 2.1 reports the quantiles of the RMSFE and SRs for the end-of-month no-change forecast relative to the monthly average no-change forecast at horizons of 1 to 12 months for the RER and REER. The gains in forecast performance are found at almost every horizon for every exchange rate measure for both directional accuracy and mean-squared precision. At the 1-month horizon in particular, gains are present for all quantiles shown, which means they were present for at least 87.5% of countries. The forecast gains are even larger and more robust for the NER and NEER, as is reported in Figure B.1 in the appendix. This indicates that the differences between the end-of-month no-change and the monthly average no-change in our sample is near universal, in addition to being substantial.

Figure 2.1: Distribution of Forecast Performance of End-of-Month Versus Monthly Average Forecasts



Note: Plot shows quantiles for 83 countries. Forecasts are compared relative to the period-average no-change forecast.

The results suggest that the loss in forecast from temporal aggregation of daily exchange rates to the monthly frequency is sizable. This is due to the high persistence of daily exchange rates and the large degree of aggregation (Amemiya and Wu 1972; Tiao 1972; Zellner and Montmarquette 1971), and is consistent with the evidence of daily data observed for other aggregated macroeconomic variables (Ellwanger and Snudden 2021, 2023a). The substantial and consistent differences in forecast accuracy show the importance of using the correct benchmark in practice. This calls into question the validity of conclusions in the existing literature derived from testing against the period-average no-change benchmark.

2.6.2 Real-time Model-Based Forecast Accuracy

We now quantify the information gains from temporally disaggregation when constructing real-time model-based forecasts of monthly average exchange rates. We evaluate forecasts from the three recursive models and the three direct models described in section 2.5. To be consistent with the last section, we begin by comparing the forecasts to the monthly average no-change forecast.

The median forecast performance of model-based forecasts of the bilateral RERs is reported in Table 2.8. When recursive and direct forecasts are estimated with monthly average data, the forecasts do worse than the monthly average no-change forecast in terms of median RMSFE at all horizons and in terms of median SR at horizons up to one year. When estimated with monthly average data, the recursive and direct forecasts are almost indistinguishable from each other in terms of RMSFE up to a 2-year horizon, and in terms of SR at all horizons. These results illustrate that, even though the monthly average no-change is an inefficient naive forecast, the inefficacy of forecasts estimated with average prices can give the perception that it is difficult to beat.

Table 2.8: Median Performance of Forecasts for Monthly Average Bilateral RER

Forecast	Model Inputs	1	3	6	12	24	36
				RMSFE Ratio			
Recursive	Month-Average	1.00	1.01	1.01	1.02	1.04	1.06
Recursive	PEPS	0.88	0.97	0.99	1.01	1.03	1.04
Recursive	Bottom-up	0.88	0.97	0.99	1.02	1.05	1.05
Direct	Month-Average	1.00	1.01	1.01	1.01	1.07	1.24
Direct	PEPS	0.88	0.96	0.98	1.00	1.07	1.24
Direct	UMIDAS	0.87	0.96	0.98	1.00	1.05	1.23
				Success Ratio			
Recursive	Month-Average	0.49	0.48	0.47	0.48	0.55	0.54
Recursive	PEPS	0.68	0.57	0.54	0.51	0.57	0.56
Recursive	Bottom-up	0.68	0.57	0.53	0.52	0.58	0.57
Direct	Month-Average	0.49	0.49	0.49	0.49	0.55	0.54
Direct	PEPS	0.68	0.57	0.53	0.52	0.56	0.54
Direct	UMIDAS	0.68	0.58	0.54	0.51	0.56	0.54

Note: Reports the median result across countries relative to the monthly average no-change forecast. Note, "PEPS" uses the end-of-month forecast as the forecast of the average. Recursive bottom-up ex-post averages daily forecasts. Direct forecasts use UMIDAS restricted to the end-of-month observation.

For both recursive and direct forecasts, using PEPS or bottom-up methods result in substantially better real-time forecast performance than month-average inputs.

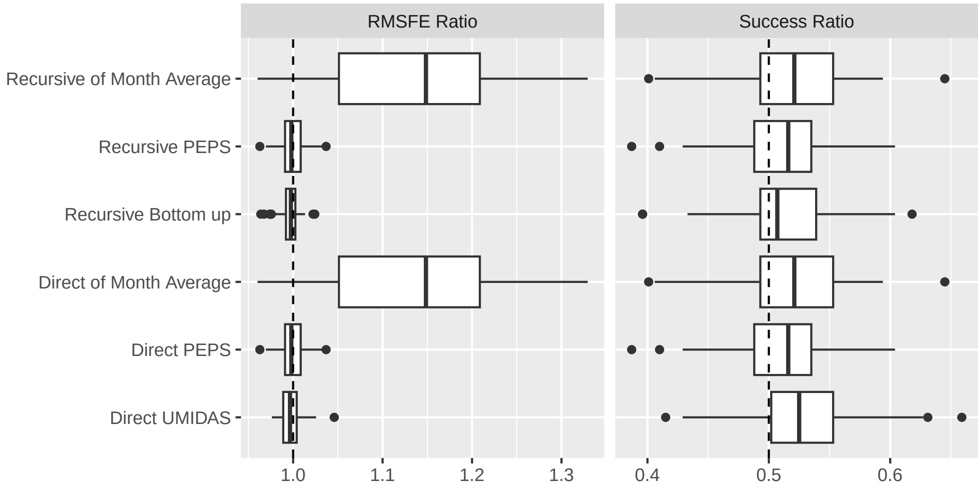
The gains are very similar to what was observed from the end-of-month no-change forecast at the one-month ahead, with 12 percent improvements in RMSFE and 36 percent improvement in the SR. For both disaggregated forecast methods, we again find that the median forecast of the recursive and direct forecasts are almost indistinguishable from each other in terms of directional accuracy, and for the RMSFE up to a 2-year horizon. These findings reinforce that the loss of forecast accuracy is substantial when model-based forecasts use aggregated daily data.

Interestingly, the results indicate that the forecasts constructed using end-of-month prices does about as well as the bottom-up forecasts constructed using daily data for real-time forecasts of the monthly average price. This evidence regarding the forecast gains from temporally disaggregated methods and the usefulness of end-of-period forecasts as a forecast of the period average is consistent with the effects of temporal aggregation of a persistent time series. It suggests that studies which have shown promise at end-of-period forecasts should be considered as candidates for period average forecasts.

We now examine the model-based forecast performance relative to the end-of-month no-change forecast. The distribution of RMSFEs and SRs from the different model-based forecasts at the 1-month-ahead for bilateral RERs is reported in Figure 2.2. For the RMSFE, the results reinforce the conclusions drawn from the exchange rate of the median country. Specifically, model-based forecasts using disaggregated methods have much lower RMSFE ratios compared to models estimated with month-average inputs. In fact, the forecasts constructed using month-average inputs are so poor that the entire interquartile range (IQR) of RSMFE ratios (indicated by the box) is above the country with the highest ratio constructed using end-of-month or daily inputs. These results show that, like the gains for no-change forecasts, the differences between the disaggregated model-based forecasts and the models estimated with monthly average data are substantial and near universal.

In contrast to the RMSFE, the SR is non-transitive, meaning that the best forecast relative to the monthly average no-change is not guaranteed to be the best forecast relative to the end-of-period no-change. This can be seen in the results

Figure 2.2: Accuracy of 1-month-ahead Forecasts for Bilateral RER relative to End-of-month No-change Benchmark



Note: The following outliers have been omitted: Brazil, Burundi, Kenya, Slovakia, Honduras

for the SRs in Figure 2.2. In particular, even though we still observe substantial deterioration in RMSFEs for model-based forecasts estimated with monthly average data, they perform similarly to the disaggregated approaches in directional accuracy. This can be seen in the overlap in the IQRs of the SRs for the three types of inputs. That is, the models have broadly similar SRs regardless of whether month-average, end-of-month or daily inputs are used. However, direct forecasts using daily inputs exhibit some of the largest and most robust forecast gains, with the lower bound of the IQR above 0.5. This suggests that mixed-frequency direct forecasts may have some advantages in forecasting directional accuracy, at least at short horizons.

Qualitatively, the results are very similar for the other exchange rates, see appendix B.4. Substantial RMSFE gains are found for all exchange rates and are robust across countries when disaggregated model-based forecasts are employed. There are only subtle differences in the best performing model at the one-month ahead. For example, for REERs, the bottom-up recursive forecasts and the end-of-month directs forecasts exhibit the best RMSFE forecast performance. Moreover, UMIDAS approaches again perform the best for the SR. That said, the real-time disaggregated forecasts for bilateral NERs are more similar and exhibit forecast performance similar to the end-of-month no-change. Overall, this suggests some robustness of real-time forecasts to out-perform the end-of-month no-change.

These results are all indicative that time-averaging introduces a loss of information for model-based forecasts of monthly average exchange rates. Integrating information from daily or end-of-month inputs into model-based forecasts substantially enhances forecast accuracy compared to specifications with month-average inputs.

2.6.3 Evidence of Real-time Predictability

We now test the real-time predictability of period average exchange rates. Specifically, we report the share of countries for which we find significant outperformance of each benchmark in terms of mean-square accuracy or directional accuracy forecasts. We examine forecasts for the six model-based forecasts and compare the forecasts to end-of-month no-change forecast as well as the erroneous period average no-change forecast, see Table 2.9.

Table 2.9: Percent of Countries with Significant One-Month Ahead Exchange Rate Forecasts

		Versus End-of-Month No-change				Versus Monthly Average No-change			
Forecast	Model-Inputs	REER	RER	NEER	NER	REER	RER	NEER	NER
		Mean-Square Accuracy							
Recursive	Month-Average	10%	5%	6%	1%	43%	10%	27%	3%
Recursive	PEPS	46%	19%	28%	3%	55%	79%	49%	74%
Recursive	Bottom-up	56%	31%	37%	3%	65%	81%	54%	74%
Direct	Month-Average	10%	5%	6%	1%	43%	10%	19%	3%
Direct	PEPS	55%	19%	28%	3%	55%	80%	49%	75%
Direct	UMIDAS	41%	19%	30%	4%	52%	77%	53%	80%
		Directional Accuracy							
Recursive	Month-Average	14%	13%	13%	10%	16%	6%	8%	7%
Recursive	PEPS	30%	6%	13%	5%	98%	91%	100%	93%
Recursive	Bottom-up	28%	10%	33%	12%	99%	87%	97%	95%
Direct	Month-Average	14%	13%	13%	9%	15%	6%	8%	7%
Direct	PEPS	98%	6%	29%	5%	98%	91%	97%	93%
Direct	UMIDAS	26%	15%	16%	18%	100%	88%	99%	94%

Note: Reports the share of countries where the forecast model is significant by the end of the forecast evaluation sample. Note, "end-of-month" inputs in model estimation follow PEPS and use the point forecast as the forecast of the average. Recursive and daily is an example of the bottom-up approach. Direct forecasts use UMIDAS restricted to the end-of-month observation.

Immediately notable is that comparisons to the monthly average no-change result in significant forecasts in the majority of cases when disaggregated methods are employed. For example, for forecasts of the monthly average NER, the disaggregated model-based forecasts significantly outperform the monthly average no-change up to

80 and 95 percent for mean-square accuracy and directional accuracy, respectively. However, this is a perfect example of spurious predictability. There is little evidence of short-term predictability of bilateral NERs when comparisons are correctly made to the end-of-month no-change forecast. For NERs, not even for five percent of countries do any of the forecasts exhibit significant predictability in RMSFE terms. This is substantial evidence that when forecasting a period-average, comparisons relative to the period-average no-change benchmark can lead to sizable type-I error.

Interestingly, evidence of real-time predictability is present for the forecasts of the other exchange rates. For bilateral RERs, significant RMSFE gains relative to the end-of-month no-change forecast are found for up to 31 percent of countries, and up to 15 percent of countries for the SR. These are slightly better for NEERs, with significant RMSFE gains relative to the end-of-month no-change forecast found for up to 37 percent of countries, and up to 33 percent of countries for the SR. By far, the most predictable exchange rate is the REERs, with up to 56 percent of countries exhibiting significant predictability and up to 98 percent for the SR. These findings of significant predictability are the first time that forecasts of period-average exchange rates have been compared against the null of no predictability.

2.7 Conclusion

Our findings have three implications. Firstly, they call into question the existent empirical literature forecasting period average exchange rates. We show the importance of comparing forecasts against the end-of-month no-change forecast to avoid spurious predictability. The substantial differences in the forecast performance of the end-of-month versus the monthly average no-change forecast suggest that such conclusions in the literature need to be re-examined. Secondly, our findings show that incorporating information from daily exchange rates results in substantial gains in real-time forecast accuracy. This holds promising potential to improve current forecasting practices, and the decisions that are reliant on these forecasts. It also shows the importance of calculating and reporting end-of-month and daily measures of exchange rates in official data sources. Finally, the evidence indicates that the period average EERs and bilateral RERs of many countries are forecastable in real-time. Developing new techniques that use end-of-period or daily inputs to forecast period-average exchange rates offers a promising avenue for future research.

3

Abolishing Imputation Credit Refunds: Evidence from an Event Study

Martin McCarthy[1]

Abstract

The theoretical literature has not reached a consensus on whether share prices are affected by taxes on dividends paid to domestic shareholders. This thesis chapter conducts an event study of the announcement of a reform to Australia's dividend imputation system. The reform would have abolished refunds to taxpayers with excess imputation credits. This would have raised taxes on dividends paid to Australian shareholders with excess imputation credits. The results provide no evidence that the announcement reduced share prices.

[1] I am grateful for comments and suggestions from audience members at Oxford's applied microeconomics workshop, the Reserve Bank of Australia and the Australian Treasury.

3.1 Introduction

The theoretical literature has not reached a consensus on whether share prices are affected by taxes on dividends paid to domestic shareholders. In some models share prices are determined by 'the' marginal investor, so dividend taxes affect prices if and only if they are imposed on this marginal investor. In other models, share prices are influenced by all investors, but there is disagreement about the relative importance of domestic and foreign investors in determining share prices.

The announcement of a reform to Australia's dividend imputation system provides an opportunity to test whether dividend taxes reduce share prices. On 13 March 2018 the Australian Labor Party announced that, if elected, it would abolish cash refunds provided to taxpayers with excess imputation credits (which are called 'franking credits' in Australia). The policy would have increased the taxation of dividends paid by domestic companies to some domestic shareholders, but would not have changed the taxation of dividends paid to foreign shareholders. Investors perceived the policy as highly likely to be implemented, as opinion polls and betting markets both suggested that Labor was the clear favourite to win the upcoming federal election.

This chapter conducts an event study of Labor's announcement. It argues that, if the announcement reduced share prices, then the reduction should be larger for companies expected to distribute more franking credits. It tests whether this is the case using a panel data method. The results provide no evidence that cumulative returns over the days following the announcement were negatively related to expected distributions of franking credits. Hence, it does not detect any effect of the reform on share prices. This is consistent with the predictions of Brennan (1970), among other models.

Understanding the effect of dividend taxes on share prices is important from a policy perspective. There is little consensus among policymakers regarding the appropriate way to tax dividends, as shown by the wide variety of systems in place. This includes classical systems that tax most dividends at full personal rates (such as the US), classical systems that tax dividends at concessional rates (such as the UK), dividend imputation systems (such as Canada), and systems that do not tax dividends at all (such as Singapore).[2] This lack of consensus about the appropriate system partly reflects disagreement about the effect of dividend taxes on investment and the incidence of dividend taxes. Whether dividend taxes reduce share prices has important implications for both of these areas of disagreement.

Firstly, if dividend taxes reduce share prices of listed companies, they may affect investment by these companies. Under the 'traditional view', the marginal source of financing is new equity, so a reduction in share prices would, all else equal, increase the cost of capital and reduce investment. Under the 'new view', the marginal source of financing is retained earnings, so even if dividend taxes reduced share prices, it would not affect the cost of capital or investment. This suggests that if dividend taxes reduce share prices, it may reduce investment among immature listed companies (who tend to rely on new equity financing), but have little effect on investment among mature listed companies (who tend rely on retained earnings) (Auerbach 2002). Hence dividend taxes may distort both the aggregate level of investment and its allocation between mature and immature listed companies.

Secondly, the economic incidence of dividend taxes depends on whether they reduce share prices. The statutory incidence of dividend taxes is on domestic investors who hold the shares on future dividend dates. However, the economic incidence of dividend taxes may differ:

- If share prices are not influenced at all by the shareholders subject to the dividend tax, then dividend taxes will not affect share prices. In this case the economic incidence matches the statutory incidence.

- If share prices are fully determined by the shareholders subject to the dividend tax, then dividend taxes will reduce share prices on the announcement date. The entire economic incidence will be on shareholders who hold the shares on the date of the announcement, whether domestic or foreign. Any domestic investor who buys the share on a future date will benefit from the lower purchase price of the share but incur the cost of higher dividend taxes, and be indifferent.

[2]For the US, see Internal Revenue Service (2020). For the UK and Canada, see Organisation for Economic Cooperation and Development (2020). For Singapore, see Deloitte (2019)

Any foreign investor who buys the share in the future will benefit from the lower purchase price but not be subject to the dividend tax, and will be better off. To the extent the reduction in share prices also reduces investment, then the future capital stock will be reduced, so the marginal product of labour will be reduced, and workers will also be harmed.

- If share prices are determined partly by shareholders subject to the tax and partly by others, then the economic incidence will fall in between these two cases.

In addition to providing evidence on the broad issue of whether dividend taxes affect share prices, this chapter provides evidence on the specific issue of dividend imputation systems. This issue remains topical, as the current Australian government has proposed new changes to the system, which have been criticised by the opposition (Australian Financial Review 2023).

The scope of this chapter is limited to listed companies, as it estimates the effect of the announced reforms on the prices of shares on the Australian Securities Exchange (ASX). Although I do not find evidence that dividend taxes affect the share prices of listed companies, it is still possible that dividend taxes affect the value of unlisted companies. Many unlisted companies will have more difficulty accessing funding from foreign investors, so the value of these companies may be more strongly influenced by domestic investors, and hence more affected by taxes on dividends paid to these investors.[3] Any policy decision regarding dividend imputation should take into account both the effects of the policy on listed companies (including the share price effects studied here) and its effects on unlisted companies (which are out of scope for this chapter).

The next section reviews the literature. Section 3.3 describes Australia's dividend imputation system and the proposed reform, and section 3.4 identifies the event window over which the effect of the reform will be evaluated. Section 3.5 argues that, if the announcement reduced share prices, then cumulative returns over the

[3] Caglio et al. (2022) has shown that large listed companies and small unlisted companies obtain different types of financing, and that they are affected differently by monetary policy as a result.

event window should be negatively related to proxies for the likelihood of a company distributing franking credits. The data, method and results are described in sections 3.6, 3.7 and 3.8. The final section concludes.

3.2 Literature Review

There are a variety of theoretical models that link the tax treatment of shareholders to asset prices. In some models share prices are determined by 'the' marginal investor, so dividend taxes affect share prices if and only if they are imposed on the marginal investor. However, the identity of this investor differs between models. In Miller and Scholes (1978) and Swan (2019) the marginal investor is a domestic financial institution, while in Boadway and Bruce (1992) it is a foreign investor.

In other models, share prices are influenced by many or all investors to some extent. In the tax-adjusted capital asset pricing model (CAPM) of Brennan (1970), the effect of dividend taxes on share prices depends on the average of the dividend tax rates faced by all investors, weighted by their wealth. This suggests that:

- Dividend taxes on US shareholders should reduce share prices in the US and abroad, as these shareholders own a large share of the world's wealth.
- Dividend taxes on shareholders in other advanced economies should have little effect on share prices in that country or abroad, as these shareholders own a small share of the world's wealth.

In contrast, Monkhouse (1993) argues that whether dividend taxes affect the share prices of a particular company may depend on the investor base of that company. National accounts data shows that Australians owned 59% of the value of Australian shares in the quarter when the announcement occurred[4], so Monkhouse's

[4] I used the December 2022 edition of 'Australian National Accounts: Finance and Wealth' published the Australian Bureau of Statistics. This provided the following figures for March 2018:

- Table 6 (Financial Assets and Liabilities of Non-Financial Corporations) reports that:
  - Amounts outstanding at end of period; Liabilities; Listed shares and other equity held by:; Total (Counterparty sectors) was $1,105.412 billion
  - Amounts outstanding at end of period; Liabilities; Listed shares and other equity held by:; Rest of world was $396.574 billion
- Table 14 (Financial Assets and Liabilities of Financial Corporations) reports that:
  - Amounts outstanding at end of period; Liabilities; Listed shares and other equity held by:; Total (Counterparty sectors) was $328.486 billion
  - Amounts outstanding at end of period; Liabilities; Listed shares and other equity held by:; Rest of world was $197.417 billion

view suggests taxes on dividends paid to Australians could affect Australian share prices significantly.

3.2.1 Evidence on the US

Empirical work supports the claim that dividend taxes on US shareholders decrease the share prices of US companies, consistent with Brennan 1970. The main source of evidence is the response of share prices to the dividend tax cuts in 2003. The tax cuts were scheduled to last until the end of 2008, and it was uncertain if they would be extended further. A company whose dividend yield is currently high will distribute more during the period when dividend taxes are definitely low. In contrast, a company whose dividend yield is currently low will distribute more in the future, when the dividend tax cut may have expired. Hence, if dividend taxes affect US share prices, then abnormal returns should be higher for companies with higher dividend yields. Auerbach and Hassett (2005), Gadarowski et al. (2007) and Amromin et al. (2008) all find that this is the case. The US dividend tax cuts also changed analysts' forecasts for corporate earnings and other variables. Dhaliwal et al. (2007) show that the ex-ante cost of equity capital implied by analysts' forecasts was lower after the tax cut than it had been before, which they view as evidence that dividend taxes reduce share prices.

Empirical work also shows that US dividend taxes reduce the share prices of foreign companies. Kenchington (2019) finds that the 2003 dividend tax cuts increased stock prices of foreign companies with high dividend yields, provided that the company was located in a country with an income tax treaty with the US.

While most empirical work is based on the dividend tax cuts of 2003, some studies present other evidence for the effect of US dividend taxes. An example is Sialm (2009), who presents time series evidence showing that dividend taxes were negatively related to share prices over the period from 1913 to 2006.

- Hence the total value of listed shares of Australian companies outstanding as of March quarter 2018 was \$1433.898 billion, of which \$593,991 billion was owned by the rest of the world. This implies that \$839.907 billion was owned by Australians, which is 59% of the total.

3.2.2 Evidence on other Advanced Economies

Brennan (1970) predicts that in advanced economies other than the US, where shareholders own a small share of the world's wealth, dividend taxes should only have a small effect on share prices. The empirical work so far has yet to reach a consensus on whether this is the case.

3.2.2.1 Dividend Drop-off Studies

One source of evidence is empirical work on countries with dividend imputation systems. In these systems, domestic shareholders can use imputation credits to reduce their tax payable on dividends, while foreign shareholders cannot. Hence, if expected distributions of imputation credits are capitalised into share prices, this is evidence that dividend taxes on domestic shareholders reduce share prices. The most popular approach to estimating the market value of imputation credits uses dividend drop off ratios.

When a company declares a dividend, they will specify that the dividend will be paid to those that hold the share on a particular date. Those who buy the share on or after the 'ex-dividend date' are not entitled to receive the dividend. Usually the share price falls when a share goes ex-dividend, but the share price fall may differ from the amount of the dividend. To quantify this, one can compute a dividend drop off ratio, defined as the ratio of the size of the price fall to the dividend amount.

There is a large empirical literature that attempts to link drop-off ratios to taxation at the shareholder level. A number of papers use drop-off ratios in dividend imputation systems to infer the extent to which imputation credits distributed to shareholders are capitalised into share prices. If the drop-off ratio of a dividends with attached imputation credits is larger than the drop-off ratio of dividends without these credits, this suggests that the market places a positive value on distributions of imputation credits. Using this method, Cannavan and Gray (2017) found that $1 of distributed imputation credits had a market value of around $0.35 in Australia. This estimate suggests share prices are partly determined by Australians and partly by foreigners.[5]

A major challenge is that, in most countries that have had dividend imputation systems, foreign shareholders have been able to benefit from imputation credits by trading around the ex-dividend date. Empirical work has found evidence of such trading in many countries with dividend imputation systems, including Germany (McDonald 2015), Finland (Rantapuska 2008) and Taiwan (Chen, Chow, et al. 2013; Tseng and Hu 2013). Trading aimed at capturing imputation credits pushes up share prices before the ex-dividend date, and pushes them back down afterwards. As a result, the drop-off ratio will mostly reflect the value placed on credits by those engaged in short-term trading around the ex-dividend date, rather than the value placed on credits by long-term investors who determine share prices at other times (Lajbcygier and Wheatley 2012). This issue is less pronounced in Australia, due to unusually strong rules intended to prevent short-term trading aimed at capturing imputation credits. The most important rule is the '45-day rule', which states that a shareholder can only use the imputation credits attached to a dividend if they held those shares 'at risk' for at least 45 days (Australian Taxation Office 2016). However, empirical evidence shows that some trading aimed at capturing imputation credits occurs in Australia despite these rules (Ainsworth et al. 2018; Grant et al. 2018; Le et al. 2020).

Some papers compare drop-off ratios before and after a tax reform. Bell and Jenkinson (2002) show that the drop-off ratio fell following a reform to the UK's dividend imputation system in 1997. Beggs and Skeels (2006) find that drop-off ratios in Australia increased after cash refunds for excess imputation credits were introduced in 2000. These studies however, do not provide much evidence on the effect of dividend taxes on share prices. Firstly, as discussed above, drop-off ratios are influenced by the value placed on imputation credits by those engaged in short-term trading around the ex-dividend date. Secondly, drop-off ratios tend to be highly

[5] An Australian should value $1 of franking credits received just as much as $1 of cash dividends, as they can redeem the $1 of franking credits for $1 in cash, and franking credits form part of the individual's taxable income in the same way as $1 of cash dividends.

volatile, making it hard to attribute changes in the ratio to a particular cause (Chetty et al. 2005). Thirdly, these studies may suffer from more specific issues. In Bell and Jenkinson (2002), for instance, the fall in the mean drop-off ratio was partly due to a greater prevalence of negative drop-off ratios, and it is hard to see how this could have been caused by the reform (Bond et al. 2007b).

3.2.2.2 Estimation of Asset Pricing Models and Event Studies

Motivated by the limitations of dividend drop-off studies, some authors have estimated the market value of imputation credits in other ways. One approach is to estimate an asset pricing model that includes imputation credits as an explanatory variable. The evidence from these models is mixed. Lajbcygier and Wheatley (2012) use a 'two-pass' method to estimate a CAPM-based model and a variety of other models. The results provide no evidence that imputation credits reduce the cost of equity capital. Using a different 'single-pass' method, Swan (2019) estimates a CAPM-based model, and finds that imputation credits substantially reduce the cost of equity, and in fact, that imputation credits are close to fully capitalised into share prices. A challenge is that over these long time periods, there will be many different influences on returns, making it hard to isolate the effect of dividend imputation specifically. This means the results will be sensitive to the precise methods used, including how they control for any other factors that influence differences in returns across companies.

Event studies can be used to study the effect of a change in dividend taxes on share prices. The literature has studied earlier reforms to Australia's dividend imputation system, but did not test if these reforms affected share prices or share ownership.[6] Bond et al. (2007b) study a change to the UK's dividend imputation system in 1997, which had the effect of increasing the taxation of dividends paid to a major class of UK shareholders (pension funds and insurance companies). They found that this reform had little effect on the share prices of UK companies when it was announced,

[6]Pattenden and Twite (2007) finds that the introduction of dividend imputation caused higher and more volatile dividend payments. Brown et al. (2020) finds that the introduction of cash refunds for excess franking credits in 2001 reduced tax avoidance, while Twite et al. (2022) finds it increased dividend payments.

though it did cause UK pension funds to reduce their holdings of UK equities.[7] The results in this chapter suggest that the results found by Bond et al. (2007b) are not due to some specific feature of the UK or the 1997 reform, but rather reflect some more general phenomenon that holds for other countries and other reforms.

Both the estimation of asset-pricing models and event studies could be affected by trading aimed at capturing imputation credits. If this trading only affects share prices in a window around the ex-dividend date, then it will only have a modest effect on the estimated models and event studies, as companies are only close to their ex-dividend date for part of the time. This would be an advantage over the dividend drop-off studies, whose estimates are based entirely on companies close to their ex-dividend date. It would be more concerning if this trading affects a company's share price even when it is not close to its ex-dividend date. If so, this would be a threat to any study that uses dividend imputation systems to test if share prices are affected by taxes on dividends paid to domestic shareholders. However, even if such trading affects share prices at all time, the effect is likely to be most pronounced in the window around the ex-dividend date, so the issue will be less severe for asset pricing models and event studies than for dividend drop-off ratios.

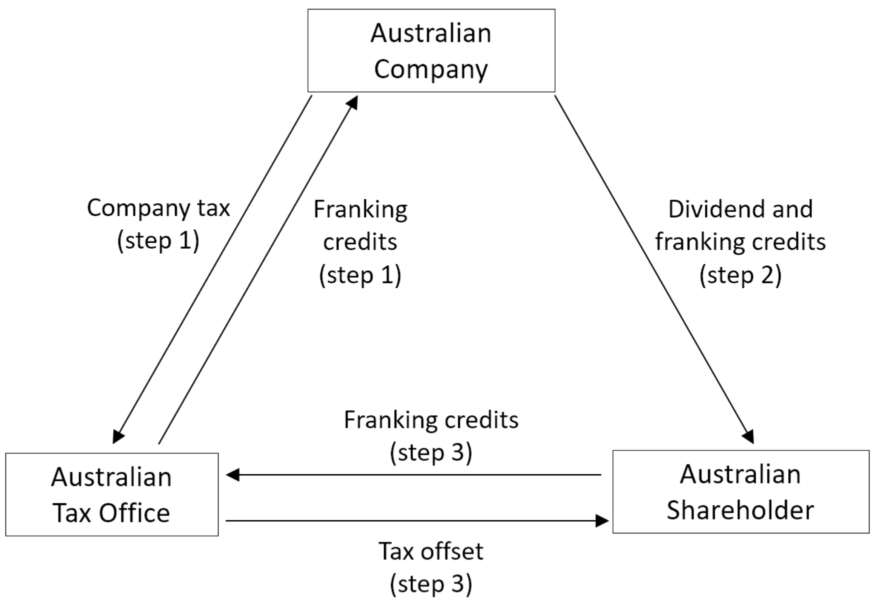
[7] A companion paper, Bond et al. (2007a), studies how the reform affected the dividend payments and investment spending of listed companies.

3.3 Dividend Imputation and the Proposed Reform

3.3.1 Description of the Reform

To understand how the proposed reform affected the taxation of dividends, one must first understand Australia's dividend imputation system. Figure 3.1 provides a simplified depiction of the system. It shows three steps. In step 1, an Australian company pays company tax to the Australian Tax Office, and receives an equal amount of franking credits. This adds to its franking account balance. In step 2, the company pays a dividend to shareholders, and attaches all of its franking credits to that dividend. This reduces the company's franking account balance back to zero. In step 3, the shareholder returns the franking credits to the Tax Office, and receives a tax offset equal to the amount of credits.[8]

Figure 3.1: A Stylised Depiction of Australia's Dividend Imputation System



Since July 2000, franking credits have been 'refundable', which means that if the credits reduce the shareholder's personal tax to zero, the excess can be received as cash.

[8]In practice the timing of these steps is flexible. For example, a company can receive franking credits in one year, which adds to its franking account balance, and attach those franking credits to a dividend many years later, which reduces its franking account balance.

The effect of the system with refundable credits is that these profits paid to Australian shareholders are only taxed once, at the shareholder's personal income tax rate.

On 12 March 2018, Labor announced that, if they formed government after the federal election, they would abolish cash refunds provided to shareholders with excess franking credits. This policy would have increased the tax on dividends paid by Australian companies to some Australian shareholders, but would not have affected the taxation of foreign shareholders.[9]

3.3.2 The Types of Shareholders Affected

Labor's policy would increase taxes paid by any taxpayer with excess franking credits. There are two classes of Australian shareholder who tend to have excess credits: self-managed superannuation funds and retired individuals.

Superannuation funds are retirement funds that are compulsory for most Australians. The tax treatment of superannuation funds is complex, but can broadly be described as a taxed-taxed-exempt ('TTE') system:

- **Contributions are taxed (T)**[10]
- **Investment earnings are usually taxed (T)**. Tax of 15% is payable on earnings of members who are not yet retired ('accumulation members'). However, no tax is payable on earnings of members who retired and receiving an income stream from their account ('pension members')
- **Benefits are usually exempt (E)**

Since superannuations funds hold Australian shares, they receive franking credits. Institutional superannuation funds, which are large funds with professional managers, will typically be able to use all or most of their franking credits to reduce their tax

[9] Unfranked dividends paid to foreign shareholders are typically subject to Australian withholding tax, while franked dividends are exempt (Australian Taxation Office 2019). Labor's proposal does not affect whether franking credits provide exemption from Australian withholding tax, so it does not affect the taxation of dividends paid to foreign shareholders.

[10] The contributions are either paid out of the member's before-personal-tax income (in which case they are subject to a 15% tax as they enter the fund), or they are paid out of the member's after-tax income.

obligations. In contrast, self-managed superannuation funds, which are funds with at most 6 members that are managed by the members themselves, are often unable to use all of their credits. Hence these funds often have excess franking credits and hence face higher dividend taxes. [11] Self-managed superannuation funds hold around a quarter of total superannuation fund assets (Australian Prudential Regulation Authority 2022).

Individual taxpayers who hold shares outside of superannuation will receive franking credits. Working-age individuals who own shares typically have substantial labour income, and hence will be able to use all of their franking credits to reduce the tax payable on that labour income. However, retired individuals typically have little labour income and the receipt of superannuation income is tax exempt, so they have little tax payable, and will often have excess franking credits.

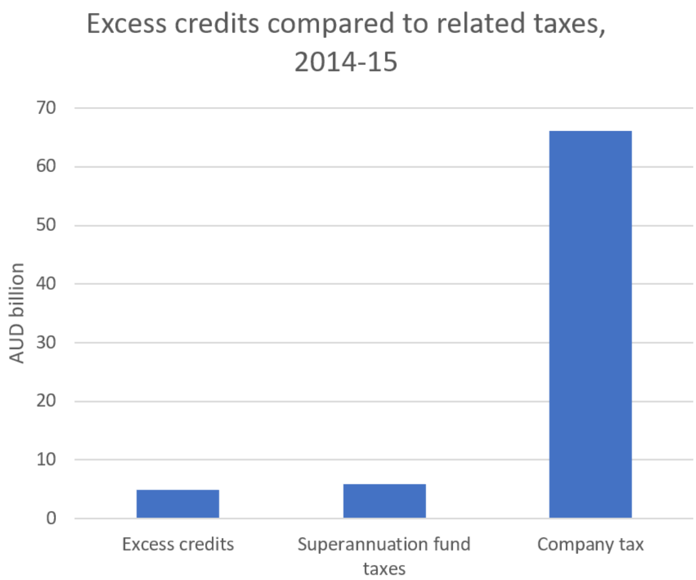
3.3.3 Quantitative Importance of the Reform

If Labor's policy were implemented, and there was no behavioural response, it would have raised revenue equal to the amount of excess franking credits. In the 2014-15 financial year, individuals claimed a total of 2.0 billion AUD of excess franking credits, self-managed superannuation funds claimed 2.6 billion AUD, while other funds claimed 0.3 billion AUD (Parliamentary Budget Office 2018). This implies that the policy would have raised 4.9 billion AUD if it had been in place in 2014-15, assuming no behavioural response. To give a sense of scale, this is equal to 84% of taxes on superannuation funds, and 7.5% of company tax (Figure 3.2). Since the policy was scheduled to take effect in the 2019-20 financial year, the amount raised would likely be larger, due to growth in nominal dividends over time.

The additional revenue raised by the policy would have been somewhat reduced by tax planning strategies. If investors expected these strategies to be very successful, then it is likely that share prices would react little to Labor's announcement, even if dividend taxes reduce share prices in general. In practice, however, these strategies were only expected to have a modest effect on tax paid.

[11] The institutional funds typically have enough working-age members to generate enough of a tax liability to use the credits. However, a self-managed superannuation fund whose members are in retirement will not have much tax liability, and will be unable to use the credits.

Figure 3.2: A comparison of excess credits to tax revenue



Data Sources: Parliamentary Budget Office (2018) Australian Government (2015)

For details on Labor's policy, including the effect of tax planning on revenue, see Appendix C.1.

3.4 Event Windows

3.4.1 Identifying Possible Windows

This paper tests whether returns during the event windows are affected by the proposed reform. The choice of event windows is therefore a key methodological choice.

This paper considers three events, which are listed in Table 3.1. The choice of these events was informed by the number of Australian news articles mentioning dividend imputation or franking credits on each day (Figure 3.3).

Table 3.1: Event dates

Event	First trading day when information is public
Announcement	13 March 2018
Pension	26 March 2018
Election	20 May 2019

The 'announcement' event is the announcement of Labor's policy to end refundability, which resulted in the first spike in news coverage. The policy was revealed in a number of news articles published after markets had closed on 12 March 2018.[12] The policy was then announced formally in a speech by the leader of the Labor party on 13 March 2018 (Shorten 2018).

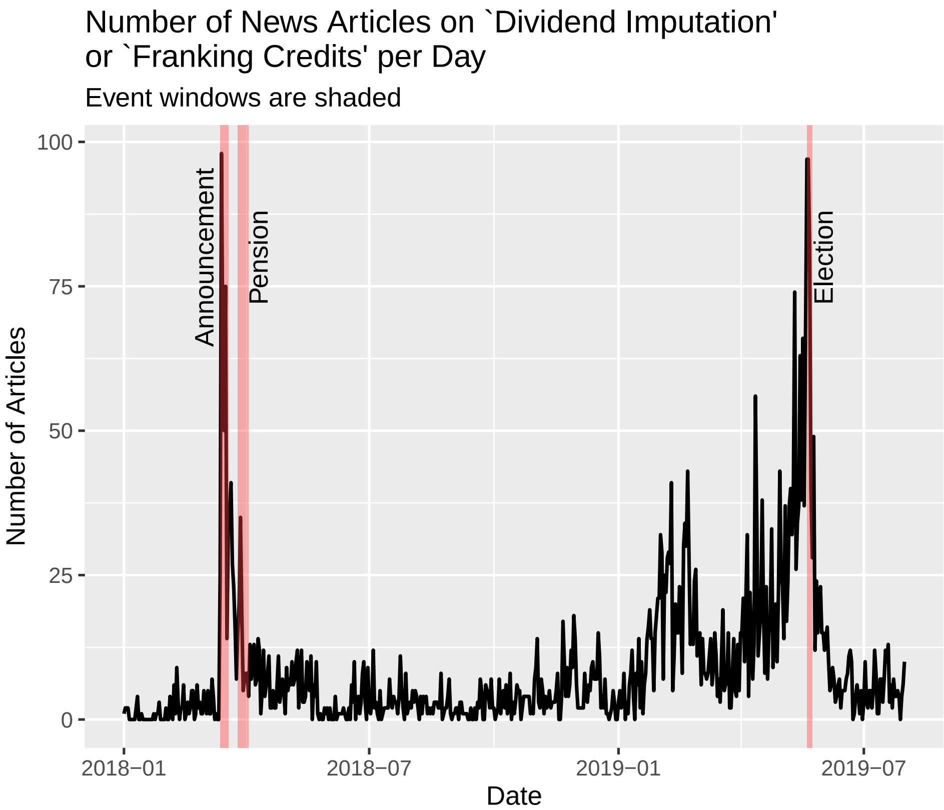
The 'pension' event is Labor's announcement that recipients of Australian Government pensions and allowances would be exempt from the policy, as would self-managed superannuation funds that had at least one such recipient as a member.

The 'election' event is the Australian federal election. In a very surprising outcome, the Coalition government was re-elected. This motivated the 'election window'.

The high number of articles in February 2019 and April 2019 were investigated, but do not correspond to suitable events for the purposes of this paper. The first surge occurred when Labor called for the resignation of a Government MP chairing a parliamentary committee on franking credits. The second related to the

[12] See, for instance, Sydney Morning Herald (2018), Australian Financial Review (2018), Australian Broadcasting Corporation (2018)).

Figure 3.3



Data Source: Nexis

debates between party leaders, where franking credits featured prominently. Neither set of articles revealed new information about the policy itself. Moreover, these developments do not appear to have had much effect on the perceived probability of the policy being implemented, as Labor retained its lead in the opinion polls and betting markets over those periods.

I use event windows of 3 trading days, starting on the day the event became public and ending two trading days later. The choice of window length trades off two considerations. If the window length is too long, then the price movements due to the event (if any) will make up a smaller share of the returns during the window, making it harder to estimate the effect of the event. On the other hand, if the window length is too short, then it may exclude some or all of the effect of the event. By choosing a window length of 3 trading days, this paper allows for the possibility that it takes some time for the event to be reflected in the cross-sectional pattern of returns, as it takes time for market participants to assess the likely franking credit distributions of different companies. Using a short window length like this one is typical in the literature; Oler et al. (2008) finding that over two thirds of event studies in management journals had a window length of between 1 and 5 trading days.[13]

3.4.2 Why the Announcement Window is Preferred

The announcement window is an ideal setting in which to test whether dividend taxes affect the cost of equity. In contrast the pension window and election window are less suitable. For this reason, this thesis chapter focusses on the announcement window, though results for the other windows are also presented.

3.4.2.1 Announcement Window

The announcement window has two advantages. The first advantage is that the announcement window covers a period when investors' beliefs about the policy changed substantially. If investor's beliefs only change slightly during the window, then any change in share prices caused by the change in these beliefs would likely be small relative to other sources of variation in the data. Before the window, shareholders attached a very low probability to such a policy being implemented. In the year preceding the announcement, there were very few news articles about dividend imputation or franking credits. Moreover, those articles were focussed on issues such as the likelihood that a particular company will pay a franked dividend, rather than speculation about possible policy changes. The announcement substantially raised the perceived probability of the policy being implemented.

One concern may be that investors were not aware of the policy. This seems unlikely, however. The announcement received extensive media attention. The

[13]Some event studies include a few days before the event in the window in case the information leaks out ahead of it becoming public. Such a 'leak' could not have occurred for the election outcome, which had not been determined until the election occurred on Saturday, 18 May 2019. Such leaks are also not a concern for the other windows. The reason is that, in most event studies, the events concern a specific company (e.g. a merger), so there is a genuine concern that an insider could influence the price of that company using this information. In contrast, the events under study affect the majority of listed companies, though to differing extents. If the announcement or pension events affected the cross-sectional pattern of returns, this will have occurred after the announcement, as it is not plausible that a few insiders could have affected the returns of so many companies so that the reform is priced in before the event date.

announcement appeared on the front page of most Australian newspapers on 13 March 2018, and was covered on the front page of the Australian Financial Review on every day of the announcement window (see Appendix C.2).

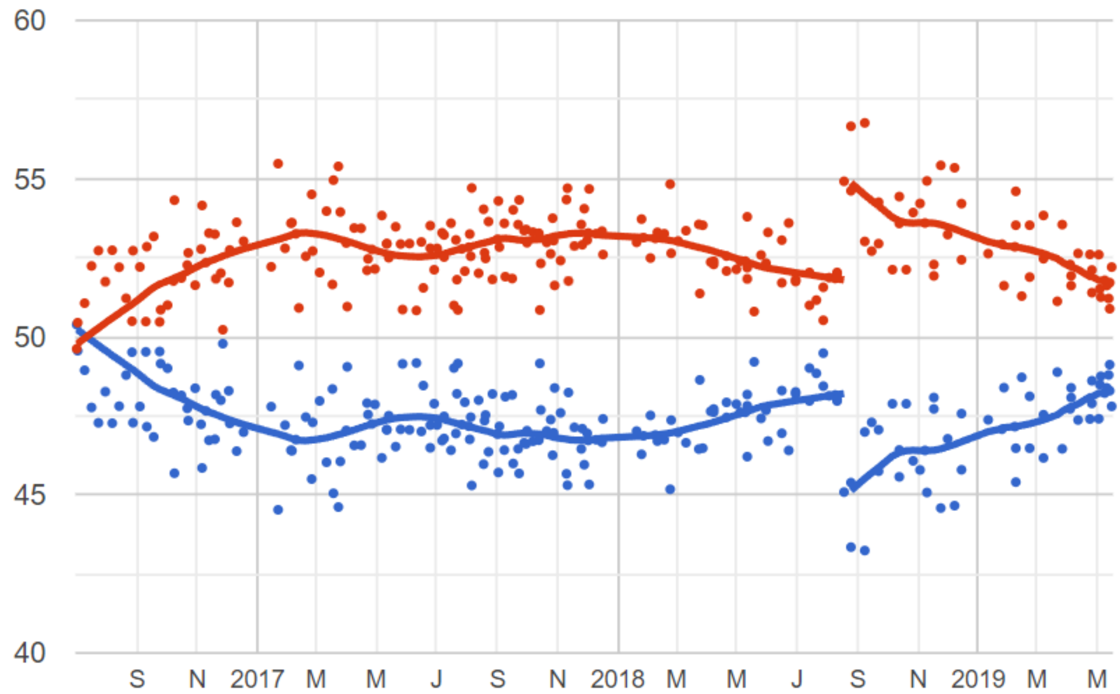
Another concern is that investors may not have expected the announced policy to actually be implemented, as it would require Labor to win government in the May 2019 federal election. However, at the time of the announcement they were the clear favourite to form government. Labor had been consistently ahead of the incumbent Liberal-National Coalition in 'two-party-preferred' polling, which asks voters to nominate either Labor or the Coalition (Figure 3.4). Two-party-preferred is the standard way of measuring electoral support in Australia (Goot 2016), which reflects the use of a preferential voting system in the House of Representatives. Consistent with polling, betting market odds at the time showed Labor was the clear favourite to form government. On the day of the announcement, the decimal odds were 1.5 for Labor and 2.6 for the Coalition (Koukoulas 2019).

The second advantage is that, although there were other developments affecting the share prices during the announcement window, it is possible to control for those developments. On the same day that the leader of Labor announced the franking credit policy, the deputy leader announced the 'Australian Investment Guarantee' (AIG). The AIG would allow companies to immediately deduct 20% of any qualifying new asset worth $20,000+, with the remainder depreciated as usual. The revenue impact of the policy was expected to be roughly half that of the dividend imputation policy.[14]

To determine if any other developments of note occurred during the announcement window, I listed the headlines of all front page stories of the Australian Financial Review on these days (Appendix C.2). There were many stories that would have affected the share price of one or two companies, but this would not have much

[14]Appendix E, Parliamentary Budget Office (2019) presents estimates of the revenue impact the policies Labor took to the election. According to these estimates, AIG's revenue impact would be 17% as large as the impact of its franking policy in 2021-22, growing to 60% as large in 2022-23. However, this is based on the revenue impact of the version of Labor's policy that exempted government pensions. In the announcement window, the relevant policy is the original policy, which had a larger revenue impact.

Figure 3.4: Two-Party-Preferred Opinion Polling Ahead of May 2019 Federal Election



Note: Figure reproduced from Poll Bludger (2019). The solid line is a an aggregate of different polls computed by the figure's creator. Each dot is a 2-party-preferred estimate from a poll by one of Newspoll, Galaxy, Ipsos, YouGov, Essential Research and ReachTEL. The jump in Labor's 2-party-preferred estimate in August 2018 occurred when the Liberal Party replaced its leader and therefore the Prime Minister.

effect on our tests of whether dividend taxes reduced Australian share prices. There was some speculation about future policy changes, but the change in beliefs due to these stories is likely low.[15] There was also some discussion about the royal commission into financial services, which may have affected the share prices of some financial institutions, but the methods used in this chapter can handle industry-specific developments.[16]

[15]'One Nation softens on company tax cuts' tells us that a minor political party seemed to lean a bit more than previously towards supporting company tax cuts. This would only have had a modest effect on investor's perceived probability of the tax cuts being implemented.

[16]As explained in section 3.7, this chapter forms a panel dataset containing the returns of individual shares on each day. It then regresses returns against interactions of dummies for the event window and proxies for future franking distributions. The test of whether Labor's announcement reduces share prices is whether the coefficient on the interaction of the dummies and proxies differs from zero. In all but the baseline specification, the coefficient of interest will not be affected by differences in returns during the event window across industries, as the explanatory variables include interactions of the event dummies and industry dummies.

In terms of Australian macroeconomic news, the announcement window was very quiet. The Australian Bureau of Statistics did not release any market-sensitive economic data.[17] The Reserve Bank of Australia did not make any monetary policy decisions nor publish any press releases on its website. Additionally, neither the federal government, nor any state or territory government, handed down their budget.

3.4.2.2 Pension window

The pension window starts on the day Labor announced that recipients of Australian Government pensions and allowances would be exempt from its policy. Self-managed superannuation funds that had at least one such recipient as a member would also be exempt. Hence the pension window covers a period during which investors beliefs regarding the nature of the policy changed. However, the change in beliefs was fairly modest. Firstly, the exemption only had a small effect on the estimated revenue impact of the policy (Parliamentary Budget Office 2018). Secondly, the policy change was much less salient than the original announcement, as shown by the smaller amount of press coverage received. Hence, even if dividend taxes do reduce share prices, the variation in returns caused by this particular policy announcement may be small, making it hard to detect.

3.4.2.3 Election window

The election window starts on the Monday that follows the Coalition's victory in the election on Saturday, 18 May 2019. Before the election, Labor was expected to win office, as discussed in section 3.4.2.1. The Coalition's surprising victory made it extremely unlikely that the franking credit policy would be implemented during the next term of government. Moreover, senior Labor figures blamed the loss partly on the franking credit policy, making it seem unlikely that the franking credit policy would be implemented even if Labor won a subsequent election.

[17] One of those releases was 'Labour Force, Rebenchmarked Estimates', which seems important at first glance, but merely takes labour force figures that have already been published and scales them to reflect new population data. Another release was 'Housing Finance'. Wallis 2020 finds no evidence that yields respond to this release, suggesting that the response of share prices to this release is likely to be small or zero. Some other releases were published, but they are not perceived as market-sensitive by investors.

A problem with the election window is that the election outcome also changed investors' beliefs regarding many other economic policies. Labor had promised higher taxes on personal incomes, housing, superannuation and trusts, and had promised increased spending on infrastructure, education and childcare (Australian Broadcasting Corporation 2019). Labor had also promised to increase the wage premium that companies must pay staff who work on Sundays or public holidays, and to introduce a policy that seemed likely to lead to a higher minimum wage (Clayton Utz 2019). This makes it more challenging to estimate the effect of the franking credit policy specifically.

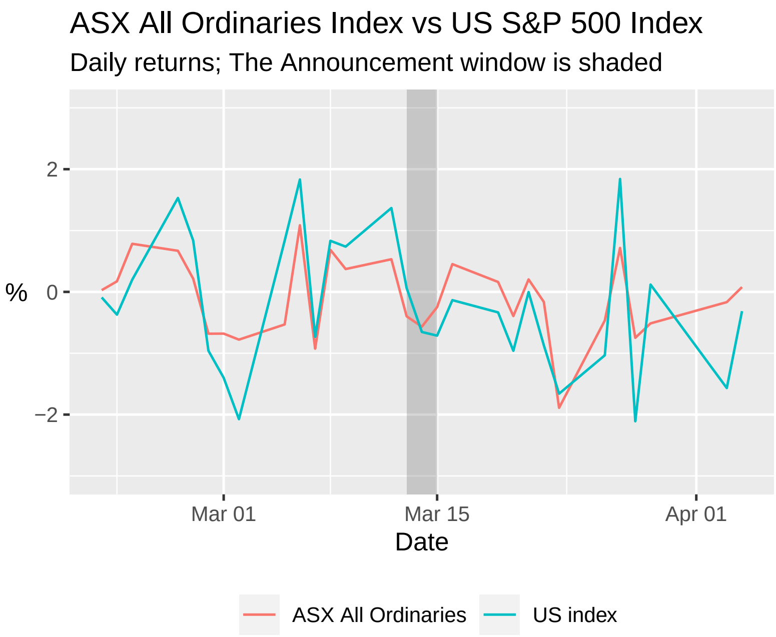
3.4.3 Aggregate Returns during the Announcement Window

The extent to which Labor's announcement affects share prices depends on the extent to which share prices are determined by different types of investors. The effect on returns would be largest if the marginal investor is an Australian with excess credits. A back of the envelope calculation suggests the announcement would have caused Australian share prices to fall by around 14.4% if this were the case (see appendix C.3.3). If all investors determined share prices to some extent, then the announcement would still reduce share prices, but to a lesser extent. The size of the fall could be substantial if the influence of each class of investors depends on the proportion of Australian shares that they own (as suggested by Monkhouse (1993)). However, the fall would be negligible if the influence of each class of ivnestor depends on their share of total world wealth (as suggested by Brennan (1970)).

During the event window, Australian share prices rose slightly. The aggregate returns evidence does seem to rule out the possibility that Labor's announcement had a large effect on share prices, such as a 14.4% fall, as this would almost certainly be visible in aggregate returns. However, it is still possible that Labor's policy reduced share prices by a smaller amount, which is not easily visible in aggregate returns due to other facts affecting share prices during that period.

A simple approach to estimating the effect of Labor's announcement would be to estimate the change in Australian share prices beyond that which would be

Figure 3.5

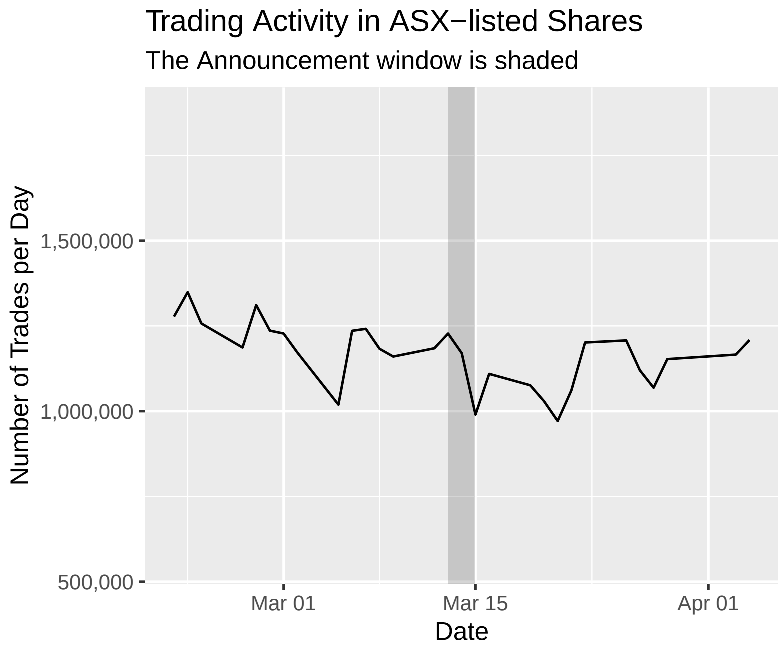


expected based on its historical correlations with other variables. For example, the Australian share market is correlated with the US share market, so we could estimate the returns in the Australian index not explained by the US index. However, this approach has two limitations. Firstly, the power of this test to detect the effect of dividend taxes would be low. Secondly, Labor's announcement may have had a direct effect (i.e. changing investor's beliefs about future dividend taxes) and indirect effects (e.g. changing investor's beliefs about how 'business friendly' a future Labor government would be more generally). The overall response of share prices would reflect both the direct and indirect effects, but for our present purposes we are only interested in the direct effect.

A better approach to estimating the effect of Labor's announcement is to focus on the heterogeneity in returns across companies during the window, as is done in the remainder of this chapter.

Trading activity during the event window was not unusually high or low (Figure 3.6). This fact by itself does not provide much evidence on whether Labor's policy affected share prices, as it is possible for prices to move substantially even with moderate trading activity.

Figure 3.6


Trading Activity in ASX−listed Shares
The Announcement window is shaded
Number of Trades per Day
1,500,000
1,000,000
500,000
Mar 01
Mar 15
Apr 01
Date

3.5 Theory

This chapter uses the returns during the event window as a source of evidence on whether Labor's franking credits policy reduced share prices. The challenge is that Labor's announcement was only one influence on returns during the window.

To formalise this, let $i \in \{1, ..., n\}$ index the set of companies. Let t be the last day of the event window and h be the length of the event window, so the first day of the event window must be day $(t-h)$. Let p_i^{t-h} and p_i^t denote the share price in dollars at the start and end of the event window respectively. Assume that the overall change in share prices in dollars during the announcement window, $\Delta p_i^t = p_i^t - p_i^{t-h}$ equals the sum of three components:

$$\Delta p_i^t = \left(\Delta p_i^t\right)^{\text{normal}} + \left(\Delta p_i^t\right)^{\text{franking}} + \left(\Delta p_i^t\right)^{\text{AIG}} \tag{3.1}$$

where $(\Delta p_i^t)^{\text{normal}}$ is the change in share prices that would have occurred absent the two policies, $(\Delta p_i^t)^{\text{franking}}$ is the contribution of changes in beliefs about franking credit policy, and $(\Delta p_i^t)^{\text{AIG}}$ is the contribution of changes in beliefs about AIG.

We define returns as changes in share prices divided by the initial share price, $r_i^t = \frac{\Delta p_i^t}{p_i^{t-h}}$. Dividing equation (3.1) by the initial price shows that the observed return in the window can be expressed as the sum of three contributions to returns:

$$r_i^t = \left(r_i^t\right)^{\text{normal}} + \left(r_i^t\right)^{\text{franking}} + \left(r_i^t\right)^{\text{AIG}} \tag{3.2}$$

During the announcement window, Labor announced both that it would abolish excess franking credit refunds and that it would introduce AIG. Section 3.5.1 shows that, if Labor's announcement reduced share prices, the contribution of franking credits to returns, $(r_i^t)^{\text{franking}}$, is likely to be more negative for companies expected to distribute more franking credits in the future. Section 3.5.2 shows that the contribution of AIG to returns, $(r_i^t)^{\text{AIG}}$, is likely to be more positive for companies expected to undertake more capex in the future. Taken together, these arguments show that if one regresses actual returns on proxies for future franking credit distributions and on proxies for future capex, then one can test if Labor's announcement reduced share prices.

While the argument below is presented in terms of the announcement window, it is straightforward to apply it to the other windows. In the pension window, Labor partly backtracked on the announced policy but said nothing about AIG, so the effect of franking credits on returns should be in the opposite direction while the effect of AIG should be zero. In the election window, Labor lost the election, so the effect of the franking credit policy and the effect of AIG should both have the opposite signs to the announcement window.

3.5.1 Contribution of Franking Credits to Returns in the Announcement Window

The simplest approach to testing if Labor's policy reduced share prices would be simply estimate the contribution of franking credits to returns, $(r_i^t)^{\text{franking}}$, and test if it is negative or not. However, it is difficult to estimate this contribution given the many influences on returns. The approach taken in this paper is to make a prediction about how $(r_i^t)^{\text{franking}}$ varies across companies, and test if this prediction holds. This subsection argues that, if Labor's policy reduced share prices, then the contribution of franking credits to returns will be more negative for companies expected to have high franking percentages on future dividend payments. I first show that this is the case if share prices are determined by 'the' marginal investor. I then show that this is also the case if share prices are determined by a weighted average of investors, provided we are willing to make an additional assumption about the weights.

3.5.1.1 Argument if share prices are determined by a marginal investor

Suppose that share prices are determined by 'the' marginal investor. If Labor's policy did not affect share prices, then the marginal investor must have been a foreigner or an Australian who did not use excess credits. In contrast, if Labor's policy did affect share prices, then the marginal investor must have been an Australian who used excess credits. Hence, if Labor's policy reduced share prices, then the price of a share should equal the present value of the expected after-tax dividends that the Australian investor with excess credits would have received. This implies that the change in the price of a share in dollars should equal the change in the present value of expected after-tax dividends in dollars. This implies that the change in share prices due to franking credit policy must equal:

$$\left(\Delta p_i^t\right)^{\text{franking}} = -\mathbb{P} \times \mathcal{F}_i^t \text{ for } i = 1, ..., n \tag{3.3}$$

where:

(a) $\mathbb{P}$ is the probability of Labor's policy being implemented.

(b) $\mathcal{F}_i^t$ is the change in dollars in the expected present value of franking credits distributed per share.

The probability $\mathbb{P}$ is the same for all companies. Hence, the share price change for a company is proportional to the expected present value of its future franking credit distributions.

$$\left(\Delta p_i^t\right)^{\text{franking}} \propto -\mathcal{F}_i^t \text{ for } i = 1, ..., n \tag{3.4}$$

Dividing through by the price level at the start of the event window, we see that the contribution of franking policy to returns will be proportional to minus the expected present value of credits divided by the price of the share.

$$\left(r_i^t\right)^{\text{franking}} \propto -\frac{\mathcal{F}_i^t}{p_i^{t-h}} \text{ for } i = 1, ..., n \tag{3.5}$$

On the right-hand-side, the numerator is the expectation on day t of the present value of franking credits, and the denominator is the expectation on day $t-h$ of the present value of after-tax dividends. This is a measure of the extent to which the company attaches franking credits to its dividends.

To aid interpretation, it's useful to note that this ratio is increasing in the 'franking percentage' of each future dividend, as shown in Appendix C.3. The 'franking percentage' of a dividend is the amount of franking credits attached as a percentage of the maximum amount. Since a company is allowed to attach $\frac{\tau}{1-\tau}$ of franking credits for every dollar of cash dividend, where τ is the statutory corporate tax rate, the franking percentage is:

$$\text{Franking percentage} = \frac{\text{Franking credits attached}}{\left(\frac{\tau}{1-\tau}\right) \times (\text{Dividend amount})} \times 100\% \tag{3.6}$$

To summarise, if Labor's announcement reduced share prices, then the percentage change in share prices should be proportional to minus the ratio $\frac{\mathcal{F}_i^t}{p_i^{t-1}}$, which is a measure of the franking percentage on all future dividend payments by the company.

3.5.1.2 Argument if share prices are determined by many investors

Suppose that share prices are determined by all investors to some extent, rather than being determined solely by some specific marginal investor. In this case, a company's share price will equal a weighted average of the expected after-tax dividends received by its investors. This implies that the change in a company's share price due to franking credits, $(\Delta p_i^t)^{\text{franking}}$, would equal:

$$\left(\Delta p_i^t\right)^{\text{franking}} = -\mathbb{P} \times w_i \times \mathcal{F}_i^t \text{ for } i = 1, ..., n \tag{3.7}$$

where $\mathbb{P}$ and $\mathcal{F}_i^t$ are defined as before, and w_i is the weight on Australian shareholders with excess credits in determining share prices.

This shows that whether Labor's policy reduced share prices depends on the weight on Australians with excess credits. The weight would be small if each investor is weighted by their share of world wealth, as in Brennan (1970). However, it could be large if each investor is weighted by their ownership of the share in question. Although Australian shareholders make up a small share of world wealth, they own around two-thirds of Australian shares (Scott et al. 2017). This suggests that Australians with excess credits (typically self-managed superannuation funds and retired individuals) own a reasonable fraction of Australian shares.

As before, we note that since the probability is common to all companies, we can remove it:

$$\left(\Delta p_i^t\right)^{\text{franking}} \propto -w_i \times \mathcal{F}_i^t \text{ for } i = 1, ..., n \tag{3.8}$$

If Labor's policy reduced share prices then the weight on Australians with excess credits in determining share prices must be non-trivial. However, to make a prediction about how the contribution of franking policy to share prices varied across firms, we need to make more specific assumptions about the weights.

One possibility is that the weights w_i are the same for all companies i. For example, Australians with excess credits may have a $\frac{1}{10}$ weight in determining the share price of each Australian company. Then the weights in equation (3.8) can be removed in the same way as the probability of the policy being implemented. This gives equation (3.5).

Another possibility is that the weights w_i are higher for firms that are expected to distribute more franking credits. One reason to expect this positive relationship between weights and franking credit distributions is that Australians are likely to choose to allocate more of their savings to companies expected to distribute the most franking credits, as they benefit from franking credits while foreigners do not.[18] This suggests that companies with the highest expected franking credit distributions $\mathcal{F}_i^t$ will tend to have more Australian ownership, suggesting the weight w_i of Australians in determining their share price is higher. In this case, equation (3.8) implies that the contribution of franking policy to share prices, $(\Delta p_i^t)^{\text{franking}}$, must be strictly decreasing in $\mathcal{F}_i^t$. Dividing through by the initial share price shows that the contribution of franking policy to returns, $(\Delta p_i^t)^{\text{franking}}$, is strictly decreasing in the ratio $\frac{\mathcal{F}_i^t}{p_i^{t-1}}$. This is similar to equation (3.5), except now one can only conclude that the relationship is negative, rather that knowing that it is linear.

To summarise, if Labor's franking policy reduced share prices, then the contribution of that policy to returns during the event window must be more negative for companies with a high ratio of expected future franking credit distributions to their share price. This ratio can be interpreted as a measure of the franking percentage on future dividends, as shown in Appendix C.3.

3.5.1.3 Franking Proxies

The previous subsection argued that, if Labor's announcement reduced share prices, then the reduction in share prices will be larger for companies expected to have high franking percentages on future dividend payments. Unfortunately, we do not directly observe investor's expectations of future franking percentages. However,

[18]There is empirical evidence that Australians overweight shares with high dividend yields (Grant et al. 2018), which is likely motivated by a desire to receive more franking credits.

it is possible to find reasonable proxies, which will enable the empirical work of the sections that follow.

Future franking percentages are determined by the amount of franking credits received by the company relative to its dividends, which will in turn depends on the amount of Australian company tax paid relative to total profits. At the time of the announcement, however, investors did not have access to useful data on Australian company tax paid.[19] However, they had access to data on a variety of other variables that are useful in forecasting future franking percentages. Two of the most useful variables are:

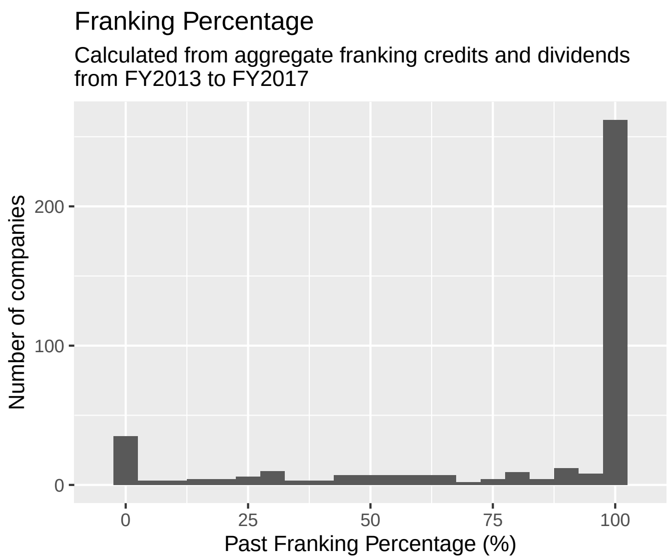
- The company's franking percentage over 5 years leading up to Labor's announcement.
- The company's franking account balance as a share of their market capitalisation

This chapter uses each of these variables as a proxy for future franking percentages. Hence, if Labor's announcement reduced share prices, then the causal effect of the announcement on cumulative returns over the window should be negatively related to each of these proxies.

The first proxy is a company's past franking percentage over the 5 years leading up to Labor's announcement. The proxy was calculated by applying the formula for the franking percentage on an individual dividend to the aggregate franking credits and aggregate dividend amounts over the past 5 years (equation (3.9)). Using data for multiple years has two potential benefits. First, it increases the sample

[19] The financial statements of listed companies must disclose company tax paid worldwide, but need not disclose the amount paid in Australia specifically. A voluntary tax transparency code was introduced in 2016, which has resulted in many companies disclosing their Australian company tax paid, but at the time of the announcement only one year of data would have been available to investors. A single observation of Australian company tax paid is not very useful for forecasting future franking percentages given the volatility in the series. For example, Qantas paid zero Australian company tax in 2016-17 due to losses carried forward, but its expected future franking percentage is likely quite high, given its track record of choosing positive franking percentages. The Australian Tax Office published data on the Australian company tax paid of the largest corporate tax entities. Many listed companies will fall under the threshold for inclusion in this dataset. Additionally, many companies comprise a set of related corporate tax entities that will be listed in the dataset separately, and it is labour intensive to aggregate the entity-level data into company-level data.

Figure 3.7



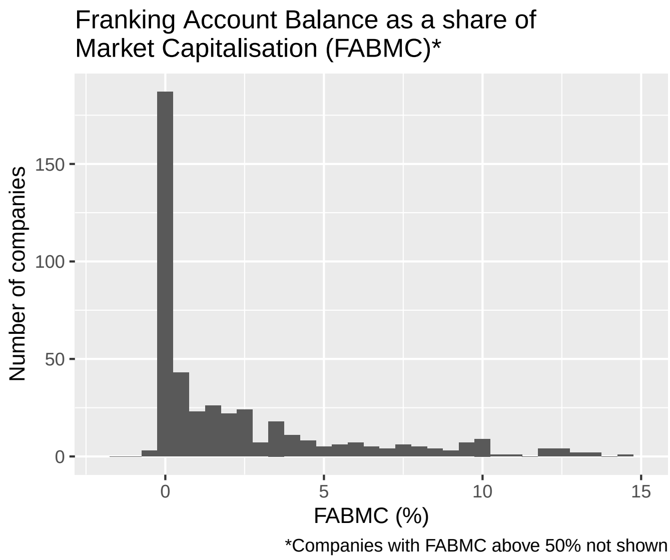
size, as it allows for a company to be included even if it hasn't paid a dividend in the past year, as long as it has paid one in the past 5 years. Second, using data for multiple years smooths out the year-to-year fluctuations, which may make it a better forecast of future franking credit distributions.

$$\text{Franking percentage proxy} = \frac{\text{Sum of franking credits distributed in past 5 years}}{\left(\frac{\tau}{1-\tau}\right) \times (\text{Sum of dividend amounts in past 5 years})} \times 100\% \quad (3.9)$$

Most companies pay a substantial amount of Australian company tax relative to their profits, so their past franking percentage is 100% (Figure 3.7). Some companies pay little or no Australian company tax, so their past franking percentage is 0%. Around a quarter of companies have a past franking percentage between these extremes.

Empirical evidence supports the use of past franking percentages as a proxy for future franking percentages. Among Australian listed companies, the franking percentage in FY2007 to FY2009 has a 0.74 correlation with their franking percentage in FY2015 to FY2017. Notably, many companies remain at the extremes. 52% of companies had a franking percentage of 100% in the earlier and later period, and 14% of companies had a franking percentage of 0% in both the earlier and later period.

Figure 3.8



The second proxy is the company's Franking Account Balance (FAB). When a company pays Australian company tax, it receives franking credits, which adds to its FAB. When it attaches franking credits to a dividend, this subtracts from its FAB. The level of a company's FAB tracks the amount of franking credits received but not yet distributed. If a company has a high franking account balance as a share of its market capitalisation (FABMC), then it is likely to be able to choose a high franking percentage on its future dividends. While most companies have distributed all the credits they have received, some have accumulated substantial balances (Figure 3.8).

Unfortunately, time series data on FAB is not available, so it is not possible to empirically test if FAB at a point in time is highly correlated with future FAB.

3.5.2 The Effect of AIG on Returns in the Announcement Window

Labor announced the AIG on the same day as it announced its franking credit policy. The AIG would have increased the value of the depreciation allowances companies can claim on capital expenditure ('capex') for the purposes of company tax. Specifically, the AIG would allow companies to immediately deduct 20% of any qualifying new asset worth \$20,000+ , with the remainder depreciated as usual. This policy increases the depreciation allowances

The contribution of AIG to a company's share price equals:

$$\left(\Delta p_i^t\right)^{\text{AIG}} = \mathbb{Q} \times \frac{\mathcal{A}_i}{s_i^{t-h}} \tag{3.10}$$

where:

- $\mathbb{Q}$ is the probability of Labor implementing the policy.
- $\mathcal{A}_i$ is the change due to AIG in the present value of depreciation allowances on a company's future capex
- s_i^{t-h} is the number of shares outstanding

Dividing through by the initial price shows that the contribution of AIG to returns equals:

$$\left(r_i^t\right)^{\text{AIG}} = \mathbb{Q} \times \frac{\mathcal{A}_i}{s_i^{t-h} p_i^{t-h}} \tag{3.11}$$

Let V_i^t denote a company's market capitalisation, which equals $s_{i,t} p_i^t$ by definition. Substituting this into equation (3.11) gives:

$$\left(r_i^t\right)^{\text{AIG}} = \mathbb{Q} \times \frac{\mathcal{A}_i}{V_i^{t-h}} \tag{3.12}$$

The probability of the policy being implemented is the same for all companies. Hence we can remove this probability and replace equality with proportionality.

$$\left(r_i^t\right)^{\text{AIG}} \propto \frac{\mathcal{A}_i}{V_i^{t-h}} \tag{3.13}$$

Suppose that the change in the present value of depreciation allowances on a company's future capex, $\mathcal{A}_i$, is proportional to the present value of the company's future capex, C_i. This assumption is likely to be approximately true, though it not exactly true as it ignores variation among companies in the proportion of capex that is eligible for the AIG. Under this assumption, the contribution of AIG to returns must be proportional to the ratio of a company's future capex to their market capitalisation.

$$\left(r_i^t\right)^{\text{AIG}} \propto \frac{\mathcal{C}_i}{V_i^{t-h}} \tag{3.14}$$

3.6 Data

3.6.1 The Panel Dataset

This chapter uses a panel dataset on listed companies in Australia. The sample period covers all trading days from 1 January 2018 to 30 June 2019, which is an 18 month period. Other event studies of dividend tax changes have used shorter sample periods, such as 12 months in Gadarowski et al. (2007) and Amromin et al. (2008), while others use longer periods, such as 21 months in Auerbach and Hassett (2005). The sample period in this chapter could not be made much shorter, as it must cover all three event windows. When results were produced for a longer sample period of 48 months, the results were little changed.

The sample comprises all companies listed on the Australian Securities Exchange (ASX), except those removed by data cleaning. For the majority of companies in the sample, their share price is the price of their ordinary shares, though for a minority, their share price is actually the price of their 'stapled securities'.[20] The sample does not include Chess Depositary Instruments[21], real estate investment trusts, or preference shares.

Data cleaning is performed in two steps. The first step of data cleaning kept only those shares that had at least 50 trading days of price data and whose price was above 10 cents on average.[22] The second step removed companies that lacked data on the proxy used in the specification being estimated.

- For past franking percentages, this results in 407 tickers.

[20] Stapled securities are sets of securities that must be sold together. For example, Transurban is a large Australian toll road operator. It issues a stapled security comprising shares in Transurban Holdings Limited, shares in Transurban International Limited, and units in Transurban Holding Trust. Stapled securities are comparable to shares for the purposes of this paper, and in particular, can distribute franking credits.

[21] For example, Resmed is a large medical devices company whose shares are traded on the New York Stock Exchange, but which also has Chess Depositary Instruments listed on the ASX. The share price of a company like Resmed is likely to be determined by foreign investors since it is listed abroad, so the behaviour of its share price provides little evidence on whether share prices in Australia are determined by Australians or foreigners. Additionally, since these instruments represent ownership in foreign companies, and only Australian companies can accumulate franking credits, they cannot distribute franking credits.

[22] Bloomberg rounds prices to the nearest cent, so if the price is low, then this rounding will introduce a material amount of error into the measured price.

- For FABMC, this resulted in 501 tickers. Of the tickers with FABMC, a further two were removed due to having FABMC above 50%.[23]

It is surprising how many companies lack data on past franking percentages or FAB, given that each ASX-listed company is required to disclose this information. One explanation is that the disclosures are not done in a standard way. For example, while all companies disclose their FAB in their annual report, some disclose it in tables while others use full sentences. It is possible that the maintainers of financial databases found it impractical to extract this information for all companies (see appendix C.4).

The panel dataset contains the following variables:

- **Share prices of each company:** I use daily closing prices. I extract a company's share price from Eikon if possible, and from Bloomberg otherwise. Eikon provides unrounded share prices, which are more preferable to the rounded share prices provided by Bloomberg.

- **Other share market data on each company:** This includes closing prices, trading volumes, and market capitalisation, which is available for each trading day. These are obtained from Bloomberg.

- **Financial statement data on each company:** This includes net fixed assets, capex, which is available semi-annually for most companies. The panel data models use the latest observation that would have been known to investors at the time of Labor's announcement. These are obtained from Bloomberg.

- **Global Industry Classification System (GICS) codes:** The GICS code of a company classifies it at four levels of aggregation: sub-industries, industries, industry groups, and sectors. I use the industry codes to compute cluster-robust standard errors, and I include dummies based on the sector codes in some of the specifications. These are obtained from Bloomberg.

[23] Recall companies with high FABMC are able to choose a high franking percentage on future dividends. However, at extreme levels of FABMC, the company is guaranteed to be able to choose a franking percentage of 100% on all dividends, so further FABMC is not informative regarding future franking percentages. If companies with extreme FABMC were included, it would bias the regressions towards finding no relationship between returns and FABMC. For this reason, the two companies with extreme FABMC are removed from the sample.

- **Past Franking Percentage:** This was calculated by dividing the total franking credits distributed by that company in the 5 years leading up to the announcement. This calculation was performed using annual data on franking credits distributed and dividends paid from the Eikon database and the S&P Capital IQ database.

- **Franking account balance:** This was difficult to obtain. This data was not available from the Eikon or S&P Capital IQ databases. Annual FAB data was available in Bloomberg and Morningstar, but there were substantial discrepancies between these sources for many companies. To investigate these discrepancies, I manually collected annual FAB data from the published financial statements of 42 listed companies, which revealed substantial errors in Bloomberg and smaller errors in Morningstar. Due to these issues, the panel dataset contains the manually collected data for the 42 companies, Morningstar data for most other companies, and Bloomberg data only if the other two data sources are unavailable. The panel data models are estimated using the FAB observation from the latest annual report that had been published by the announcement date. The annual report publication dates were obtained from company press releases. The construction of the panel dataset, including the manual collection of FAB data, is described in detail in Appendix C.4.[24]

- **Constituents of ASX Indexes:** A list of the companies that comprise the S&P/ASX 200, S&P/ASX 300 and S&P/ASX All Ordinaries index were obtained from S&P Capital IQ.[25]

[24]The difficulty involved in obtaining FAB data suggests some investors may have found it difficult to incorporate information about FAB into their trading decisions during the announcement window. However, investors do not need to explicitly rely on FAB data for it to be a suitable proxy. All that is required is that FAB be correlated with investor's expectations about future franking credit distributions. This could occur if, for instance, investors form their expectations about franking credit distributions based on the public statements of companies, and those companies take FAB into account when making those statements.

[25]Unfortunately, S&P does not make available the list of constituents of the index as of previous dates, so I used the list of constituents as at the time I accessed this data (23 April 2023). These lists of constituents will be slightly different, as companies are occasionally added to or removed from the list.

Figure 3.9

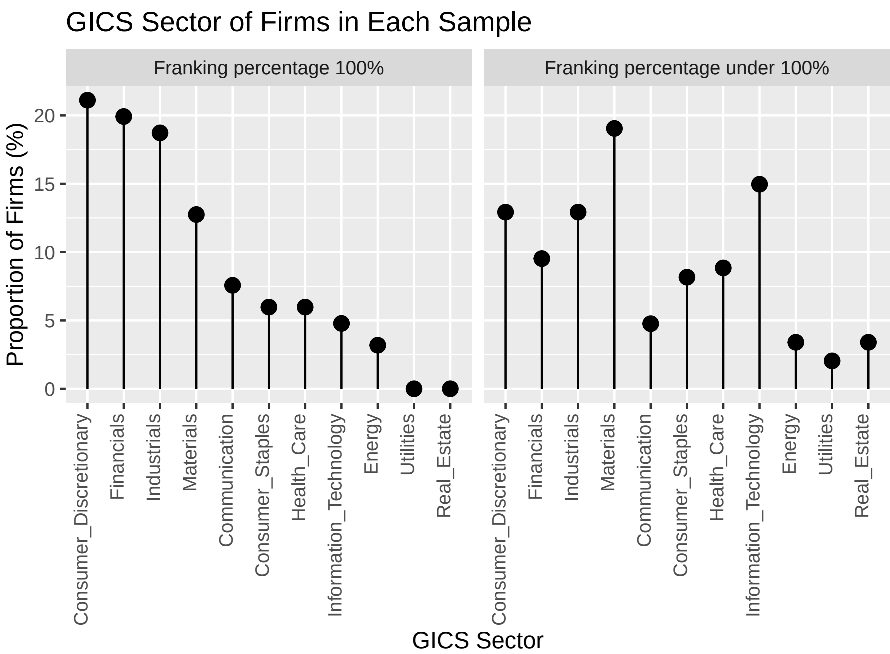


Table 3.2 presents descriptive statistics on the companies in the sample. The first few rows present statistics for the set of all companies in the dataset, even though data on the proxies was not available for all companies. The other rows contain statistics for sub-samples of companies formed based on each of the franking proxies. This table shows that companies with high franking percentages tend to have lower market capitalisation than other companies. On the other hand, it shows that companies with positive FAB tend to be considerably larger than companies with zero FAB.

It is also useful to compare the GICS sectors of companies in each sub-sample. companies with a franking percentage of 100% are particularly likely to be in the Financials or Consumer Discretionary sector, while companies with lower franking percentages are unusually likely to be in the Materials or Information Technology sectors (Figure 3.9).

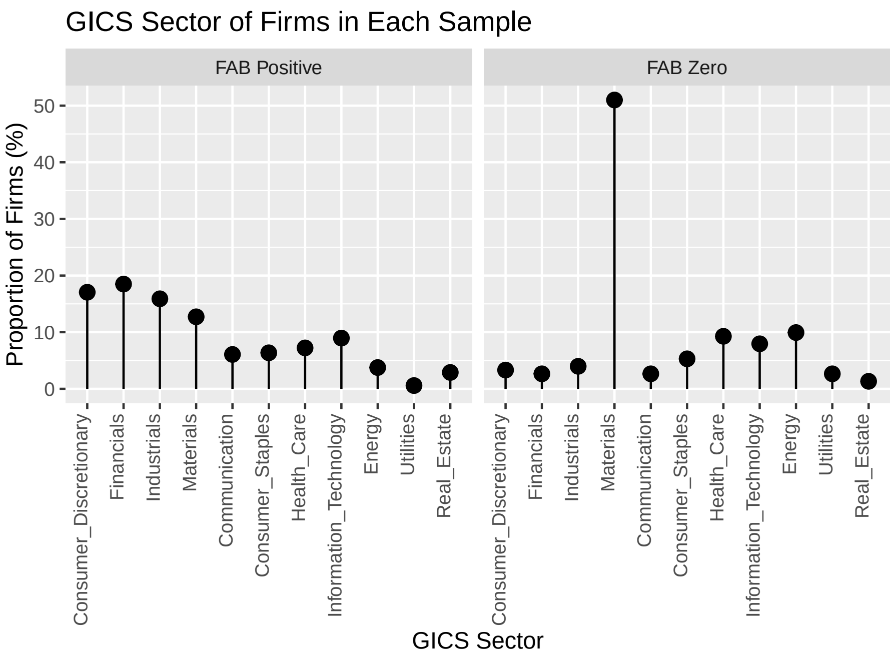
The difference between sub-samples formed on FABMC is much more stark (Figure 3.10). The sample of companies with positive FAB contains a mix of many different sectors. However, the sub-sample of companies with zero FAB is unusual in that almost half are in the Materials sector, of which most are in the mining industry.[26]

[26]It is not obvious why this is the case. One possible explanation is that many mining companies

Table 3.2: Descriptive Statistics on All Companies and on Sub-samples

	Mean	25th percentile	Median	75th percentile
All firms (N = 969)				
Market Cap ($m)	1865.8	36.7	113.4	466.1
Net Fixed Assets / Market Cap	28.8	0.8	6.1	31.3
Capex / Market Cap	2.6	0.1	0.7	2.3
Franking Percentage	79.4	64.7	100.0	100.0
FABMC	4.9	0.0	1.0	4.7
Franking percentage 100% (N = 259)				
Market Cap ($m)	3449.2	133.3	329.3	1431.1
Net Fixed Assets / Market Cap	25.9	2.0	10.0	32.3
Capex / Market Cap	2.2	0.2	1.0	2.3
Franking Percentage	100.0	100.0	100.0	100.0
FABMC	7.8	1.7	4.1	10.0
Franking percentage under 100% (N = 148)				
Market Cap ($m)	5386.3	106.8	572.9	3875.8
Net Fixed Assets / Market Cap	33.4	1.3	10.9	46.6
Capex / Market Cap	2.8	0.2	1.0	2.7
Franking Percentage	43.4	6.5	45.5	76.9
FABMC	2.9	0.0	0.3	1.3
Franking Percentage Missing (N = 564)				
Market Cap ($m)	220.5	24.2	55.8	156.2
Net Fixed Assets / Market Cap	29.0	0.3	3.1	23.2
Capex / Market Cap	2.7	0.0	0.3	2.1
FABMC	2.4	0.0	0.0	0.2
FAB Positive (N = 348)				
Market Cap ($m)	4071.5	123.1	392.9	1928.3
Net Fixed Assets / Market Cap	24.2	1.3	7.6	31.1
Capex / Market Cap	2.0	0.2	1.0	2.2
Franking Percentage	88.1	92.7	100.0	100.0
FABMC	7.0	0.8	2.6	8.2
FAB Zero (N = 153)				
Market Cap ($m)	462.8	22.1	73.7	215.6
Net Fixed Assets / Market Cap	31.7	1.1	9.0	33.8
Capex / Market Cap	2.8	0.1	0.7	2.7
Franking Percentage	48.8	0.0	48.1	99.0
FABMC	0.0	0.0	0.0	0.0
FAB Missing (N = 468)				
Market Cap ($m)	684.4	26.0	60.4	184.5
Net Fixed Assets / Market Cap	32.8	0.3	3.2	28.7
Capex / Market Cap	3.1	0.0	0.3	2.1
Franking Percentage	57.6	0.0	81.7	100.0

Figure 3.10



3.6.2 The Returns on Other Asset Classes

Daily nominal returns on a variety of asset classes were used as controls for other sources of variation in individual share prices. The returns on Australian share indexes were obtained from Eikon. The returns on other asset classes were obtained from Bloomberg.

In the panel dataset, the return on the shares of individual companies is calculated as the percentage change from market close to market close. To maximise the extent to which the returns on other asset classes are useful in explaining variation in Australian shares, I calculate the returns in other assets over a similar time interval. Due to differences in time zones, this requires calculating returns in some assets from close to close, and calculating the returns in other assets from open to open.

Table 3.4 lists the asset classes included, and the period over which their prices are measured.

operate at a loss for many years, in the hope of making a discovery and becoming extremely profitable in the future. These companies will typically have paid little or no company tax, and hence have accumulated zero FAB.

Table 3.4: Asset Classes Used as Controls

Asset Class	Variable Used	Currency	Time when Prices are Measured
Australian shares	ASX All Ordinaries Index	Australian dollars	Close
US shares	S&P 500 Index	US dollars	Open
Europe shares	S&P Europe 350 Index	Euros	Open
China shares	S&P China Broad Market Index	US dollars	Close
Japan shares	Nikkei 225 Index	Japanese Yen	Close
Crude oil	Brent futures contract	US dollars	Open
Metallurgical coal	SGX TSI Australia Premium Coking Coal Futures	US dollars	Close
Thermal coal	Thermal coal grade 5500 kcal/kg	Chinese Renminbi	Close
Iron ore	Iron Ore Australian Fines	US dollars	Close

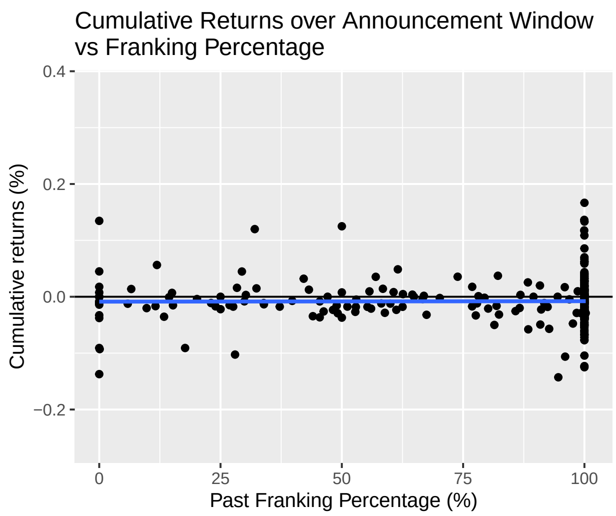
3.7 Method

3.7.1 Cross-sectional Relationship of Returns and Franking Proxies

To isolate the effect of the dividend imputation change, I focus on heterogeneity in returns across companies. Section 3.5 decomposed a company's returns during the announcement window into three components: normal returns; the contribution of franking policy; and the contribution of AIG (equation (3.2)). It showed that if Labor's franking policy reduced share prices, then the contribution of franking policy to a company's returns will be negatively related to the expected future franking percentages of that company. It also showed that the contribution of AIG should be positively related to a company's expected present value of a company's market capitalisation relative to its market capitalisation.

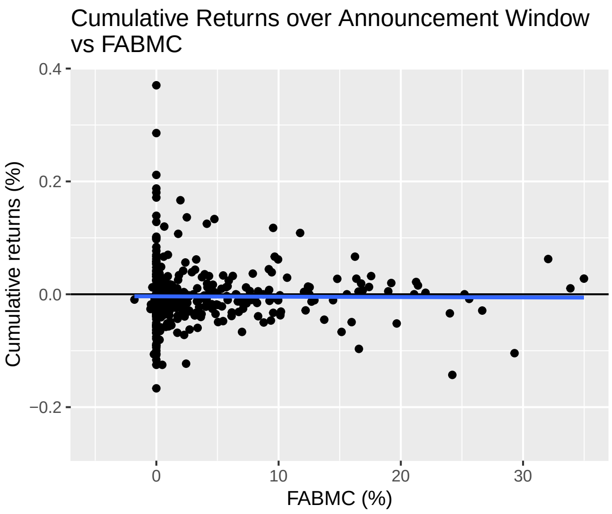
As a starting point, assume that observed returns during the event window, $r_{i,t}$ are linearly related to an intercept, a franking proxy and an error term. I estimated this regression separately for the first proxy (past franking percentage) and then for the second proxy (FABMC). The results provide no evidence of a negative relationship between cumulative returns and either proxy, as illustrated by figures 3.11 and 3.12.

Figure 3.11



Unfortunately, the simple approach suffers from two major limitations. The first is that the error and regressor may not be orthogonal, in which case ordinary least

Figure 3.12



squares (OLS) would not be consistent. The error term contains both normal returns and the contribution of AIG. Orthogonality will fail if the sum of normal returns and the contribution of AIG are correlated with the proxy. This could occur for many reasons. For example, the companies with high franking proxies may also have high capital expenditure on average, and hence have positive contributions of AIG, and hence a positive error term. The second limitation is that the error is likely to be correlated across companies due to factors that affect many companies simultaneously. For example, an increase in iron ore prices would increase the returns of many different mining companies. The usual formulas for standard errors assume that the errors for different observations are uncorrelated with each other.

3.7.2 Cross-sectional Relationship of Abnormal Returns and Proxies

This chapter uses a panel data method similar to that used in Auerbach and Hassett (2005). Unlike the cross-sectional regression, the panel regressions will control for normal returns and the contribution of AIG. This provides better estimates of how the contribution of the franking policy to returns varies with the franking proxies, and hence a cleaner test of whether Labor's policy reduced share prices.

To illustrate the method, suppose there is a single event window that comprises a single event date t^*. Consider the following 'CAPM-inspired' regression:[27]

$$r_{i,t} = \mu_i + m_t\beta_i + D_t\eta_i + v_{i,t} \quad \text{for all } i, t$$

where:

- $r_{i,t}$ is the return on company i on day t
- μ_i is a company i fixed effect
- m_t is the return on the share market as a whole
- β_i is the coefficient on the market return
- D_t is a dummy that equals 1 on the event date and 0 otherwise
- η_i is the coefficient on the dummy
- $v_{i,t}$ is an error term

To interpret the coefficient on the dummy, rearrange the CAPM-inspired specification and set the date to the event date t^*. This shows that, for each company i, the coefficient η_i is the return on the event date r_{i,t^*} less the company fixed effect, the response of the company's share price to the overall market return on that date, and the error.

$$\eta_i = r_{i,t^*} - (\mu_i + m_{t^*}\beta_i) + v_{i,t} \quad \text{for all } i \tag{3.15}$$

If the event has a negligible effect on the market return, then $(\mu_i + m_t\beta_i + v_{i,t})$ is the return the company would have earned absent the event, so it is called the 'normal return'. This implies that the coefficient η_i is the difference between the expected return of the company with the event and the normal return, so it is called the 'abnormal' return.

The franking credits announcement could affect a large number of companies, so might have affected the market return. Hence if Labor's announcement reduced share prices, it will have had two offsetting effects on each company's abnormal returns.

[27]This model does not include the risk-free rate. It is therefore closer to the Black CAPM, which does not have any special role for the risk-free rate, rather than the Sharpe-Litner CAPM, where the risk-free rate appears on both sides of the equation (Fama and French 2004).

1. The announcement directly reduces actual returns r_{i,t^*}, reducing abnormal returns

2. The announcement reduces the market return m_{t^*}, reducing the normal return $(\mu_i + m_{t^*}\beta_i + v_{i,t})$, raising abnormal returns.

Due to these offsetting effects, the average level of abnormal returns is not informative about the effect of the event. In fact, it could be zero even if the event substantially reduced all company's share prices. However, the variation in abnormal returns across companies is still informative regarding which companies were more or less affected.

To illustrate, suppose an event affects companies representing half of market capitalisation but does not affect other companies. If the event causes actual returns of affected companies to be 10 percentage points (ppt) lower, then it will cause the market return to be 5 ppt lower. If each company's β_i is 1, then their normal return changes one-for-one with the market return. Hence, the affected companies will experience a 10 ppt reduction in actual returns, a 5 ppt increase in normal returns, so abnormal returns will be minus 5 ppt. The unaffected companies will experience no change in actual returns, a 5 ppt increase in normal returns, so will have abnormal returns of plus 5 ppt. The cross-company average of abnormal returns will be zero, but the abnormal returns will still be informative about which companies were more or less affected.

To formalise this, assume the abnormal return η_i for a company is linearly related to the company's proxy w_i. The coefficient γ, is the abnormal return of a company for whom the company characteristic is zero. The coefficient λ measures how much higher (or lower) the abnormal return is when the company characteristic is one unit higher. To test if Labor's announcement reduced share prices, we test if the slope λ is zero (since cross-company variation in abnormal returns is informative), but do not test the intercept (since the level of abnormal returns is not informative).

$$\eta_i = \gamma + w_i\lambda \quad \text{for all i}$$

Substituting this assumed relationship between abnormal returns and the franking proxy into equation (3.15) gives the following regression specification. Our method is to estimate regressions like this one and test the sign of λ. Section 3.8 will present results where the proxy is the past franking percentage and results where the proxy is FABMC.

$$r_{i,t} = \mu_i + m_t\beta_i + D_t\gamma + D_t w_i\lambda + v_{i,t}$$

3.7.3 Specifications

The baseline specification, 3.16, used in this chapter is similar to the equation above, except the announcement window lasts for 3 days, so there is one dummy for each day, D_t^1, D_t^2, D_t^3.

The company characteristics specification, 3.17, builds on the baseline specification by adding interaction terms of a vector of other company characteristics, z_i, and the announcement dummies.

Finally, the full specification, 3.18, extends the company characteristics specification by adding a vector of returns on other asset classes, q_t.

$$r_{i,t} = \mu_i + m_t\beta_i + \sum_{j=1}^{3} D_t^j\gamma^j + \sum_{j=1}^{3} D_t^j w_i\lambda^j + v_{i,t} \tag{3.16}$$

$$r_{i,t} = \mu_i + m_t\beta_i + \sum_{j=1}^{3} D_t^j\gamma^j + \sum_{j=1}^{3} D_t^j w_i\lambda^j + \sum_{j=1}^{3} D_t^j z_i\delta + v_{i,t} \tag{3.17}$$

$$r_{i,t} = \mu_i + m_t\beta_i + \sum_{j=1}^{3} D_t^j\gamma^j + \sum_{j=1}^{3} D_t^j w_i\lambda^j + \sum_{j=1}^{3} D_t^j z_i\delta + q_t'\eta_g + v_{i,t} \tag{3.18}$$

The purpose of the vector of company characteristics z_i is to reduce the risk of omitted variable bias due to the AIG.

Section 3.5 showed that the contribution of AIG on a company's returns is proportional to the ratio of the expected present value of future capex to market capitalisation (equation (3.14)). We include three proxies for this ratio:

- **Capex as a share of market capitalisation**. I use the last annual observation known on the date Labor's franking policy was announced

- **Net fixed assets as a share of market capitalisation**. The amount of net fixed assets depends on the cumulated amount of past capex, so it may be correlated with future capex and hence the benefit of AIG. I also use the last annual observation known as of the franking policy announcement date

- **Dummies for each of the 11 GICS sectors**. These dummies could proxy for differences in the extent to which different industries are expected to invest in assets that qualify for AIG.

The inclusion of these company characteristics may also help mitigate omitted variable bias arising from other sources. For example, a number of news stories about misconduct in the banking industry were published during the announcement window (see Appendix C.2). The industry dummies should control for the returns of banks being lower due to these news stories, which stops these lower returns being erroneously attributed to the tendency of Australian banks to have large FAB.

The purpose of including the returns on other asset classes q_t is to control for common sources of variation in the returns of many different shares (see section 3.6.2) for a list of assets). If this is not done, the errors of companies in different industries may be correlated, in which case the standard errors may not be appropriate, leading to incorrect inferences. The vector of coefficients on these other assets, η_g, is allowed to differ by industry. This allows for the possibility that, for instance, an increase in crude oil prices raises the returns of companies in the materials sector companies but reduces the returns of companies in the consumer discretionary sector.

3.7.4 Estimation and Inference

3.7.4.1 Estimation of the Coefficients

The assumptions made about the panel data models inform the chosen methods for estimating the coefficients and computing the standard errors. Assume:

- Strict exogeneity. i.e. The conditional expectation of the error given the regressors is zero.

- Errors are clustered at the industry level. This allows for heteroscedasticity, serial correlation of the error for a given company, and correlation between the errors of different companies in the same industry.

- Fixed effects[28]

All three specifications were estimated by the within groups estimator. That is, each specification was transformed by expressing variables as deviations from their company-specific means, and then the transformed specification was estimated by OLS.

The model contains individual-invariant regressors m_t, D_t, but does not contain time fixed effects. A model with both individual-invariant regressors and time effects would have perfect multi-collinearity problems (Biorn 2016).

3.7.4.2 Computation of the Standard Errors

Cluster-robust standard errors are computed, with clusters defined at the level of the 53 GICS industries. The GICS classifies companies at varying level of aggregation, ranging from sub-industries up to sectors, meaning that larger or smaller cluster sizes are possible. Choosing larger clusters has both benefits and costs:

- The benefit of allowing for larger clusters is that it represents a weaker assumption about the error covariance matrix. In economic terms, it allows for the possibility that companies that are more distinct from each other have correlated errors. For example, errors clustered at the GICS sub-industry level allow for errors to be correlated across companies in 'Oil & Gas Drilling' sub-industry, while errors clustered at the GICS industry level allow for errors to be correlated across the 'Energy Equipment & Services' industry, which is broader.

[28] As there are many time periods in the sample, T, the loss of degrees of freedom due to having to estimate N fixed effects is small relative to the total sample size NT. This loss of degrees of freedom is a price worth paying to avoid making the restrictive assumption that the individual-specific effects $\mu_1, \ldots, \mu_N$ are uncorrelated with the regressors. As a robustness check, the baseline specification was also estimated with random effects. This did not change the key finding, which is that the cumulative abnormal returns do not have a statistically significant relationship with FABMC or past franking percentage.

- The cost of larger clusters is that it reduces the number of clusters. My chosen method of computing the standard errors is consistent as the number of clusters approaches infinity. If the number of clusters is too small, the standard errors may be very imprecise.

It is not easy to trade off these considerations. Cameron and Miller (2015) say that one reasonable approach is to compute the standard errors on progressively more aggregated levels, and then stop when the standard errors stop changing substantially. When I performed this exercise, the standard errors changed noticeably as I moved from clustering at the company level to the GICS sub-industry level, and again as I moved to clustering at the GICS industry level. However, the standard errors were little affected by clustering at higher levels of aggregation. Hence, I have decided to compute standard errors clustered at the level of the GICS industries. The sample used in this chapter contains data on companies in 53 GICS industries.

Analysis on 'few' clusters can be a source of econometric difficulties. The first issue is the computation of the standard errors. The 'uncorrected' formula for cluster-robust standard errors tends to be downwardly biased when there are few clusters. To remedy this, I used the formula with the finite-sample correction in equation (12) of Cameron and Miller (2015). The second issue relates to hypothesis testing. Whenever I test the hypothesis that an individual coefficient is zero, I assume the Wald statistic has a student t distribution with $(G-1) = (53-1)$ degrees of freedom. If I had instead used a student t distribution with degrees of freedom based on the number of observations, the distribution would have had thinner tails, leading to understated p-values. Although the methods in this chapter were implemented in R, both of these choices mimic the defaults in Stata.

3.7.4.3 Cumulative Abnormal Returns

To assess the total effect of the announcement on share prices over the announcement window, one should calculate cumulative abnormal return (CARs). The CAR from day 1 to j is the sum of the abnormal returns from those days. Define:

- $\Gamma^1 = \gamma^1$, $\Gamma^2 = \gamma^1 + \gamma^2$ and $\Gamma^3 = \gamma^1 + \gamma^2 + \gamma^3$. This is the CAR of a company for which the proxy is zero.

- $\Lambda^1 = \lambda^1$, $\Lambda^2 = \lambda^1 + \lambda^2$ and $\Lambda^3 = \lambda^1 + \lambda^2 + \lambda^3$ for each j. This is the increase in CAR associated with a 1 percentage point increase in the proxy.

The point estimates of Λ^j can be given by adding the point estimates of λ^j. To compute the standard errors one must account not only for the variances of each λ^j, but also the correlations between them. Salinger (1992) suggests rearranging the regression equation before estimation, so that when the rearranged equation is estimated, it provides estimates of Λ (along with appropriate standard errors) rather than λ. The results in this chapter were produced by rearranging the specifications in this way.

3.8 Results

If the announcement reduced share prices, then CARs should be negatively related to the proxies for future franking percentage. To test this, the three specifications were estimated for the first proxy, past franking percentage, and then separately for the second proxy, FABMC. We also estimate specifications with both proxies included simultaneously. All of these specification contain hundreds of coefficients, so I only report the coefficients on the interaction of CARs and the proxy $\Lambda^1, \Lambda^2, \Lambda^3$. The focus is on Λ^3, which is the increase in the CAR over the 3-day event window associated with a 1 percentage point increase in the proxy. The results do not provide any evidence that these coefficients are negative, and hence do not provide evidence that Labor's announcement reduced share prices. Appendix C.5 present a variety of robustness checks.

3.8.1 Results using Past Franking Percentage as a Proxy

Across all three specifications, the point estimates of each of the coefficients of $\Lambda^1, \Lambda^2, \Lambda^3$ were close to zero (Table 3.5). Moreover, none of the coefficients were different from zero at a 5% level of significance. These results do not provide any evidence that a company's CARs were negatively related to a company's past franking percentage over the announcement window. Hence they also do not any provide evidence for the alternative hypothesis that dividend taxes reduce share prices.

The confidence intervals are consistent with the policy having zero effect or a small effect in either direction. For example, the 95% confidence interval for Λ^3 in the full specification is $[-0.009, 0.033]$. The lower bound estimate of $\Lambda^3 = -0.009$ suggests that, if a company had a franking percentage of 100% rather than 0%, then its returns would have been 0.9 percentage points lower during the window. However, these methods do make it unlikely that Australians with excess credits determine share prices in proportion to their ownership of their shares. They even more clearly rule out that Australians with excess credits are 'the' marginal investor, as the share price of a company with 100% franking percentage would be expected to fall around 17.1% in this case (see appendix C.3.3).

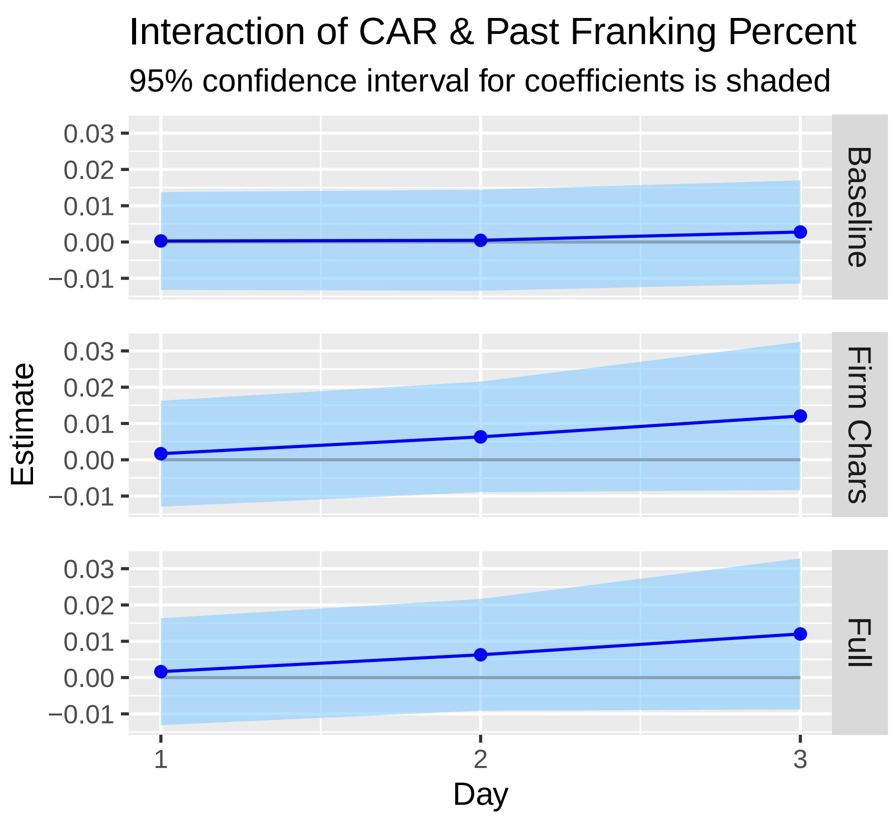
Table 3.5

Estimated Relationship between Cumulative Abnormal Returns and Past Franking Percent

	Baseline		Firm Characteristics		Full	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Day 1	0	0.007	0.002	0.007	0.002	0.007
Day 2	0	0.007	0.006	0.008	0.006	0.008
Day 3	0.003	0.007	0.012	0.010	0.012	0.010

Note: * is significant at 5%, ** is significant at 1%, *** is significant at 0.1%

Figure 3.13



3.8.2 Results using FABMC as a Proxy

The previous section did not find any evidence that CARs were negatively related to the first proxy, past franking percentages. This section shows that there is no evidence that CARs are negatively related to FABMC either. In two of these specifications, the coefficient on the interaction of the CAR of day 1 and FABMC is positive and significant at the 5% level, though this positive relationship disappears from day 2 onwards. I do not view this as evidence that Labor's announcement genuinely increased returns, as otherwise one would not expect the positive relationship to be unwound from day 2 onwards. In any case, these results do not provide any evidence of a negative relationship, and hence no evidence for the alternative hypothesis that dividend taxes reduce share prices.

Once again, the confidence intervals are consistent with the policy having zero effect or a small effect in either direction. The 95% confidence interval for Λ^3 is $[-0.088, 0.094]$. If Λ^3 were equal to the lower estimate of -0.088, then a company whose FAB was 10% (which would be well above average, as shown in 3.2) would have a return that is 0.88 percentage points lower than otherwise than if their FAB were zero.

Table 3.7

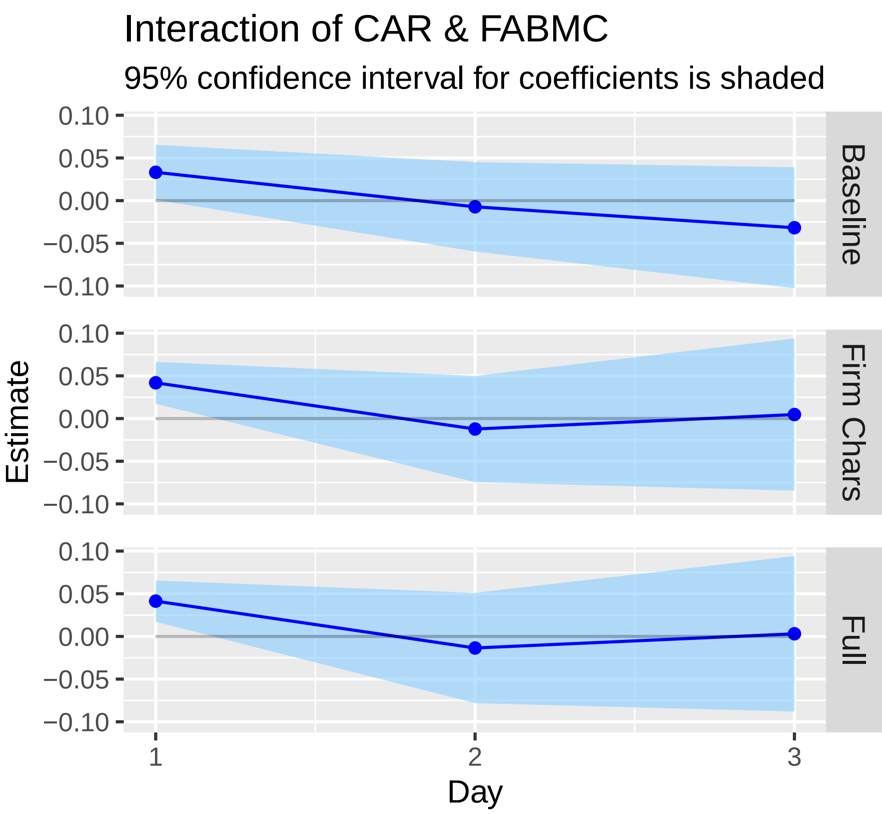
Estimated Relationship between Cumulative Abnormal Returns and FABMC

	Baseline		Firm Characteristics		Full	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Day 1	0.033*	0.016	0.042*	0.012	0.041*	0.012
Day 2	-0.007	0.026	-0.012	0.031	-0.014	0.032
Day 3	-0.032	0.035	0.005	0.044	0.003	0.045

Note: * is significant at 5%, ** is significant at 1%, *** is significant at 0.1%

.

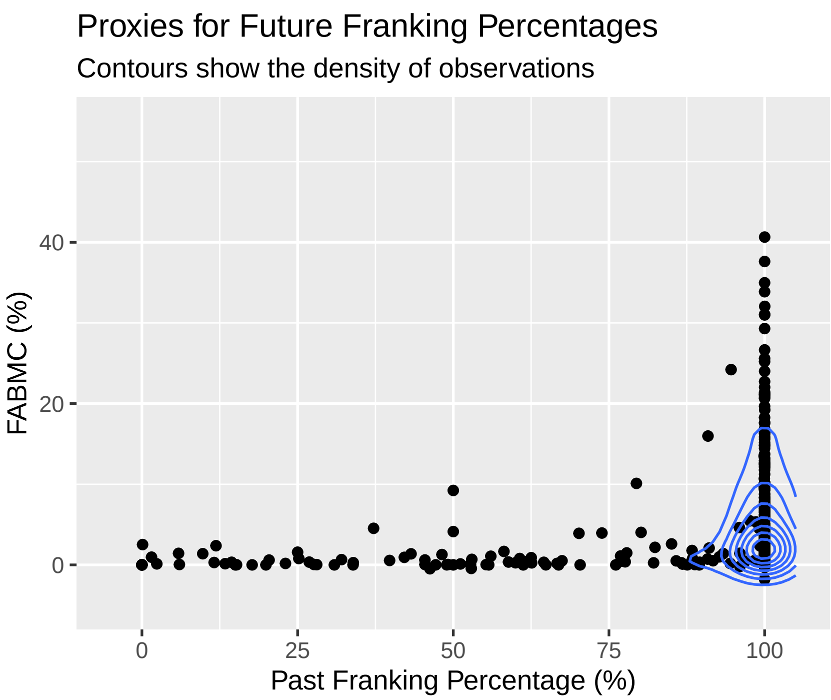
Figure 3.14



3.8.3 Results using Both Proxies

Section 3.5 showed that the two franking proxies are each likely to provide useful information about future franking percentages. Until now, each proxy has been included separately. However, it's likely that the two proxies each contain useful information. To see why, one must understand how the proxies are related. Typically, when a company pays a dividend they choose a franking percentage that is as high as possible. This ensures that, for a given sequence of dividend payments, the company distributes its FAB as quickly as possible. Distributing franking credits quickly is desirable because the company does not earn any nominal returns on its FAB. Consistent with this, companies with a past franking percentage below 100% almost always have a zero or near-zero FABMC, while companies with a franking percentage of 100% vary greatly in their FABMC (Figure 3.15).

Figure 3.15



FABMC is likely to be useful in explaining variation in future franking percentages among companies whose past franking percentage is 100%. To illustrate, consider two companies:

- Company A pays Australian company tax on half of its worldwide profits, and pays out half of its profits as dividends. This company will have a franking percentage of 100% and FABMC of 0. If it then decides to pay out its retained earnings as dividends, it will not have any franking credits, so its future franking percentage will be 0%.

- Company B pays Australian company tax on all of its worldwide profits, and pays out half its profits as dividends. This company will have a franking percentage of 100% and a substantial FABMC. If it decides to pay out its retained earnings as dividends, it will have substantial franking credits to attach, so its future franking percentage will be high.

The above reasoning suggests estimating a regression with both proxies. Figure 3.16 shows the estimated slope coefficients for a regression with past franking percentage and FABMC as proxies. Since FABMC is likely to be informative only for those companies with a past franking percentage of 100%, figure 3.17 shows the estimated coefficients for a regression where the first proxy is past franking percentage and the second proxy is the interaction of a dummy for a 100% past franking percentage and FABMC. In both of these specifications, there is no statistically significant relationship between CARs and either proxy.

Figure 3.16

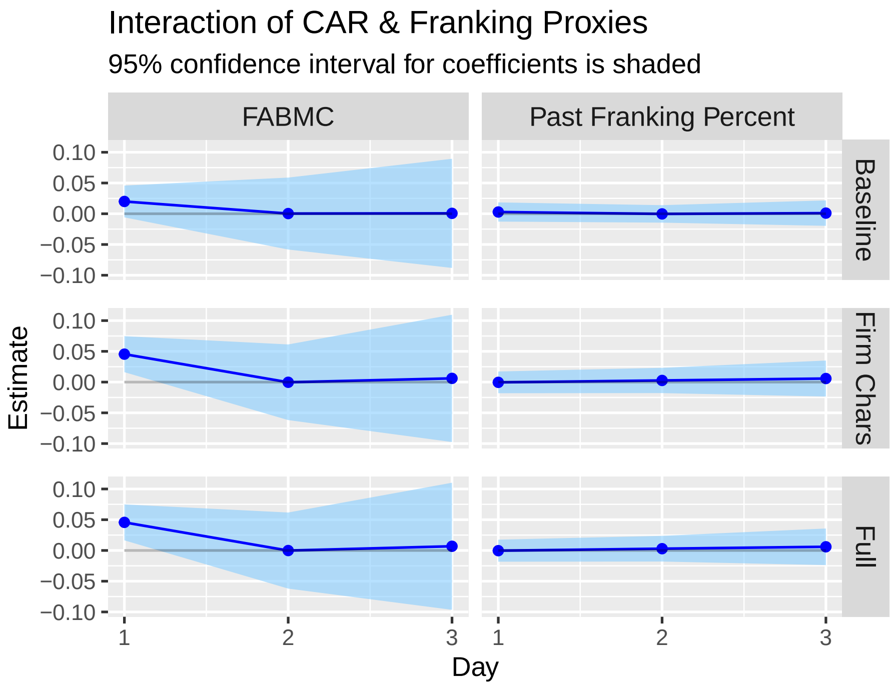
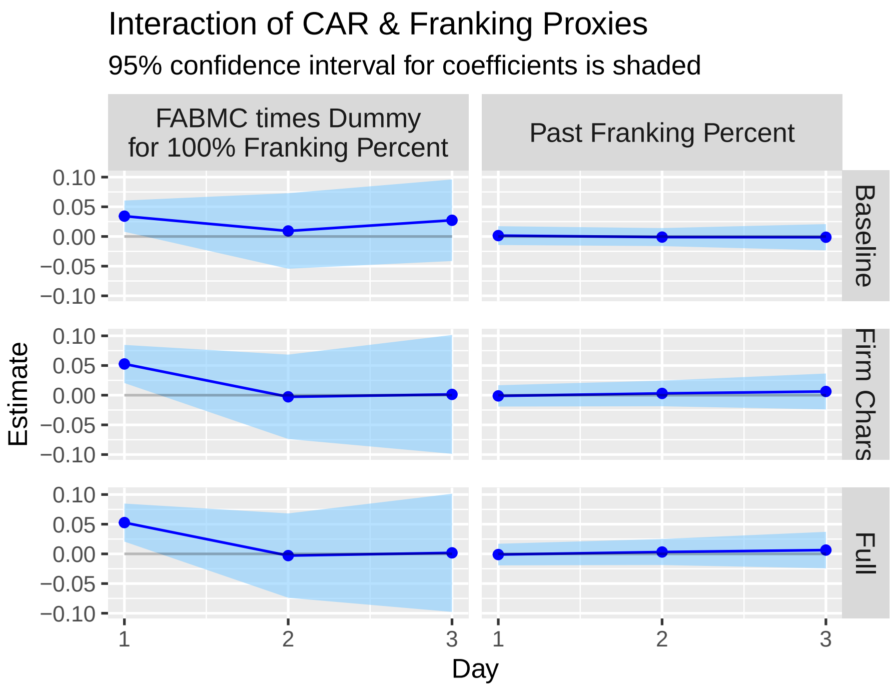


Figure 3.17



3.9 Conclusion

This chapter studies whether dividend taxes reduce share prices. To do this, it conducts an event study of the announcement of a major reform to Australia's dividend imputation system. The reform would have increased the taxation of dividends paid by Australian companies to Australian shareholders, but would not have affected the taxation of dividends paid to foreign shareholders. If the reform had reduced share prices, then companies' CARs during the announcement window should be negatively related to their expected future franking percentage. The results provide no evidence that CARs are negatively related to two reasonable proxies for the future franking percentage. Hence the chapter does not find any evidence that the reform reduced share prices.

The results are consistent with models in which dividend taxes imposed on domestic shareholders in small economies have a negligible effect on prices, such as models where the marginal investor is a foreign shareholder, or the tax-adjusted CAPM model of Brennan (1970). However, the results are also compatible with models in which the dividend taxes have a modest effect on share prices, as the methods in this chapter have low power to detect small violations of the null. One way to improve the method is to collect past franking percentages from the Bloomberg and Eikon databases, which typically cover a larger sample of companies than the S&P Capital IQ database. A larger sample size may result in narrower confidence intervals, enhancing the ability of the method to distinguish between competing models.

Further work could consider the effect of the announcement on companies' dividend behaviour. The franking credit policy would have weakened the preference of domestic investors for receiving dividends over capital gains. Domestic investors may be able to influence companies to behave in ways that do not maximise their share prices (Bond et al. 2007b). This suggests that if the franking credit policy were implemented, then the changing tax preference of domestic investors could cause companies to reduce their dividend yields. If investors believed this would result in a change in dividend yields, then it may have affected returns during the announcement window.

List of Abbreviations

AIG . . . . . . Australian Investment Guarantee

AR(1) . . . . . Autoregressive Model of Order 1

ASX . . . . . . Australian Securities Exchange

BIS . . . . Bank for International Settlements

CAPM . . . . Capital Asset Pricing Model

CAR . . . . . . Cumulative Abnormal Return

CPI . . . . . . Consumer Price Index

DGP . . . . . . Data-generating Process

DM . . . . . . Diebold-Mariano

EC . . . . European Commission

EER . . . . . . Effective Exchange Rates

FAB . . . . . . Franking Account Balance

FABMC . . . . Franking Account Balance as a share of Market Capitalisation

FIRE . . . . . Full Information Rational Expectations

GDP . . . . . . Gross Domestic Product

GEP . . . . . . Global Economic Prospects

GFC . . . . . . Global Financial Crisis

GICS . . . . . Global Industry Classification Standard

HAC . . . . . . Hetersocedasticity and Autocorrelation Consistent

IEO . . . . . . Independent Evaluation Office

IMF . . . . . . International Monetary Fund

IQR . . . . . . Interquartile Range

OECD . . . . . Organisation for Economic Cooperation and Development

MONA . . . . Monitoring of Fund Arrangements

NEER . . . . . Nominal Effective Exchange Rate

NER . . . . . . Nominal Exchange Rate

OLS . . . . . . Ordinary Least Squares

PEPS . . . . . Period-end Price Sampling

pdf . . . . Probability density function

ppt . . . . Percentage points

REER . . . . . Real Effective Exchange Rates

RER . . . . . . Real Exchange Rate

RMSFE . . . . Root Mean Square Forecast Error

S&P . . . . . . Standard & Poors'

SR . . . . Success Ratio

UMIDAS . . . Unrestricted Mixed Data Sampling

WEO . . . . . World Economic Outlook

Appendices

A

Appendices for Real GDP Forecasts by International Organisations

A.1 Fiscal Years that differ from Calendar Years

First day of Year over which Forecaster reports Real GDP

Country	WEO	GEP	OECD	EC	Consensus
Bangladesh	1 July	1 July	NA	NA	1 July
Bhutan	1 July	1 July	NA	NA	NA
Egypt	1 July	1 July	NA	NA	1 July
Ethiopia	1 July	8 July	NA	NA	NA
Haiti	1 October	1 October	NA	NA	NA
India	1 April	1 April	1 April	NA	1 April
Iran	1 April	21 March	NA	NA	NA
Lesotho	1 April	1 January	NA	NA	NA
Marshall Islands	1 October	1 October	NA	NA	NA
Micronesia	1 October	1 October	NA	NA	NA
Myanmar	1 October	1 October	NA	NA	1 October
Nauru	1 July	1 January	NA	NA	NA
Nepal	1 August	16 July	NA	NA	NA
Pakistan	1 July	1 July	NA	NA	1 July
Palau	1 October	1 October	NA	NA	NA
Puerto Rico	1 July	1 January	NA	NA	NA
Samoa	1 July	1 June	NA	NA	NA
South Sudan	1 January	1 July	NA	NA	NA
Tonga	1 July	1 July	NA	NA	NA
Uganda	1 January	1 July	NA	NA	NA

A.2 Comparing Accuracy across Forecasters

A.2.1 Proof for Section 1.4.3

Proof of Theorem 1. The loss differential can be given the following multiplicative decomposition:

$$
\begin{aligned}
d_{t+h|t} &= \left(v^c_{t+h|t}\right)^2 - \left(v^b_{t+h|t}\right)^2 \\
&= \left(y_{t+h} - y^c_{t+h|t}\right)^2 - \left(y_{t+h} - y^b_{t+h|t}\right)^2 \\
&= \left(y^2_{t+h} - 2y_{t+h}y^c_{t+h|t} + \left(y^c_{t+h|t}\right)^2\right) - \left(y^2_{t+h} - 2y_{t+h}y^b_{t+h|t} + \left(y^b_{t+h|t}\right)^2\right) \\
&= -2y_{t+h}\left(y^c_{t+h|t} - y^b_{t+h|t}\right) + (y^c_{t+h|t})^2 - (y^b_{t+h|t})^2 \\
&= -2y_{t+h}\left(y^c_{t+h|t} - y^b_{t+h|t}\right) + \left(y^c_{t+h|t} - y^b_{t+h|t}\right)\left(y^c_{t+h|t} + y^b_{t+h|t}\right) \\
&= -2\left(y^c_{t+h|t} - y^b_{t+h|t}\right)\left(y_{t+h} - \frac{1}{2}\left(y^c_{t+h|t} + y^b_{t+h|t}\right)\right)
\end{aligned}
\tag{A.1}
$$

Hence the absolute loss differential can be decomposed as:

$$
\left|d_{t+h|t}\right| = 2\left|y_{t+h} - \frac{1}{2}\left(y^c_{t+h|t} + y^b_{t+h|t}\right)\right|\left|y^c_{t+h|t} - y^b_{t+h|t}\right|
$$

□

A.2.2 Estimation Method for Section 1.4.3.1

We combine equations (1.2) and (1.3) into a single equation by defining a new variable by $z^k_{t+h|h}$, where k is one of l (for 'loss differential'), e (for 'error in average forecast') or d (for 'disagreement')

$$
\begin{aligned}
z^k_{t+h|t} &= \alpha_1\mathbb{I}(k=l) + \alpha_2 h\mathbb{I}(k=l) \\
&\quad + \beta_1\mathbb{I}(k=e) + \beta_2 h\mathbb{I}(k=e) \\
&\quad + \gamma_1\mathbb{I}(k=d) + \gamma_2 h\mathbb{I}(k=d) \quad \text{for all } t,h,k
\end{aligned}
\tag{A.2}
$$

Substituting in the constraints gives:

$$
\begin{aligned}
z^k_{t+h|t} &= \alpha_1\mathbb{I}(k=l) + \alpha_2 h\mathbb{I}(k=l) \\
&\quad + \beta_1\mathbb{I}(k=e) + \beta_2 h\mathbb{I}(k=e) \\
&\quad + (\alpha_1 - log(2) - \beta_1)\mathbb{I}(k=d) + (\alpha_2 - \beta_2)h\mathbb{I}(k=d) \quad \text{for all } t,h,k
\end{aligned}
\tag{A.3}
$$

Rearranging gives the following equation. We estimate it with OLS.

$$\begin{aligned} & z^k_{t+h|t} + log(2)\mathbb{I}(k=d) \\ & \quad = \alpha_1(\mathbb{I}(k=l) + \mathbb{I}(k=d)) + \alpha_2 h(\mathbb{I}(k=l) \ + \mathbb{I}(k=d)) \\ & \quad + \beta_1(\mathbb{I}(k=e) - \mathbb{I}(k=d)) + \beta_2 h(\mathbb{I}(k=e) - \mathbb{I}(k=d)) \quad \text{for all } t,h,k \end{aligned} \tag{A.4}$$

Using the OLS estimates of $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2$. we compute $\hat{\gamma}_1 = \hat{\alpha}_1 - log(2) - \hat{\beta}_1$ and $\hat{\gamma}_2 = \hat{\alpha}_2 - \hat{\beta}_2$. To test a hypothesis about γ_1 or γ_2, we can perform a Wald test of the appropriate restriction on $(\alpha_1, \alpha_2, \beta_1, \beta_2)$, using a HAC estimator of the covariance matrix of $(\alpha_1, \alpha_2, \beta_1, \beta_2)$.

A.2.3 Full Results of Comparisons of Accuracy

Accuracy of 1-year-ahead forecasts relative to WEO

Blue and red cells indicate significant over or underperformance at a 5% level of significance

	EC		OECD		GEP		Consensus	
	RMSFE Ratio	DM p-val	RMSFE Ratio	DM p-val	RMSFE Ratio	DM p-val	RMSFE Ratio	DM p-val
Afghanistan	NA	NA	NA	NA	0.956	0.370	NA	NA
Albania	0.961	0.374	NA	NA	0.848	0.255	0.914	0.174
Algeria	NA	NA	NA	NA	1.026	0.721	NA	NA
Angola	NA	NA	NA	NA	1.040	0.418	NA	NA
Argentina	NA	NA	1.026	0.538	1.070	0.145	0.953	0.089
Armenia	NA	NA	NA	NA	0.999	0.963	0.964	0.286
Australia	NA	NA	0.815	0.222	NA	NA	0.870	0.047
Austria	0.827	0.012	1.051	0.493	NA	NA	1.006	0.829
Azerbaijan	NA	NA	NA	NA	1.077	0.459	1.005	0.965
Bahrain	NA	NA	NA	NA	0.998	0.928	NA	NA
Bangladesh	NA	NA	NA	NA	1.263	0.234	1.072	0.428
Belarus	NA	NA	NA	NA	1.088	0.278	1.030	0.760
Belgium	1.116	0.373	0.989	0.930	NA	NA	0.988	0.680
Belize	NA	NA	NA	NA	0.929	0.040	NA	NA
Benin	NA	NA	NA	NA	0.998	0.973	NA	NA
Bhutan	NA	NA	NA	NA	0.868	0.203	NA	NA
Bolivia	NA	NA	NA	NA	1.015	0.252	1.014	0.692
Bosnia And Herzegovina	NA	NA	NA	NA	NA	NA	1.030	0.362
Botswana	NA	NA	NA	NA	1.040	0.386	NA	NA
Brazil	NA	NA	1.018	0.717	1.026	0.762	0.895	0.004
Bulgaria	1.206	0.258	NA	NA	1.094	0.128	0.985	0.484
Burkina Faso	NA	NA	NA	NA	0.962	0.479	NA	NA
Burundi	NA	NA	NA	NA	0.840	0.046	NA	NA
Cabo Verde	NA	NA	NA	NA	1.024	0.262	NA	NA
Cambodia	NA	NA	NA	NA	1.012	0.331	NA	NA
Cameroon	NA	NA	NA	NA	0.926	0.776	NA	NA
Canada	1.015	0.587	1.032	0.210	NA	NA	0.966	0.073
Central African Republic	NA	NA	NA	NA	0.083	NA	NA	NA
Chad	NA	NA	NA	NA	1.060	0.324	NA	NA
Chile	NA	NA	1.074	0.138	1.203	0.137	0.977	0.266
China	NA	NA	1.642	0.304	1.107	0.342	0.957	0.297
Colombia	NA	NA	0.976	0.788	1.016	0.548	0.974	0.209
Comoros	NA	NA	NA	NA	0.852	0.368	NA	NA
Costa Rica	NA	NA	1.051	0.212	0.978	0.633	0.953	0.152
Croatia	0.814	0.023	NA	NA	0.984	0.047	1.013	0.409
Cyprus	0.968	0.220	NA	NA	NA	NA	1.054	0.202
Czech Republic	0.952	0.130	1.126	0.389	NA	NA	0.967	0.213
Denmark	0.920	0.278	0.961	0.499	NA	NA	0.931	0.305
Djibouti	NA	NA	NA	NA	1.110	0.421	NA	NA
Dominica	NA	NA	NA	NA	1.061	NA	NA	NA

Accuracy of 1-year-ahead forecasts relative to WEO

Blue and red cells indicate significant over or underperformance at a 5% level of significance

	EC		OECD		GEP		Consensus	
	RMSFE Ratio	DM p-val	RMSFE Ratio	DM p-val	RMSFE Ratio	DM p-val	RMSFE Ratio	DM p-val
Dominican Republic	NA	NA	NA	NA	1.040	0.549	0.980	0.353
Ecuador	NA	NA	NA	NA	1.065	0.161	0.951	0.471
Egypt	NA	NA	NA	NA	1.068	0.199	1.212	0.073
El Salvador	NA	NA	NA	NA	1.043	0.306	1.006	0.869
Equatorial Guinea	NA	NA	NA	NA	1.410	0.047	NA	NA
Eritrea	NA	NA	NA	NA	1.769	NA	NA	NA
Estonia	0.992	0.889	1.080	0.521	NA	NA	1.018	0.495
Eswatini	NA	NA	NA	NA	0.788	0.060	NA	NA
Fiji	NA	NA	NA	NA	1.114	0.401	NA	NA
Finland	1.005	0.888	1.005	0.967	NA	NA	1.004	0.854
France	1.005	0.121	1.134	0.465	NA	NA	0.973	0.227
Gabon	NA	NA	NA	NA	1.021	0.862	NA	NA
Gambia, The	NA	NA	NA	NA	0.984	0.745	NA	NA
Georgia	NA	NA	NA	NA	0.968	0.442	0.977	0.299
Germany	0.909	0.128	1.082	0.646	NA	NA	0.924	0.062
Ghana	NA	NA	NA	NA	1.066	0.358	NA	NA
Greece	0.958	0.078	1.110	0.090	NA	NA	1.008	0.679
Grenada	NA	NA	NA	NA	1.037	0.060	NA	NA
Guatemala	NA	NA	NA	NA	0.882	0.096	0.952	0.130
Guinea	NA	NA	NA	NA	1.068	0.262	NA	NA
Guinea-Bissau	NA	NA	NA	NA	1.153	0.157	NA	NA
Guyana	NA	NA	NA	NA	1.027	0.333	NA	NA
Haiti	NA	NA	NA	NA	0.927	0.109	NA	NA
Honduras	NA	NA	NA	NA	1.008	0.522	0.997	0.882
Hong Kong Sar	NA	NA	NA	NA	NA	NA	0.953	0.107
Hungary	0.918	0.056	1.131	0.256	0.950	0.114	0.867	0.169
Iceland	1.187	0.025	0.886	0.036	NA	NA	NA	NA
India	NA	NA	1.068	0.254	1.022	0.302	0.990	0.596
Indonesia	NA	NA	0.983	0.511	1.040	0.239	0.942	0.187
Iraq	NA	NA	NA	NA	1.138	0.224	NA	NA
Ireland	1.005	0.793	1.207	0.225	NA	NA	1.025	0.075
Israel	NA	NA	1.075	0.401	NA	NA	0.928	0.105
Italy	0.998	0.854	1.045	0.229	NA	NA	0.975	0.208
Jamaica	NA	NA	NA	NA	1.039	0.106	NA	NA
Japan	1.010	0.920	1.191	0.311	1.006	0.835	0.976	0.331
Jordan	NA	NA	NA	NA	0.815	0.291	NA	NA
Kazakhstan	NA	NA	NA	NA	1.081	0.395	0.983	0.630
Kenya	NA	NA	NA	NA	0.965	0.422	NA	NA
Korea	0.990	0.824	0.917	0.549	NA	NA	0.932	0.085
Kosovo	NA	NA	NA	NA	1.070	0.054	NA	NA

Accuracy of 1-year-ahead forecasts relative to WEO

Blue and red cells indicate significant over or underperformance at a 5% level of significance

	EC		OECD		GEP		Consensus	
	RMSFE Ratio	DM p-val	RMSFE Ratio	DM p-val	RMSFE Ratio	DM p-val	RMSFE Ratio	DM p-val
Kuwait	NA	NA	NA	NA	1.114	0.025	NA	NA
Kyrgyz Republic	NA	NA	NA	NA	1.021	0.692	NA	NA
Latvia	0.981	0.520	1.009	0.948	NA	NA	1.057	0.392
Lebanon	NA	NA	NA	NA	1.015	0.552	NA	NA
Liberia	NA	NA	NA	NA	1.061	0.286	NA	NA
Lithuania	0.965	0.125	1.112	0.304	NA	NA	0.985	0.642
Luxembourg	0.991	0.928	1.087	0.285	NA	NA	NA	NA
Madagascar	NA	NA	NA	NA	1.010	0.661	NA	NA
Malawi	NA	NA	NA	NA	1.056	0.612	NA	NA
Malaysia	NA	NA	NA	NA	1.006	0.778	0.928	0.048
Maldives	NA	NA	NA	NA	0.982	0.235	NA	NA
Mali	NA	NA	NA	NA	0.854	0.345	NA	NA
Malta	0.954	0.023	NA	NA	NA	NA	NA	NA
Mauritania	NA	NA	NA	NA	0.843	0.063	NA	NA
Mauritius	NA	NA	NA	NA	1.081	0.296	NA	NA
Mexico	1.000	0.953	1.022	0.661	1.012	0.692	0.987	0.488
Moldova	NA	NA	NA	NA	1.058	0.106	0.992	0.798
Mongolia	NA	NA	NA	NA	0.942	0.220	NA	NA
Montenegro, Rep. Of	1.052	0.357	NA	NA	NA	NA	NA	NA
Morocco	NA	NA	NA	NA	1.013	0.692	NA	NA
Mozambique	NA	NA	NA	NA	0.710	0.085	NA	NA
Myanmar	NA	NA	NA	NA	0.999	0.770	NA	NA
Namibia	NA	NA	NA	NA	0.954	0.220	NA	NA
Netherlands	1.015	0.876	1.208	0.183	NA	NA	0.939	0.175
New Zealand	1.244	NA	0.869	0.401	NA	NA	0.863	0.015
Nicaragua	NA	NA	NA	NA	1.046	0.778	0.953	0.493
Niger	NA	NA	NA	NA	0.967	0.665	NA	NA
Nigeria	NA	NA	NA	NA	1.073	0.432	0.909	0.164
North Macedonia	0.997	0.977	NA	NA	1.080	0.186	0.992	0.701
Norway	1.077	0.450	0.883	0.150	NA	NA	1.070	0.228
Oman	NA	NA	NA	NA	0.811	0.290	NA	NA
Pakistan	NA	NA	NA	NA	1.207	0.062	0.964	0.648
Panama	NA	NA	NA	NA	1.042	0.349	0.970	0.262
Papua New Guinea	NA	NA	NA	NA	0.773	0.101	NA	NA
Paraguay	NA	NA	NA	NA	1.058	0.684	0.994	0.944
Peru	NA	NA	NA	NA	1.009	0.825	0.963	0.138
Philippines	NA	NA	NA	NA	1.034	0.287	1.005	0.796
Poland	1.035	0.309	1.410	0.229	1.054	0.328	0.885	0.203
Portugal	1.012	0.366	1.061	0.078	NA	NA	1.015	0.380
Qatar	NA	NA	NA	NA	0.972	0.489	NA	NA

Accuracy of 1-year-ahead forecasts relative to WEO

Blue and red cells indicate significant over or underperformance at a 5% level of significance

	EC		OECD		GEP		Consensus	
	RMSFE Ratio	DM p-val	RMSFE Ratio	DM p-val	RMSFE Ratio	DM p-val	RMSFE Ratio	DM p-val
Romania	NA	NA	1.176	NA	0.970	0.414	0.972	0.258
Russia	NA	NA	1.138	0.458	0.915	0.015	0.975	0.479
Rwanda	NA	NA	NA	NA	1.000	0.995	NA	NA
Saudi Arabia	NA	NA	NA	NA	0.961	0.679	1.010	0.785
Senegal	NA	NA	NA	NA	1.007	0.794	NA	NA
Serbia	1.040	0.492	NA	NA	1.008	0.905	0.976	0.611
Seychelles	NA	NA	NA	NA	0.950	0.492	NA	NA
Sierra Leone	NA	NA	NA	NA	1.028	0.889	NA	NA
Singapore	NA	NA	NA	NA	NA	NA	0.916	0.181
Slovak Republic	0.881	0.068	1.063	0.445	NA	NA	0.995	0.881
Slovenia	0.862	0.001	1.067	0.351	NA	NA	0.980	0.534
Solomon Islands	NA	NA	NA	NA	1.002	0.961	NA	NA
South Africa	NA	NA	1.019	0.540	1.001	0.936	1.070	0.108
Spain	0.894	0.064	1.133	0.307	NA	NA	1.009	0.548
Sri Lanka	NA	NA	NA	NA	1.022	0.670	0.979	0.619
St. Lucia	NA	NA	NA	NA	1.089	0.361	NA	NA
Sudan	NA	NA	NA	NA	1.629	0.234	NA	NA
Suriname	NA	NA	NA	NA	1.056	0.423	NA	NA
Sweden	0.926	0.445	1.169	0.073	NA	NA	0.934	0.163
Switzerland	0.960	0.615	1.112	0.311	NA	NA	0.930	0.209
Taiwan Province Of China	NA	NA	NA	NA	NA	NA	0.900	0.207
Tajikistan	NA	NA	NA	NA	0.856	0.028	NA	NA
Tanzania	NA	NA	NA	NA	0.855	0.588	NA	NA
Thailand	NA	NA	NA	NA	1.071	0.074	0.935	0.158
Timor-Leste	NA	NA	NA	NA	0.946	0.154	NA	NA
Togo	NA	NA	NA	NA	1.167	0.002	NA	NA
Tunisia	NA	NA	NA	NA	1.045	0.296	NA	NA
Turkey	1.033	0.487	0.984	0.848	0.948	0.362	0.956	0.105
Turkmenistan	NA	NA	NA	NA	1.006	0.925	0.943	0.455
Ukraine	NA	NA	NA	NA	1.053	0.494	1.007	0.845
United Arab Emirates	NA	NA	NA	NA	0.931	0.254	NA	NA
United Kingdom	1.648	0.222	0.983	0.834	NA	NA	0.972	0.119
United States	0.985	0.728	0.923	0.416	0.831	0.250	0.924	0.242
Uruguay	NA	NA	NA	NA	1.139	0.085	0.944	0.248
Uzbekistan	NA	NA	NA	NA	1.256	0.282	1.020	0.850
Venezuela	NA	NA	NA	NA	1.219	0.125	1.006	0.913
Vietnam	NA	NA	NA	NA	1.013	0.767	1.013	0.886
Zambia	NA	NA	NA	NA	0.962	0.606	NA	NA
Zimbabwe	NA	NA	NA	NA	0.917	0.079	NA	NA

A.3 Proofs for Section 1.6.1

Proof of Theorem 2. The rational forecast is the sum of the current practice forecast and an extra term.

$$\begin{aligned}
\underbrace{E\left[y_{t+h}\right]}_{\text{Rational Forecast}} &= E\left[r(x_{t+h})\right] \\
&= \int_{-\infty}^{\infty} r(x)p(x)dx \\
&= \int_{-\infty}^{\infty} r(\mu)p(x)dx + \int_{-\infty}^{\infty} (r(x) - r(\mu))p(x)dx \\
&= r(\mu) + \int_{-\infty}^{\infty} (r(x) - r(\mu))p(x)dx \\
&= \underbrace{E\left[y|\mu\right]}_{\text{Current Practice Forecast}} + \underbrace{\int_{-\infty}^{\infty} (r(x) - r(\mu))p(x)dx}_{\text{Extra Term}}
\end{aligned} \tag{A.5}$$

If the future explanatory variable has a degenerate distribution, it is guaranteed to equal μ, so the rational forecast and current practice forecast are equal. Suppose that the future explanatory variable has a non-degenerate distribution. We will consider cases (a) and (b) separately.

Case (a)

In the first case, the response function is $r(x) = a + bx$. The extra term is zero, so the rational forecast equals the current practice forecast regardless of the distribution of x.

$$\begin{aligned}
\int_{-\infty}^{\infty} (r(x) - r(\mu))p(x)dx &= \int_{-\infty}^{\infty} (a + bx - a - b\mu)p(x)dx \\
&= b\int_{-\infty}^{\infty} xp(x)dx - b\int_{-\infty}^{\infty} \mu p(x)dx \\
&= b\mu - b\mu = 0
\end{aligned}$$

Case (b)

In the second case, the extra term is also zero. The first equality is given by performing integration by substitution with $u(x) = x - \mu$, so $u'(x) = 1$. The third equality is given by performing integration by substitution on the first term with $v(x) = -u(x)$, so $v'(x) = -1$. The fourth equality holds by the symmetry of the conditional pdf around the mean, $p(-v + \mu) = p(v + \mu)$. The final equality holds by the rotational symmetry of r around the point $(\mu, r(\mu))$.

$$\begin{aligned}
&\int_{-\infty}^{\infty}(r(x)-r(\mu))p(x)dx \\
&\quad=\int_{-\infty}^{\infty}(r(u+\mu)-r(\mu))p(u+\mu)du \\
&\quad=\int_{0}^{\infty}(r(u+\mu)-r(\mu))p(u+\mu)du+\int_{-\infty}^{0}(r(u+\mu)-r(\mu))p(u+\mu)du \\
&\quad=\int_{0}^{\infty}(r(u+\mu)-r(\mu))p(u+\mu)du+\int_{0}^{\infty}(r(-v+\mu)-r(\mu))p(-v+\mu)dv \\
&\quad=\int_{0}^{\infty}(r(u+\mu)-r(\mu))p(u+\mu)du+\int_{0}^{\infty}(r(-v+\mu)-r(\mu))p(v+\mu)dv \\
&\quad=\int_{0}^{\infty}(r(u+\mu)-r(\mu))p(u+\mu)du+\int_{0}^{\infty}-(r(v+\mu)-r(\mu))p(v+\mu)dv \\
&\quad=0
\end{aligned}$$

$\square$

Proof of Theorem 3. If the response function attains a strict global minimum at μ, the 'extra term' from equation (A.5) will be positive, so the rational forecast will exceed the current forecast. To show this, we integrate over two subsets of the real line: $A = \{x \in \mathbb{R} : x \neq \mu \text{ and } p(x) > 0\}$ and its complement $A^c = \{x \in \mathbb{R} : x = \mu \text{ or } p(x) = 0\}$. The set A is non-empty because the future explanatory variable has a non-degenerate distribution, and hence has some probability density on values of x other than its mean μ. The integral on A is strictly positive[1], while the integral over A^c is zero[2], so the extra term must be strictly positive.

$$\begin{aligned}
&\int_{-\infty}^{\infty}(r(x)-r(\mu))p(x)dx \\
&\quad=\underbrace{\int_{A}(r(x)-r(\mu))p(x)dx}_{>0}+\underbrace{\int_{A^c}(r(x)-r(\mu))p(x)dx}_{=0} \\
&\quad>0
\end{aligned}$$

If the response function attains a strict global maximum at μ, the extra term is negative by the same argument, and hence the rational forecast is lower than the current practice forecast.

[1]The integrand $(r(x) - r(\mu))p(x)$ is strictly positive on A because $r(x) - r(\mu) > 0$ on A (because r attains a strict global minimum at μ, and $x \neq \mu$ in A), and because $p(x) > 0$ by definition of A.

[2]The integrand is zero for all x in A^c because at least one of $x = \mu$ or $p(x) = 0$ must be true by definition of A.

Suppose the response function $r(x)$ is strictly convex. Jensen's inequality states that, for any strictly convex function $\varphi : \mathbb{R} \to \mathbb{R}$

$$E[\varphi(x)] > \varphi E[x]$$

The sum of a strictly convex function and a constant is strictly convex, so the function $\varphi(x) \equiv r(x) - r(\mu)$ must be strictly convex. Hence we can show that the 'extra term' is strictly positive as follows, where the strict inequality is due to Jensen's inequality. Since the extra term is positive, the rational forecast is strictly greater than the current practice forecast.

$$\begin{aligned}
&\int_{-\infty}^{\infty} (r(x) - r(\mu))p(x)dx \\
&\quad = E\left[r(x) - r(\mu)\right] \\
&\quad > r(E[x]) - r(\mu) \\
&\quad = 0
\end{aligned}$$

If the response function $r(x)$ is strictly concave, we can make the same argument to show that the extra term is strictly negative, and hence that the rational forecast is strictly less than the current practice forecast. $\square$

B

Appendices for Temporal Aggregation of Exchange Rates

B.1 Proofs for Section 2.3

Before the proof of theorem 4, we introduce two useful lemmas. These lemmas show that the terms $(Z_m - A_m)$ and $(A_{m+h} - Z_m)$ are linear combinations of the random walk errors of different months. Since the errors are independent of one another, these terms must be independent of each other. Independence implies that the sign of the former term is not useful in predicting the sign of the latter term, which plays a key role in the proof that follows.

Lemma 1. *Suppose the daily data is a random walk, as in equation (2.5). The difference between the current end-of-month level and current month-average level, $(Z_m - A_m)$, is a weighted sum of errors in month m.*

$$Z_m - A_m = \sum_{j=1}^{n} e_{(m-1)n+j} - \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{i} e_{(m-1)n+j}\right) \tag{B.1}$$

Lemma 2. *Suppose the daily data is a random walk, as in equation (2.5). The difference between the future month-average and current end-of-month, $(A_{m+h} - Z_m)$, is a weighted sum of errors in months $(m+1)$ to $(m+h)$.*

$$A_{m+h} - Z_m = \sum_{j=n+1}^{hn} e_{(m-1)n+j} + \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{i} e_{(m+h-1)n+j}\right) \tag{B.2}$$

Proof of lemmas 1 and 2. The daily data is a random walk (equation (2.5)), so we can write it as:

$$\underbrace{D_{(m-1)n+i}}_{\text{Level on day } i \text{ of month } m} = \underbrace{D_{(m-1)n}}_{\text{Level on last day of month } (m-1)} + \underbrace{\sum_{j=1}^{i} e_{(m-1)n+j}}_{\text{Errors in month } m} \tag{B.3}$$

This allows us to rewrite end-of-month and month-average observations in terms of the errors.

- The end-of-month level in month m is:

$$Z_m = D_{(m-1)n} + \sum_{j=1}^{n} e_{(m-1)n+j} \tag{B.4}$$

- The month-average level in month m is:

$$A_m = \frac{1}{n}\sum_{i=1}^{n}\left(D_{(m-1)n} + \sum_{j=1}^{i} e_{(m-1)n+j}\right) = D_{(m-1)n} + \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{i} e_{(m-1)n+j}\right) \tag{B.5}$$

- The month-average level in month $(m+h)$ is:

$$A_{m+h} = \frac{1}{n}\sum_{i=1}^{n}\left(D_{(m+h-1)n} + \sum_{j=1}^{i} e_{(m+h-1)n+j}\right) \tag{B.6}$$

$$= D_{(m+h-1)n} + \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{i} e_{(m+h-1)n+j}\right) \tag{B.7}$$

$$= D_{(m-1)n} + \sum_{j=1}^{n} e_{(m-1)n+j} + \sum_{j=n+1}^{hn} e_{(m-1)n+j} + \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{i} e_{(m+h-1)n+j}\right) \tag{B.8}$$

The difference between the current end-of-month level and current month-average level, $(Z_m - A_m)$, is a weighted sum of errors in month m. Equations (B.5) and (B.4) imply:

$$\begin{aligned} &Z_m - A_m \\ &= \left[D_{(m-1)n} + \sum_{j=1}^{n} e_{(m-1)n+j}\right] - \left[D_{(m-1)n} + \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{i} e_{(m-1)n+j}\right)\right] \\ &= \underbrace{\sum_{j=1}^{n} e_{(m-1)n+j} - \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{i} e_{(m-1)n+j}\right)}_{\text{Errors in } m} \end{aligned} \tag{B.9}$$

Similarly, the difference between the future month-average and current end-of-month, $(A_{m+h} - Z_m)$, is a weighted sum of errors in months $(m+1)$ to $(m+h)$. Equations (B.8) and (B.4) imply:

$$\begin{aligned}
&A_{m+h} - Z_m \\
&= \left[D_{(m-1)n} + \sum_{j=1}^{n} e_{(m-1)n+j} + \sum_{j=n+1}^{hn} e_{(m-1)n+j} + \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{i} e_{(m+h-1)n+j}\right)\right] \\
&- \left[D_{(m-1)n} + \sum_{j=1}^{n} e_{(m-1)n+j} \right] \\
&= \underbrace{\sum_{j=n+1}^{hn} e_{(m-1)n+j}}_{\text{Errors in } (m+1) \text{ to } (m+h-1)} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{i} e_{(m+h-1)n+j}\right)}_{\text{Errors in } m+h}
\end{aligned}$$

□

We now prove each part of theorem 4, making use of the lemmas above.

Proof of Theorem 4(a). For convenience, define terms u and v by:

$$u \equiv \mathbb{P}[(A_{m+h} - Z_m) > 0 \cap (Z_m - A_m) > 0]$$

$$v \equiv \mathbb{P}[(A_{m+h} - A_m) > 0 \cap (Z_m - A_m) > 0]$$

To prove theorem 4(a), we will show that:

$$E[DA_{m,h}] = 2v > 2u = 2 \times \frac{1}{4} = \frac{1}{2}$$

First, we prove the first equality in equation (B.1), which is that directional accuracy equals $2v$. When the candidate is end-of-month no-change and the benchmark is month-average no-change, directional accuracy can be written:

$$\begin{aligned}
E[DA_{m,h}] &= E\left[\mathbb{1}\left\{sgn(A_{m+h} - A_m) = sgn(Z_m - A_m)\right\}\right] \\
&= \mathbb{P}[sgn(A_{m+h} - A_m) = sgn(Z_m - A_m)] \\
&= \mathbb{P}[(A_{m+h} - A_m) > 0 \cap (Z_m - A_m) > 0] + \mathbb{P}[(A_{m+h} - A_m) \leq 0 \cap (Z_m - A_m) \leq 0] \\
&= \mathbb{P}[(A_{m+h} - A_m) > 0 \cap (Z_m - A_m) > 0] + \mathbb{P}[(A_{m+h} - A_m) < 0 \cap (Z_m - A_m) < 0] \qquad \text{(B.10)} \\
&= 2 \times \mathbb{P}[(A_{m+h} - A_m) > 0 \cap (Z_m - A_m) > 0] = 2v \qquad \text{(B.11)}
\end{aligned}$$

Equality (B.10) holds because the terms $(A_{m+h} - A_m)$ and $(Z_m - A_m)$ are continuous random variables, as they are the sum of errors (by lemmas 1 and 2), and the errors are continuous. Equality (B.11) holds because the terms $(A_{m+h} - Z_m)$ and $(Z_m - A_m)$ are symmetric about zero, since they are sums of errors that are symmetric about zero.

Second, we prove the inequality in equation (B.1), which is $2v > 2u$, or equivalently, $v > u$.

$$\begin{aligned}
v &\equiv \mathbb{P}[(A_{m+h} - A_m) > 0 \cap (Z_m - A_m) > 0] \\
&= \mathbb{P}[((A_{m+h} - Z_m) + (Z_m - A_m)) > 0 \cap (Z_m - A_m) > 0] \\
&= \mathbb{P}[((A_{m+h} - Z_m) + (Z_m - A_m)) > 0 | (Z_m - A_m) > 0] \times \mathbb{P}[(Z_m - A_m) > 0] \\
&> \mathbb{P}[(A_{m+h} - Z_m) > 0 | (Z_m - A_m) > 0] \times \mathbb{P}[(Z_m - A_m) > 0] \qquad \text{(B.12)} \\
&= \mathbb{P}[(A_{m+h} - Z_m) > 0] \times \mathbb{P}[(Z_m - A_m) > 0] \\
&= u
\end{aligned}$$

Inequality (B.12) holds because $\mathbb{P}\left[x + y > 0 | y > 0\right] > \mathbb{P}\left[x > 0 | y > 0\right]$ for any numbers x and y.

Third, we show that u equals $\frac{1}{4}$.

$$\begin{aligned}
u &= \mathbb{P}[(A_{m+h} - Z_m) > 0] \times \mathbb{P}[(Z_m - A_m) > 0] \qquad \text{(B.13)} \\
&= \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} \qquad \text{(B.14)}
\end{aligned}$$

Equality (B.13) holds because the terms $(A_m - Z_m)$ and $(A_{m+h} - Z_m)$ are independent of each other. We know the terms are independent because the former term is the sum of errors in m (see lemma 1), the latter is the sum of errors in $(m+1)$ to $(m + h)$, and $\{e\}$ is an independent sequence. Equality (B.14) holds because the terms $(A_m - Z_m)$ and $(A_{m+h} - Z_m)$ are symmetric about zero, since they are linear combinations of errors that are symmetric about zero. □

Proof of Theorem 4(b). The benchmark is the end-of-month no-change. Hence, expected directional accuracy can be written:

$$E\left[DA_{m,h}\right] \equiv E\left[\mathbb{1}\left\{sgn(A_{m+h} - Z_m) = sgn\left(\hat{A}^{candidate}_{m+h|m} - Z_m\right)\right\}\right]$$
$$= \mathbb{P}\left[sgn(A_{m+h} - Z_m) = sgn\left(\hat{A}^{candidate}_{m+h|m} - Z_m\right)\right]$$
$$= \mathbb{P}\left[(A_{m+h} - Z_m) > 0 \cup \left(\hat{A}^{candidate}_{m+h|m} - Z_m\right) > 0\right] + \mathbb{P}\left[(A_{m+h} - Z_m) \leq 0 \cup \left(\hat{A}^{candidate}_{m+h|m} - Z_m\right) \leq 0\right]$$
$$= \mathbb{P}[(A_{m+h} - Z_m) > 0] \times \mathbb{P}\left[\left(\hat{A}^{candidate}_{m+h|m} - Z_m\right) > 0\right] + \mathbb{P}[(A_{m+h} - Z_m) \leq 0] \times \mathbb{P}\left[\left(\hat{A}^{candidate}_{m+h|m} - Z_m\right) \leq \tag{B.15}$$
$$= \frac{1}{2} \times \mathbb{P}\left[\left(\hat{A}^{candidate}_{m+h|m} - Z_m\right) > 0\right] + \frac{1}{2} \times \mathbb{P}\left[\left(\hat{A}^{candidate}_{m+h|m} - Z_m\right) \leq 0\right] \tag{B.16}$$
$$= \frac{1}{2}$$

Equality (B.15) holds because $(A_{m+h} - Z_m)$ and $(\hat{A}^{candidate}_{m+h|m} - Z_m)$ are independent. This is true because:

- the former term depends solely on errors from $(m+1)$ to $(m+h)$, as shown in lemma 2
- the latter term depends solely on the daily exchange rate's initial level and errors up to month m, and the levels of other variables up to month m.[1]
- The errors from $(m+1)$ to $(m+h)$ are independent of the errors up to month m and the past levels of all other variables, by the random walk assumption (2.5).

Equality (B.16) holds because the term $(A_{m+h} - Z_m)$ is the sum of the errors of a random walk (see lemma 2). The term is continuous and symmetric about zero because the errors are continuous and symmetric about zero. Hence, the probability that the term is strictly positive is half, and the probability it is negative or zero is also a half. □

[1]A candidate forecast $\hat{A}^{candidate}_{m+h|m}$ only depends on variables up to the time it is made, so it only depends on the daily exchange rate's initial level and errors up to m, and on the levels of other variables up to m. The end-of-period level, $Z_m = D_{(m-1)n} + \sum_{j=1}^{n} e_{(m-1)n+j}$, depends only on initial level and errors up to month m.

B.2 Literature on Point-in-Time Sampled Nominal Bilateral Exchange Rates

Table B.1: Summary of Literature Focusing on Point-in-Time Sampled Nominal Bilateral Exchange Rates

Paper	Level or Return	Frequency	Forecast Target	Benchmark	Model Estimation	Real or Nominal	Real-time
Edwards (1983)	Level	M	EoP	EoP	EoP	Nominal	N
Meese and Rogoff (1983a)	Level	M	EoP	EoP	EoP	Nominal	N
Meese and Rogoff (1983b)	Level	M	EoP	EoP	EoP	Nominal	N
Fama (1984)	Return	M	EoP	EoP	EoP	Nominal	N
Boughton (1987)	Both	M	EoP*	EoP*	EoP*	Nominal	N
Schinasi and Swamy (1987)	Level	M	EoP*	EoP*	EoP*	Nominal	N
Wolff (1987)	Level	M	EoP	EoP	EoP	Nominal	N
Hodrick (1989)	Return	M	EoP	EoP	EoP	Nominal	N
Diebold and Nason (1990)	Return	W	MoP	MoP	MoP	Nominal	N
Engel and Hamilton (1990)	Return	Q	EoP	EoP	EoP	Nominal	N
Chinn (1991)	Level	Q	EoP	EoP	EoP	Nominal	N
Meese and Rose (1991)	Level	M	EoP*	EoP*	EoP*	Nominal	N
Mizrach (1992)	Return	D	EoP	EoP	EoP	Nominal	N
Canova (1993)	Level	W	MoP	MoP	MoP	Nominal	N
Krager and Kruger (1993)	Return	W	EoP	EoP	EoP	Nominal	N
Macdonald and Taylor (1993)	Level	M	EoP	EoP	EoP	Nominal	N
Throop (1993)	Return	Q	EoP*	EoP*	EoP*	Nominal	N
Diebold et al. (1994)	Return	D	SoP	SoP	SoP	Nominal	N
Engel (1994)	Return	Q	EoP	EoP	EoP	Nominal	N
Chinn and Meese (1995)	Level	M	EoP	EoP	EoP	Nominal	N
Mark (1995)	Return	Q	EoP	EoP	EoP	Nominal	N
Clarida and Taylor (1997)	Both	W	EoP	EoP	EoP	Nominal	Y
Groen (1999)	Return	M	EoP	EoP	EoP	Nominal	N
Kilian (1999)	Return	Q	EoP*	EoP*	EoP*	Nominal	N
Berkowitz and Giogianni (2001)	Return	Q	EoP	EoP	EoP	Nominal	N
Clements and Smith (2001)	Level	W	EoP	EoP	EoP	Nominal	N
Mark and Sul (2001)	Return	Q	EoP	EoP	EoP	Nominal	N
Rapach and Wohar (2002)	Level	A	EoP*	EoP*	EoP*	Nominal	N
Clarida et al. (2003)	Return	W	EoP*	EoP*	EoP*	Nominal	Y
Faust et al. (2003)	Return	Q	EoP	EoP	EoP	Nominal	Y
Qi and Wu (2003)	Level	M	EoP	EoP	EoP	Nominal	N
Rapach and Wohar (2004)	Both	Q	EoP*	EoP*	EoP*	Nominal	N
Abhyankar et al. (2005)	Return	M	EoP	EoP	EoP	Nominal	N
Cheung et al. (2005)	Both	Q	EoP	EoP	EoP	Nominal	N
Engel and West (2005)	Level	Q	EoP	EoP	EoP	Nominal	N
Evans and Lyons (2005)	Return	D	EoP	EoP	EoP	Nominal	N
Groen (2005)	Level	Q	EoP	EoP	EoP	Nominal	N
Rossi (2005)	Level	Q	EoP*	EoP*	EoP*	Nominal	N
Clark and West (2006)	Return	M	EoP	EoP	EoP	Nominal	N
Rossi (2006)	Return	M	EoP	EoP	EoP	Nominal	N
Alquist and Chinn (2008)	Level	Q	EoP	EoP	EoP	Nominal	N
Engel et al. (2008)	Return	Q	EoP	EoP	EoP	Nominal	N

Note: Papers whose "Forecast Target" are effective exchange rates. "*" is used in cases where the paper did not provide information whether exchange rates are average or point sampled, and so point-in-time sampling was assumed. "Benchmark" refers to the no-change forecast that the forecast was compared against. "Model Estimation" refers to the data used in estimation. "EoP", "MoP", and "SoP" refer to end-, middle-, and start-of-period sampling, respectively.

Table B.2: Summary of Literature Focusing on Point in Time Sampled Nominal Bilateral Exchange Rates, continued

Paper	Level or Return	Frequency	Forecast Target	Benchmark	Model Estimation	Real or Nominal	Real-time
Adrian et al. (2009)	Return	M	EoP	EoP	EoP	Nominal	N
Della Corte et al. (2009)	Return	M	EoP*	EoP*	EoP*	Nominal	N
Sarno and Valente (2009)	Return	Q	EoP*	EoP*	EoP*	Nominal	Y
Wang and Wu (2009)	Return	M	EoP	EoP	EoP	Nominal	N
Altavilla and De Frauwe (2010)	Both	Q	EoP*	EoP*	EoP*	Nominal	N
Bacchetta et al. (2010)	Return	M	EoP	EoP	EoP	Nominal	N
Cerra and Saxena (2010)	Level	A	EoP	EoP	EoP	Nominal	N
Chen et al. (2010)	Return	Q	EoP	EoP	EoP	Nominal	N
Rime et al. (2010)	Return	D	EoP	EoP	EoP	Nominal	N
Lopez-Suarez et al. (2011)	Return	Q	EoP	EoP	EoP	Nominal	N
Molodtsova et al. (2011)	Return	Q	MoP	MoP	MoP	Nominal	Y
Pacelli et al. (2011)	Level	D	EoP*	EoP*	EoP*	Nominal	N
Bianco et al. (2012)	Return	W	EoP	EoP	EoP	Nominal	N
Chinn and Moore (2012)	Return	M	EoP	EoP	EoP	Nominal	N
Della Corte et al. (2012)	Return	Q	EoP*	EoP*	EoP*	Nominal	Y
Molotsova and Papell (2012)	Return	Q	EoP	EoP	EoP	Nominal	Y
Rossi and Inoue (2012)	Level	M	EoP*	EoP*	EoP*	Nominal	N
Wang and Wu (2012)	Return	M	EoP	EoP	EoP	Nominal	N
Bashar and Kabir (2013)	Level	Q	EoP*	EoP*	EoP*	Nominal	N
Chen and Tsang (2013)	Return	M	EoP	EoP	EoP	Nominal	N
Morales-Arias and Moura (2013)	Return	M	EoP*	EoP*	EoP*	Nominal	N
Park and Park (2013)	Both	Q	EoP*	EoP*	EoP*	Nominal	N
Rossi (2013)	Both	M, Q	EoP*	EoP*	EoP*	Nominal	N
Berge (2014)	Return	M	EoP	EoP	EoP	Nominal	N
Garratt Mise (2014)	Return	Q	EoP	EoP	EoP	Nominal	N
Ince (2014)	Return	Q	EoP	EoP	EoP	Nominal	Y
Engel et al. (2015)	Return	Q	EoP	EoP	EoP	Nominal	N
Ferraro et al. (2015)	Return	D	EoP	EoP	EoP	Nominal	Y
Ferraro et al. (2015)	Return	M	SoP	SoP	SoP	Nominal	Y
Ferraro et al. (2015)	Return	Q	MoP	MoP	MoP	Nominal	Y
Li et al. (2015)	Return	M	EoP	EoP	EoP	Nominal	N
Beckman and Schussler (2016)	Return	M	EoP	EoP	EoP	Nominal	Y
Byrne et al. (2016)	Return	Q	EoP	EoP	EoP	Nominal	N
Kohlscheen et al. (2016)	Return	D	EoP*	EoP*	EoP*	Nominal	Y
Zhang et al. (2016)	Return	D	EoP*	EoP*	EoP*	Nominal	N
Kouwenberg et al (2017)	Return	Q	EoP*	EoP*	EoP*	Nominal	Y
Cheung el al. (2018)	Both	Q	EoP	EoP	EoP	Nominal	N
Engel et al. (2019)	Return	M	EoP	EoP	EoP	Nominal	N
Kremens and Martins (2019)	Return	M	MoP	MoP	MoP	Nominal	N
Beckmann et al. (2020)	Return	M	EoP	EoP	EoP	Nominal	Y
Ca'Zorzi and Rubaszek (2020)	Return	M	EoP	EoP	EoP	Nominal	N
Bork et al. (2022)	Return	M	EoP	EoP	EoP	Nominal	N
Lilley et al. (2022)	Return	M	EoP*	EoP*	EoP*	Nominal	N
Liu and Shaliastovich (2022)	Return	M	EoP	EoP	EoP	Nominal	N
Engel and Wu (2023a)	Return	M	EoP	EoP	EoP	Nominal	N
Engel and Wu (2023b)	Return	M	EoP	EoP	EoP	Nominal	N

Note: Papers whose "Forecast Target" are effective exchange rates. "*" is used in cases where the paper did not provide information whether exchange rates are average or point sampled, and so point-in-time sampling was assumed. "Benchmark" refers to the no-change forecast that the forecast was compared against. "Model Estimation" refers to the data used in estimation. "EoP", "MoP", and "SoP" refer to end-, middle-, and start-of-period sampling, respectively.

B.3 Inputs into Bilateral RER and REER Calculations

Section 2.4 describes how we constructed real-time vintages of bilateral RERs and REERs. This appendix provides detail on each of the inputs into these calculations. I.e. Daily nominal exchange rates (NERs); daily consumer price index (CPI) levels; and trade weights.

B.3.1 Daily Nominal Exchange Rates

B.3.1.1 IMF Nominal Exchange Rates

We extracted daily NERs from an internal IMF database called 'Global Data Source'.[2] We extracted all of the series from the 'live' versions of IMF databases in October 2022. Earlier vintages were not available.

Table B.3: IMF Data used as Inputs into EER Calculations

Indicator code	Indicator name	Units	Frequency	Earliest period	Latest period	Countries available
EDNA	Exchange rate for EER, period average	USD per national currency	Daily	1 Jan 1993	21 Oct 2022	165
Weights	Weights	Percent	Ocassional	1990-1995	2016-Latest	193

The EDNA data provides a single time series for each country. Where a country has adopted a new currency during the sample period, the exchange rates of the two currencies are spliced together so that EDNA does not contain a level shift when the new currency is adopted.[3]

[2]The Global Data Source database contains two similar series: EDNA and EDNA_EER. For some countries, EDNA_EER only reports exchange rates on trading days, and reports N/A on other days. EDNA reports rates on all days, because on weekends and public holidays it carries forward the observation from the last trading day. The two series are otherwise identical. We use EDNA_EER, but since we carry forward the observation from the last trading day this is equivalent to using EDNA.

[3]For example, the EDNA data contains a single series for Austria from 1 January 1993 onwards, even though Austria switched from the Austrian Schilling to the Euro on 1 January 1999. For days before the adoption of a new currency, EDNA reports the Schilling/USD exchange rate. From 1 January 1999, EDNA starts at the Schilling/USD rate and is then grown based on the euro/USD exchange rate. Splicing exchange rates in this way avoids a jump in EDNA, which avoids a jump in RERs or REERs.

B.3.1.2 Splicing on Eikon Nominal Exchange Rates

The IMF NERs start on 1 January 1993 for some countries, and later for others. For many countries, Eikon NERs are available from an earlier date. For many countries, we splice the Eikon NERs onto the IMF NERs, resulting in a longer time series of NERs, and increasing the estimation sample for our models.

We perform the splicing in stages.

1. For each country with an IMF NER, we guess the currency they used before the start of the IMF NERs. This is needed because each Eikon NER series refers to a currency, while each IMF NER series refers to a country.

2. Check if Eikon has data on the currency of interest. This is not the case for some discontinued currencies.

3. Check if the Eikon NERs start earlier than the IMF NERs. This is not the case for currencies introduced relatively recently.

4. Check that the Eikon and IMF NERs are the same during any period when both series are available. If this were not the case, it would suggest that we have guessed the currency incorrectly, or that the Eikon and IMF NERs are not comparable for some other reason.

5. Splice the series if the previous checks are met. We use the IMF series on each day it is available, and the Eikon series otherwise.

Using the above process, we are able to splice Eikon and IMF NERs for 62 countries (Table B.4).

To guess which exchange rate each country used before the start of the IMF data, we rely on the IMF exchange arrangements and exchange restrictions dataset. This dataset lists the currency that each currency used in each year as early as 1999. We assume that a country did not introduce a new currency before 1999 if it did not withdraw a currency before 2004. We allow for this 5-year gap between introducing and withdrawing a currency because countries sometimes introduce a

Table B.4: Countries where Splicing was Possible

Situation	Number.of.Countries
No splice as pre-1993 currency not guessed	55
No splice as Eikon lacks data on pre-1993 currency	13
No splice as Eikon NERs start no earlier than IMF NERs	35
No splice as Eikon and IMF differ on overlapping days	11
Splice made	51

new currency and withdraw the old one a few years later.[4] We determine if the country withdrew a currency before 2002 using the list of discontinued currencies that accompanies the ISO-4217 standard for currency codes.[5]

There are 11 countries where splicing was not possible because the Eikon and IMF NERs differ during an overlapping period. This check could, in principle, detect cases where the country's currency has been guessed incorrectly. However, the series tend to be broadly similar, suggesting that the currency has been guessed correctly, but Eikon and IMF providing different exchange rates for the same currency, such as a black market rate versus an official rate.

B.3.2 Monthly Consumer Price Index Levels

B.3.2.1 World Bank Dataset of Monthly CPI Levels

The World Bank CPI dataset provides a variety of inflation measures for a large set of countries since 1970. We use the monthly headline CPI indexes. These are available for 171 countries in total, though individual countries drop in and out of the sample. The dataset is described in Ha, Kose, et al. 2021. As the authors do not specify whether the data is seasonally adjusted, we assume that it is non-seasonally adjusted.

[4]For example, the IMF exchange arrangements dataset lists France as using the euro in all years from 1999 onwards. If I did no allow for the gap, I would erroneously conclude that France had not introduced any new currency before the end of 1999, and hence that before 1999 it had always used the currency the IMF lists it as using in 1999, which was the euro. Similarly, in 1998 Russia replaced the old Russian Ruble (ISO code RUR) with the new Russian Ruble (ISO code RUB), but the old Russian Ruble is listed as being withdrawn in 2004.

[5]`https://www.six-group.com/en/products-services/financial-information/data-standards.html`

As such, any seasonal pattern in the CPI index levels will translate into a seasonal pattern in the RERs. Our main estimates rely on these non-seasonally adjusted CPIs.

By restricting ourselves to the monthly dataset, we exclude countries for which only quarterly indexes are available. However, these tend to be the countries that also have shorter histories of nominal exchange rates, with the notable exceptions of Australia and New Zealand.

B.3.2.2 Constructing Real-time Vintages of Monthly CPIs

To determine the latest CPI outcomes known to forecasters at the time of their forecast. To determine this, we need to know the 'publication lag', which is the number of months it takes for the statistical agency to publish a country's CPI after the relevant month.

We estimate the publication lag using the World Bank dataset. Typically, the World Bank dataset reports the latest CPI outcome available when they compiled the dataset. We know the dataset was compiled in January 2023.[6] The latest month for which data is available varies by country (Table B.5). For many countries, the latest observation is December 2022, so the publication lag is estimated to be 1 month. Similarly, for countries where the latest observation is November 2022, October 2022 or September 2022, we estimate the publication lag to be 2, 3 or 4 months respectively.

There are some countries where the latest observation is even earlier than September 2022. Taken at face value, this suggest a publication lag of 5 months or more, which seems implausible. In some of these countries, such as Ghana, the latest observation in the World Bank dataset is not actually the latest outcome published by the statistical agency. In other countries, such as Afghanistan, the statistical agency has suspended its CPI. This means the latest observation is far in the past, but prior to the suspension of the CPI series, the publication lag may have been much shorter. We take the pragmatic approach of setting the publication lag to 4 months wherever the latest observation was before September 2022.

[6]We use the January 2023 vintage of the dataset. The webpage for the dataset says it was last updated on 2 February 2023. Either the dataset was made available on this date, or it was made available slightly earlier than the webpage was updated in some other way on 2 February 2023.

Table B.5: World Bank Monthly CPI Dataset

Latest Observation	Number of Countries	Apparent Publication Lag
Dec 2022	60	1
Nov 2022	36	2
Oct 2022	13	3
Sep 2022	18	4
Earlier	44	5

To construct the real-time CPI vintages for a country, we extract subsets of the latest vintage of CPI outcomes using our estimated publication lag. Each CPI vintage is intended to contain the data available at the end of a specified month. For example, the July 2020 vintage of Belarusian CPI is intended to contain Belarusian CPI available at the end of July 2020. Since Belarus's publication lag is 2 months, we make this vintage by extracting Belarusian CPI levels up to May 2020.

Instead of constructing our own real-time vintages from World Bank data, we could have used the real-time vintages of the OECD's Main Economic Indicators, both those provided on the OECD website and those compiled by the Dallas Fed. This would avoid the need to estimate the publication lags, removing one source of error in our estimates. We decided against this for two reasons. Firstly, some vintages are missing.[7] Second, the OECD vintages are only available for 35 countries, most of which use the Euro or have a floating exchange rate, limiting our ability to evaluate forecasts for real exchange rates governed by other exchange rate regimes.

B.3.2.3 Extrapolating Monthly CPI Levels

The CPI vintage for a particular month contains the data available at the end of that month. We will use that CPI vintage to compute an RER up to the end of the month. Hence, we need to extrapolate the CPI data from the latest observation

[7] The Dallas Fed provides vintages up to Q4 1998, while the OECD website provides vintages from January 2000 onwards, so neither provides vintages for 1999. Additionally, the vintages that the OECD website lists as relating to April 2021, January 2020 and August 2017 are actually duplicates of the vintages for other months.

to the end of the specified month. For example, given the July 2020 Belarusian vintage, we need to extrapolate from the latest observation of May 2020 to the end of July 2020. The number of months by which we need to extrapolate the series is the publication lag, so it varies from 1 to 4 months depending on the country.

Our approach is to use linear extrapolation. i.e. We compute the rate of change for the log CPI from the second-latest month to the latest month, and then extrapolate that forward as far as needed. Since this interpolation does not affect the actual outcomes, just the inputs we provide to our forecasting methods, the quality of our approach to extrapolation should ultimately be judged by the performance of the forecasts.

B.3.3 Trade Weights

An effective exchange rate of a country aggregates together information about that country's trading partners. To do this, we need weights that each country places on its trading partners. We use the trade weights produced by the IMF.[8] The IMF has published eight sets of weights, each referring to a different time period, ranging from 1979-1989 to 2016-2018 (Table B.6). The weights are available for almost all countries. For a given reporting country (i.e. the country whose EER is being calculated), the number of partners with weights varies. For example, in 1979-1989, China has weights for 20 partners, while Iraq only has weights for 11 partners. We use the IMF weights because they cover a longer time period and a larger number of countries than alterantive sources of weights, such as those published by the BIS. The IMF's method for computing these weights is described in Bayoumi et al. 2006.

When constructing real-time vintages of REERs, we assume that the set of trade weights for a period only become available with a 5 year delay. Historically, the delay between the end of a weight reference period and the IMF publishing new weights has varied over time. We assume a 5 year delay to approximate the IMF's current practice.

[8]These weights are contained in two internal databases: the information notice system (INS) and global data source (GDS). The IMF intends for these databases to have the same weights, but at the time I extracted weights from these databases, the IMF had implemented an update to the weights in INS but had yet to update the weights in GDS. For this reason, we use the weights in INS.

Table B.6: Descriptive Statistics for IMF Weights

Period	Number of reporting countries	Average number of partner countries
1979-1989	155	18
1990-1995	187	17
1996-2003	187	20
2004-2006	190	30
2007-2009	191	31
2010-2012	192	24
2013-2015	192	27
2016-2018	192	29

For example, the January 2000 vintage is the first to have access to the 1990-1995 weights.[9] As the weights are published with a lag, the REERs for the latest days must be calculated with the weights for an earlier period. For example, in the January 2022 vintage, the daily REERs from 1 January 2019 to 31 January 2022 must be computed with the 2016-2018 weights, as these are the latest available at the time.[10]

Ideally, our real-time vintages of REERs would not only account for the fact that each set of weights to be published with a lag, but would also account for the fact that a given set of weights are revised over time. For example, in March 2019 the IMF revised the weights for 2004-2006, which had been published some time ago.[11]. Unfortunately, previous vintages of weights are not available, so it is not possible to account for this.

Although our real-time vintages of REERs take into account the tendency of

[9]The aim of our paper is to provide evidence on how useful different methods of temporal aggregation would be if adopted today. For that purpose it is better to provide the forecasting models with data that mimics the delays we expect to see in the future, which is achieved by choosing a 5 year delay. If we instead constructed the vintages using the longer delays that were used historically, our results would be less informative to forecasters choosing a temporal aggregation method today.

[10]The IMF follows the same practice. For this reason, they refer to the '2016-2018' weights as the '2016-Latest' weights. We use the term '2016-2018' weights to emphasise that these weights are based only on trade data for these three years, and will eventually be followed by weights for later periods, such as '2019-2021'.

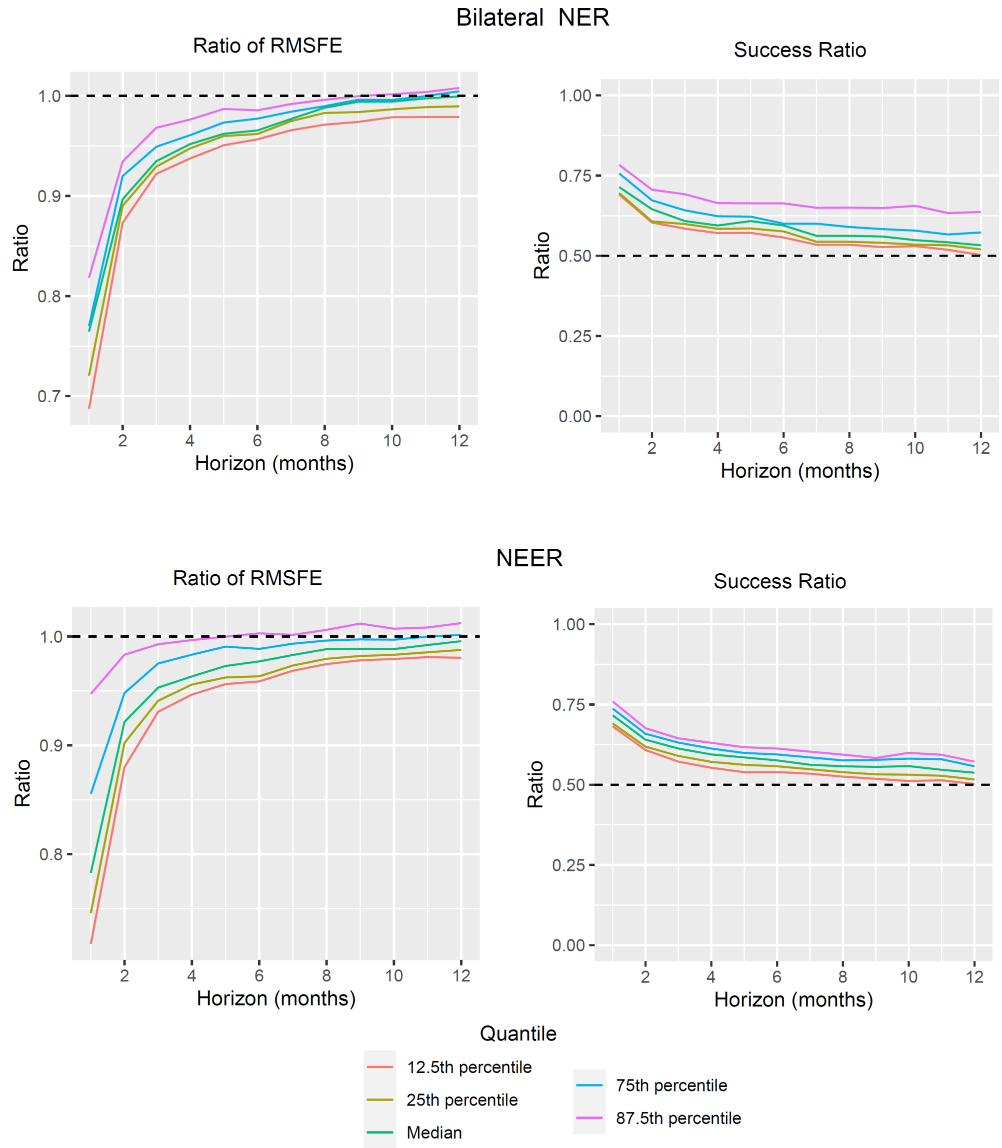
[11]`https://www.imf.org/en/News/Articles/2019/03/26/pr1993-the-imf-updates-the-effective-exchange`

the IMF to publish weights with a lag, they don't take into account the tendency of the IMF to revise the weights over time.

B.4 Real-time Forecast Accuracy for Other Exchange Rates

B.4.1 No-Change Forecasts

Figure B.1: Distribution of Forecast Performance of End-of-Month Versus Monthly Average Forecasts



Note: Plot shows quantiles for 83 countries. Relative to period average no-change forecast.

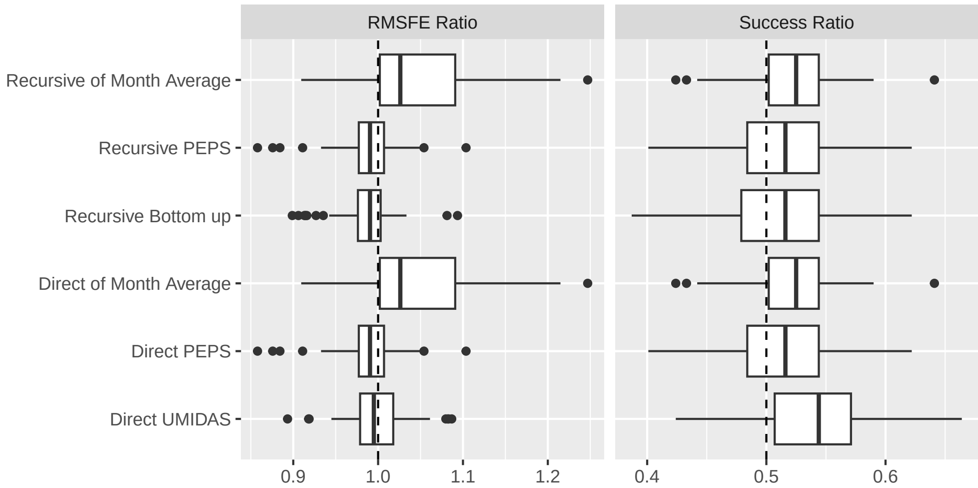
B.4.2 Model-Based Forecasts

Table B.7: Median Monthly Average Real Effective Exchange Rate Forecasts

Forecast	Model Inputs	1	3	6	12	24	36
				RMSFE Ratio			
Recursive	Month-Average	1.00	0.99	0.99	1.00	1.01	1.02
Recursive	PEPS	0.96	0.98	0.98	1.01	1.02	1.01
Recursive	Bottom-up	0.95	0.97	0.99	1.03	1.04	1.04
Direct	Month-Average	1.00	0.98	0.99	0.99	1.03	1.09
Direct	PEPS	0.96	0.97	0.98	0.99	1.04	1.10
Direct	UMIDAS	0.97	0.97	0.98	0.99	1.03	1.09
				Success Ratio			
Recursive	Month-Average	0.51	0.51	0.51	0.53	0.56	0.54
Recursive	PEPS	0.68	0.57	0.54	0.54	0.57	0.55
Recursive	Bottom-up	0.67	0.57	0.53	0.52	0.56	0.53
Direct	Month-Average	0.51	0.51	0.52	0.53	0.56	0.54
Direct	PEPS	0.68	0.57	0.54	0.53	0.56	0.53
Direct	UMIDAS	0.69	0.58	0.54	0.54	0.56	0.53

Note: Reports the median across relative to the monthly average no-change forecast. Note, “end-of-month" inputs in model estimation and uses the point forecast as the forecast of the average. Recursive and daily is an example of the bottom-up approach. Direct forecasts use UMIDAS restricted to the end-of-month observation.

Figure B.2: Accuracy of 1-month-ahead Forecasts for REER relative to End-of-month No-change Benchmark



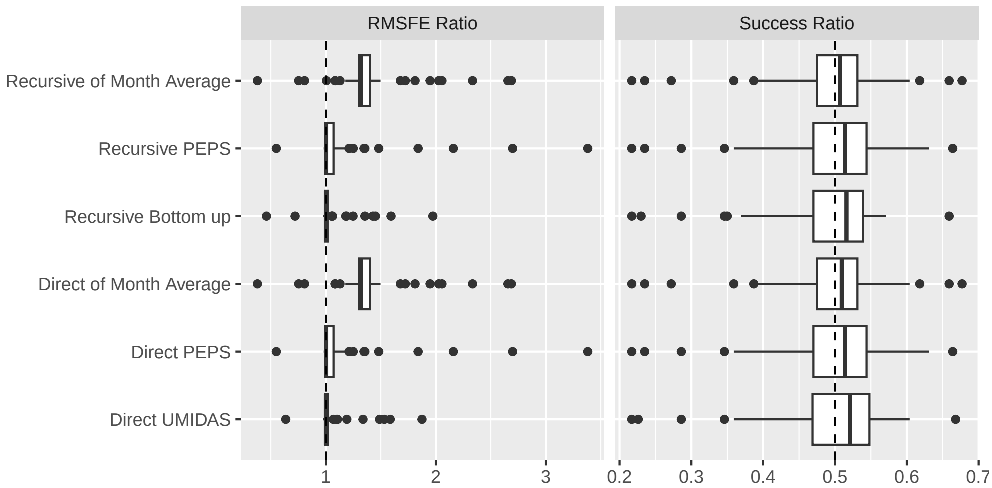
Note: The following outliers have been omitted: Brazil, Tunisia

Table B.8: Median Monthly Average Nominal Bilateral Exchange Rate Forecasts

Forecast	Model Inputs	1	3	6	12	24	36
				RMSFE Ratio			
Recursive	Month-Average	1.00	1.01	1.01	1.02	1.05	1.07
Recursive	PEPS	0.77	0.95	0.99	1.02	1.03	1.07
Recursive	Bottom-up	0.76	0.95	0.98	1.01	1.02	1.06
Direct	Month-Average	1.00	1.01	1.01	1.02	1.08	1.32
Direct	PEPS	0.77	0.95	0.98	1.00	1.07	1.31
Direct	UMIDAS	0.76	0.95	0.98	1.00	1.07	1.31
				Success Ratio			
Recursive	Month-Average	0.51	0.45	0.47	0.50	0.57	0.57
Recursive	PEPS	0.70	0.56	0.53	0.51	0.56	0.58
Recursive	Bottom-up	0.69	0.57	0.52	0.51	0.52	0.55
Direct	Month-Average	0.51	0.45	0.48	0.51	0.57	0.56
Direct	PEPS	0.69	0.57	0.52	0.53	0.57	0.57
Direct	UMIDAS	0.69	0.58	0.53	0.53	0.58	0.57

Note: Reports the median across relative to the monthly average no-change forecast. Note, “end-of-month" inputs in model estimation and uses the point forecast as the forecast of the average. Recursive and daily is an example of the bottom-up approach. Direct forecasts use UMIDAS restricted to the end-of-month observation.

Figure B.3: Accuracy of 1-month-ahead Forecasts for Bilateral NER relative to End-of-month No-change Benchmark



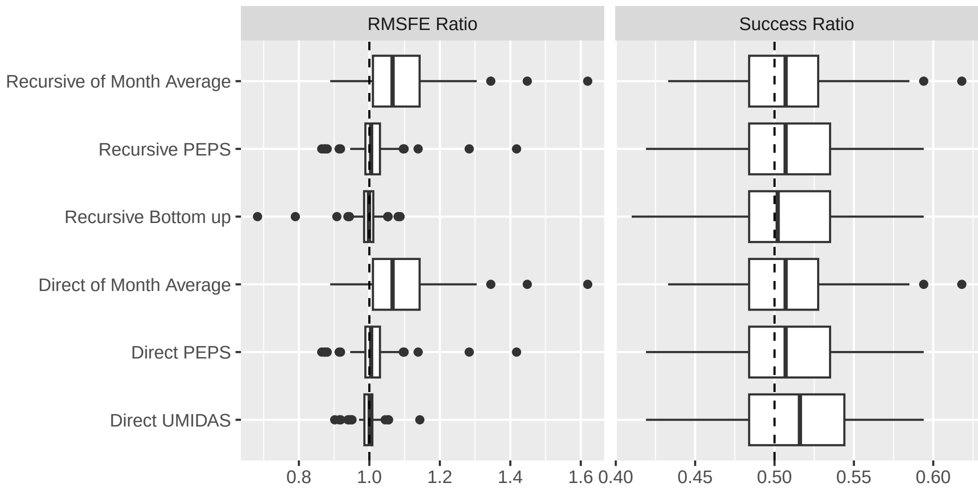
Note: The following outliers have been omitted: Burundi, Jamaica, Romania, The Bahamas

Table B.9: Median Monthly Average Nominal Effective Exchange Rate Forecasts

Forecast	Model Inputs	1	3	6	12	24	36
				RMSFE Ratio			
Recursive	Month-Average	1.00	1.01	1.02	1.05	1.12	1.15
Recursive	PEPS	0.95	0.99	1.01	1.04	1.10	1.11
Recursive	Bottom-up	0.95	0.99	1.04	1.07	1.13	1.14
Direct	Month-Average	1.00	1.01	1.03	1.06	1.14	1.20
Direct	PEPS	0.95	0.99	1.01	1.06	1.14	1.21
Direct	UMIDAS	0.95	0.98	1.01	1.06	1.14	1.19
				Success Ratio			
Recursive	Month-Average	0.50	0.49	0.49	0.50	0.51	0.49
Recursive	PEPS	0.69	0.57	0.53	0.53	0.51	0.48
Recursive	Bottom-up	0.70	0.56	0.52	0.52	0.50	0.48
Direct	Month-Average	0.50	0.50	0.50	0.52	0.51	0.47
Direct	PEPS	0.69	0.57	0.53	0.52	0.51	0.47
Direct	UMIDAS	0.71	0.57	0.53	0.52	0.51	0.48

Note: Reports the median across relative to the monthly average no-change forecast. Note, “end-of-month" inputs in model estimation and uses the point forecast as the forecast of the average. Recursive and daily is an example of the bottom-up approach. Direct forecasts use UMIDAS restricted to the end-of-month observation.

Figure B.4: Accuracy of 1-month-ahead Forecasts for NEER relative to End-of-month No-change Benchmark



Note: The following outliers have been omitted: Romania, Tunisia, Turkey, Fiji

C

Appendices for Abolishing Imputation Credit Refunds

C.1 Detail on Labor's proposed policy

This chapter claims that Labor's proposal would have increased the taxes on domestic shareholders, while leaving taxes on foreign shareholders unchanged. This appendix provides a detailed explanation of why. This appendix begins by describing Australia's dividend imputation system. It then explains how the policy would have affected individual taxpayers and superannuation funds. Finally, it discusses the quantitative importance of the policy, taking into account tax planning strategies.

C.1.1 Illustration of Australia's Dividend Imputation System

C.1.1.1 Distribution of Franking Credits by a Company

To illustrate, suppose the company makes Australian profits of $\pi > 0$, and pays a statutory rate of $\tau \in [0, 1]$. Then:

$$\text{(Franking credits added)} = \text{(Company tax paid)} = \tau\pi$$

Suppose the company distributes all the remaining profits as a cash-dividend. Then:

$$\text{(Dividend amount)} = \text{(Profits after company tax)} = (1 - \tau)\pi$$

The company would like to 'fully frank' the dividend, which means it is attaching as many franking credits to the dividend as is allowed under the rules. The maximum amount allowed is:

$$\text{(Maximum amount of credits)} = \text{(Dividend amount)} \times \frac{\tau}{1-\tau} = (1-\tau)\pi \times \frac{\tau}{1-\tau} = \tau\pi$$

That is, if the company fully franks the dividend, it attaches franking credits equal to the amount of company tax paid on the profits that funded that dividend.

Companies can attach franking credits to some other types of payments, such as off-market share buy-backs. The principle behind attaching franking credits to these other payments is similar, but the details are not relevant for this chapter.

C.1.1.2 Use of the credits by the Australian shareholder

Now suppose the company has one shareholder, who is an Australian resident. Suppose that, in addition to receiving the franked dividend, the individual receives income of ≥ 0 from other sources. The individual must include both the cash amount of the dividend and the amount of any franking credits in their taxable income. Hence, their taxable income is:

$$\text{(Taxable income)} = \underbrace{(1-\tau)\pi}_{\text{Dividend amount}} + \underbrace{\tau\pi}_{\text{Franking credits}} + \underbrace{y}_{\text{Other income}} = \pi + y$$

The individual's personal income tax payable is given by:

$$\text{(Income tax payable)} = f\text{(Taxable income)} - \text{(Franking credit offset)} = f(\pi+y) - \tau\pi \tag{C.1}$$

where $f : \mathbb{R} \to \mathbb{R}$ is a function that calculates the tax payable on a given amount of taxable income.

C.1.1.3 Total tax on profits

The effect of the system is that these profits are subject to personal tax, but not subject to company tax at all. To see this, note that the total tax collected on these profits is:

$$\begin{aligned}
\text{(Total tax on profits)} &= \text{(Company tax)} + \text{(Income tax with dividends)} \\
&\quad - \text{(Income tax without dividends)} \\
&= \tau\pi + f(\pi+y) - \tau\pi - f(y) = f(\pi+y) - f(y)
\end{aligned}$$

C.1.2 The Effect of Labor's Policy on Individuals

C.1.2.1 Effect on Income Tax Payable

In the status quo the income tax payable by the Australian shareholder is given by equation (C.1). For most individual taxpayers, the franking credit offset reduces income tax payable, but to zero. These shareholders do not have excess credits to redeem for cash. However, for some shareholders the offset reduces income tax payable to zero, leaving excess credits, which can then be redeemed for cash.

Labor's proposal was to make the offset 'non-refundable'. i.e. Shareholders would be entitled to use franking credits to reduce income tax payable, but could no longer redeem excess credits for cash. Under this policy, income tax payable would be:

$$
\begin{aligned}
(\text{Income tax payable}) &= max\,\{f(\text{Taxable income}) - (\text{Franking credit offset}), 0\} \\
&= max\,\{f(\pi + y) - \tau\pi, 0\}
\end{aligned}
$$

Most individuals do not have excess credits, so they would not be directly affected. However, individuals with excess credits will see their income tax payable increase. With refundability, their income tax payable is negative, $(f(\pi + y) - \tau\pi) < 0$, but without refundability income tax payable is zero.

C.1.2.2 Effect on total tax on profits

Labor's policy increases total tax payable on profits distributed to shareholders with low taxable incomes. This is because it leaves the company tax on these profits unchanged, while increasing income tax payable from a negative number to zero. Figure C.1 shows the total tax on profits in the case where the shareholder has no income other than franked dividends, y=0. Whenever the total profits are less than \$200,000, the Australian shareholder has excess franking credits. Consequently, whenever profits are below this level, the total tax on profits is higher when these credits are not refundable.

Many shareholders will have substantial other income, $y > 0$. For example, a shareholder of working-age will likely have labour income. These shareholders will

Figure C.1

have fewer excess credits. Consequently, the tax on profits distributed to these shareholders will not increase as much as suggested by Figure C.1.

C.1.3 The Effect of Labor's Policy on Superannuation Funds

The explanation so far has focussed on individual shareholders. However, Labor's policy would also increase the tax paid on profits distributed to superannuation funds, which are a type of retirement saving account popular in Australia. The amount of assets held is very large, making it the world's fourth largest pension market in 2018 (Willis Towers Watson 2019).

The vast majority of superannuation assets are held by defined contribution funds. In these funds, the member and their employer makes contributions to the fund over the member's life. The fund invests the member's contributions according to the member's chosen asset allocation. Upon retirement, the member's account is used to provide a lump-sum, an income stream, or some combination of the two.[1]

As discussed in section 3.3, the tax treatment of Australia's superannuation system is complex, but is roughly a taxed-taxed-exempt ('TTE') system. the superannuation fund will have tax liabilities arising from the contributions and investment earnings of many of its members, and will also receive franking credits on behalf of any members holding Australian shares. The fund can apply franking credits of one member to reduce tax liabilities of another member, provided it compensates the first member.

Funds that have a high proportion of pension members tend to have small tax liabilities relative to the franking credits received, as the earnings of pension members are not taxed. These funds often have substantial excess credits. Usually these are 'self-managed superannuation funds', which are funds with at most four members. Occasionally they are Australian Prudential Regulation Authority regulated superannuation funds, which can have an unlimited number of members.

C.1.4 Quantitative Importance of the Policy

If Labor's policy were implemented, and there were no behavioural response, it would raise revenue equal to the amount of excess franking credits. In the 2014-15 financial year, individuals claimed a total of 2.0 billion AUD of excess franking credits, while superannuation funds claimed 2.9 billion (Parliamentary Budget Office 2018). Hence, absent a behavioural response the total amount that would have been raised in that year was 4.9 billion, which is equal to 84% of taxes on superannuation funds, and 7.5% of company tax in that year (Figure 3.2).

The actual amount of revenue raised is likely smaller due to two types of behavioural responses:

- Tax planning responses reduce the tax on a given amount of profit distributed by Australian companies to Australian shareholders

[1] A minority of Australians have defined benefit funds, where the member makes contributions, and the fund commits to pay benefits according to some formula (say, based on years of service and salary in last year of work). The fund invests the contributions in various assets, and uses those assets to pay the benefits. I do not discuss defined benefit funds in this chapter.

- Real behavioural responses reduce the amount of such profit, for example by companies increasing their reliance on debt financing rather than equity financing

I want to understand the extent to which Labor's policy increases the tax paid on a given amount of profit distributed by Australian companies to Australian shareholders. For this purpose, I want to take into account tax planning responses, but not real behavioural responses.

The amount of excess credits provides an estimate of the revenue increase absent any behavioural response. This suggests a revenue increase in 2014-15 of 4.9 billion. The Parliamentary Budget Office produced estimates of the revenue increase in future years in the presence of both tax planning responses and real behavioural responses. They said that their estimates would have been about 15% higher if they had not allowed for these responses. Applying this 15% figure to the amount of excess credits in 2014-15 suggests a revenue increase in 2014-15 of 4.3 billion. Hence, the revenue increase in 2014-15 if one took into account tax planning responses, but not real behavioural responses, is somewhere in the range of 4.3 to 4.9 billion. These estimates suggest that Labor's policy would increase the taxation of domestic shareholders by a substantial amount, even when tax planning responses are taken into account.

C.2 Newspaper Headlines

I manually recorded the headlines on the front page of each issue of the Australian Financial Review during the announcement window. I classified the headlines into stories related to dividend imputation (indicated in bold) and other stories.

Date	Headline stories (stories related to franking credits shown in bold)
Tue, 13 Mar 2018	**'Labor to cut dividends cash refund'** 'ASEAN in struggle to be relevant' 'What partner gender gap, asks PwC' 'Red hot Emeco talks sale with shareholder' 'Cattle rustling: Terra Firma to offload old Packer empire for $1 billion'
Wed, 14 Mar 2018	**'Labor sets up income tax fight'** 'NAB bankers took cash bribes to falsify loans and earn bonuses'
Thu, 15 Mar 2018	**'Dividend row risk to Labor seat'** **'Super fund boost in share shake-up'** 'Mike Tilley gets behind local cryptoexchange' 'China hardliner to replace Rex Tillerson' 'Some cricket games could go pay TV way' 'Master of the universe: Stephen Hawking'
Fri, 16 Mar 2018	**'Shorten plan to calm share grab backlash'** 'Widodo seeks deeper closer ties with Australia' 'On the campaign trail with Jokowi' 'PM to boost ASEAN as foil against China' 'Singapore doubts US commitment to Asia' 'Brokers obsessed by loan commissions'
Sat, 17 Mar 2018	**'Retirement Collision'** 'Wesfarmers throws Coles off the table' 'One Nation softens on company tax cuts' 'NAB's big blue' 'Russian dossier: The true story of an investigation too hot to handle'
Sun, 18 Mar 2018	No issue published on Sundays
Mon, 19 Mar 2018	**'Batman win emboldens Labor on shares grab'** 'ASEAN's rival China plan' 'ASX tightens listing rules after scandals' 'New visa scheme aims to lure global talent' 'US Fed's Powell to lift above RBA's 1.5pc' 'Deadly sins of the banks'

C.3 Interpretation of $\left(\frac{\mathcal{F}_i^t}{p_i^{t-1}}\right)$

Section 3.5 claimed that the ratio $\left(\frac{\mathcal{F}_i^t}{p_i^{t-1}}\right)$ could be interpreted as a measure of the franking percentages on future dividends. This appendix justifies that interpretation.

C.3.1 Simplifying assumption about $\mathcal{F}_i^t$

The fraction $\frac{\mathcal{F}_i^t}{p_i^{t-1}}$ is the ratio of the expected present value of franking credit distributions on day t and the price of the share on day $t-1$. The numerator is an expectation given information on day t while the denominator is an expectation given information on day $t-1$. To avoid the complications resulting from different information sets in the numerator and denominator, this appendix assumes that the expected present value of franking credit distributions does not change on the day of the announcement, $\mathcal{F}_i^{t-1} = \mathcal{F}_i^t$. This assumption is likely to be approximately satisfied:

- The incentive of companies to engage in tax planning to reduce their Australian company tax bill may be reduced by the dividend imputation system, as the company receives franking credits equal to the company tax they pay, and expected franking credit distributions may be partly or fully capitalised into their share price. If Labor's announcement reduces the extent to which the company's shareholder's value franking credits, the companies may engage in more tax planning than before, reducing company tax paid, and reducing expected franking credit distributions. In practice this effect is likely to be small. Australian companies already have an incentive to engage in substantial tax planning due to Australia's high statutory rate (30% for the largest companies) relative to other countries. If companies minimise their Australian company tax payable as much as is possible given legal and reputational constraints, then their company tax payable and hence franking credit distributions will be unchanged.

- Companies with a positive franking account balance have an incentive to fully frank their dividends, regardless of whether Labor's policy is implemented or not. The reason is that shareholders would rather receive a given amount of franking credits today than the same amount of franking credits later, as they can earn nominal returns on the cash they receive when redeeming credits. Hence, for a company that receives a given amount of credits, the timing with which franking credits are distributed should be unchanged by Labor's announcement.

With the simplifying assumption, the ratio is equal to $\frac{\mathcal{F}_i^{t-1}}{p_i^{t-1}}$. This simplifies the algebra considerably, as it avoids the presence of expectations computed with the information sets of different periods.

C.3.2 Algebra

The appendix now shows that the ratio $\frac{\mathcal{F}_i^{t-1}}{p_i^{t-1}}$ is a strictly increasing function of the franking percentage of a dividend payment on any future date.

Model setup

Let d_t denote the dividend payment (if any) made on date t, and let f_t denote the amount of franking credits attached to dividend payments on that date. Let τ denote the statutory tax rate applicable to the company. Then the franking percentage on a given day, denoted ρ_t, is defined:

$$\rho_t \equiv \frac{f_t}{(\frac{\tau}{1-\tau}) \times d_t}$$

With this definition, the day $t-1$ expected present value of franking credits distributed by the firm, $\mathcal{F}_i^t$, is:

$$\mathcal{F}_i^{t-1} = E\left[\sum_{j=t-1}^{\infty} \frac{1}{(1+r)^j} \left(\frac{\tau}{1-\tau}\right) d_j \rho_j\right]$$

where r is a constant nominal discount rate.

The day $t-1$ price of one share in company i will equal the present value of any future dividends and franking credits paid by that share, where the representative investor also places a value of $\theta \in [0,1]$ per dollar of franking credits received.[2]

$$p_i^{t-1} = E\left[\sum_{j=t-1}^{\infty} \frac{1}{(1+r)^j}\left(d_j + \theta\left(\frac{\tau}{1-\tau}\right)d_j\rho_j\right)\right]$$

The ratio $(\frac{\mathcal{F}_i^{t-1}}{p_i^{t-1}})$ is:

$$\frac{\mathcal{F}_i^{t-1}}{p_i^{t-1}} = \frac{E\left[\sum_{j=t-1}^{\infty}\frac{1}{(1+r)^j}\left(\frac{\tau}{1-\tau}\right)d_j\rho_j\right]}{E\left[\sum_{j=t-1}^{\infty}\frac{1}{(1+r)^j}\left(d_j+\theta\left(\frac{\tau}{1-\tau}\right)d_j\rho_j\right)\right]} \tag{C.2}$$

Claim: This ratio is strictly increasing in the franking percentage ρ_k of any given future period $t = k$.

Proof: Differentiating with respect to ρ_k gives:

$$\frac{\partial\left(\frac{\mathcal{F}_i^{t-1}}{p_i^{t-1}}\right)}{\partial\rho_k} = \frac{p_i^{t-1}.\frac{\partial\mathcal{F}_i^{t-1}}{\partial\rho_k} - \mathcal{F}_i^{t-1}.\frac{\partial p_i^{t-1}}{\partial\rho_k}}{(p_i^{t-1})^2} \tag{C.3}$$

To compute this, I first compute the following two derivatives. Calculating both of these derivatives required interchanging differentiation and the expectation operator, which is possible due to Leibniz's rule.

$$\begin{aligned}\frac{\partial\mathcal{F}_i^{t-1}}{\partial\rho_k} &= \frac{\partial}{\partial\rho_k}E\left[\sum_{j=t-1}^{\infty}\frac{1}{(1+r)^j}\left(\frac{\tau}{1-\tau}\right)d_j\rho_j\right]\\ &= E\left[\frac{\partial}{\partial\rho_k}\sum_{j=t-1}^{\infty}\frac{1}{(1+r)^j}\left(\frac{\tau}{1-\tau}\right)d_j\rho_j\right]\\ &= E\left[\frac{1}{(1+r)^k}\left(\frac{\tau}{1-\tau}\right)d_k\right]\end{aligned} \tag{C.4}$$

and:

$$\begin{aligned}\frac{\partial p_i^{t-1}}{\partial\rho_k} &= \frac{\partial}{\partial\rho_k}E\left[\sum_{j=t-1}^{\infty}\frac{1}{(1+r)^j}\left(d_j+\theta\left(\frac{\tau}{1-\tau}\right)d_j\rho_j\right)\right]\\ &= E\left[\frac{\partial}{\partial\rho_k}\sum_{j=t-1}^{\infty}\frac{1}{(1+r)^j}\left(d_j+\theta\left(\frac{\tau}{1-\tau}\right)d_j\rho_j\right)\right]\\ &= E\left[\frac{1}{(1+r)^k}\theta\left(\frac{\tau}{1-\tau}\right)d_k\right]\end{aligned} \tag{C.5}$$

[2]Brennan (1970) implies that θ is close to 0, as share prices are determined almost entirely by foreign investors. Monkhouse (1993) argued that θ is not too close to either 0 or 1.

Comparing equation C.4 and C.5 shows that:

$$\frac{\partial \mathcal{F}_i^{t-1}}{\partial \rho_k} = E\left[\frac{1}{(1+r)^t}\left(\frac{\tau}{1-\tau}\right)d_k\right] \geq E\left[\frac{1}{(1+r)^t}\theta\left(\frac{\tau}{1-\tau}\right)d_k\right] = \frac{\partial p_i^{t-1}}{\partial \rho_k} \quad \text{for all} \quad \theta \in [0,1]$$

Similarly, comparing the equations for p_i^{t-1} and $\mathcal{F}_i^{t-1}$ shows that:

$$p_i^{t-1} > \mathcal{F}_i^{t-1} \quad \text{for all} \quad \theta \in [0,1]$$

Together these statements imply that:

$$p_i^{t-1}.\frac{\partial \mathcal{F}_i^{t-1}}{\partial \rho_k} > \mathcal{F}_i^{t-1}.\frac{\partial p_i^{t-1}}{\partial \rho_k}$$

Together with C.3, this implies that an increase in the franking percentage of year $t = k$ results in an increase in the ratio, $\frac{\mathcal{F}_i^{t-1}}{p_i^{t-1}}$.

C.3.3 The Magnitude of the Effect of the Announcement on the Market as a Whole

One can estimate how much the announcement would have affected the share market as a whole if share prices are determined by a marginal investor who receives excess credits. The change in the price of one share in company i on announcement day is given by equation (3.3). If S_i^t denotes the number of shares of company i outstanding on day $t-1$, and no new shares are issued on day t, then the change in market capitalisation of the ASX as a whole is:

$$\sum_{i=1}^{N} S_i^{t-1} \Delta p_i^t = -\mathbb{P} \sum_{i=1}^{N} S_i^{t-1} \mathcal{F}_i^t \tag{C.6}$$

i.e. The change in the total market capitalisation of the ASX is minus the probability of the announced policy being implemented $\mathbb{P}$ times the expected present value of all franking credit distributions by listed companies $\sum_{i=1}^{N} S_i^{t-1}\mathcal{F}_i^t$.

Assume that:

- The statutory tax rate is $\tau = 0.3$. This was chosen as it is the statutory rate for large companies in Australia.

- Expected franking credit distributions are fully capitalised into share prices, $\theta = 1$. This follows from the assumption that share prices are determined by a marginal investor who receives excess credits.

- The probability of the policy being implemented is $\mathbb{P} = 0.57$. On the day of the announcement, betting markets had decimal odds of 1.5 for Labor and 2.6 for the Coalition forming government at the next federal election(Koukoulas 2019). This implies a 0.63 probability of winning the upcoming federal election. Assuming a 0.9 probability of Labor implementing its policy conditional on forming government, the probability of the policy being implemented is $0.63 \times 0.9 = 0.57$.

Imposing these values on equation (C.2) gives:

$$\begin{aligned}
\frac{\mathcal{F}_i^{t-1}}{p_i^{t-1}} &= \frac{E\left[\sum_{j=t-1}^{\infty} \frac{1}{(1+r)^j} \left(\frac{0.3}{0.7}\right) \rho_j d_j\right]}{E\left[\sum_{j=t-1}^{\infty} \frac{1}{(1+r)^j} \left(d_j + \left(\frac{0.3}{0.7}\right) \rho_j d_j\right)\right]} \\
&= \frac{\left(\frac{0.3}{0.7}\rho_j\right) E\left[\sum_{j=t-1}^{\infty} \frac{1}{(1+r)^j} d_j\right]}{\left(1 + \frac{0.3}{0.7}\rho_j\right) E\left[\sum_{j=t-1}^{\infty} \frac{1}{(1+r)^j} d_j\right]} \\
&= \frac{\left(\frac{0.3}{0.7}\rho_j\right)}{\left(1 + \frac{0.3}{0.7}\rho_j\right)}
\end{aligned} \tag{C.7}$$

If we assume $\mathcal{F}_i^{t-1} = \mathcal{F}_i^t$ (see section C.3.1), then:

$$\mathcal{F}_i^{t-1} = p_i^{t-1} \frac{\left(\frac{0.3}{0.7}\rho_j\right)}{\left(1 + \frac{0.3}{0.7}\rho_j\right)}$$

$\mathcal{F}_i^t = 0.253 p_i^{t-1}$. Substituting this into C.6, the change in market capitalisation is:

$$\sum_{i=1}^{N} S_i^{t-1} \Delta p_i^t = -\mathbb{P} \frac{\left(\frac{0.3}{0.7}\rho_j\right)}{\left(1 + \frac{0.3}{0.7}\rho_j\right)} \sum_{i=1}^{N} S_i^{t-1} p_i^{t-1} = -0.57 \frac{\left(\frac{0.3}{0.7}\rho_j\right)}{\left(1 + \frac{0.3}{0.7}\rho_j\right)} \sum_{i=1}^{N} S_i^{t-1} p_i^{t-1}$$

Hence the percentage change in market capitalisation on the announcement day is:

$$100 \times \frac{\sum_{i=1}^{N} S_i^{t-1} \Delta p_i^t}{\sum_{i=1}^{N} S_i^{t-1} p_i^{t-1}} = -100 \times 0.57 \frac{\left(\frac{0.3}{0.7}\rho_j\right)}{\left(1 + \frac{0.3}{0.7}\rho_j\right)}$$

This shows that:

- For a company whose franking percentage is $\rho_j = 1$ in all future years, the percentage change in market capitalisation is -17.1%

- For a company whose franking percentage is $\rho_j = 0.79$ in all future years, the percentage change in market capitalisation is -14.4%. Since this is the average franking percentage across all listed companies (see Table 3.2), this is an estimate of the overall return on the share market if the marginal investor is an Australian with excess credits.

C.4 Detail on Panel Dataset

C.4.1 Data on Individual Australian Shares

For each ticker in the sample, I extracted a variety of variables from Bloomberg. First I extracted the Global Industry Classification Standard (GICS) codes. I collected the codes at all four levels of aggregation: sub-industry; industry; industry group; and sector. Second, I extracted daily time series data on each of these tickers. I extracted a variety of different measures of the concepts of interest so that I can compare them. Whenever I extracted data, I ensured it was denominated in Australian dollars. I collected:

- Closing prices. The variable used was BLOOMBERG CLOSE PRICE. I could have used PX CLOSE 1D, but this variable was identical except lagged by one day. I inspected a variety of other measures of closing prices, but they are not available for Australian companies.
- Number of trades. I used NUM TRADE
- Market capitalisation. I used CUR MKT CAP
- Net fixed assets. I used BS NET FIXED ASSET. This is defined as 'Gross fixed Assets less amounts of Accumulated Depreciation.'
- Capital expenditure. I used CAPITAL EXPEND. This is defined as expenditure on tangible fixed assets, though it includes intangible assets for some companies.

The measures of net fixed assets and capital expenditure chosen were those that seemed closest in scope to the capital goods subject to the AIG. This ensures these variables are as effective as possible in controlling for potential omitted variable bias due to the influence of AIG on returns during the announcement window.

C.4.2 Past Franking Percentages

One of our franking proxies is the company's franking percentage over the 5 years leading up to Labor's announcement. This proxy is computed with equation (3.9). I compute the franking percentage with Eikon data and separately with Capital IQ data. If a company has data in both databases, I use the average of the franking percentage based on Eikon data and the franking percentage computed with Capital IQ data. For all companies, I use the franking percentage based on whichever database has data on that company.

C.4.2.1 Eikon Data

The Eikon database provides data on each individual dividend paid by a listed company. For each dividend paid in the 5 years up to the day before Labor's announcement (12 March 2018), I extracted the three variables in Table C.1. In Eikon, a gross dividend amount is defined as the sum of a net dividend amount and the attached franking credits.[3] Hence, in equation (3.9), franking credits are given by Eikon gross dividends less Eikon net dividends, and dividend amounts are Eikon net dividends.

Table C.1: Eikon Data on Franking Percentages

Code	Title
TR.DivFrankedPercent	Franked percent
TR.DivUnadjustedGross	Gross dividend amount
TR.DivUnadjustedNet	Net dividend amount

C.4.2.2 Capital IQ Data

The S&P Capital IQ database provides data on total dividend payments and franking credit distributions in each financial year, rather than providing data on individual dividends. I use the data on for financial years 2013 to 2017. Unlike the Eikon data, it is not possible to make use of data from the 2018 financial year, as there is no way

[3]I confirmed this interpretation by using the gross dividend and net dividend variables to replicate the franking percentages.

to separate data on payments made before 12 March 2018 (which would have been known on the announcement date) and payments made afterwards. The Capital IQ database provides six useful variables, which are listed in Table C.2. As in the Eikon data, a gross dividend amount is defined as the sum of a net dividend amount and the attached franking credits. The first three variables refer to ordinary dividends only.[4] Hence in 3.9, I computed the franking credits distributed by adding the IQ DPS IMPUTED CRD AMT and IQ DPS SPEC IMPUTED CRD AMT, and I computed the dividend amount by adding the IQ DPS NET and IQ DPS SPEC NET.

I identified one extreme observation, which I removed. Blue Sky Alternative Investments Limited is recorded as having distributed $74,235 franking credits per share in FY2010, which is impossible given that it distributed less than 5 cents in dividends per share that year.

Table C.2: Capital IQ Data on Franking Percentages

Code	Title
IQ DPS IMPUTED CRD AMT	Imputation credit amount
IQ DPS GROSS	Gross dividends per share
IQ DPS NET	Net dividends per share
IQ DPS SPEC IMPUTED CRD AMT	Special dividend imputation credit amount
IQ DPS SPEC GROSS	Special dividend gross dividends per share
IQ DPS SPEC NET	Special dividend net dividends per share

C.4.3 Franking Account Balances

Each ASX-listed company must publish a Corporations Act annual report, which must include FAB data. ASX-listed companies must also make other disclosures, such as the Appendix 4E half-year report. However, companies are not required to include FAB in these other disclosures, and in practice, few companies choose to do so.

When companies disclose FAB they typically include three measures:

[4]The variable titles do not specify they refer to ordinary dividends only, but I was able to verify this interpretation by comparing them to the published financial statements of a number of listed companies

- The franking credit balance on the last date of their financial year (the 'unadjusted' balance)

- Additions to or subtractions from franking credit balance that had not occurred by the end of their financial year, but occurred afterwards or are anticipated to occur afterwards

- The franking credit balance available for future years (the 'adjusted' balance). This equals the unadjusted balance plus the adjustments.

In this chapter, FAB is used as a proxy for a company's expected franking credit distributions. For this purpose, the adjusted balance is most appropriate

I collected FAB data from Bloomberg (which covers 670 companies), from Morningstar (which covers 400 companies), and manual collection from the annual reports of listed companies (for 42 companies). Comparing these data sources revealed some large errors in the Bloomberg data, and some smaller errors in the Morningstar data.

There were a few tickers for which Bloomberg reported FAB that was 1000 times higher than Morningstar or manual collection. One such ticker is Harvey Norman. To show this occurred, an excerpt from their annual report is presented here (Table C.3). The first row contains unadjusted balances, the second and third rows contain adjustments, and the last row contains adjusted balances. The adjusted balance for June 2017 is 543 million AUD, which differs from Bloomberg which says 543134 million AUD. This is an error by a factor of 1000, which likely arises because Bloomberg thought the figures in the table were in millions rather than in thousands. However, the adjusted balance for June 2018 in Bloomberg is correct.

Another source of errors is the fact that some companies disclose their FAB in a non-standard way. Consider this disclosure in Rio Tinto's FY 2018 report:

"The approximate amount of the Rio Tinto Limited consolidated tax group's retained profits and reserves that could be distributed as dividends and franked out of available credits that arose from net payments of income tax in respect of periods

Table C.3: Excerpt from Harvey Norman's Annual Report for FY2018

	CONSOLIDATED June 2018 \$000	CONSOLIDATED June 2017 \$000
Franking Account Balance:		
The amount of franking credits available for the subsequent financial years are:		
- franking account balance as at the end of the financial year at 30%	**590,529**	564,369
- franking credits that will arise from the payment of income tax payable as at the end of the financial year	**4,900**	36,008
- franking credits that will be utilised in the payment of proposed final dividend	**(85,952)**	(57,243)
The amount of franking credits available for future reporting years	**509,477**	543,134

up to 31 December 2018 (after deducting franking credits expected to be utilised on the 2018 final dividend declared) is US\$6,178 million."

The manually collected data performs two adjustments to calculate Rio Tinto's FAB. Rio Tinto discloses that its FAB is just sufficient to attach the maximum amount of credits to a US\$6,178 million distribution. To calculate its FAB, this amount must be multiplied by $\frac{\tau}{1-\tau} = \frac{0.3}{1-0.3}$, which implies its FAB must be US\$2,647 million. This amount must then be converted into AUD using the exchange rate prevailing at the end of the FY2018 financial year, which implies it is A\$3,751 million. Due to these kinds of difficulties, Morningstar does not report a FAB for Rio. Bloomberg does report a FAB, but the figure is incorrect as they have not applied the $\frac{0.3}{1-0.3}$ factor.

In this chapter I use manually collected data for the 42 companies. If this is unavailable, I use Morningstar data. If Morningstar is also unavailable, I use Bloomberg. This ensures I use data that is as accurate as possible. A further advantage is that, by manually collecting the data, I can ensure I use adjusted balances rather than unadjusted balances.

C.4.4 Metadata on Annual Reports

As discussed above, the FAB data in Bloomberg, Morningstar is largely derived from annual reports, and the manually collected data is also derived from annual reports. To make use of this data, I had to collect metadata variables on annual report.

Annual report publication dates If investors took FAB into account then making trading decisions during the event windows, they will have based their decisions on the latest available FAB data known at the time. Determining what FAB data was known at the time is not trivial, as each company can choose its own financial year, and each can choose when to publish its annual report. To determine the latest annual report that had been published by each company before the announcement date, I collected data on annual report publication dates. Strictly speaking, I needed data on the publication dates of Corporations Act annual reports, not other periodic disclosures such as Appendix 4D preliminary final reports, because companies typically only disclose their FAB in the Corporations Act annual report. Unfortunately, after investigating the variables in the Bloomberg and Eikon databases, it became that neither database contained a variable that reported the publication date of the Corporations Act annual report specifically. For this reason, I collected these dates from company press releases announcing the publication of this report. These press releases are available on an Australian website for retail investors called Hot Copper.

Presentation currency A challenge when manually collecting FAB data is that some companies present their financial statements in currencies other than Australian dollars. To facilitate manual collection, I needed data on the 'presentation currency' of each company's annual reports. I collected data on the presentation currency of individual companies from Bloomberg, Reuters, and manual collection from company reports. I found the three measures were usually the same. Differences arose when a company changed their presentation currency from one report to the next. To my knowledge, Bloomberg and Eikon do not allow users to specify a particular year and extract the presentation currency for that year. For this reason, I manually collected the presentation currencies for the specific companies and years needed.

Financial year end dates When a company reports its FAB as at the end of financial year in foreign currency, it will have converted the FAB amount in AUD to the FAB amount in that currency using the exchange rate prevailing on the end of financial year. To reverse this calculation, I need to know the financial year end date of each company. I collected financial year end dates from Bloomberg, Eikon and manual collection. The three data sources agreed in almost all cases.

The differences seem to arise when a company changes its financial year from one year to the next. For this reason, I use manually collected financial year end dates where possible, since this allows me to ensure I use the financial year end date for the specific annual report of interest.

C.5 Robustness Checks

C.5.1 Did Returns Differ between Large and Small Companies?

The results of previous sections may mask important heterogeneity between large and small companies. Foreign investors tend to invest predominantly in larger listed companies, and rarely invest in smaller ones. This suggests that larger listed companies may have their share prices determined by foreign investors (as in Brennan (1970)) while smaller listed companies may have their share prices determined by domestic investors. If this is the case, then Labor's announcement should have no effect on the share prices of larger companies (whose prices are determined by foreigners who cannot use franking credits), but reduce the share prices of smaller companies (whose prices are determined by Australians who can use franking credits). This motivates two exercises. Firstly, did large companies have higher returns than small ones? Section C.5.1 finds little evidence that this was the case. Secondly, does the relationship between returns and franking proxies differ between large companies and small companies? Section C.5.2 finds no evidence or weak evidence for this claim, depending on the specification considered.

C.5.1.1 Aggregate Evidence

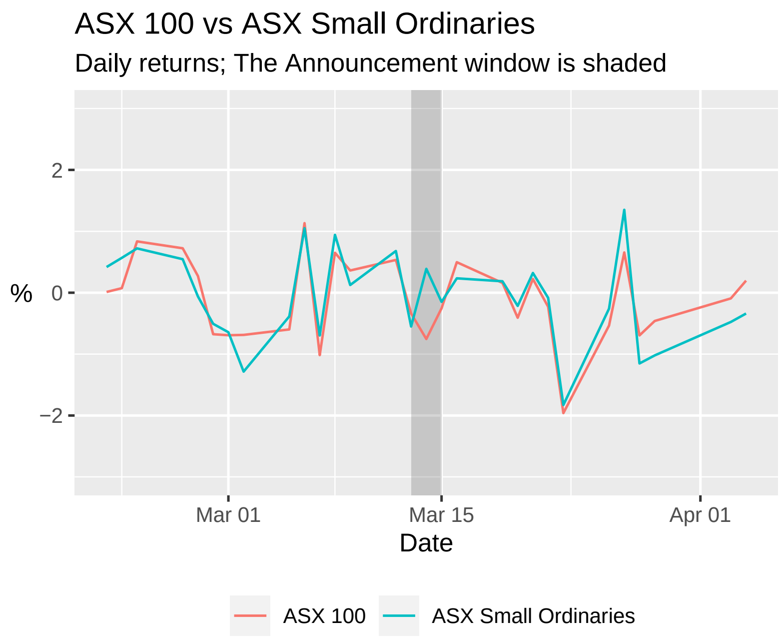
I compare share price indexes for large companies to indexes for small companies. For the present purpose, a 'large' company is one that foreign investors are likely to invest in. However, it is not obvious where the line should be drawn.

First, I adopt a narrow definition of 'large company' by comparing:

- **The S&P/ASX 100:** "The index measures the performance of the 100 largest stocks listed on the ASX by float-adjusted market capitalization" (S&P Dow Jones Indices 2023b)
- **The S&P/ASX Small Ordinaries:** "The index measures the performance of companies included in the S&P/ASX 300, but not in the S&P/ASX 100" (S&P Dow Jones Indices 2023b)

If Labor's announcement reduced the returns of small companies relative to large ones, then the index for small companies should have lower returns than large ones during the event window. For these indexes, the returns for smaller companies are actually slightly higher than for large ones (Figure C.2) on average during the announcement window. Hence this aggregate evidence provides no evidence for this particular story.

Figure C.2

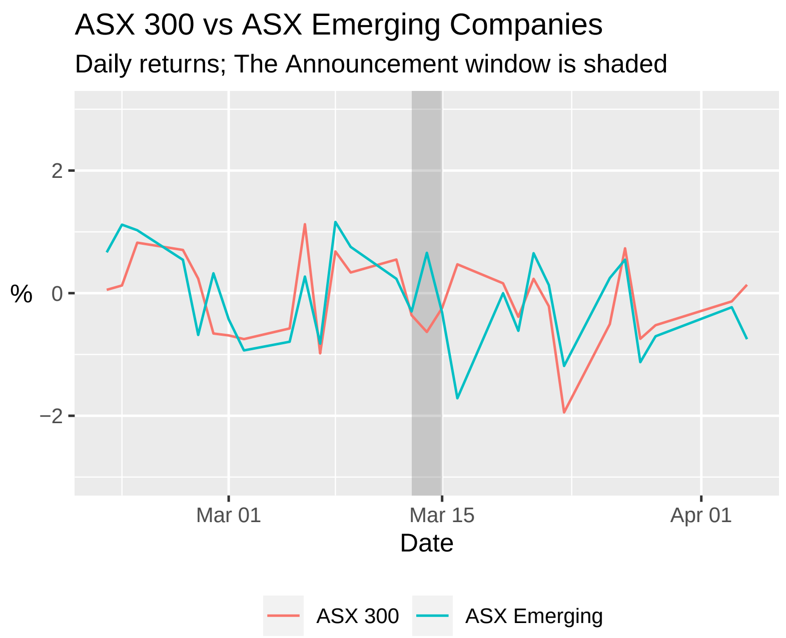


Second, I adopt a broad definition of large company by comparing. Once again, we find that the returns on small companies is higher than for large companies (Figure C.3).

- **S&P/ASX 300:** "The index measures the performance of 300 of the largest, highly liquid securities listed on the ASX by float-adjusted market capitalization." (S&P Dow Jones Indices 2023b)

- **S&P/ASX Emerging Companies:** The index measures the performance of up to 200 companies that are ranked between 350 and 600 in market capitalisation (S&P Dow Jones Indices 2023a)

Figure C.3



The ideal definition of large company may be the S&P ASX 200, which is the "key institutional benchmark for the Australian market" (S&P Dow Jones Indices 2023b). This definition would be between the broad and narrow definitions adopted above. Unfortunately, there is no index measuring companies outside the ASX 200.

A limitation of the aggregate evidence is that the returns in each index will be affected by many factors other than the announcement. For this reason, it is useful to present evidence on abnormal returns from panel regressions, as I do now.

C.5.1.2 Panel Data Evidence

To complement the aggregate evidence, I estimated panel data regressions. Specifically, I estimated the baseline, firm characteristics and full specifications (equations (3.16), (3.17), (3.18)), but the franking proxy w_i was replaced by a dummy for whether the company is in the ASX 200. If Labor's announcement had no effect on the prices of large companies but reduced the prices of smaller ones, then the abnormal returns of large companies should be higher than for small ones. However, there is no evidence this is the case (Figure C.4). The results are similar if we use a dummy for whether the company is in the ASX 200, which as suggested above, is likely a better definition of whether a company is 'large' for the present purposes (Figure C.5

Figure C.4

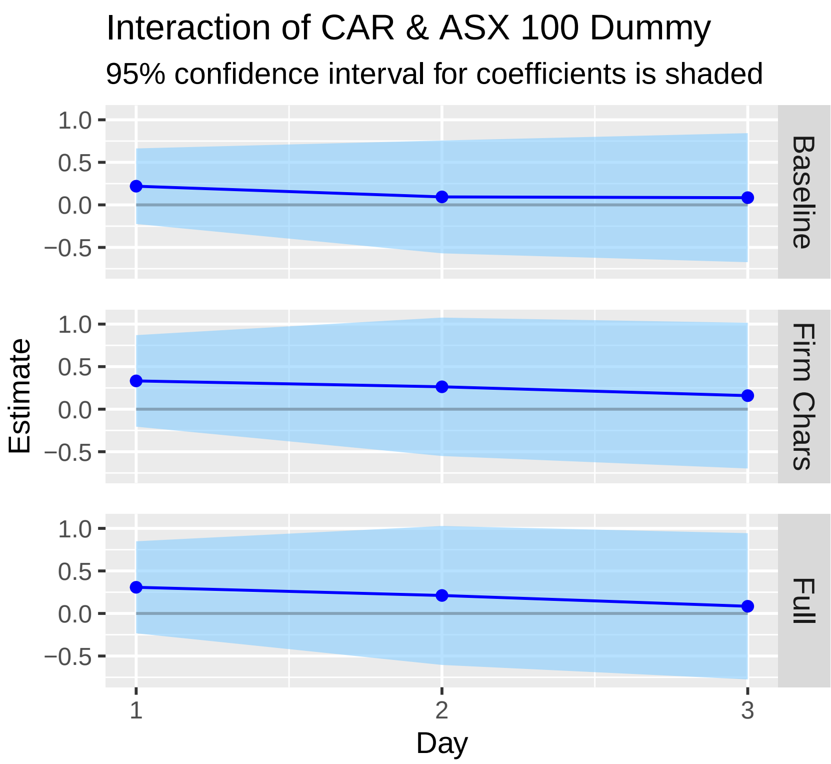
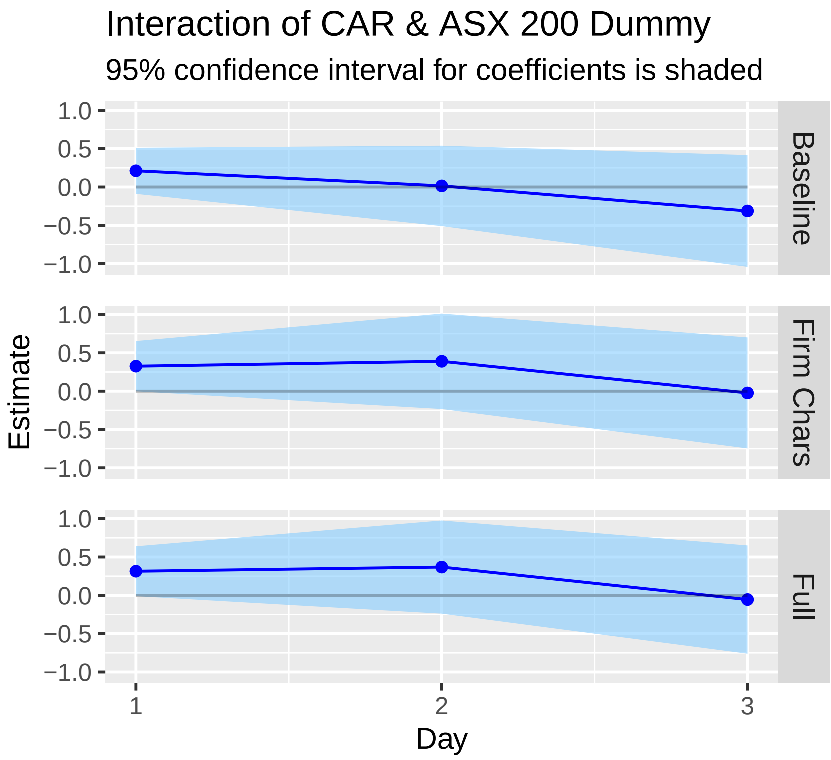


Figure C.5



C.5.2 Does the Relationship between Returns and Franking Proxies Differ between Large and Small Companies?

If Labor's policy reduced the share prices of small companies but not of large companies, then abnormal returns of small companies during the event window should be negatively related to franking proxies, while the abnormal returns of large companies should be unrelated to franking proxies.

To test this, I estimated panel data regressions similar to the baseline, firm characteristics and full specifications (equations (3.16), (3.17) and (3.18)). The only change is that in addition to including the franking proxy w_i, I included the ASX 200 dummy and an interaction of the franking proxy and ASX dummy. The coefficients $\lambda^{j,\text{proxy}}$ measure the association between abnormal returns and the proxy for small companies.

$$r_{i,t} = \mu_i + m_t\beta_i + \sum_{j=1}^{5} D_t^j\gamma^j + \sum_{j=1}^{5} D_t^j\left((ASX200dummy)\right)\lambda^{j,\text{ASX}} + \sum_{j=1}^{5} D_t^j w_i \lambda^{j,\text{proxy}}$$
$$+ \sum_{j=1}^{5} D_t^j w_i\left((ASX200dummy)\right)\lambda^{j,\text{interaction of proxy and ASX}} + \sum_{j=1}^{5} D_t^j z_i\delta + q_t'\eta_g + v_{i,t}$$

I then estimated an identical regression, but with a 'Not ASX' dummy instead of an 'ASX dummy'. In this regression, the coefficients on the proxy measures the association between abnormal returns and the proxy for large companies.

If Labor's policy reduced the returns of small companies but not large ones, then there should be a negative relationship between returns and the proxy among small companies, but not among large ones. For the past franking percentage, there is no evidence of a negative relationship for large companies or for small ones (Figure C.6). Similarly, there is no evidence of a negative relationship between returns and the FABMC proxy for large companies or for small ones (Figure C.7).

Figure C.6

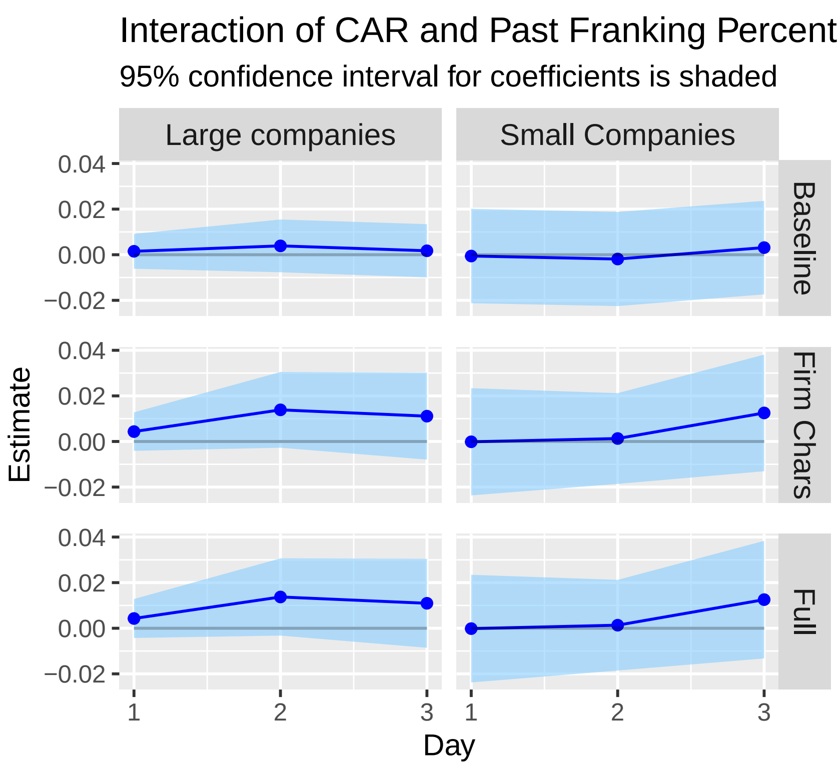
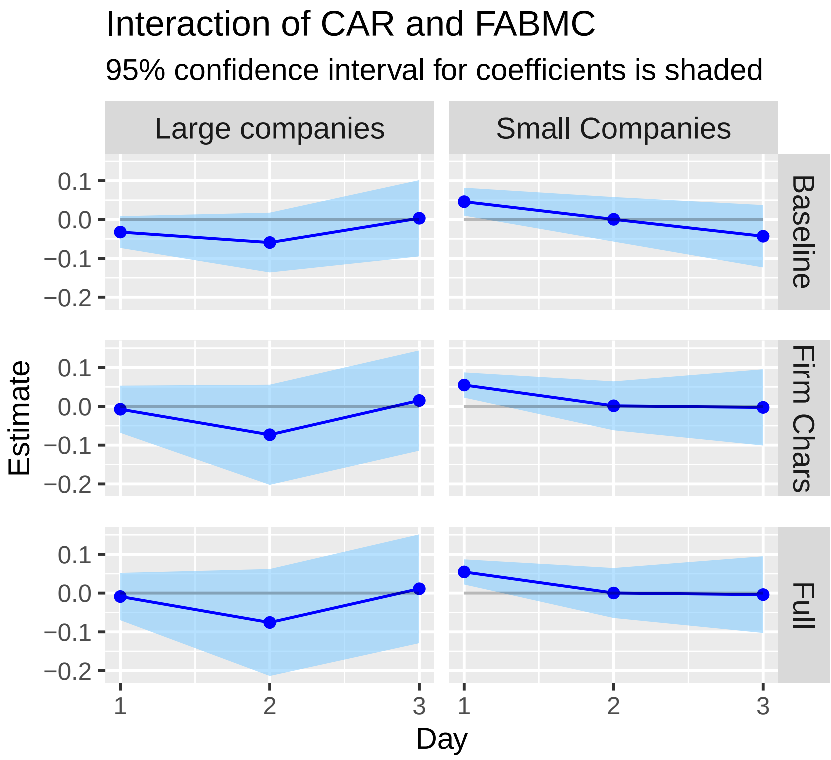


Figure C.7



C.5.3 Results for Other Event Windows

The previous sections presented results for the announcement window, as this window is the ideal setting in which to test whether the franking credit policy affected share prices (see section 3.4). However, it is possible to extend the methods used earlier to simultaneously estimate the CARs for all three windows. This is done by simply adding dummies to the previous specifications. For example, the baseline specification can be written:

$$r_{i,t} = \alpha_i + m_t\beta_i + \sum_{j=1}^{3} D_t^{annt,j}\gamma^{annt,j} + \sum_{j=1}^{3} D_t^{pension,j}\gamma^{pension,j} + \sum_{j=1}^{3} D_t^{elect,j}\gamma^{elect,j}$$
$$+ \sum_{j=1}^{3} D_t^{annt,j} w_i \lambda^{annt,j} + \sum_{j=1}^{3} D_t^{pension,j} w_i \lambda^{pension,j} + \sum_{j=1}^{3} D_t^{elect,j} w_i \lambda^{elect,j} + v_{i,t}$$

where:

- $D_t^{annt,1}, D_t^{annt,2}, D_t^{annt,3}$ are dummies for each day in the announcement window
- $D_t^{pension,1}, D_t^{pension,2}, D_t^{pension,3}$ are dummies for each day in the pension window

- $D_t^{elect,1}, D_t^{elect,2}, D_t^{elect,3}$ are dummies for each day in the election

A usual, if dividend taxes reduce share prices, then CARs over the announcement window should be negatively related to the proxy, $\Lambda^{annt,3} < 0$. However, the pension window covers a period in which Labor announced an exemption for pensioners. If dividend taxes reduce share prices, then CARs over the pension window should be *positively* related to the proxy. Similarly, the election window covers a time in which the probability of the franking credit policy fell dramatically. Hence if dividend taxes reduce share prices, then CARs over the election window should also be positively related to the proxy.

The estimated model when the proxy is the past franking percentage is shown in Table C.4 and Figure C.8. The estimated model when the proxy is FABMC are in Table C.6 and Figure C.9.

For the announcement window and pension window, there is no statistically significant relationship between CARs and either proxy. The absence of a relationship for the announcement window is consistent with previous sections. The lack of relationship for the pension window is, arguably, unsurprising. As discussed in 3.4.2.2, a limitation of the pension window is that the proposed change would have had a small effect on dividend tax revenue. As such, even if dividend taxes reduce share prices in general, the effect of the pension announcement specifically might have been small and difficult to detect.

For the election window, thee point estimates suggest a positive relationship between CARs and past franking percentage and FABMC, and this relationship is statistically significant for at least some days for each specification. On the face of it, this suggests that share prices rose because the election made investor's expect lower dividend taxes in the future. Unfortunately, the election window does not provide a good setting in which to test the effect of dividend taxes. That is because the Coalition's surprising election victory changed investor's beliefs about a wide variety of other policies, and these other changes will have affected the cross-sectional pattern of returns and the franking proxies. The May 2019 election was characterised, by Australian standards, as having unusually large differences in the policy positions of the two parties. As mentioned in section 3.4.2.3, Labor had promised higher taxes on incomes, housing, superannuation funds and trusts. The Coalition's victory caused investor's to expect these taxes to be lower. This likely increased the returns of many shares, but the increase was likely largest for companies that make most of their profits inside Australia. However, since these firms will pay more Australian company tax as a share of their profits, they will have higher past franking percentages and higher FABMC. Hence the election would be expected to cause a positive correlation between returns and FABMC, even if dividend taxes do not affect share prices.

A natural concern is that, if omitted variable bias is thought to affect the estimated relationship between CARs and the franking proxy, could this be a problem in the other windows too? Fortunately, the announcement window is likely to provide a much cleaner test of the effect of Labor's announced changes. As discussed in section 3.4.2.1, Australia did not have any signifnicant macroeconomic data releases, policy decisions or elections during that time. Moreover, the financial press did not report any major market-moving stories during that period (appendix C.2).

Table C.4

Estimated Relationship between Cumulative Abnormal Returns and Past Franking Percent

	Baseline		Firm Characteristics		Full	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Announcement Day 1	0	0.007	0.002	0.007	0.002	0.007
Announcement Day 2	0	0.007	0.006	0.008	0.006	0.008
Announcement Day 3	0.003	0.007	0.012	0.010	0.012	0.010
Pension Day 1	-0.004	0.006	-0.002	0.005	-0.002	0.005
Pension Day 2	-0.002	0.007	0.004	0.006	0.004	0.006
Pension Day 3	-0.008	0.009	0.003	0.007	0.003	0.007
Election Day 1	0.019*	0.004	0.019*	0.006	0.02*	0.006
Election Day 2	0.024*	0.008	0.015	0.012	0.016	0.013
Election Day 3	0.027*	0.008	0.018	0.013	0.019	0.013

Note: * is significant at 5%, ** is significant at 1%, *** is significant at 0.1%

Table C.6

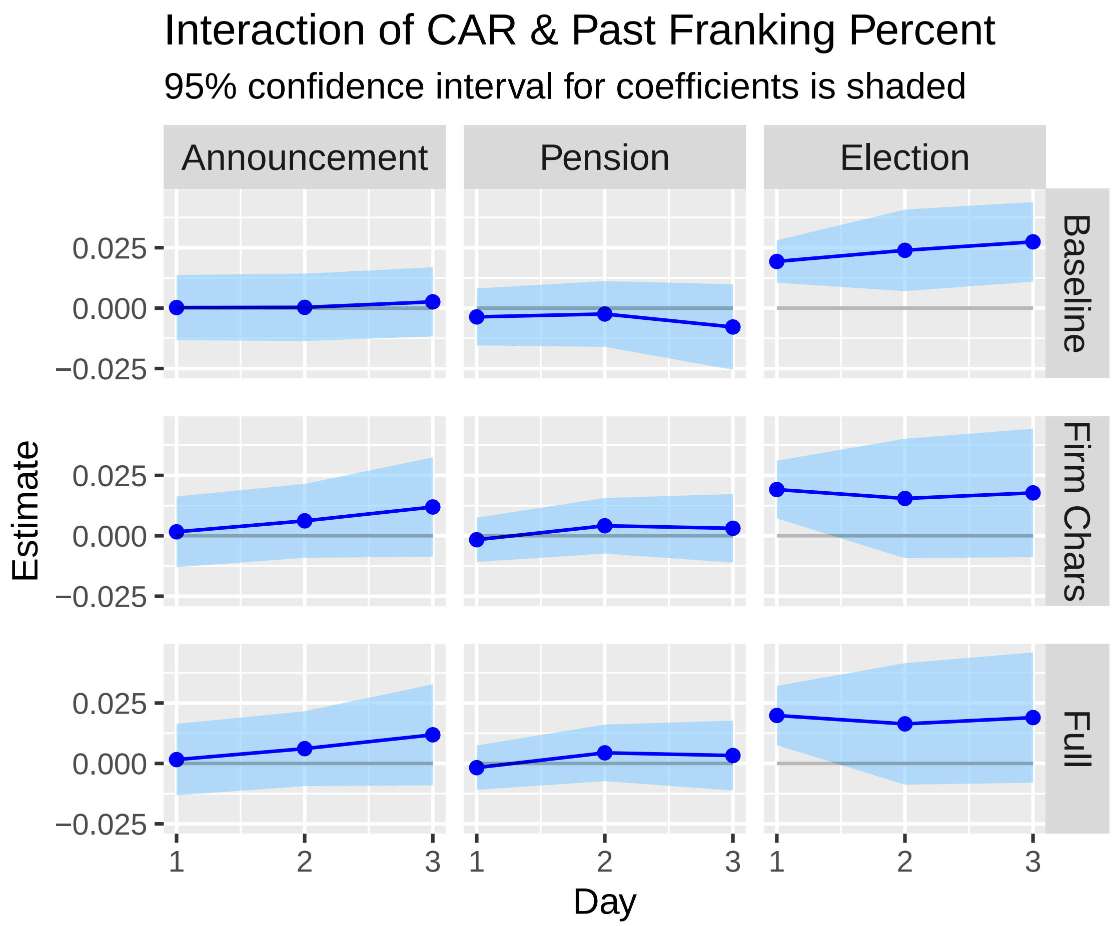
Estimated Relationship between Cumulative Abnormal Returns and FABMC

	Baseline		Firm Characteristics		Full	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Announcement Day 1	0.033*	0.016	0.042*	0.012	0.041*	0.012
Announcement Day 2	-0.007	0.026	-0.013	0.031	-0.014	0.032
Announcement Day 3	-0.031	0.035	0.004	0.044	0.003	0.045
Pension Day 1	-0.003	0.028	-0.041	0.032	-0.041	0.033
Pension Day 2	-0.011	0.034	-0.03	0.042	-0.025	0.041
Pension Day 3	0.012	0.044	-0.029	0.056	-0.021	0.057
Election Day 1	0.082*	0.025	0.093*	0.021	0.098*	0.021
Election Day 2	0.098*	0.031	0.085*	0.030	0.094*	0.030
Election Day 3	0.176*	0.077	0.178	0.098	0.193	0.101

Note: * is significant at 5%, ** is significant at 1%, *** is significant at 0.1%

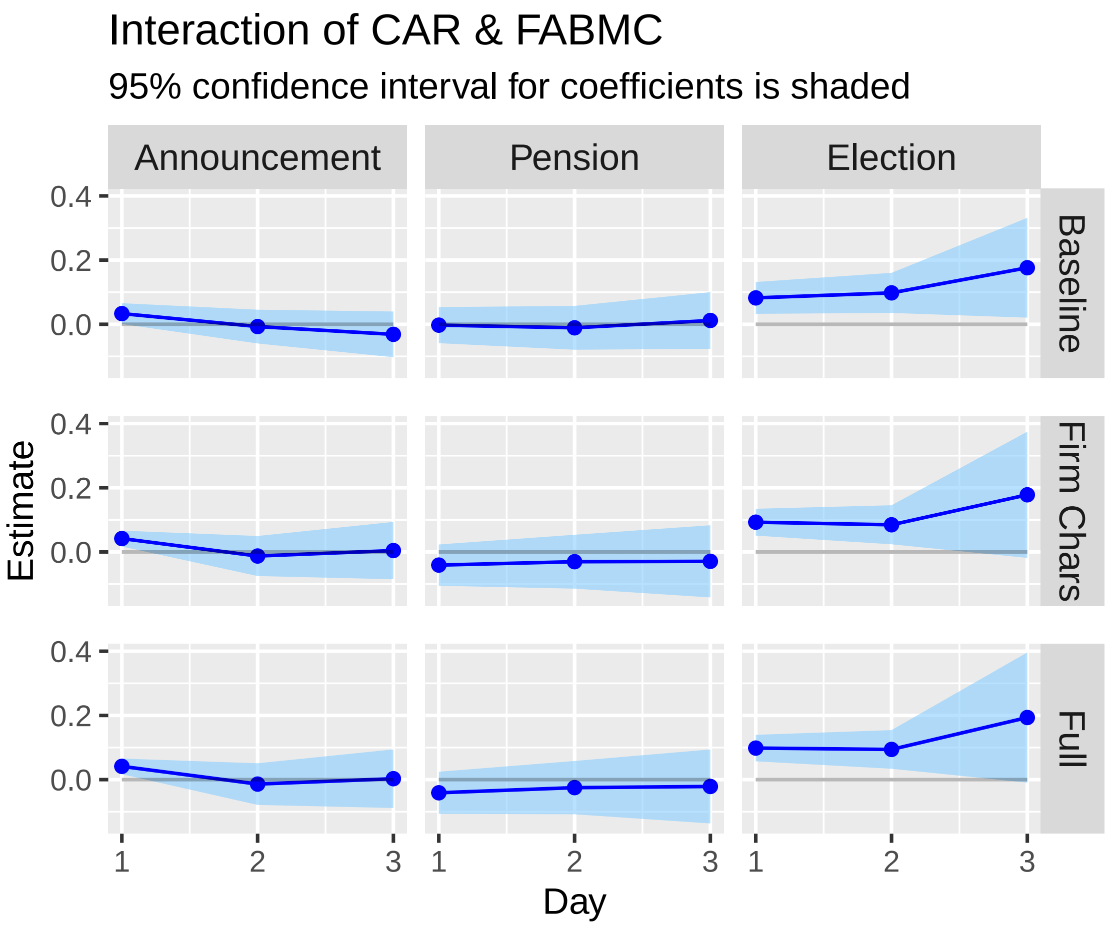
-

Figure C.8


Interaction of CAR & Past Franking Percent
95% confidence interval for coefficients is shaded
Announcement
Pension
Election
Baseline
Firm Chars
Full
Estimate
0.025
0.000
−0.025
1
2
3
Day

-

Figure C.9



C.5.4 Long Estimation Sample

We estimated the panel data models over an 18-month sample. One concern might be that the estimation sample is too short, leading to imprecision in the estimated coefficients on the company fixed-effects and time-varying controls, contributing to imprecision in the estimated coefficients on the interaction between CARs and franking proxies. Figures C.10 and C.11 show estimated slope coefficients when the regressions are estimated on a 36 month sample, which are qualitatively similar to the results shown for the 18 month sample in section 3.8.

Figure C.10

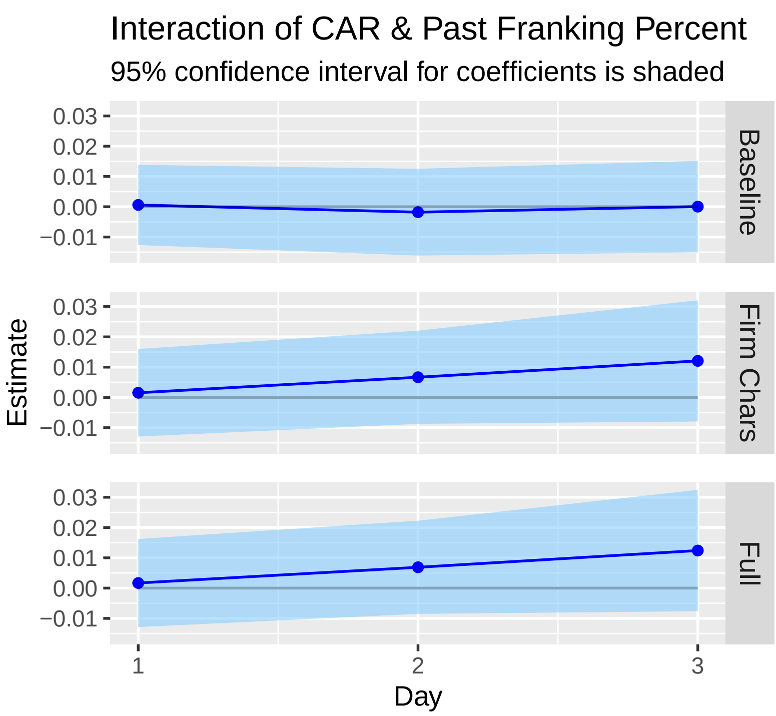
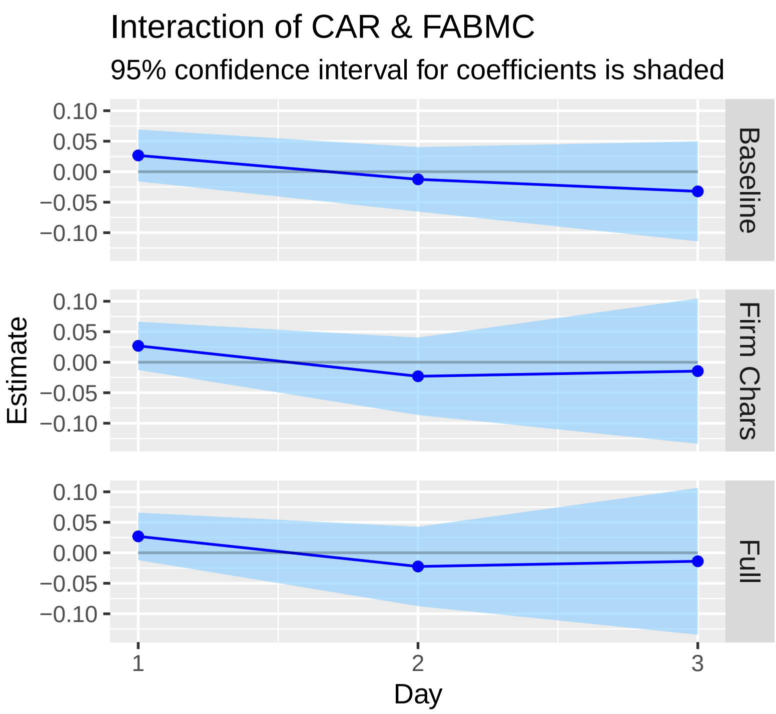


Figure C.11



References

Abbate, Angela and Massimiliano Marcellino (2018). “Point, interval and density forecasts of exchange rates with time varying parameter models”. In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 181.1, pp. 155–179.

Abreu, Ildeberta (July 2011). *International organisations’ vs. private analysts’ forecasts: an evaluation*. Working Papers w201120. Banco de Portugal, Economics and Research Department.

Afonso, Antonio (2010). “Long-term government bond yields and economic forecasts: evidence for the EU”. In: *Applied Economics Letters* 17.15, pp. 1437–1441.

Ainsworth, Andrew B. et al. (2018). “Taxes, Order Imbalance and Abnormal Returns around the ex-Dividend day”. In: *International Review of Finance* 18.3, pp. 379–409. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/irfi.12155`.

Alessi, Lucia et al. (Oct. 2014). “Central Bank Macroeconomic Forecasting During the Global Financial Crisis: The European Central Bank and Federal Reserve Bank of New York Experiences”. In: *Journal of Business & Economic Statistics* 32.4, pp. 483–500.

Amemiya, Takeshi and Roland Y Wu (1972). “The effect of aggregation on prediction in the autoregressive model”. In: *Journal of the American Statistical Association* 67.339, pp. 628–632.

Amromin, Gene, Paul Harrison, and Steven Sharpe (2008). “How Did the 2003 Dividend Tax Cut Affect Stock Prices?” In: *Financial Management* 37.4, pp. 625–646. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1755-053X.2008.00028.x`.

Artis, Michael J. (1996). “How Accurate Are the Imf’s Short-Term Forecasts? Another Examination of the World Economic Outlook”. In: *Staff Studies for the WEO*.

Ashiya, Masahiro (2006). “Testing the rationality of forecast revisions made by the IMF and the OECD”. In: *Journal of Forecasting* 25.1, pp. 25–36.

Auerbach, Alan J. (2002). “Taxation and corporate financial policy”. In: *Handbook of Public Economics*. Ed. by A. J. Auerbach and M. Feldstein. Vol. 3. Handbook of Public Economics. Elsevier. Chap. 19, pp. 1251–1292.

Auerbach, Alan J. and Kevin A. Hassett (July 2005). *The 2003 Dividend Tax Cuts and the Value of the Firm: An Event Study*. NBER Working Papers 11449. National Bureau of Economic Research, Inc.

Australian Broadcasting Corporation (Mar. 2018). “Labor wants to scrap a policy which means some who pay no tax get 2.5 million refund”. In: URL: `https://www.abc.net.au/news/2018-03-13/labor-plan-scrap-5-billion-year-shareholder-refund-policy/9541016`.

— (2019). “Labor’s policy revolution finally detailed, but will voters buy it?” In: URL: `https://www.abc.net.au/news/2019-05-10/federal-election-2019-labor-reveals-costings/11100352?nw=0`.

Australian Financial Review (Mar. 2018). “Labor to end share dividend cash perk in 59b grab”. In: URL: `https://www.afr.com/politics/labor-to-end-share-dividend-cash-perk-in-59b-grab-20180312-h0xcxp`.

— (June 2023). “Government to consider amending franking credit changes”. In: URL: `https://www.afr.com/policy/tax-and-super/government-to-consider-amending-franking-credit-changes-20230602-p5ddh1`.

Australian Government (2015). *Final Budget Outcome 2014-15*. URL: https://archive.budget.gov.au/2014-15/fbo/FBO-2014-15-Consolidated.pdf.

Australian Prudential Regulation Authority (Aug. 2022). *Quarterly superannuation performance statistics - September 2004 to June 2022*. URL: https://www.apra.gov.au/quarterly-superannuation-statistics.

Australian Taxation Office (Dec. 2016). *Franking credit trading*. URL: https://www.ato.gov.au/Business/Imputation/Integrity-rules/Franking-credit-trading/.

— (May 2019). *Dividends paid or credited to non-resident shareholders*. URL: https://www.ato.gov.au/Forms/You-and-your-shares-2019/?page=13.

Bank of England (2023). *Monetary Policy Report - February 2023*. London, United Kingdom.

Baqir, Reza, Rodney Ramcharan, and Ratna Sahay (2005). "IMF programs and growth: Is optimism defensible?" In: *IMF Staff Papers* 52.2, pp. 260–286.

Bauer, Michael D. and Glenn D. Rudebusch (Oct. 2016). "Monetary Policy Expectations at the Zero Lower Bound". In: *Journal of Money, Credit and Banking* 48.7, pp. 1439–1465.

Baumgartner, Josef (Dec. 2002). "Evaluation of Macro-economic Forecasts for Austria in the 1980s and 1990s". In: *Austrian Economic Quarterly* 7.4, pp. 191–206.

Bayoumi, Tamim, Jaewoo Lee, and Sarma Jayanthi (2006). "New Rates from New Weights". In: *IMF Staff Papers* 53.2, pp. 1–4.

Beach, W. (1999). "How Reliable Are IMF Economic Forecasts?" In: *Heritage Foundation Report*.

Beaudry, Paul and Tim Willems (Jan. 2022). "On the Macroeconomic Consequences of Over-Optimism". In: *American Economic Journal: Macroeconomics* 14.1, pp. 38–59.

Beggs, David J. and Christopher L. Skeels (2006). "Market Arbitrage of Cash Dividends and Franking Credits*". In: *Economic Record* 82.258, pp. 239–252. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-4932.2006.00337.x.

Bekaert, Geert and Alexander Popov (Dec. 2019). "On the Link Between the Volatility and Skewness of Growth". In: *IMF Economic Review* 67.4, pp. 746–790. URL: https://link.springer.com/article/10.1057/s41308-019-00092-2.

Bell, Leonie and Tim Jenkinson (2002). "New Evidence of the Impact of Dividend Taxation and on the Identity of the Marginal Investor". In: *The Journal of Finance* 57.3, pp. 1321–1346. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1540-6261.00462.

Benmoussa, Amor Aniss, Reinhard Ellwanger, and Stephen Snudden (2020). "Carpe Diem: Can daily oil prices improve model-based forecasts of the real price of crude oil?" In: *Bank of Canada Staff Working Paper, 2020-39*.

Bickel, David R. and Rudolf Frühwirth (2006). "On a fast, robust estimator of the mode: Comparisons to other robust estimators with applications". In: *Computational Statistics & Data Analysis* 50.12, pp. 3500–3530. URL: https://www.sciencedirect.com/science/article/pii/S0167947305001581.

Biorn, Erik (2016). *Econometrics of Panel Data: Methods and Applications*. OUP Catalogue 9780198753445. Oxford University Press. Chap. 5.

Boadway, Robin and Neil Bruce (1992). "Problems with integrating corporate and personal income taxes in an open economy". In: *Journal of Public Economics* 48.1, pp. 39–66. URL: http://www.sciencedirect.com/science/article/pii/004727279290041D.

Bond, Stephen R., Michael P. Devereux, and Alexander Klemm (2007a). "Dissecting Dividend Decisions: Some Clues about the Effects of Dividend Taxation from Recent UK Reforms". In: *Taxing Corporate Income in the 21st Century*. Ed. by Alan J. Auerbach, James R. Hines Jr., and Joel Slemrod. Cambridge University Press, pp. 41–75.

— (Aug. 2007b). *The Effects of Dividend Taxes on Equity Prices; A Re-examination of the 1997 U.K. Tax Reform*. IMF Working Papers 07/204. International Monetary Fund.

Bordalo, Pedro, Nicola Gennaioli, Spencer Yongwook Kwon, et al. (2021). "Diagnostic bubbles". In: *Journal of Financial Economics* 141.3, pp. 1060–1077.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, et al. (Sept. 2020). "Overreaction in Macroeconomic Expectations". In: *American Economic Review* 110.9, pp. 2748–2782.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (Feb. 2018). "Diagnostic Expectations and Credit Cycles". In: *Journal of Finance* 73.1, pp. 199–227.

Brennan, Michael J. (1970). "Taxes, Market Valuation and Corporate Financial Policy". In: *National Tax Journal* 23.4, pp. 417–427. URL: `http://www.jstor.org/stable/41792223`.

Broer, Tobias and Alexandre N. Kohlhas (June 2022). "Forecaster (Mis-)Behavior". In: *The Review of Economics and Statistics*, pp. 1–45. eprint: `https://direct.mit.edu/rest/article-pdf/doi/10.1162/rest\_a\_01210/2032374/rest\_a\_01210.pdf`. URL: `https://doi.org/10.1162/rest%5C_a%5C_01210`.

Brown, Rodney, Youngdeok Lim, and Chris Evans (2020). "The impact of full franking credit refundability on corporate tax avoidance". In: *eJournal of Tax Research* 17.2, pp. 134–167.

Burgess, Matthew G. et al. (July 2021). *Optimistically biased economic growth forecasts and negatively skewed annual variation*. SocArXiv vndqr. Center for Open Science.

Ca'Zorzi, Michele et al. (2022). "The Reliability of Equilibrium Exchange Rate Models: A Forecasting Perspective". In: *International Journal of Central Banking* 18.3, pp. 229–280.

Caglio, Cecilia, Matt Darst, and Sebnem Kalemli-Özcan (Apr. 2022). *Risk-Taking and Monetary Policy Transmission: Evidence from loans to SMEs and large firms*. CEPR Discussion Papers 17175. C.E.P.R. Discussion Papers.

Cameron, Colin and Douglas L. Miller (2015). "A Practitioner's Guide to Cluster-Robust Inference". In: *Journal of Human Resources* 50.2, pp. 317–372.

Cannavan, Damien and Stephen Gray (2017). "Dividend drop-off estimates of the value of dividend imputation tax credits". In: *Pacific-Basin Finance Journal* 46, pp. 213–226. URL: `http://www.sciencedirect.com/science/article/pii/S0927538X17301580`.

Carriero, Andrea, George Kapetanios, and Massimiliano Marcellino (2009). "Forecasting exchange rates with a large Bayesian VAR". In: *International Journal of Forecasting* 25.2, pp. 400–417.

Chen, Hung-Ling, Edward H. Chow, and Cheng-Yi Shiu (2013). "Ex-dividend prices and investor trades: Evidence from Taiwan". In: *Pacific-Basin Finance Journal* 24, pp. 39–65. URL: `http://www.sciencedirect.com/science/article/pii/S0927538X13000188`.

Chen, Shu-Ling, John D Jackson, et al. (2014). "What drives commodity prices?" In: *American Journal of Agricultural Economics* 96.5, pp. 1455–1468.

Chetty, Raj, Joseph Rosenberg, and Emmanuel Saez (July 2005). *The Effects of Taxes on Market Responses to Dividend Announcements and Payments: What Can we Learn from the 2003 Dividend Tax Cut?* Working Paper 11452. National Bureau of Economic Research. URL: http://www.nber.org/papers/w11452.

Clarida, Richard H, Lucio Sarno, et al. (2003). "The out-of-sample success of term structure models as exchange rate predictors: a step beyond". In: *Journal of International Economics* 60.1, pp. 61–83.

Clarida, Richard H and Mark P Taylor (1997). "The term structure of forward exchange premiums and the forecastability of spot exchange rates: correcting the errors". In: *Review of Economics and Statistics* 79.3, pp. 353–361.

Clayton Utz (2019). "Federal Election 2019 – key workplace issues 101: Minimum wages and penalty rates". In: URL: https://www.claytonutz.com/knowledge/2019/march/federal-election-2019-key-workplace-issues-01-minimum-wages-and-penalty-rates.

Clements, Michael P. (1997). "Evaluating the Rationality of Fixed-event Forecasts". In: *Journal of Forecasting* 16.4, pp. 225–239.

— (Mar. 2018). "Do Macroforecasters Herd?" In: *Journal of Money, Credit and Banking* 50.2-3, pp. 265–292.

— (Mar. 2022). "Forecaster Efficiency, Accuracy, and Disagreement: Evidence Using Individual-Level Survey Data". In: *Journal of Money, Credit and Banking* 54.2-3, pp. 537–568.

Clements, Michael P. and David I. Harvey (Mar. 2009). "Forecast Combination and Encompassing". In: *Palgrave Handbook of Econometrics*. Ed. by Terence C. Mills and Kerry Patterson. Palgrave Macmillan Books. Palgrave Macmillan. Chap. 4, pp. 169–198.

Coibion, Olivier and Yuriy Gorodnichenko (2012). "What Can Survey Forecasts Tell Us about Information Rigidities?" In: *Journal of Political Economy* 120.1, pp. 116–159.

— (Aug. 2015a). "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts". In: *American Economic Review* 105.8, pp. 2644–2678.

Davies, Gavyn (Nov. 19, 2017). "Loeys' laws on asset allocation". In: *The Financial Times*. URL: https://www.ft.com/content/1faa00d1-d16c-3f37-aa50-0083ba073ca4.

Debelle, Guy (Oct. 2017). *Uncertainty*. URL: https://www.rba.gov.au/speeches/2017/sp-dg-2017-10-26.html.

Deloitte (2019). *International Tax: Singapore Highlights*. URL: https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Tax/dttl-tax-singaporehighlights-2019.pdf.

Dhaliwal, Dan, Linda Krull, and Oliver Zhen Li (2007). "Did the 2003 Tax Act reduce the cost of equity capital?" In: *Journal of Accounting and Economics* 43.1, pp. 121–150. URL: http://www.sciencedirect.com/science/article/pii/S0165410106000711.

Diebold, Francis X. (Jan. 2015). "Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests". In: *Journal of Business & Economic Statistics* 33.1, pp. 1–1.

Dison, Will and David Elliott (Dec. 2015). *Mean and modal Bank Rate expectations*. URL: https://bankunderground.co.uk/2015/12/03/mean-and-modal-bank-rate-expectations/.

Dornbusch, Rüdiger (1987). "Exchange Rates and Prices". In: *American Economic Review* 77.1, pp. 93–106.

Eicher, Theo S. et al. (2019). “Forecasts in times of crises”. In: *International Journal of Forecasting* 35.3, pp. 1143–1159.

Elliott, Graham, Ivana Komunjer, and Allan Timmermann (Mar. 2008). “Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss?” In: *Journal of the European Economic Association* 6.1, pp. 122–157.

Ellwanger, Reinhard and Stephen Snudden (Sept. 2021). “Seize the Last Day: Period-End-Price Sampling for Forecasts of Temporally Aggregated Data”. In: *LCERPA Working Papers* bm0127.

— (2023a). “Forecasts of the Real Price of Oil Revisited: Do they Beat the Random Walk?” In: *Journal of Banking and Finance* https://doi.org/10.1016/j.jbankfin.2023.106962.

Engel, Charles, Nelson C Mark, Kenneth D West, et al. (2007). “Exchange rate models are not as bad as you think [with comments and discussion]”. In: *NBER Macroeconomics Annual* 22, pp. 381–473.

Engelke, Carola, Katja Heinisch, and Christoph Schult (2019). *How forecast accuracy depends on conditioning assumptions.* IWH Discussion Papers 18/2019. Halle Institute for Economic Research.

Fama, Eugene F. and Kenneth R. French (Sept. 2004). “The Capital Asset Pricing Model: Theory and Evidence”. In: *Journal of Economic Perspectives* 18.3, pp. 25–46. URL: `https://www.aeaweb.org/articles?id=10.1257/0895330042162430`.

Faust, Jon, John H Rogers, and Jonathan H Wright (2003). “Exchange rate forecasting: the errors we’ve really made”. In: *Journal of International Economics* 60.1, pp. 35–59.

Faust, Jon and Jonathan H. Wright (Oct. 2008). “Efficient forecast tests for conditional policy forecasts”. In: *Journal of Econometrics* 146.2, pp. 293–303.

Fioramanti, Marco et al. (Mar. 2016). *European Commission’s Forecasts Accuracy Revisited: Statistical Properties and Possible Causes of Forecast Errors.* European Economy - Discussion Papers 027. Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.

Forbes, Kristin, Ida Hjortsoe, and Tsvetelina Nenova (2018). “The shocks matter: improving our estimates of exchange rate pass-through”. In: *Journal of International Economics* 114, pp. 255–275.

Foroni, Claudia, Massimiliano Marcellino, and Christian Schumacher (2015). “Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials”. In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 178.1, pp. 57–82.

Frankel, Jeffrey A and Andrew K Rose (1995). “Empirical research on nominal exchange rates”. In: *Handbook of International Economics* 3, pp. 1689–1729.

Froot, Kenneth A and Tarun Ramadorai (2005). “Currency returns, intrinsic value, and institutional-investor flows”. In: *The Journal of Finance* 60.3, pp. 1535–1566.

Fuhrer, Jeffrey C. (May 2018). *Intrinsic expectations persistence: evidence from professional and household survey expectations.* Working Papers 18-9. Federal Reserve Bank of Boston.

Gadarowski, Christopher et al. (2007). “Dividend Tax Cut and Security Prices: Examining the Effect of the Jobs and Growth Tax Relief Reconciliation Act of 2003”. In: *Financial Management* 36.4, pp. 89–106. URL: `http://www.jstor.org/stable/30129813`.

Genberg, Hans and Andrew Martinez (2014a). “On the Accuracy and Efficiency of IMF Forecasts: A Survey and Some Extensions”. In: *IEO Background Paper.*

Genberg, Hans and Andrew Martinez (2014b). “User Perspectives on IMF Forecasts: Survey Methodology and Results”. In: *IEO Background Paper*.

Genberg, Hans, Andrew Martinez, and Michael Salemi (2014). “The IMF/WEO Forecast Process”. In: *IEO Background Paper*.

Glas, Alexander and Katja Heinisch (2023). “Conditional macroeconomic survey forecasts: Revisions and errors”. In: *Journal of International Money and Finance* 138, p. 102927. URL: `https://www.sciencedirect.com/science/article/pii/S0261560623001286`.

Goldberg, Pinelopi K and Michael M Knetter (1996). *Goods prices and exchange rates: What have we learned?*

Goot, Murray (Mar. 2016). “The Transformation of Australian Electoral Analysis: The Two-Party Preferred Vote - Origins, Impacts, and Critics”. In: *Australian Journal of Politics and History* 62, pp. 59–86.

Grant, Andrew R., P. Joakim Westerholm, and Winston Wu (2018). “Imputation Credits and Trading Around Ex-Dividend Day: New Evidence in Australia”. Mimeo, Available at SSRN. URL: `https://ssrn.com/abstract=3221900%20or%20http://dx.doi.org/10.2139/ssrn.3221900`.

Ha, Jongrim, Ayhan Kose, and Franziska Lieselotte Ohnsorge (July 2021). *One-Stop Source : A Global Database of Inflation*. Policy Research Working Paper Series 9737. The World Bank.

Hodrick, Robert J and Edward C Prescott (Feb. 1997). “Postwar U.S. Business Cycles: An Empirical Investigation”. In: *Journal of Money, Credit and Banking* 29.1, pp. 1–16.

Hong, Pingfan and Zhibo Tan (2014). “A comparative study of the forecasting performance of three international organizations”. In: *Journal of Policy Modeling* 36.4. Rapid Growth or Stagnation in the U.S. and World Economy?, pp. 745–757. URL: `https://www.sciencedirect.com/science/article/pii/S0161893814000532`.

Hyndman, Rob, George Athanasopoulos, et al. (2022). *forecast: Forecasting functions for time series and linear models*. R package version 8.19. URL: `https://pkg.robjhyndman.com/forecast/`.

IMF (Apr. 2021). *World Economic Outlook: April 2021*.

Independent Evaluation Office (Jan. 2008). *Structural Conditionality in IMF-Supported Programs*.

Internal Revenue Service (2020). *Publication 550 Investment Income and Expenses*, p. 19.

International Monetary Fund (2023). “Assumptions and Conventions”. In: *World Economic Outlook, April 2023*. International Monetary Fund.

Jerić, Silvija Vlah, Davor Zoričić, and Denis Dolinar (Jan. 2020). “Analysis of forecasts of GDP growth and inflation for the Croatian economy”. In: *Economic Research-Ekonomska Istraživanja* 33.1, pp. 310–330.

Kangur, Alvar et al. (Sept. 2019). *How Informative Are Real Time Output Gap Estimates in Europe?* IMF Working Papers 2019/200. IMF.

Kenchington, David G. (2019). “Does a change in dividend tax rates in the U.S. affect equity prices of non-U.S. stocks?” In: *Review of Accounting Studies* 24.2, pp. 593–628.

Kiefer, Nicholas M., Timothy J. Vogelsang, and Helle Bunzel (May 2000). “Simple Robust Testing of Regression Hypotheses”. In: *Econometrica* 68.3, pp. 695–714.

Kilian, Lutz and Mark P Taylor (2003). “Why is it so difficult to beat the random walk forecast of exchange rates?” In: *Journal of International Economics* 60.1, pp. 85–107.

Klau, Marc and San Sau Fung (Mar. 2006). “The new BIS effective exchange rate indices”. In: *BIS Quarterly Review*.

Kohler, Marion (May 2023). *The Why, How and What of Forecasting.* Address to the Committee for Economic Development of Australia. URL: https://www.rba.gov.au/speeches/2023/sp-so-2023-05-03.html#r15.

Kohlhas, Alexandre N. and Donald Robertson (Sept. 2022). "Cautious Expectations". URL: http://perseus.iies.su.se/~akohl/limited_attention.pdf.

Kohlhas, Alexandre N. and Ansgar Walther (Sept. 2021). "Asymmetric Attention". In: *American Economic Review* 111.9, pp. 2879–2925.

Kohlscheen, Emanuel, Fernando H Avalos, and Andreas Schrimpf (2017). "When the walk is not random: commodity prices and exchange rates". In: *International Journal of Central Banking* 13.2, pp. 121–158.

Kohn, Robert (1982). "When is an aggregate of a time series efficiently forecast by its past?" In: *Journal of Econometrics* 18.3, pp. 337–349.

Koukoulas, Stephen (Mar. 2019). *Tweet.* URL: https://twitter.com/TheKouk/status/973374002319142914?s=20.

Lajbcygier, Paul and Simon M. Wheatley (2012). "Imputation Credits and Equity Returns*". In: *Economic Record* 88.283, pp. 476–494. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-4932.2012.00833.x.

Le, Nguyen Ngoc Anh, Xiangkang Yin, and Jing Zhao (Dec. 2020). "Effects of investor tax heterogeneity on stock prices and trading behaviour around the ex-dividend day: the case of Australia". In: *Accounting and Finance* 60.4, pp. 3775–3812.

Lewis, Christine and Nigel Pain (2014). "Lessons from OECD forecasts during and after the financial crisis". In: *OECD Journal: Economic Studies* 1, pp. 9–39.

Luna, Francesco (2014). "IMF Forecasts in the Context of Program Countries". In: *IEO Background Paper.*

Lütkepohl, Helmut (1986). "Forecasting temporally aggregated vector ARMA processes". In: *Journal of Forecasting* 5.2, pp. 85–95.

Mankiw, N. Gregory and Ricardo Reis (Nov. 2002). "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve*". In: *The Quarterly Journal of Economics* 117.4, pp. 1295–1328. eprint: https://academic.oup.com/qje/article-pdf/117/4/1295/5304341/117-4-1295.pdf. URL: https://doi.org/10.1162/003355302320935034.

Mankiw, N. Gregory, Ricardo Reis, and Justin Wolfers (Mar. 2004). "Disagreement about Inflation Expectations". In: *NBER Macroeconomics Annual 2003, Volume 18.* NBER Chapters. National Bureau of Economic Research, Inc, pp. 209–270.

Marcellino, Massimiliano (1999). "Some consequences of temporal aggregation in empirical analysis". In: *Journal of Business & Economic Statistics* 17.1, pp. 129–136.

McDonald, Robert L. (June 2015). "Cross-Border Investing with Tax Arbitrage: The Case of German Dividend Tax Credits". In: *The Review of Financial Studies* 14.3, pp. 617–657. eprint: https://academic.oup.com/rfs/article-pdf/14/3/617/24432194/140617.pdf. URL: https://doi.org/10.1093/rfs/14.3.617.

Melander, A., G. Sismanidis, and D. Grenouilleau (Oct. 2007). *The track record of the Commission's forecasts - an update.* European Economy - Economic Papers 2008 - 2015 291. Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.

Miller, Merton and Myron Scholes (1978). "Dividends and taxes". In: *Journal of Financial Economics* 6.4, pp. 333–364. URL: http://www.sciencedirect.com/science/article/pii/0304405X78900090.

Mincer, Jacob A. and Victor Zarnowitz (Apr. 1969). "The Evaluation of Economic Forecasts". In: *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*. NBER Chapters. National Bureau of Economic Research, Inc, pp. 3–46.

Molodtsova, Tanya, Alex Nikolsko-Rzhevskyy, and David H Papell (2008). "Taylor rules with real-time data: A tale of two countries and one exchange rate". In: *Journal of Monetary Economics* 55, S63–S79.

Monkhouse, Peter H.L. (1993). "The Cost of Equity under the Australian Dividend Imputation Tax System". In: *Accounting & Finance* 33.2, pp. 1–18. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-629X.1993.tb00321.x.

Musso, Alberto and Steven Phillips (2002). "Comparing Projections and Outcomes of IMF-Supported Programs". In: *IMF Working Paper* 49.1, pp. 22–48.

Newey, Whitney K and Kenneth D West (1987). "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". In: *Econometrica* 55.3, pp. 703–708.

Nordhaus, William D. (1987). "Forecasting Efficiency: Concepts and Applications". In: *The Review of Economics and Statistics* 69.4, p. 667.

Oler, Derk K, Jeffrey S. Harrison, and Matthew R. Allen (2008). "The danger of misinterpreting short-window event study findings in strategic management research: an empirical illustration using horizontal acquisitions". In: *Strategic Organization* 6(2), pp. 151–184.

Organisation for Economic Cooperation and Development (2020). *OECD Tax Database Explanatory Annex, Part II, Taxation of corporate and capital income*.

Page, Adrian and Kyriacos Lambrias (2019). "The performance of the Eurosystem/ECB staff macroeconomic projections since the financial crisis". In: *Economic Bulletin Articles* 8.

Parliamentary Budget Office (2018). *Attachment C of Policy costing: Dividend imputation credit refunds*. URL: https://www.aph.gov.au/-/media/05_About_Parliament/54_Parliamentary_Depts/548_Parliamentary_Budget_Office/Costings/Publicly_released_costings/Dividend_imputation_credit_refunds_-_PDF.pdf?la=en&hash=C527527FE69933D485232D161820AADA062A6E81.

— (2019). *2019 Post-election report*. URL: https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Budget_Office/General_elections/2019_general_election/2019_Post-election_report.

Pattenden, Kerry R. and Garry J. Twite (2007). "Taxes and Dividend Policy Under Alternative Tax Regimes". Mimeo, Available at SSRN. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=622003.

Pesaran, M Hashem and Allan Timmermann (1992). "A simple nonparametric test of predictive performance". In: *Journal of Business & Economic Statistics* 10.4, pp. 461–465.

— (2009). "Testing dependence among serially correlated multicategory variables". In: *Journal of the American Statistical Association* 104.485, pp. 325–337.

Petropoulos, Fotios, Daniele Apiletti, Vassilios Assimakopoulos, Mohamed Zied Babai, Devon K. Barrow, Souhaib Ben Taieb, et al. (2022b). "Forecasting: theory and practice". In: *International Journal of Forecasting* 38.3, pp. 705–871.

Poll Bludger (2019). *BludgerTrack 2019*. URL: https://www.pollbludger.net/bludgertrack2019/.

Poncet, Paul (2019). *modeest: Mode Estimation.* R package version 2.4.0. URL: https://CRAN.R-project.org/package=modeest.

Rantapuska, Elias (2008). "Ex-dividend day trading: Who, how, and why?: Evidence from the Finnish market". In: *Journal of Financial Economics* 88.2, pp. 355–374. URL: http://www.sciencedirect.com/science/article/pii/S0304405X08000299.

Ravn, Morten O. and Harald Uhlig (2002). "On adjusting the Hodrick-Prescott filter for the frequency of observations". In: *The Review of Economics and Statistics* 84.2, pp. 371–375.

Rogoff, Kenneth (1996). "The purchasing power parity puzzle". In: *Journal of Economic Literature* 34.2, pp. 647–668.

Rossi, Barbara (2013). "Exchange rate predictability". In: *Journal of economic literature* 51.4, pp. 1063–1119.

Rossi, Barbara and Tatevik Sekhposyan (Apr. 2016). "Forecast Rationality Tests in the Presence of Instabilities, with Applications to Federal Reserve and Survey Forecasts". In: *Journal of Applied Econometrics* 31.3, pp. 507–532.

Rudebusch, Glenn D. (Jan. 2008). "Publishing FOMC economic forecasts". In: *FRBSF Economic Letter.*

S&P Dow Jones Indices (2023a). *Factsheet: S&P/ASX 100.* URL: https://www.spglobal.com/spdji/en/idsenhancedfactsheet/file.pdf?calcFrequency=M&force_download=true&hostIdentifier=48190c8c-42c4-46af-8d1a-0cd5db894797&indexId=124511.

— (2023b). *Factsheet: S&P/ASX Small Ordinaries.* URL: https://www.spglobal.com/spdji/en/idsenhancedfactsheet/file.pdf?calcFrequency=M&force_download=true&hostIdentifier=48190c8c-42c4-46af-8d1a-0cd5db894797&indexId=124560.

Salinger, Michael (1992). "Standard Errors in Event Studies". In: *Journal of Financial and Quantitative Analysis* 27.1, pp. 39–53.

Scott, Brian J et al. (Jan. 2017). *The global case for strategic asset allocation and an examination of home bias.* Tech. rep. Vanguard Research. URL: https://intl.assets.vgdynamic.info/intl/australia/documents/research/rs016_paper-strategic-asset-allocation.pdf.

Shambaugh, Jay (2008). "A new look at pass-through". In: *Journal of International Money and Finance* 27.4, pp. 560–591.

Shorten, Bill (Mar. 2018). *A fairer tax system: Dividend imputation reform.* URL: https://www.billshorten.com.au/news/bill-s-media-releases/a-fairer-tax-system-dividend-imputation-reform-tuesday-13-march-2018/.

Sialm, Clemens (2009). "Tax Changes and Asset Pricing". In: *American Economic Review* 99.4, pp. 1356–83. URL: https://www.aeaweb.org/articles?id=10.1257/aer.99.4.1356.

Sims, Christopher A. (2003). "Implications of rational inattention". In: *Journal of Monetary Economics* 50.3. Swiss National Bank/Study Center Gerzensee Conference on Monetary Policy under Incomplete Information, pp. 665–690. URL: https://www.sciencedirect.com/science/article/pii/S0304393203000291.

Stekler, Herman (2004). "The Rationality and Efficiency of Individuals' Forecasts". In: *A Companion to Economic Forecasting.* Ed. by Michael P. Clements and David F. Hendry. Blackwell Publishing Ltd. Chap. 10.

Swan, Peter L. (2019). "Investment, the Corporate Tax Rate, and the Pricing of Franking Credits". In: *Economic Record* 95.311, pp. 480–496. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1475-4932.12493.

Sydney Morning Herald (Mar. 2018). "Shorten hits shareholderes with plan for 59 billion revenue grab". In: URL: https://www.smh.com.au/politics/federal/shorten-hits-shareholders-with-plan-for-59-billion-revenue-grab-20180312-p4z40d.html.

Takagi, Shinji and Halim Kucur (2006). "Testing the Accuracy of IMF Macroeconomic Forecasts, 1994-2003". In: *IEO Background Paper*.

Tiao, George C (1972). "Asymptotic behaviour of temporal aggregates of time series". In: *Biometrika* 59.3, pp. 525–531.

Timmermann, Allan (2007). "An Evaluation of the World Economic Outlook Forecasts". In: *IMF Staff Papers* 54.1, pp. 1–33. URL: http://www.jstor.org/stable/30036001 (visited on 09/25/2023).

Tseng, Yun-lan and Shing-yang Hu (2013). "Tax reform and the identity of marginal traders around ex-dividend days". In: *Pacific-Basin Finance Journal* 25, pp. 181–199. URL: http://www.sciencedirect.com/science/article/pii/S0927538X13000528.

Tsuchiya, Yoichi (2016). "Assessing macroeconomic forecasts for Japan under an asymmetric loss function". In: *International Journal of Forecasting* 32.2, pp. 233–242.

— (2023). "Assessing the World Bank's growth forecasts". In: *Economic Analysis and Policy* 77.C, pp. 64–84.

Turner, Phillip and Jozef Van 't dack (Nov. 1993). *Measuring International Price and Cost Competitiveness*. BIS Economic Papers 39. Bank for International Settlements.

Twite, Garry J., Jin Roc Lv, and Emma Schultz (2022). "Investor-Level Taxes and Corporate Dividend Policy: Evidence from a Quasi-Natural Experiment". Mimeo, Available at SSRN. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3984631.

Vartia, Yrjö and Pentti Vartia (1984). "Descriptive Index Number Theory and the Bank of Finland Currency Index". In: *The Scandinavian Journal of Economics* 86.3, pp. 352–364. URL: http://www.jstor.org/stable/3439868 (visited on 03/15/2023).

Vogel, Lukas (Sept. 2007). *How do the OECD Growth Projections for the G7 Economies Perform?: A Post-Mortem*. OECD Economics Department Working Papers 573. OECD Publishing.

Wei, William WS (1978). "Some consequences of temporal aggregation in seasonal time series models". In: *Seasonal analysis of economic time series*. NBER, pp. 433–448.

Weiss, Andrew A (1984). "Systematic sampling and temporal aggregation in time series models". In: *Journal of Econometrics* 26.3, pp. 271–281.

West, Kenneth D and Michael W McCracken (Nov. 1998). "Regression-Based Tests of Predictive Ability". In: *International Economic Review* 39.4, pp. 817–840.

Wieland, Volker and Maik Wolters (2013). "Chapter 5 - Forecasting and Policy Making". In: *Handbook of Economic Forecasting*. Ed. by Graham Elliott and Allan Timmermann. Vol. 2. Handbook of Economic Forecasting. Elsevier, pp. 239–325. URL: https://www.sciencedirect.com/science/article/pii/B9780444536839000050.

Willis Towers Watson (2019). "Global Pension Assets Study 2019". In: URL: https://www.willistowerswatson.com/en-US/news/2019/02/global-dc-pension-assets-exceed-db-assets-for-the-first-time.

Woodford, Michael (2003). "Imperfect Common Knowledge and the Effects of Monetary Policy". In: *Knowledge, Information, and Expectations in Modern Macroeconomics*. Princeton University Press. Chap. 1.

Working, Holbrook (1960). "Note on the correlation of first differences of averages in a random chain". In: *Econometrica* 28.4, pp. 916–918.

Wright, Jonathan H (2008). "Bayesian model averaging and exchange rate forecasts". In: *Journal of Econometrics* 146.2, pp. 329–341.

Yetman, James (Feb. 2018). *The perils of approximating fixed-horizon inflation forecasts with fixed-event forecasts*. BIS Working Papers 700. Bank for International Settlements.

Zeileis, Achim (2004). "Econometric Computing with HC and HAC Covariance Matrix Estimators". In: *Journal of Statistical Software* 11.i10.

Zellner, Arnold and Claude Montmarquette (Nov. 1971). "A study of some aspects of temporal aggregation problems in econometric analyses". In: *The Review of Economics and Statistics* 53.4, pp. 335–342.

Zhang, Hui Jun, Jean-Marie Dufour, and John W Galbraith (2016). "Exchange rates and commodity prices: Measuring causality at multiple horizons". In: *Journal of Empirical Finance* 36, pp. 100–120.

Ziembińska, Paulina (Dec. 2021). "Quality of Tests of Expectation Formation for Revised Data". In: *Central European Journal of Economic Modelling and Econometrics* 13.4, pp. 405–453.

Papers Surveyed in Section 2.2

Abhyankar, Abhay, Lucio Sarno, and Giorgio Valente (2005). "Exchange rates and fundamentals: evidence on the economic value of predictability". In: *Journal of International Economics* 66.2, pp. 325–348.

Adrian, Tobias, Erkko Etula, and Hyun Song Shin (2010). "Risk appetite and exchange rates". In: *FRB of New York Staff Report* 361.

Alquist, Ron and Menzie D Chinn (2008). "Conventional and unconventional approaches to exchange rate modelling and assessment". In: *International Journal of Finance & Economics* 13.1, pp. 2–13.

Altavilla, Carlo and Paul De Grauwe (2010). "Forecasting and combining competing models of exchange rate determination". In: *Applied Economics* 42.27, pp. 3455–3480.

Amano, Robert A and Simon Van Norden (1998a). "Exchange rates and oil prices". In: *Review of International Economics* 6.4, pp. 683–694.

— (1998b). "Oil prices and the rise and fall of the US real exchange rate". In: *Journal of International Money and Finance* 17.2, pp. 299–316.

Bacchetta, Philippe, Eric Van Wincoop, and Toni Beutler (2010). "Can Parameter Instability Explain the Meese-Rogoff Puzzle?" In: *NBER International Seminar on Macroeconomics*. Vol. 6. 1. The University of Chicago Press Chicago, IL, pp. 125–173.

Bashar, OKMR and Sarkar Humayun Kabir (2013). "Relationship between commodity prices and exchange rate in light of global financial crisis: evidence from Australia". In: *International Journal of Trade, Economics and Finance* 4.5, pp. 265–269.

Beckmann, Joscha, Gary Koop, et al. (2020). "Exchange rate predictability and dynamic Bayesian learning". In: *Journal of Applied Econometrics* 35.4, pp. 410–421.

Beckmann, Joscha and Rainer Schüssler (2016). "Forecasting exchange rates under parameter and model uncertainty". In: *Journal of International Money and Finance* 60, pp. 267–288.

Berge, Travis J (2014). "Forecasting disconnected exchange rates". In: *Journal of Applied Econometrics* 29.5, pp. 713–735.

Berkowitz, Jeremy and Lorenzo Giorgianni (2001). "Long-horizon exchange rate predictability?" In: *Review of Economics and Statistics* 83.1, pp. 81–91.

Bianco Dal, Marcos, Maximo Camacho, and Gabriel Perez Quiros (2012). "Short-run forecasting of the euro-dollar exchange rate with economic fundamentals". In: *Journal of International Money and Finance* 31.2, pp. 377–396.

Boughton, James M (1987). "Tests of the performance of reduced-form exchange rate models". In: *Journal of International Economics* 23.1-2, pp. 41–56.

Byrne, Joseph P, Dimitris Korobilis, and Pinho J Ribeiro (2016). "Exchange rate predictability in a changing world". In: *Journal of International Money and Finance* 62, pp. 1–24.

Ca'Zorzi, Michele, Jakub Muck, and Michal Rubaszek (2016). "Real exchange rate forecasting and PPP: This time the random walk loses". In: *Open Economies Review* 27, pp. 585–609.

Ca'Zorzi, Michele and Michał Rubaszek (2020). "Exchange rate forecasting on a napkin". In: *Journal of International Money and Finance* 104, p. 102168.

Canova, Fabio (1993). "Modelling and forecasting exchange rates with a Bayesian time-varying coefficient model". In: *Journal of Economic Dynamics and Control* 17.1-2, pp. 233–261.

Chen, ShiuSheng and HungChyn Chen (2007). “Oil prices and real exchange rates”. In: *Energy Economics* 29.3, pp. 390–404.

Chen, Yuchin and Kenneth Rogoff (2003). “Commodity currencies”. In: *Journal of International Economics* 60.1, pp. 133–160.

Chen, YuChin, Kenneth S Rogoff, and Barbara Rossi (2010). “Can exchange rates forecast commodity prices?” In: *The Quarterly Journal of Economics* 125.3, pp. 1145–1194.

Chen, Yuchin and Kwok Ping Tsang (2013). “What does the yield curve tell us about exchange rate predictability?” In: *Review of Economics and Statistics* 95.1, pp. 185–205.

Cheung, YinWong, Menzie D Chinn, and Antonio Garcia Pascual (2005). “Empirical exchange rate models of the nineties: Are any fit to survive?” In: *Journal of International Money and Finance* 24.7, pp. 1150–1175.

Cheung, YinWong, Menzie D Chinn, Antonio Garcia Pascual, and Yi Zhang (2019). “Exchange rate prediction redux: New models, new data, new currencies”. In: *Journal of International Money and Finance* 95, pp. 332–362.

Chinn, Menzie D and Richard A Meese (1995). “Banking on currency forecasts: how predictable is change in money?” In: *Journal of International Economics* 38.1-2, pp. 161–178.

Chinn, Menzie D and Michael J Moore (2011). “Order flow and the monetary model of exchange rates: Evidence from a novel data set”. In: *Journal of Money, Credit and Banking* 43.8, pp. 1599–1624.

Chinn, Menzie David (1991). “Some linear and nonlinear thoughts on exchange rates”. In: *Journal of International Money and Finance* 10.2, pp. 214–230.

Clark, Todd E and Kenneth D West (2006). “Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis”. In: *Journal of Econometrics* 135.1-2, pp. 155–186.

Clements, Kenneth W and Renée Fry (2008). “Commodity currencies and currency commodities”. In: *Resources Policy* 33.2, pp. 55–73.

Clements, Michael P and Jeremy Smith (2001). “Evaluating forecasts from SETAR models of exchange rates”. In: *Journal of International Money and Finance* 20.1, pp. 133–148.

Della Corte, Pasquale, Lucio Sarno, and Giulia Sestieri (2012). “The predictive information content of external imbalances for exchange rate returns: How much is it worth?” In: *Review of Economics and Statistics* 94.1, pp. 100–115.

Della Corte, Pasquale, Lucio Sarno, and Ilias Tsiakas (2009). “An economic evaluation of empirical exchange rate models”. In: *The Review of Financial Studies* 22.9, pp. 3491–3530.

Diebold, Francis X, Javier Gardeazabal, and Kamil Yilmaz (1994). “On cointegration and exchange rate dynamics”. In: *The Journal of Finance* 49.2, pp. 727–735.

Diebold, Francis X and James A Nason (1990). “Nonparametric exchange rate prediction?” In: *Journal of International Economics* 28.3-4, pp. 315–332.

Edwards, Sebastian (1983). “Floating exchange rates, expectations and new information”. In: *Journal of Monetary Economics* 11.3, pp. 321–336.

Eichenbaum, Martin S, Benjamin K Johannsen, and Sergio T Rebelo (2021). “Monetary policy and the predictability of nominal exchange rates”. In: *The Review of Economic Studies* 88.1, pp. 192–228.

Engel, Charles (1994). “Can the Markov switching model forecast exchange rates?” In: *Journal of International Economics* 36.1-2, pp. 151–165.

Engel, Charles, Dohyeon Lee, et al. (2019). "The uncovered interest parity puzzle, exchange rate forecasting, and Taylor rules". In: *Journal of International Money and Finance* 95, pp. 317–331.

Engel, Charles, Nelson C Mark, and Kenneth D West (2015). "Factor model forecasts of exchange rates". In: *Econometric Reviews* 34.1-2, pp. 32–55.

Engel, Charles and Kenneth D West (2005). "Exchange rates and fundamentals". In: *Journal of Political Economy* 113.3, pp. 485–517.

— (2006). "Taylor Rules and the Deutschmark: Dollar Real Exchange Rate". In: *Journal of Money, Credit and Banking* 38.5, pp. 1175–1194.

Engel, Charles and Steve Pak Yeung Wu (2023). "Liquidity and exchange rates: An empirical investigation". In: *The Review of Economic Studies* 90.5, pp. 2395–2438.

Engle, Charles and James D Hamilton (1990). "Long swings in the dollar: Are they in the data and do markets know it?" In: *American Economic Review* 80.4, pp. 689–713.

Evans, Martin D D and Richard K Lyons (2005). "Meese-Rogoff redux: Micro-based exchange-rate forecasting". In: *American Economic Review* 95.2, pp. 405–414.

Fama, Eugene F (1984). "Forward and spot exchange rates". In: *Journal of Monetary Economics* 14.3, pp. 319–338.

Ferraro, Domenico, Kenneth Rogoff, and Barbara Rossi (2015). "Can oil prices forecast exchange rates? An empirical analysis of the relationship between commodity prices and exchange rates". In: *Journal of International Money and Finance* 54, pp. 116–141.

Fratzscher, Marcel et al. (2015). "The scapegoat theory of exchange rates: the first tests". In: *Journal of Monetary Economics* 70, pp. 1–21.

Garratt, Anthony and Emi Mise (2014). "Forecasting exchange rates using panel model and model averaging". In: *Economic Modelling* 37, pp. 32–40.

Giacomini, Raffaella and Barbara Rossi (2010). "Forecast comparisons in unstable environments". In: *Journal of Applied Econometrics* 25.4, pp. 595–620.

Gourinchas, PierreOlivier and Helene Rey (2007). "International financial adjustment". In: *Journal of Political Economy* 115.4, pp. 665–703.

Groen, Jan JJ (1999). "Long horizon predictability of exchange rates: Is it for real?" In: *Empirical Economics* 24, pp. 451–469.

— (2005). "Exchange rate predictability and monetary fundamentals in a small multi-country panel". In: *Journal of Money, Credit and Banking*, pp. 495–516.

Hodrick, Robert J (1989). "Risk, uncertainty, and exchange rates". In: *Journal of Monetary Economics* 23.3, pp. 433–459.

Hooper, Peter and John Morton (1982). "Fluctuations in the dollar: A model of nominal and real exchange rate determination". In: *Journal of International Money and Finance* 1, pp. 39–56.

Ince, Onur (2014). "Forecasting exchange rates out-of-sample with panel methods and real-time data". In: *Journal of International Money and Finance* 43, pp. 1–18.

Islam, M Faizul and Mohammad S Hasan (2006). "The monetary model of the Dollar-Yen exchange rate determination: A cointegration approach". In: *International Journal of Business and Economics* 5.2, p. 129.

Issa, Ramzi, Robert Lafrance, and John Murray (2008). "The turning black tide: energy prices and the Canadian dollar". In: *Canadian Journal of Economics* 41.3, pp. 737–759.

Jorion, Philippe and Richard J Sweeney (1996). "Mean reversion in real exchange rates: evidence and implications for forecasting". In: *Journal of International Money and Finance* 15.4, pp. 535–550.

Kilian, Lutz (1999). “Exchange rates and monetary fundamentals: what do we learn from long-horizon regressions?” In: *Journal of Applied Econometrics* 14.5, pp. 491–510.

Kouwenberg, Roy et al. (2017). “Model uncertainty and exchange rate forecasting”. In: *Journal of Financial and Quantitative Analysis* 52.1, pp. 341–363.

Kräger, Horst and Peter Kugler (1993). “Non-linearities in foreign exchange markets: a different perspective”. In: *Journal of International Money and Finance* 12.2, pp. 195–208.

Kremens, Lukas and Ian Martin (2019). “The quanto theory of exchange rates”. In: *American Economic Review* 109.3, pp. 810–843.

Li, Jiahan, Ilias Tsiakas, and Wei Wang (2015). “Predicting exchange rates out of sample: Can economic fundamentals beat the random walk?” In: *Journal of Financial Econometrics* 13.2, pp. 293–341.

Lilley, Andrew et al. (2022). “Exchange rate reconnect”. In: *Review of Economics and Statistics* 104.4, pp. 845–855.

Liu, Yang and Ivan Shaliastovich (2022). “Government policy approval and exchange rates”. In: *Journal of Financial Economics* 143.1, pp. 303–331.

Lopez-Suarez, Carlos Felipe and Jose Antonio Rodriguez-Lopez (2011). “Nonlinear exchange rate predictability”. In: *Journal of International Money and Finance* 30.5, pp. 877–895.

MacDonald, Ronald (1998). “What determines real exchange rates?: The long and the short of it”. In: *Journal of International Financial Markets, Institutions and Money* 8.2, pp. 117–153.

MacDonald, Ronald and Mark P Taylor (1993). “The monetary approach to the exchange rate: rational expectations, long-run equilibrium, and forecasting”. In: *Staff Papers* 40.1, pp. 89–107.

Mark, Nelson C (1995). “Exchange rates and fundamentals: Evidence on long-horizon predictability”. In: *American Economic Review*, pp. 201–218.

Mark, Nelson C and Donggyu Sul (2001). “Nominal exchange rates and monetary fundamentals: evidence from a small post-Bretton Woods panel”. In: *Journal of International Economics* 53.1, pp. 29–52.

Meese, Richard A and Kenneth Rogoff (1983b). “The out-of-sample failure of empirical exchange rate models: sampling error or misspecification?” In: *Exchange rates and international macroeconomics*. University of Chicago Press, pp. 67–112.

— (1988). “Was it real? The exchange rate-interest differential relation over the modern floating-rate period”. In: *Journal of Finance* 43.4, pp. 933–948.

Meese, Richard A and Andrew K Rose (1991). “An empirical assessment of non-linearities in models of exchange rate determination”. In: *The Review of Economic Studies* 58.3, pp. 603–619.

Mizrach, Bruce (1992). “Multivariate nearest-neighbour forecasts of EMS exchange rates”. In: *Journal of Applied Econometrics* 7.S1, S151–S163.

Molodtsova, Tanya, Alex Nikolsko-Rzhevskyy, and David H Papell (2011). “Taylor rules and the euro”. In: *Journal of Money, Credit and Banking* 43.2-3, pp. 535–552.

Molodtsova, Tanya and David H Papell (2013). “Taylor rule exchange rate forecasting during the financial crisis”. In: *NBER International Seminar on Macroeconomics*. Vol. 9. 1. University of Chicago Press Chicago, IL, pp. 55–97.

Morales-Arias, Leonardo and Guilherme V Moura (2013). “Adaptive forecasting of exchange rates with panel data”. In: *International Journal of Forecasting* 29.3, pp. 493–509.

Mumtaz, Haroon and Laura Sunder-Plassmann (2013). "Time-varying dynamics of the real exchange rate: An empirical analysis". In: *Journal of Applied Econometrics* 28.3, pp. 498–525.

Pacelli, Vincenzo, Vitoantonio Bevilacqua, and Michele Azzollini (2011). "An artificial neural network model to forecast exchange rates". In: *Journal of Intelligent Learning Systems and Applications* 3.02, pp. 57–69.

Park, Cheolbeom and Sookyung Park (2013). "Exchange rate predictability and a monetary model with time-varying cointegration coefficients". In: *Journal of International Money and Finance* 37, pp. 394–410.

Qi, Min and Yangru Wu (2003). "Nonlinear prediction of exchange rates with monetary fundamentals". In: *Journal of Empirical Finance* 10.5, pp. 623–640.

Rapach, David E and Mark E Wohar (2002). "Testing the monetary model of exchange rate determination: new evidence from a century of data". In: *Journal of International Economics* 58.2, pp. 359–385.

— (2004). "Testing the monetary model of exchange rate determination: a closer look at panels". In: *Journal of International Money and Finance* 23.6, pp. 867–895.

— (2006). "The out-of-sample forecasting performance of nonlinear models of real exchange rate behavior". In: *International Journal of Forecasting* 22.2, pp. 341–361.

Rime, Dagfinn, Lucio Sarno, and Elvira Sojli (2010). "Exchange rate forecasting, order flow and macroeconomic information". In: *Journal of International Economics* 80.1, pp. 72–88.

Rogoff, Kenneth S and Vania Stavrakeva (2008). *The continuing puzzle of short horizon exchange rate forecasting.* Tech. rep. National Bureau of Economic Research.

Rossi, Barbara (2005). "Testing long-horizon predictive ability with high persistence, and the Meese–Rogoff puzzle". In: *International Economic Review* 46.1, pp. 61–92.

— (2006). "Are exchange rates really random walks? Some evidence robust to parameter instability". In: *Macroeconomic Dynamics* 10.1, pp. 20–38.

Rossi, Barbara and Atsushi Inoue (2012). "Out-of-sample forecast tests robust to the choice of window size". In: *Journal of Business & Economic Statistics* 30.3, pp. 432–453.

Sarantis, Nicholas (1999). "Modeling non-linearities in real effective exchange rates". In: *Journal of International Money and Finance* 18.1, pp. 27–45.

Sarno, Lucio and Giorgio Valente (2009). "Exchange rates and fundamentals: Footloose or evolving relationship?" In: *Journal of the European Economic Association* 7.4, pp. 786–830.

Schinasi, Garry J and Paravastu Ananta Venkata Bhattanadha Swamy (1989). "The out-of-sample forecasting performance of exchange rate models when coefficients are allowed to change". In: *Journal of International Money and Finance* 8.3, pp. 375–390.

Siddique, Akhtar and Richard J Sweeney (1998). "Forecasting real exchange rates". In: *Journal of International Money and Finance* 17.1, pp. 63–70.

Taylor, Mark P, David A Peel, and Lucio Sarno (2001). "Nonlinear mean-reversion in real exchange rates: toward a solution to the purchasing power parity puzzles". In: *International Economic Review* 42.4, pp. 1015–1042.

Throop, Adrian W (1993). "A generalized uncovered interest parity model of exchange rates". In: *Economic Review-Federal Reserve Bank of San Francisco* 2, p. 3.

Wang, Jian and Jason J Wu (2012). "The Taylor rule and forecast intervals for exchange rates". In: *Journal of Money, Credit and Banking* 44.1, pp. 103–144.

Wolff, Christian CP (1987). "Time-varying parameters and the out-of-sample forecasting performance of structural exchange rate models". In: *Journal of Business & Economic Statistics* 5.1, pp. 87–97.