Advances in Forecasting Under Instability*

Barbara Rossi

ICREA, UPF, CREI and BGSE

Prepared for the *Handbook of Economic Forecasting*,

G. Elliott and A. Timmermann (eds.),

Elsevier Publications.

This Revision: June 30, 2012

*Acknowledgments. I am extremely grateful to Jonathan Wright for sharing his Matlab codes and to Lily Liu for excellent research assistance. The chapter benefitted from detailed comments or suggestions from the editors, two anonymous referees, Todd Clark, James Mitchell, Raffaella Giacomini, Lutz Kilian, Tatevik Sekhposyan, Allan Timmermann as well as participants at the 2011 St. Louis Conference. Financial support from NSF Grant # 0647627 is gratefully acknowledged.

1 Introduction

The forecasting literature has identified two important, broad issues (see e.g. Stock and Watson, 1996, 2003, for a discussion). The first stylized fact is that there are several predictors of output growth and inflation that have substantial and statistically significant predictive content, although that is apparent only sporadically, at some times and in some countries. Whether this predictive content can be reliably exploited is unclear. In fact, finding predictors that work well in one period is no guarantee that such predictors will maintain their usefulness in subsequent periods. **That is, the predictive content is unstable over time.** This lack of stability is mainly established using parameter instability tests (such as Andrews' (1993) QLR test) in Granger-causality regressions as well as by evaluating out-of-sample forecasts over two sub-samples and noting that the good (poor) forecasting ability of a predictor in one sub-sample seems totally uncorrelated with whether the same predictor will have a good (poor) forecasting ability in the other sub-sample.

A second important finding concerns the relationship between in-sample fit and out-of-sample forecasting ability. Researchers typically identify predictors on the basis of in-sample Granger-causality tests. In-sample Granger-causality tests assess the significance of the proposed predictors in a regression of the dependent variable (say y_{t+h}) onto the lagged predictors (say, x_t), where h is the forecast horizon. That is, the Granger-causality test is a simple F-test on the parameter vector β_h, where:

$$y_{t+h} = \beta_h' x_t + \gamma_h' z_t + \varepsilon_{t+h}, \ t = 1, ..., T \tag{1}$$

and z_t are other control variables (for example, lags of y: $y_{t,}$ $y_{t-1}, ...$). Given that time series are typically serially correlated and possibly heteroskedastic and the data are overlapping, the error term might be both serially correlated and heteroskedastic, and the F-test requires HAC-robust variance estimates (Newey and West, 1987). The researcher deems regressors to be suitable predictors when the statistical tests reject the null hypothesis that the regressor is insignificant (that is, when the F-test for testing the hypothesis $\beta_h = 0$ rejects at standard significance levels). However, empirical results in the literature find that significant Granger-causality statistics contain little or no information about whether the predictor is reliable out-of-sample. **Indeed, in-sample predictive content does not necessarily translate into out-of-sample predictive ability, nor ensures the stability of the predictive relation over time.** This is a well-known, although disconcerting, empirical stylized fact; one of the earliest examples dates back to Meese and Rogoff (1983a,b, 1988), who found that successful in-sample fit of exchange rate models does not always translate into out-of-sample predictive ability – see also Swanson and White (1995), who similarly found in-sample predictive ability in the term structure of interest rates but no out-of-sample forecasting ability, and Stock and Watson (2003), who found similar results for a much broader set of macroeconomic time series.

Why do instabilities matter for forecasting? Clearly, if the predictive content is not stable over time, it will be very difficult to exploit it to improve forecasts. In addition, if the regressors are selected according to in-sample Granger-causality tests, and the latter are not indicative of true out-of-sample predictive power, this practice may result in even poorer forecasts. In fact, the empirical evidence that we discuss has documented large swings in parameter magnitudes and signs, which can potentially affect forecasts in practice.

In a comprehensive analysis, Stock and Watson (2003) focus attention on forecasting output growth (measured by the rate of growth of Gross Domestic Product, GDP) and inflation (measured by the percentage change of the consumer price index or the implicit GDP deflator) in U.S. data, and consider a multitude of predictors one-at-a-time, in particular asset prices such as interest rates, term spreads, default spreads, stock prices, dividend yields, as well as non-financial indicators such as unemployment, money growth and the output gap, and find that the two issues above are widespread in their database up to the early 2000's. Is it the case also when considering the last decade of data? And do these results hold in other databases? In what follows, we review the empirical evidence on forecasting in the presence of instabilities, and show that the same two findings emerge in the recent literature as well as in other databases: there is clear empirical evidence of instabilities in the predictive relationships as well as poor correlation between in-sample and out-of-sample predictive content.

The objective of this chapter is to understand what we have learned about forecasting in the presence of instabilities, especially regarding the two questions above. The empirical evidence raises a multitude of questions. If in-sample tests provide poor guidance to out-of-sample forecasting ability, what should researchers do? If there are statistically significant instabilities in the Granger-causality relationships, how do researchers establish whether there is any Granger-causality at all? If there is substantial instability in predictive relationships, how do researchers establish which models is the "best" forecasting model? And finally, if a model forecasts poorly, why is that and how should researchers proceed to improve the forecasting models? In this chapter, we answer these questions by discussing various methodologies for inference as well as estimation that have been recently proposed in the literature. The last question is the hardest one, as improving models' forecasts has been proven to be difficult empirically, although the literature does provide partial answers, which we overview.

This chapter is divided in three parts. Section 2 analyzes whether predictive content is unstable over time and, if that is the case, the tools that researchers can use to assess predictive ability or improve models' estimation in the presence of instabilities. Section 3 focuses on the relationship between in-sample fit and out-of-sample forecasting ability; in particular, it provides theoretical results on why the two may differ, and reviews statistical tests to assess whether that is the case in practice, and what are the causes of the divergence. Section 4 provides an empirical analysis of whether these issues are important in practice by focusing on an empirical analysis. We focus on the same database as Stock and Watson (2003) and test whether the predictive content is unstable, which estimation methods are most successful in practice, and whether the in-sample fit is indicative of out-of-sample forecasting performance, and what are the likely reasons of the discrepancy.

Throughout the chapter we focus on conditional mean forecasts in linear models, given their importance in practice.[1] This allows us to clearly expose the main concepts with simple notation, while at the same time be consistent with the empirical application in Section 4. When results are applicable to more general models, we note so and refer readers to the relevant references. Finally, note that the chapter focuses on recent contributions on forecast evaluation and estimation in the presence of instabilities (including several of the author's own and related works). The chapter does not cover in-sample instability tests nor on in-sample estimation of models with breaks (unless the estimation is explicitly shown to improve models' out-of-sample forecasting ability)[2]. Finally, the discussion focuses on frequentist methods; Bayesian techniques for handling model instability receive mention but a less detailed attention.

[1] For a review of forecasting in non-linear models see Terasvirta (2009) and Calhoun and Elliott (2012) for an analysis of the relative advantages of linear versus non-linear models; for a review of forecasting with trending data, see Elliott (2009); and for a review of volatility forecasting see Andersen et al. (2009).

[2] For a review of tests of structural breaks, see Stock (1994).

2 Is the Predictive Content Unstable Over Time?

The goal of this section is to determine whether the predictive content in typical economic relationships is unstable over time. In particular, which tools are available to researchers for assessing whether that is the case, and which models they should use for forecasting. First, we review the empirical evidence on instabilities in predictive regressions and databases of interest to economists. The literature suggests that the predictive content of several time series predictors is indeed unstable over time in macroeconomics, finance and international finance. Second, given the empirical findings, we then review the tools that researchers have at their disposal for evaluating forecasting ability in the presence of instabilities. Typically, researchers are interested in the following questions: (i) does a vector of time series Granger-cause a variable of interest (e.g. inflation or output growth)? (ii) which, among two models, forecasts the best? (iii) are forecasts rational? Typical Granger-causality tests as well as tests for out-of-sample forecast comparisons and tests for forecast rationality are inconsistent in the presence of instabilities. The chapter provides guidance on which tools are available to researchers who are interested in answering these questions when there are instabilities in the data. Third, is it possible to exploit instabilities to improve the out-of-sample forecasting ability of existing models? There are several approaches taken in the literature, from methods that identify historic breaks and impose them in the estimation to the estimation of time-varying parameter models. The chapter provides guidance to practitioners by focusing on methods that have been developed with the clear aim of improving forecasting ability and have been empirically successful.[3] In Section 4, we select several of these methodologies and evaluate their usefulness for forecasting inflation and output growth using a large database of macroeconomic predictors in an empirical exercise similar to Stock and Watson (2003).

2.1 Is the Predictive Content Unstable Over Time? The Empirical Evidence

The literature in the last decade has shown that Stock and Watson's (2003) empirical stylized facts have been echoed in several other databases. For example, in finance, instabilities have been found when forecasting stock returns. Goyal and Welch (2003) is one of the early studies that reports instabilities in stock return predictability, whereas Ang and Bekaert (2004) find

[3] Due to space limitations, we will not overview the literature that focuses strictly on in-sample tests for structural breaks or in-sample estimation in the presence of structural changes.

a deterioration in stock return predictability in the 1990s. Rapach and Wohar (2005) find several breaks in both real interest rates as well as inflation for 13 industrialized countries. Rapach and Wohar (2006) document the existence of structural breaks in the predictive ability of several variables (such as the dividend price ratio and the default spread) and S&P 500; the results are similar when predicting CRSP equal-weighted real stock returns. Similarly, Paye and Timmermann (2006) find structural breaks in predicting stock returns using the lagged dividend yield, short term interest rates and the term spread, among other predictors. Interestingly, they note that the timing of the break is not uniform over time: several countries experience breaks at different times. They also find that, in the majority of the cases, the predictable component in stock returns has diminished following the most recent break. Timmermann (2008) concludes that "most of the time the forecasting models perform rather poorly, but there is evidence of relatively short-lived periods with modest return predictability. The short duration of the episodes where return predictability appears to be present and the relatively weak degree of predictability even during such periods makes predicting returns an extraordinarily challenging task." See also the chapter by Rapach and Zhou in this Handbook. Another area in finance where instabilities seem very important is firm and industry-level CAPM betas, see Blume (1975) and Fama and French (1997) for classic references.[4]

A second area of research where instabilities in forecasting performance are important is exchange rate prediction. Schinasi and Swamy (1989) and Wolff (1987) are among the first papers that found instabilities in exchange rate models and their forecasting ability. Rossi (2006) considers traditional models of exchange rate dynamics based on macroeconomic fundamentals, such as interest rates, money or output differentials using the Granger-causality tests robust to the presence of parameter instability discussed later in this Section. She shows that for some countries it is possible to reject the hypothesis that exchange rates are random walks in favor of the existence of a time-varying relationship between exchange rate and fundamentals. Her findings raise the possibility that economic models were previously rejected in favor of an a-theoretical random walk model not because the fundamentals are completely unrelated to exchange rate fluctuations, but because the relationship is unstable over time and, thus, difficult to capture by Granger-causality tests or by forecast comparisons. She also analyzes forecasts that exploit time variation in the parameters and finds that, in some cases, they can improve economic models' forecasts relative to the random walk. Rogoff

[4] As a referee points out, in fact instabilities are so important that it is common practice to limit monthly CAPM regressions to 3-5 years of historical data.

and Stavrakeva (2008) point out that the predictive ability of macroeconomic fundamentals strongly depends on the sample split chosen for forecasting, also suggesting that instabilities are very important. Giacomini and Rossi (2010a) document that the relative forecasting performance of the models is time-varying: economic fundamentals do have forecasting ability in the late Eighties, but the predictive ability disappears in the Nineties. Beckmann et al. (2011) consider instabilities in the relationship between the Deutschmark/U.S. dollar exchange rate and macroeconomic fundamentals using a time-varying coefficient model. They show that fundamentals are important explanatory variables for exchange rates, although their impact greatly differs over time. Sarno and Valente (2009) consider forecasting five major U.S. dollar exchange rates using a time-varying coefficient model. They conclude that the poor out-of-sample forecasting ability of exchange rate models may be caused by the poor performance of in-sample model selection criteria, and that the difficulty in selecting the best predictive model is largely due to frequent shifts in the fundamentals. Bacchetta and van Wincoop (2009) and Rime et al. (2010) provide theoretical explanations for the instabilities found in the relationship between exchange rates and macroeconomic fundamentals: the former rely on unstable expectations, and the latter on learning about the state of the economy.

A third area where researchers have found evidence of instability is macroeconomic variables' predictions, for example forecasting output growth using the term spread. Giacomini and Rossi (2006) consider the relationship between the lagged term spread and output growth and find empirical evidence of the existence of a relationship between the term spread and output growth, although it is unstable over time. Bordo and Haubrich (2008) show that the spread between corporate bonds and commercial paper predicts future output growth over the period 1875-1997 although the predictive ability varies over time, and has been strongest in the post-World War II period. Schrimpf and Wang (2010) examine the predictive ability of the yield curve in four major developed countries (Canada, Germany, the United Kingdom, and the United States). They find strong evidence of instabilities in the relationship between the yield spread and output growth by using structural break tests; they also find that the yield curve has been losing its edge as a predictor of output growth in recent years. See also Wheelock and Wohar (2009) for an overview of the usefulness of the spread for predicting economic activity across countries and over time. More broadly, Stock and Watson (2007), D'Agostino, Giannone and Surico (2008) and Rossi and Sekhposyan (2010) have documented a change in the forecastability of inflation as well as output growth over time; in particular, a decrease in predictability. The same decrease in predictive ability is apparent also when comparing the forecasting performance of structural models, as Edge and Gurkaynak (2011) demonstrate.

2.2 Testing When the Predictive Content Is Unstable Over Time

As discussed in the introduction, forecasters are interested in several questions, among which: (i) does a potential predictor Granger-cause an economic variable of interest? (ii) which one between two competing models forecasts the best? (iii) are forecasts rational (or optimal)? In this section, we review techniques that allow forecasters to answer these questions in unstable environments.

2.2.1 How Can Researchers Establish Granger-causality in the Presence of Instabilities?

In the presence of instabilities, traditional Granger-causality tests are inconsistent: in fact, Rossi (2005) showed that traditional Granger-causality tests may have no power in the presence of instabilities. To understand why, consider the following example, which is a special case of eq. (1). The data are generated by: $y_{t+h} = \beta_t x_t + \varepsilon_{t+h}$, $t = 1, 2, ..., T$, where, for simplicity, x_t and ε_{t+h} are both univariate random draws from i.i.d. standard Normal distributions, and they are independent of each other. We assume that the prediction horizon, h, is fixed. The parameter changes over time, and this is formalized by allowing the parameter to have a time-subscript: β_t. Let

$$\beta_t = 1\,(t \leq T/2) - 1(t > T/2). \tag{2}$$

A traditional Granger-causality test in this example would be a t-test for testing the null hypothesis that the Ordinary Least Squares (OLS) parameter estimate in a regression of y_{t+h} onto x_t equals zero. In this example, the full-sample OLS parameter estimate is:

$$\begin{aligned}\left(\sum_{t=1}^{T} x_t^2\right)^{-1} \sum_{t=1}^{T} x_t y_{t+h} &= \left(\sum_{t=1}^{T} x_t^2\right)^{-1} \sum_{t=1}^{T} x_t \varepsilon_{t+h} + \left(\sum_{t=1}^{T} x_t^2\right)^{-1} \sum_{t=1}^{T} x_t^2 \beta_t \\ &= \left(T^{-1} \sum_{t=1}^{T} x_t^2\right)^{-1} T^{-1} \sum_{t=1}^{T} x_t \varepsilon_{t+h} \\ &\quad + \left(T^{-1} \sum_{t=1}^{T} x_t^2\right)^{-1} T^{-1} \left[\sum_{t=1}^{T/2} x_t^2 + \sum_{t=T/2+1}^{T} x_t^2\,(-1)\right] \underset{p}{\rightarrow} 0,\end{aligned}$$

since $T^{-1}\sum_{t=1}^{T} x_t^2 \underset{p}{\rightarrow} E\left(x_t^2\right) = 1$ and $T^{-1}\sum_{t=1}^{T} x_t\varepsilon_{t+h} \underset{p}{\rightarrow} 0$. Thus, instabilities are such that the estimate of the Granger-causality parameter is negligible, leading to a failure to reject the no Granger-causality hypothesis[5] even if the regressor does Granger-cause y_t in reality. The problem is that the predictive ability is unstable over time, which does not satisfy the stationarity assumption underlying traditional Granger-causality tests. While this example is extremely simplified, it can be generalized to instabilities other than eq. (2); by varying the time of the break and the magnitude of the parameters before and after the break it is possible to find similar results. The main conclusion is that traditional Granger-causality tests are inconsistent if there are instabilities in the parameters. Note that this problem is empirically relevant: quite often, parameter estimates change substantially in sign and magnitude through time. See for example Goyal and Welch (2008) for suggestive plots of time variation in sum of squared residuals of equity premium returns predictors, or the dramatic swings over time in the sign of the coefficients in exchange rate models (Rossi, 2005).[6]

What should researchers do in such situations? Rossi (2005) proposes tests for evaluating whether the variable x_t has no predictive content for y_t in the situation where the parameter β_t might be time-varying.[7] Her procedure is based on testing jointly the significance of the predictors and their stability over time. Among the various forms of instabilities that she considers, we focus on the case in which β_t may shift from β_1 to $\beta_2 \neq \beta_1$ at some unknown point in time, τ. That is, $\beta_t = \beta_1 \cdot 1\left(t \leq \tau\right) + \beta_2 \cdot 1(t > \tau)$.[8] Note that, although the parameter may have parameter instability, the null hypothesis is not just parameter stability: the main objective of the test is to capture predictive ability, *even though* the

[5] Recall that the null hypothesis of a Granger-causality test is the absence of predictive ability.

[6] Note also that even if there are no dramatic swings in the coefficient signs but swings in the coefficient magnitudes, and the traditional test is consistent, yet the finite sample power of the traditional test is likely to be inferior to that of a test that is robust to instabilities, such as the one we discuss below.

[7] Rossi (2005) relaxes these conditions. She considers the general case of testing possibly nonlinear restrictions in models estimated with Generalized Method of Moments (GMM). Here, we specialize the description for the simple case of no Granger-causality restrictions in models whose parameters are consistently estimated with OLS, such as Granger-causality regressions. She also considers the case of tests on subsets of parameters, that is, in the case of Granger-causality regressions, tests on whether x_t Granger-causes y_t in the model $y_{t+h} = x_t'\beta_t + z_t'\gamma + \varepsilon_{t+h}$.

[8] Note that the test is designed to have power in situations where there is at most a one-time break in the parameters. However, by construction, since the test uses a sup-type procedure, in the presence of multiple breaks in predictive ability the test will pick up the largest break, and it is therefore robust to the presence of multiple breaks.

predictive ability may potentially appear only in a sub-sample. As such, the null hypothesis involves the irrelevance of the predictor while allowing the relationship between the predictor and the target (dependent) variable to be possibly time-varying.

The test is implemented as follows. Let $\widehat{\beta}_{1\tau}$ and $\widehat{\beta}_{2\tau}$ denote the OLS estimators before and after the break:

$$\begin{aligned}\widehat{\beta}_{1\tau} &= \left(\frac{1}{\tau}\sum_{t=1}^{\tau} x_t x_t'\right)^{-1}\left(\frac{1}{\tau}\sum_{t=1}^{\tau} x_t y_{t+h}\right), \\ \widehat{\beta}_{2\tau} &= \left(\frac{1}{T-\tau}\sum_{t=\tau+1}^{T} x_t x_t'\right)^{-1}\left(\frac{1}{T-\tau}\sum_{t=\tau+1}^{T} x_t y_{t+h}\right).\end{aligned}$$

The test builds on two components: $\frac{\tau}{T}\widehat{\beta}_{1\tau}+\left(1-\frac{\tau}{T}\right)\widehat{\beta}_{2\tau}$ and $\widehat{\beta}_{1\tau}-\widehat{\beta}_{2\tau}$. The first is simply the full-sample estimate of the parameter, $\frac{\tau}{T}\widehat{\beta}_{1\tau}+\left(1-\frac{\tau}{T}\right)\widehat{\beta}_{2\tau} = \left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} x_t y_{t+h}\right)^{-1}$; a test on whether this component is zero is able to detect situations in which the parameter β_t is constant and different from zero. However, if the regressor Granger-causes the dependent variable in such a way that the parameter changes but the average of the estimates equals zero (as in the example previously discussed), then the first component would not be able to detect such situations. The second component is introduced to perform this task. It is the difference between the parameters estimated in the two sub-samples; a test on whether this component is zero is able to detect situations in which the parameter changes. Rossi (2005) proposes several test statistics, among which the following:

$$QLR_T^* = \sup_{\tau=[0.15T],\dots,[0.85T]} \Phi_T^* \tag{3}$$

$$Exp-W_T^* = \frac{1}{T}\sum_{\tau=[0.15T]}^{[0.85T]} \frac{1}{0.7}\exp\left\{\left(\frac{1}{2}\right)\Phi_T^*\right\} \tag{4}$$

$$Mean-W_T^* = \frac{1}{T}\sum_{\tau=[0.15T]}^{[0.85T]} \frac{1}{0.7}\Phi_T^* \tag{5}$$

where:[9] $\Phi_T^* \equiv \left(\left(\widehat{\beta}_{1\tau}-\widehat{\beta}_{2\tau}\right)' \quad \left(\frac{\tau}{T}\widehat{\beta}_{1\tau}+\left(1-\frac{\tau}{T}\right)\widehat{\beta}_{2\tau}\right)' \right)\widehat{V}^{-1}\begin{pmatrix}\left(\widehat{\beta}_{1\tau}-\widehat{\beta}_{2\tau}\right) \\ \left(\frac{\tau}{T}\widehat{\beta}_{1\tau}+\left(1-\frac{\tau}{T}\right)\widehat{\beta}_{2\tau}\right)\end{pmatrix}$,

$$\widehat{V} = \begin{pmatrix} \frac{\tau}{T}S_{xx}'\widehat{S}_1^{-1}S_{xx} & 0 \\ 0 & \frac{T-\tau}{T}S_{xx}'\widehat{S}_2^{-1}S_{xx}\end{pmatrix},$$

[9] The necessity to trim the set of values for τ such that $\tau = [0.15T], ..., [0.85T]$ derives from the fact that one needs a sufficient number of observations to estimate $\widehat{\beta}_{1\tau}$ and $\widehat{\beta}_{2\tau}$ – see Andrews (1993) for example.

$$S_{xx} \equiv \frac{1}{T}\sum_{t=1}^{T} x_t x_t'$$

$$\widehat{S}_1 = \left(\frac{1}{\tau}\sum_{t=1}^{\tau} x_t \widehat{\varepsilon}_{t+h}\widehat{\varepsilon}_{t+h} x_t'\right) + \sum_{j=1}^{\tau}\left(1-\left|\frac{j}{\tau^{1/3}}\right|\right)\left(\frac{1}{\tau}\sum_{t=j+1}^{\tau} x_t \widehat{\varepsilon}_{t+h}\widehat{\varepsilon}_{t+h-j} x_{t-j}'\right), \quad (6)$$

$$\widehat{S}_2 = \left(\frac{1}{T-\tau}\sum_{t=\tau+1}^{T} x_t \widehat{\varepsilon}_{t+h}\widehat{\varepsilon}_{t+h} x_t'\right) \quad (7)$$
$$+\sum_{j=\tau+1}^{T}\left(1-\left|\frac{j}{(T-\tau)^{1/3}}\right|\right)\left(\frac{1}{T-\tau}\sum_{t=j+1}^{T} x_t \widehat{\varepsilon}_{t+h}\widehat{\varepsilon}_{t+h-j} x_{t-j}'\right),$$

for $\widehat{\varepsilon}_{t+h} \equiv y_{t+h} - x_t'\widehat{\beta}$ and (6, 7) are HAC estimates of the relevant variances. If there is no serial correlation in the data, only the first component in (6) and (7) is relevant. Under the null hypothesis of no Granger-causality at any point in time, ($\beta_t = \beta = 0,\ \forall t$), QLR_T^*, $Mean - W_T^*$ and $Exp - W_T^*$ have asymptotic distributions whose critical values depend on the number of predictors, p, and are tabulated in Rossi's (2005) Table B1. For convenience, a subset of the table is reproduced in Table A.1 in Appendix 1.

The Granger-causality test robust to parameter instabilities has been shown to be useful in practice. For example, it was used by Rapach and Wohar (2006) to provide empirical evidence on predictive ability of asset returns, by Giacomini and Rossi (2006) to demonstrate that the term structure Granger-causes future output growth, and by Chen, Rogoff and Rossi (2010) to provide empirical evidence that exchange rates Granger-cause commodity prices. Note that the tests (3, 4, 5) detect in-sample Granger-causality that appeared at some point in the historical sample, which is in many ways similar to pseudo-out-of-sample forecast evaluation procedures whose goal is to evaluate whether, historically, there was forecasting ability; one might instead be interested in detecting whether Granger-causality currently exists, to exploit it for forecasting. An example of the latter is Pesaran and Timmermann's (2002) ROC procedure, discussed in Section 2.3.2.

2.2.2 If There Are Instabilities in Predictive Relationships, How Do Researchers Establish Which Model Forecasts the "Best"?

A second, important series of tools commonly used by practitioners to evaluate forecasts are out-of-sample forecast comparison tests. Typically, they involve comparing two $h-$step ahead forecasts for the variable y_t, which we assume for simplicity to be a scalar.

We assume that the researcher has divided the sample of size, $T+h$, into an in-sample portion of size R and an out-of-sample portion of size P,[10] and obtained two competing sequences of $h-$step ahead out-of-sample forecasts. Let the first model be characterized by parameters θ_1 and the second model by parameters θ_2. For a general loss function $L(.)$, we thus have a sequence of P out-of-sample forecast loss differences, $\left\{\Delta L_{t+h}\left(\widehat{\theta}_{t,R}\right)\right\}_{t=R}^{T}$ $\equiv \left\{L^{(1)}(y_{t+h},\widehat{\theta}_{1,t,R}) - L^{(2)}(y_{t+h},\widehat{\theta}_{2,t,R})\right\}_{t=R}^{T}$, which depend on the realizations of the variable and on the in-sample parameter estimates for each model, $\widehat{\theta}_{t,R} \equiv [\widehat{\theta}'_{1,t,R}, \widehat{\theta}'_{2,t,R}]'$. These parameters are typically estimated only once, using a sample including data indexed $1, ..., R$ (fixed scheme) or re-estimated at each $t = R, ..., T$ over a window of R data including data indexed $t - R + 1, ..., t$ (rolling scheme) or re-estimated at each $t = R, ..., T$ over a window of R data including data indexed $1, ..., t$ (recursive scheme). See Section 2.3.1 for more details. In this section, we assume that the researcher is using either a rolling scheme with a fixed window size R or a fixed scheme, and discuss the recursive window scheme as a special case. Also, here and in the rest of the chapter, we simplify notation, and denote the sequence of out-of-sample forecast error loss differences $\left\{\Delta L_{t+h}\left(\widehat{\theta}_{t,R}\right)\right\}_{t=R}^{T}$ as:

$$\{\Delta L_{t+h}\}, \text{ for } t = R, R+1, ..., T. \tag{8}$$

For example, in the case of a quadratic loss function, eq. (8) is the sequence of the difference between the two models' squared forecast errors, and their average is the Mean Squared Forecast Error, or MSFE.

Typically, researchers establish which model forecasts the best by looking at the average out-of-sample forecast error loss difference. For example, the statistic proposed by Diebold and Mariano (1995) and West (1996), which we refer to as DMW_P or $MSE - t$, is:

$$DMW_P = \widehat{\sigma}^{-1} P^{-1/2} \sum_{t=R}^{T} \Delta L_{t+h}, \tag{9}$$

where $\widehat{\sigma}^2$ is a HAC estimator of

$$\sigma^2 = \lim_{T\to\infty} var\left(P^{-1/2}\sum_{t=R}^{T} \Delta L_{t+h}\right). \tag{10}$$

The limiting distribution of DMW_P is typically obtained under stationarity assumptions.

The implications of structural instability for forecast evaluation have not been formally investigated in the literature until relatively recently.[11] Giacomini and Rossi (2010a and

[10] P is such that $R + P + h = T + h$.

[11] The typical approach to forecast evaluation, based on assessing the expected loss of some forecast relative

2010b) focus on the relative evaluation of two models.[12] In particular, Giacomini and Rossi (2010a) test the null hypothesis:

$$E\left(\Delta L_{j+h}\right)=0,\ \forall t=R,...,T, \tag{11}$$

They introduce two classes of methods, which depend on whether one considers a smooth change or a one-time change at an unknown date under the alternative hypothesis. Note that the conventional tests by Diebold and Mariano (1995), West (1996), Clark and McCracken (2001), Giacomini and White (2005) and Clark and West (2006, 2007) would assume that $E\left(\Delta L_{j+h}\right)$ is constant over time ($E\left(\Delta L_{j+h}\right)=\mu$), and test the hypothesis that $\mu=0$ by a standard t-test. Note that the latter tests differ on the specification of the null hypotheses and on the treatment of parameter estimation error. We refer to West (2006) as well as the chapter by Clark and McCracken (in this Handbook) for an extensive review of conventional tests of predictive ability.[13] The methodologies proposed by Giacomini and Rossi (2010a) can be implemented no matter which of the latter approaches the researcher prefers.

Smooth change in relative performance. In this scenario, the models' relative performance is estimated by a kernel estimator, which, for the rectangular kernel, amounts to computing

---

to that of a benchmark, starts with the premise that there exists a forecast that is "globally best", in the sense that its performance is superior to that of its competitors at all time periods. From an econometric point of view, this means assuming that the expectation in the the measure of performance, $E\left[L\left(\cdot\right)\right]$, is constant over time, and can therefore be estimated by the average loss computed over the entire out-of-sample period. In the applied literature, some authors more or less explicitly acknowledge that this might be a restrictive assumption by computing the average loss over subsamples that are chosen in an arbitrary way (e.g., the 1980s and the 1990s) (e.g., Stock and Watson, 2003, and D'Agostino et al., 2008). The typical finding of these studies is that the performance varies widely across subsamples. Whereas the analysis in this applied literature is informal, Giacomini and Rossi (2009, 2010a and 2010b) have recently introduced formal methods for forecast evaluation in the presence of instability. Giacomini and Rossi (2009) focused on "absolute measures" of accuracy, wheareas Giacomini and Rossi (2010a and 2010b) considered "relative measures".

[12] Giacomini and Rossi (2010b) consider the following local-level model for in-sample loss differences: $\Delta L_t=\mu_t+\varepsilon_t$, $t=1,...,T$, and propose a method for testing the hypothesis of equal performance at each point in time: $H_0:\mu_t=0$ for all t.

[13] Note that the Giacomini and White (2006) test requires forecasts to satisfy a mixing assumption. Thus, the test is robust to "small" structural changes that satisfy the mixing assumption, but not robust to breaks that generate non-stationarity.

rolling average losses:[14]

$$\widehat{\mu}_t = m^{-1} \sum_{j=t-m/2}^{t+m/2-1} \Delta L_{j+h}, \ t = R + m/2, ..., T - m/2 + 1. \tag{12}$$

In practice, their test involves computing the sequence of statistics:[15]

$$F_t = \widehat{\sigma}^{-1} m^{-1/2} \sum_{j=t-m/2}^{t+m/2-1} \Delta L_{j+h}, \ t = R + m/2, ..., T - m/2 + 1, \tag{13}$$

where $\widehat{\sigma}^2$ is a HAC estimator of (10), e.g.,

$$\widehat{\sigma}^2 = \sum_{s=-\widetilde{q}+1}^{\widetilde{q}-1} (1 - |s/\widetilde{q}|) P^{-1} \sum_{t=R}^{T} \Delta L_{t+h} \Delta L_{t+h-s}, \tag{14}$$

and $\widetilde{q}$ is an appropriately chosen bandwidth (see e.g., Andrews, 1991 and Newey and West, 1987).[16] To test the null hypothesis of equal predictive ability at each point in time against the alternative that one of the two models forecasts the best at least one point in time, Giacomini and Rossi (20010a) propose the following Fluctuation test statistic:

$$F_P = \max_t |F_t| \,. \tag{15}$$

The null hypothesis is rejected at the $100\alpha\%$ significance level against the two-sided alternative for some t when $\max_t |F_t| > k_\alpha^{GR}$, where k_α^{GR} is the appropriate critical values. The critical values depend on δ, and are reported in their Table 1.[17] Selected values are reproduced in Table A.2 in the Appendix for convenience. Critical values for testing H_0 against the one-sided alternative $E(\Delta L_{j+h}) > 0$ for some t are reported as well in Table A.2 for various choices of δ, in which case the null is rejected when $\max_t F_t > k_\alpha^{GR}$.

[14] Here we use a rectangular kernel estimator centered at time $j + h$; one-sided kernels could alternatively be used.

[15] To test the null hypothesis, one has two options: either considering the standard nonparametric approximation which assumes that the bandwidth m/P goes to zero at an appropriate rate as $m, P \to \infty$, or to consider a different asymptotic approximation that assumes m/P to be fixed and equal to δ as $m, P \to \infty$. That is, $\underset{T\to\infty}{lim} \frac{m}{P} = \delta$. Giacomini and Rossi (2010b) show that in the former case one could use uniform confidence bands to construct a test, but that the procedure has poor finite sample properties.

[16] Alternatively, the variance can be estimated in each of the rolling windows, $\widehat{\sigma}_j^2 = \sum_{s=-\widetilde{q}+1}^{\widetilde{q}-1} (1 - |s/\widetilde{q}|)$ $m^{-1} \sum_{j=t-m/2}^{t+m/2-1} \Delta L_{j+h} \Delta L_{j+h-s}$, , and the test be constructed as: $F_t = m^{-1/2} \sum_{j=t-m/2}^{t+m/2-1} \widehat{\sigma}_j^{-1} \Delta L_{j+h}$, $t = R + m/2, ..., T - m/2 + 1$.

[17] Under the null hypothesis (**??**), Giacomini and Rossi (2009) show that the asymptotic distribution of F_t is a functional of Brownian motions.

The test statistic F_t in (13) is equivalent to Diebold and Mariano's (1995) and Giacomini and White's (2006) (unconditional) test statistic, computed over rolling out-of-sample windows of size m. Giacomini and Rossi (2010a) show that their approach can be generalized to allow for any other commonly used for out-of-sample predictive ability comparisons, as long as their asymptotic distribution is Normal. In particular, one could use the test statistics proposed by West (1996) or by Clark and West (2006, 2007), which are respectively applicable to non-nested and nested models.[18] The adoption of West's (1996) framework involves replacing $\widehat{\sigma}$ in (14) with an estimator of the asymptotic variance that reflects the contribution of estimation uncertainty (see Theorem 4.1 of West (1996)). For the nested case, the use of the Clark and West (2006, 2007) test statistic in practice amounts to replacing ΔL_{j+h} in (13) with Clark and West's (2006, 2007) corrected version).

Also note that West's (1996) approach allows the parameters to be estimated using a recursive scheme, in addition to a rolling or fixed scheme. In that case, let $\left\{W_t^{OOS}\right\}$ denote a sequence of West's (1996) test statistics for h-steps ahead forecasts calculated over recursive windows (with an initial window of size R) for $t = R + h + m/2, ..., T - m/2 + 1$. Giacomini and Rossi (2010a) show that the null hypothesis of equal predictive ability is rejected when $\max_t \left|W_t^{OOS}\right| > k_\alpha^{rec}\sqrt{\frac{T-R}{t}}\left(1 + 2\frac{t-R}{T-R}\right)$, where (α, k_α^{rec}) are $(0.01, 1.143)$, $(0.05, 0.948)$ and $(0.10, 0.850)$.

One-time reversal in the relative forecasting performance at unknown date. In this scenario, the alternative hypothesis postulates a one-time change in relative performance at an unknown date.[19] The test is performed as follows:

(i) Consider the test statistic

$$\begin{aligned} QLR_P &= \sup_t \Phi(t), t \in \{[0.15P], ... [0.85P]\}, \\ \Phi(t) &= LM_1 + LM_2(t), \end{aligned}$$

[18] The fundamental difference between these approaches and Giacomini and White (2006) is that they test two different null hypotheses: the null hypothesis in West (1996) and Clark and West (2006, 2007) concerns forecast losses that are evaluated at the population parameters, whereas in Giacomini and White (2006) the losses depend on estimated in-sample parameters. This reflects the different focus of the two approaches on comparing forecasting models (West, 1996, and Clark and West, 2006, 2007) versus comparing forecasting methods (Giacomini and White, 2006).

[19] Note that the test against one-time change in the relative performance of the models will have power against multiple reversals since it would capture the largest reversal. It might also be interesting to extend the approach to multiple breaks following Bai and Perron (1998).

where

$$LM_1 = \widehat{\sigma}^{-2} P^{-1} \left[\sum_{t=R}^{T} \Delta L_{t+h} \right]^2 \tag{16}$$

$$LM_2(t) = \widehat{\sigma}^{-2} P^{-1} (t/P)^{-1} (1 - t/P)^{-1} [\sum_{j=R}^{t} \Delta L_{j+h} - (t/P) \sum_{j=R}^{T} \Delta L_{j+h}]^2,$$

and $\widehat{\sigma}^2$ is as in (14). Reject the null hypothesis $H_0 : E[\Delta L_{t+h}] = 0$, for every $t = R, ..., T$ when $QLR_P > k_\alpha$, where (α, k_α) are, e.g., $(0.01, 13.4811)$, $(0.05, 9.8257)$ and $(0.10, 8.1379)$.

(ii) If the null is rejected, compare LM_1 and $\sup_t LM_2(t)$, $t \in \{[0.15P], ... [0.85P]\}$ with the critical values: $(2.71, 7.17)$ for $\alpha = 0.10$, $(3.84, 8.85)$ for $\alpha = 0.05$, and $(6.63, 12.35)$ for $\alpha = 0.01$. If only LM_1 rejects, conclude that one model is constantly better than its competitor. If only LM_2 rejects, conclude that there are instabilities in the relative performance of the two models but neither is constantly better over the full sample. If both reject, then it is not possible to attribute the rejection to a unique source.

(iii) Estimate the time of the change by $t^* = \arg\max_{t \in \{0.15P,...,0.85P\}} LM_2(t)$.

(iv) Estimate the path of relative performance as

$$\begin{cases} \frac{1}{t^*} \sum_{j=R}^{t^*} \Delta L_{j+h}, & \text{for } t < t^* \\ \frac{1}{(P-t^*)} \sum_{j=t^*+1}^{T} \Delta L_{j+h}, & \text{for } t \geq t^* \end{cases}.$$

Note that the Fluctuation and the One-time reversal tests capture changes other than in the conditional mean (such as changes in the variance of the forecast error), whereas Rossi's (2005) test does not.[20]

One might think that the problem of time variation in models' relative forecasting performance is minor. On the contrary, substantial time-variation in models' relative predictive ability of inflation, for example, has been documented since Stock and Watson (2007). They notice that the root mean squared error (RMSE) of univariate benchmark inflation forecasts (obtained using either autoregressive or random walk models) has declined sharply during the period 1990s-early 2000 relative to the 1970s or early 1980s since inflation (like many other time series) has been much less volatile. This implies that inflation has been easier to forecast in the former period. However, on the other hand, the relative improvement of standard

[20]It might be interesting to directly model relative out-of-sample forecast error losses as following a regime switching process. To the extent that there is cyclical behavior in relative performance, and that it can be captured using a regime switching model, adopting a specification that accommodates this variation might improve inference.

multivariate forecasting models (e.g. Phillips curve models) over the univariate benchmark model has decreased in the 1990s-early 2000 relative to the previous period. Therefore, in a sense, it is also true that inflation has become harder to forecast. Stock and Watson (2007) propose a time-varying trend-cycle model for univariate inflation which will be reviewed in details in Section 2.3.2. According to their model, during the 1970s the inflation process was well approximated by a low-order autoregression (AR) with a substantial permanent component (akin to a trend) whose variance was large; thus, the estimation of such permanent component provided large gains relative to simple univariate benchmark models, even though this resulted in a large MSFE. However, the coefficients of the AR model changed since 1984, and the AR model has become since then a less accurate approximation to the inflation process.

Stock and Watson's (2007) findings are consistent with recent results by Rossi and Sekhposyan (2010), which we discuss at length as they are related to the empirical analysis in this chapter. Rossi and Sekhposyan (2010) use Giacomini and Rossi's (2010a) Fluctuation test to empirically investigate whether the relative performance of competing models for forecasting U.S. industrial production growth and consumer price inflation has changed over time. They focus on the same models considered by Stock and Watson (2003), but use monthly data. Their predictors include interest rates, measures of real activity (such as unemployment and GDP growth), stock prices, exchange rates and monetary aggregates. Their benchmark model is the autoregressive model. Using both fully revised and real-time data, they find sharp reversals in the relative forecasting performance. They also estimate the time of the reversal in the relative performance, which allows them to relate the changes in the relative predictive ability to economic events. In particular, when forecasting output growth, interest rates and the spread were useful predictors in the mid-1970s, but their performance worsened at the beginning of the 1980s.[21] When forecasting inflation, the empirical evidence in favor of predictive ability is weaker than that of output growth, and the predictive ability of most variables breaks down around 1984, which dates the beginning of the Great Moderation. Such predictors include employment and unemployment measures, among others, thus implying that the predictive power of the Phillips curve disappeared around the time of the Great Moderation. Section 4 revisits this empirical evidence using data up to 2010 (whereas Rossi and Sekhposyan's (2010) sample ended in 2005) and using quarterly data.[22]

[21] Similar results hold for money growth (M2), the index of supplier deliveries, and the index of leading indicators.

[22] Rossi and Sekhposyan (2010) also document the robustness of their results to the use of real-time data

D'Agostino et al. (2008) also consider sub-samples identified by the Great Moderation and show a similar decrease in predictive ability of factor models as well as forecast combinations for inflation. They also find a decrease in predictive ability in GDP growth at the time of the Great Moderation. Their timing of the reversal in predictive ability seems to be at odds with Rossi and Sekhposyan (2010), who estimated the largest break to be around the mid-1970s; however, if it were in the mid-1970s, one would still find a decrease in predictive ability by looking at the two sub-samples before and after the Great Moderation.

Similar findings hold for other databases. Carstensen et al. (2010) evaluate the predictive ability of seven leading indicators for euro area industrial production. They implement Giacomini and Rossi's (2010a) Fluctuation test to evaluate the forecasting stability of each indicator over time, especially during booms and recessions. They find that a simple autoregressive benchmark is difficult to beat in normal times whereas the indicators have advantageous predictive ability in booms and recessions. A similar result is found by Diaz and Leyva (2008) for forecasting inflation in Chile. Additional examples of time variation in the relative performance of financial models over time and linked to the business cycle include Rapach, Strauss and Zhou (2010) and Henkel, Martin and Nardari (2011), who find that stock return predictability concentrates during recessions, and Paye and Vol (2011), who find that the ability of macroeconomic variables to improve long horizon volatility forecasts concentrates around the onset of recessions. A series of papers have also built on the empirical evidence of a breakdown in the ability of forecasting models to predict U.S. inflation and output: see Castelnuovo et al. (2008) for a regime-switching model in Taylor rules which finds a switch towards active monetary policy at the time of the Great Moderation. Fernandez, Koenig and Nikolosko-Rzhevskyy (2010) evaluate how well several alternative Taylor rule specifications describe Federal Reserve policy decisions in real time. Giacomini and Rossi (2010a) evaluate the instability in the predictive ability of fundamental-based models of exchange rates. They argue that, as shown by Rossi (2006), the estimates of exchange rate models with economic fundamentals are plagued by parameter instabilities, and so might the resulting exchange rate forecasts. They show that conventional out-of-sample forecast comparison tests do find some empirical evidence in favor of models with economic fundamentals for selected countries. However, the Fluctuation test indicates that the relative forecasting performance has changed over time. In fact, their procedures indicate that the Deutsche Mark and the

---

(Croushore and Stark, 2001): the evidence in favor of predictive ability in the early part of the sample is slightly weaker for a few series when using real-time data; however, their main qualitative conclusions are robust to the use of real-time data.

British Pound exchange rates were predictable in the late Eighties, but such predictability has disappeared in the Nineties. Conventional out-of-sample tests would have been unable to uncover such evidence in favor of models with economic fundamentals. Finally, Galvão (2011) considers a smooth transition regression to model regime changes in high frequency variables for predicting low frequency variables using a MIDAS framework.[23] She focuses on real-time forecasts of U.S. and U.K. output growth using daily financial indicators. The Fluctuation test reveals strong evidence of instability in the predictive content of financial variables for forecasting output growth. In addition, she finds evidence that the inclusion of nonlinearities (captured by the smooth transition model) may improve predictive ability.

2.2.3 If There Are Instabilities In Forecasting Performance, How Should Researchers Determine Whether Forecasts Are Optimal?

Under a MSFE loss function, optimal forecasts have several properties: they should be unbiased, the one step ahead forecast errors should be serially uncorrelated, and h-steps ahead forecast errors should be correlated at most of order $h-1$. A large literature has focused on empirically testing whether forecasts are actually optimal – see Granger and Newbold, 1986, Diebold and Lopez, 1996, Patton and Timmermann (2011), among others.

However, traditional tests for forecast optimality are subject to the same issues as the other tests previously discussed: they are potentially inconsistent in the presence of instabilities. In a recent paper, Rossi and Sekhposyan (2011b) have developed methodologies for implementing forecast rationality and forecast optimality tests robust to instabilities. They follow the general framework developed in West and McCracken (1998). Let's assume one is interested in the (linear) relationship between the prediction error and a vector of variables known at time t. Let the h-steps ahead forecast made at time t be denoted by $y_{t+h|t}$ and let a $(p \times 1)$ vector of variables known at time t to be denoted by g_t. The variables in g_t are not used to produce the forecast; rather, they will be used to study whether their correlation with the forecast error is zero; in fact, if the forecasts are optimal, the forecast error should be uncorrelated with any information available at the time the forecasts are made. Finally, let the forecast error of a model evaluated at the true parameter value, θ^*, be denoted by v_{t+h}, and its estimated value be denoted by $\widehat{v}_{t+h}$.

[23] MIDAS models are designed for modeling variables that are available at different frequencies; for a discussion of MIDAS regressions, see Andreou, Ghysels and Kourtellos (2010).

Consider the regression:

$$v_{t+h} = g_t' \cdot \phi + \eta_{t,h}, \text{ for } t = R, ..., T, \tag{17}$$

where ϕ is a $(p \times 1)$ parameter vector. The null hypothesis of interest is $H_0 : \phi = \phi_0$, where typically $\phi_0 = 0$. For example, in forecast rationality tests (Mincer and Zarnowitz, 1969), v_{t+h}, $g_t = [1, y_{t+h|t}]$, $\phi = [\phi_1, \phi_2]'$, and typically a researcher is interested in testing whether ϕ_1 and ϕ_2 are jointly zero.[24] For forecast unbiasedness, $g_t = 1$, for forecast encompassing g_t is the forecast of the encompassed model, and for serial uncorrelation $g_t = v_t$. We will refer to all these tests as "tests for forecast optimality". To test forecast optimality, one typically uses the following re-scaled Wald test:

$$\mathcal{W}_T = \widehat{\phi}' \, \widehat{V}_\phi^{-1} \widehat{\phi}, \tag{18}$$

where $\widehat{V}_\phi$ is a consistent estimate of the long run variance of the parameter vector obtained following West and McCracken (1998). West and McCracken (1998) have shown that it is necessary to correct eq. (18) for parameter estimation error in order to obtain test statistics that have good size properties in small samples, and proposed a general variance estimator as well as adjustment procedures that take into account estimation uncertainty.

Rossi and Sekhposyan (2011b) propose the following procedure, inspired by Giacomini and Rossi (2010a). Let $\widehat{\phi}_t$ be the parameter estimate in regression (17) computed over centered rolling windows of size m (without loss of generality, we assume m to be an even number). That is, consider estimating regression (18) using data from $t - m/2$ up to $t + m/2 - 1$, for $t = m/2, ..., P - m/2 + 1$. Also, let the Wald test in the corresponding regressions be defined as:

$$\mathcal{W}_{t,m} = \widehat{\phi}_t' \, \widehat{V}_{\phi,t}^{-1} \widehat{\phi}_t, \text{ for } t = m/2, ..., P - m/2 + 1, \tag{19}$$

where $\widehat{V}_{\phi,t}$ is a consistent estimator of the asymptotic variance of the parameter estimates in the rolling windows obtained following West and McCracken (1998). Rossi and Sekhposyan (2011b) refer to $\mathcal{W}_{t,m}$ as the Fluctuation optimality test. The test rejects the null hypothesis $H_0 : E\left(\widehat{\phi}_t\right) = 0$ for all $t = m/2, ..., P - m/2 + 1$ if $\max_t \mathcal{W}_{t,m} > k_{\alpha,p}^{RS}$, where $k_{\alpha,p}^{RS}$ are the critical values at the $100\alpha\%$ significance level. The critical values are reported in their Table 1 for various values of $\mu = [m/P]$ and the number of restrictions, p.[25] The critical values at 5% significance level are reproduced in Table A.3 in Appendix 1 for convenience for the cases

[24] This is similar to testing whether the slope is one and the intercept is zero in a regression of y_{t+h} onto a constant and $y_{t+h|t}$.

[25] Here we assume that the researcher is interested in jointly testing whether all the ϕ are equal to zero,

of one and two regressors (that is, the cases of forecast unbiasedness and Mincer-Zarnowitz's (1969) regressions).

A simple, two-sided t-ratio test on the s-th parameter, $\phi^{(s)}$, can be obtained as $\widehat{\phi}_t^{(s)}$ $\widehat{V}_{\phi^{(s)},t}^{-1/2}$, where $\widehat{V}_{\phi^{(s)},t}$ is element in the s-th row and s-th column of $\widehat{V}_{\phi,t}$; then, reject the null hypothesis $H_0 : E\left(\widehat{\phi}_t^{(s)}\right) = \phi_0^{(s)}$ for all $t = m/2, ..., P - m/2 + 1$ at the $100\alpha\%$ significance level if $\max_t \left|\widehat{\phi}_t^{(s)} \widehat{V}_{\phi^{(s)},t}^{-1/2}\right| > k_\alpha^{GR}$, where k_α^{GR} are the critical values provided by Giacomini and Rossi (2010a) – see Table A.2 in Appendix 1.

Rossi (2011) considers the robustness of forecast rationality tests to instabilities in Federal Reserve "Greenbook" forecasts of quarter-over-quarter rates of change in GDP and the GDP deflator, the same database considered in Faust and Wright (2009) and Patton and Timmermann (2011). Using both heuristic empirical evidence of time variation in the rolling estimates of the coefficients of forecast rationality regressions as well as the Fluctuation optimality test, she rejects forecast rationality. The Fluctuation optimality test, eq. (19), is also applied to the Patton and Timmermann's (2011) optimal revision regression tests, which shows that forecast rationality is not rejected for the GDP deflator, whereas it is rejected for GDP growth mainly in the late 1990's. Rossi and Sekhposyan (2011b) use the same technique to test whether the Federal Reserve has an information advantage in forecasting inflation beyond what is known to the private forecasters. They find evidence that the Federal Reserve has an informational advantage relative to the private sector's forecasts, although it deteriorated after 2003.[26]

2.3 Estimation When the Predictive Content Is Unstable Over Time

Given the widespread empirical evidence of instabilities in the data, established in the previous section, it is reasonable to ask whether it is possible to exploit such instabilities to improve the estimation of forecasting models. For example, one might expect that, in the presence of a one-time break in the parameters, it might be possible to improve models' estimation by determining the time of the break and then use only the observations after the

and hence the number of restrictions is p. Alternatively, one might be interested in testing whether a subsets of ϕ are equal to zero, in which case the test statistic should consider only a subset of ϕ and the degrees of freedoms should be adjusted accordingly to be equal to the subset dimension.

[26] They also find empirical evidence against rationality in the Money, Market and Services (MMS) survey forecasts once instabilities are taken into account.

breaks for forecasting, as this would provide unbiased parameter estimates. However, this intuition might be misleading. First, it might be very difficult to constructively utilize break dates to improve forecasts in practice because the time of the break might be imprecisely estimated. As shown by Elliott and Muller (2007), paradoxically, even in a simple model with a single, one-time break, it is more difficult to determine the exact break date than it is to determine whether there was a break or not in the data. Elliott and Muller (2007) also show that standard methods for constructing confidence intervals for the break date have poor coverage rates, and propose a new methodology that accurately captures the uncertainty in the estimated break date. Second, even if one were able to estimate the time of the break with sufficient precision, Pesaran and Timmermann (2002) show the existence of a trade-off between bias and variance in the evaluation of MSFE, which might favor estimation using more data than just the observations after the break. In a nutshell, while the detection of structural breaks and their type are clearly important for econometric modeling, it is difficult to use that information productively to improve forecasts.

Overall, several estimation procedures have been proposed:

(i) *Ad-hoc estimation methods,* such as rolling or recursive estimation schemes, discounted least squares, and exponential smoothing. They provide an agnostic, non-parametric way to sequentially update the parameter vector. But which one should be used? Should we give all the observations the same weight (as the rolling estimation window does, for example), or should we give more weight to recent observations and discount the older ones (as discounted least squares does)? And how should researchers choose the size of the estimation window? Researchers have also suggested to improve forecasts by averaging across window sizes (Pesaran and Timmermann, 2007), as well as forecast evaluation methods whose conclusions are robust to the estimation window size (Inoue and Rossi, 2010, and Hansen and Timmermann, 2011).

(ii) *Estimate historic breaks*, by either testing for breaks (e.g. using Andrews, 1993, Bai and Perron, 1998, Elliott and Mueller, 2006, among others), or by adapting the estimation window to the latest break (Pesaran and Timmermann, 2002), or by explicitly modeling the size and duration of the breaks process, either via time-varying parameter models (with a change point every period, as in Stock and Watson, 2007) or models with multiple discrete breaks (Pesaran, Pettenuzzo and Timmermann, 2006, and Koop and Potter, 2007), or intercept corrections (Clemens and Hendry, 1996).

(iii) *Combine forecasts*, either by using equal weights or by using time-varying weights estimated using either frequentist procedures or Bayesian model averaging.

In what follows, we review each of these approaches. Section 4 provides an evaluation of how several of these estimation methodologies perform in practice.

2.3.1 If There Are Instabilities, Do Ad-Hoc Estimation Methods Help in Forecasting?

Ad-hoc forecasting methods are not based on any parametric model. They are simple to implement and still used by practitioners. There are several such ad-hoc methods, differing according to the weight that they give to observations.

(i) Simple exponentially weighted moving average (EWMA, or exponential smoothing). The EWMA forecasts made at time t for predicting y_{t+h} are:

$$y_{t+h|t}^{ES,f} = \alpha_t y_t + (1-\alpha_t)\, y_{t|t-h}^{ES,f} \tag{20}$$

where α_t is the adaptive parameter. α_t can be fixed a-priori or estimated by minimizing the sum of squared forecast errors; a large estimated value of α_t is a signal that the series is close to a random walk. The initial value for the recursion can be the initial observation.[27] Holt (1957) and Winters (1960) generalized the approach to include a local linear trend. See Harvey (1989, sec. 2.2.2).

(ii) Discounted least squares. A general version of the simple discounted least squares method (DLS, Brown, 1963) in the model with exogenous regressors such as eq. (1) implies choosing parameters estimates that minimizes the discounted sum-of-squared residuals. For simplicity of exposition, consider the simplified model: $y_{t+h} = \beta_h' x_t + \varepsilon_{t+h}$, $t = 1, ..., T$. Let $\underline{y}_{t+h} \equiv [y_{t+h-R+1}, ..., y_{t+h}]'$, $\underline{X}_{t,R} \equiv \left[x_{t-R+1}', ..., x_t'\right]'$, and $\underline{W}_t \equiv diag\left(\delta^{R-1}, ..., \delta, 1\right)$ be the matrix of weights to discount past observations. Then, DLS estimates the parameters at time t as (see Agnew, 1982):[28]

$$\widehat{\beta}_{h,t}^{DLS} = \left(\underline{X}_{t,R}' \underline{W}_t \underline{X}_{t,R}\right)^{-1} \underline{X}_{t,R}' \underline{W}_t \underline{y}_{t+h}, \tag{21}$$

and

$$y_{t+h|t}^{DLS,f} = \widehat{\beta}_{h,t}' x_t.$$

[27]For $h = 1$ and α_t constant, the EWMA corresponds to a forecast that is a weighted average of previous observations, where the weights are declining exponentially: $y_{t+1|t}^{ES,f} = \sum_{j=0}^{t-1} \omega_j y_{t-j}$, and $\omega_j = \lambda (1-\lambda)^j$.

[28]When $h = 1$ and the model includes only a constant, the formula simplifies to: $y_{t+1|t}^{DLS,f} = \left(\sum_{j=0}^{t-1} \delta^j\right)^{-1} \left(\sum_{j=0}^{t-1} \delta^j y_{t-j}\right)$, as in Brown (1963).

The weights can be either imposed a-priori or estimated.[29] Typically, one might prefer to give higher weight to more recent observations and lower weight to more distant observations, which would be a successful strategy if later observations reflect more accurately the most recent data generating process.

(iii) Rolling and recursive window estimation schemes. Note that several estimation weighting schemes that have become popular in the forecasting literature are special cases of eq. (21). For example, the recursive window estimation scheme is such that $\delta = 1$, that is all observations are weighted equally, and $R = t$, that is all observations in the sample up to time t are used in the estimation:

$$\widehat{\beta}_{h,t}^{REC} = \left(\underline{X}_{t,t}'\underline{X}_{t,t}\right)^{-1}\underline{X}_{t,t}'\underline{y}_{t+h} = \left(\sum_{j=1}^{t} x_j x_j'\right)^{-1}\left(\sum_{j=1}^{t} x_j y_{j+h}\right), \tag{22}$$

whereas the rolling window estimation scheme with window size R is such that:

$$\widehat{\beta}_{h,t}^{ROL} = \left(\underline{X}_{t,R}'\underline{X}_{t,R}\right)^{-1}\underline{X}_{t,R}'\underline{y}_{t+h} = \left(\sum_{j=t-m+1}^{t} x_j x_j'\right)^{-1}\left(\sum_{j=t-m+1}^{t} x_j y_{j+h}\right). \tag{23}$$

Rolling or recursive window estimation procedures are agnostic, non-parametric ways to update the parameter vector. But which one should be used? Pesaran and Timmermann (2002) show that, when regressors are strictly exogenous, in the presence of a structural break in the parameters OLS estimates based on post-break data are unbiased. Including pre-break data always increases the bias; thus, there is always a trade-off between a larger squared bias and a smaller variance of the parameter estimates as more pre-break information is used. In particular, rolling estimation is advantageous in the presence of big and recurrent breaks whereas recursive estimation is advantageous when such breaks are small or non-existent. Pesaran and Timmermann (2002) use this trade-off to optimally determine the window size. On the other hand, Pesaran and Timmermann (2005) show that the situation can be very different in autoregressive models, for which the coefficients inherit a small sample bias. They show that when the true coefficient declines after a break, both the bias and the forecast error variance can be reduced using pre-break data in the estimation. Thus, in these cases, rolling windows could perform worse than recursive windows even in the presence of breaks. This might explain why, in some cases, recursive window forecasts perform better than rolling window forecasts. As discussed in Pesaran and Timmermann (2005), the choice

[29] E.g. one might estimate $\widetilde{\beta}_{h,\tau} = \arg\min_{\beta} \sum_{t=\tau-R+1}^{\tau} \omega_t^2 \left(y_{t+h} - \beta_h' x_t\right)^2$, where ω_t^2 are weights and are typically constrained to be between zero and one and to sum to unity.

of the window size depends on the nature of the possible model instability and the timing of the breaks. A large window is preferable if the data generating process is stationary, but comes at the cost of lower power since there are fewer observations in the evaluation window. Similarly, a shorter window may be more robust to structural breaks, although it may not provide as precise estimation as larger windows if the data are stationary.

Pesaran and Timmermann (2007) find that the optimal length of the observation window is weakly decreasing in the magnitude of the break, the size of any change in the residual variance, and the length of the post-break period. They also consider model combinations as a competitor to the optimal choice of the observation window. Their approach is to determine the window size that guarantees the best forecasting performance, especially in the presence of breaks. They propose several methods in practice. Among the methods they propose, several are available if the researcher possesses an estimate of the break, in which case, using either only the post-break window data to estimate the parameter or a combination of pre- and post-break data according to weights that trade-off bias against reduction in parameter estimation error, might improve forecasting performance. A difficulty in the latter methods is the fact that, in practice, it may be difficult to have a precise estimate of the time and magnitude of the break. Thus, rather than selecting a single window, it might be convenient to combine forecasts based on several estimation windows. A very simple way to combine forecasts based on several estimation windows is to simply average them using equal weights. That is, imagine that the researcher is interested in estimating the parameters of the models using the latest R available observations, and imagine that the researcher's minimum number of observations to be used for estimation is $\underline{R}$. Denote the forecast for the target variable h-steps into the future made at time t based on data from the window size R (that is data from time $t-R+1$ to t) by $y^{f}_{t+h|t}(R)$. Then the average ("Ave") forecast proposed by Pesaran and Timmermann (2007) is:

$$y^{AVE,f}_{t+h|t} = \left(T - \underline{R} + 1\right)^{-1} \sum_{R=t-\underline{R}}^{t} y^{f}_{t+h|t}(R) \tag{24}$$

Pesaran and Timmermann (2007) demonstrate, via Monte Carlo simulations, that in the case of many breaks, forecast combinations obtained in eq. (24) perform quite well, especially when the magnitude of the break is very small and thus the break is more difficult to detect.

It is also possible that better forecasts be obtained by combining rolling and recursive forecasts. Clark and McCracken (2009) show that there is a bias-variance trade-off between rolling and recursive forecasts in the presence of model instability. By analyzing the trade-off, they analytically derive the optimal estimation window. Let $y_{t+h} = \beta_t + \varepsilon_{t+h}$, where $\varepsilon_{t+h} \sim iid\,(0, \sigma^2)$, $\beta_t = \beta^* + T^{-1/2}1\,(t \geq t^*)\,\Delta$ and $\tau^* \equiv t^*/T$. Note that the breakpoint is local-to-zero, which allows Clark and McCracken (2009) to emphasize the importance of the observation window in situations where structural break tests may have little power. The OLS parameter estimate based on rolling windows of size R_t will be $\widehat{\beta}_{roll,t} = R_t^{-1} \sum_{j=t-R_t+1}^{t} y_j$ and the one based on recursive windows will be $\widehat{\beta}_{rec,t} = t^{-1} \sum_{j=1}^{t} y_j$. Note that the rolling window parameter estimates are based on a partial sample whose size (R_t) is allowed to change as forecasting moves forward in time. Let $\widehat{t^*}$ be an estimate of the time of the break, $\widehat{\sigma}$ be an estimate of σ, and $\widehat{\Delta}$ be an estimate of the size of the break in the parameter. Clark and McCracken (2009) show that the optimal window to use in the rolling scheme is $R^* = t - \widehat{\tau}$ (so that the optimal window uses only data after the break) and that the forecast that minimizes the MSFE is a weighted average of the rolling and recursive parameter estimates:

$$\alpha_t^* \widehat{\beta}_{rec,t} + (1 - \alpha_t^*)\,\widehat{\beta}_{roll,t}, \text{ where } \alpha_t^* = \left(1 + \frac{t}{\widehat{\sigma}} \widehat{\Delta}^2 \frac{\widehat{t^*}}{T} \left(1 - \frac{\widehat{t^*}}{T}\right)\right)^{-1}. \quad (25)$$

The result in eq. (25) can be explained, again, by noting that using data before the break in the estimation of the parameter value after the break would lead to a bias in the parameter estimate and in their forecast, which results in an increase in the MSFE of the recursive forecast relative to the rolling; on the other hand, reducing the sample by choosing a window of data that starts after the break causes an increase in the variance of the parameter estimates, which results in an increase in the MSFE of the rolling forecast relative to the recursive. How much more weight we should put on the recursive (rolling) forecast thus depends on the values of the parameters. For example, the larger the estimated size of the break in the parameter, $\widehat{\Delta}$, the higher the weight on the rolling window forecast. Similarly, a higher variance of the error (σ^2) leads to more imprecise parameter estimates for any given sample, thus leading to a higher optimal weight on the recursive forecast. Finally, the closer the break to the middle of the sample $(\frac{\widehat{t^*}}{T} \simeq \frac{1}{2})$, the lower the weight on the recursive forecast; in fact, if the break is at the very beginning or the very end of the sample, it is optimal to use as many observations as possible to estimate the parameters. The fact that such values might be imprecisely estimated might adversely affect the forecasting improvements provided by eq. (25).

An alternative approach is suggested by Inoue and Rossi (2010) and Hansen and Timmermann (2011). While Pesaran and Timmermann's (2007) and Clark and McCracken's

(2009) objective is to improve the model's out-of-sample forecasts, the objective of Inoue and Rossi (2010) and Hansen and Timmermann (2011) is different. They are not interested in improving the forecasting model nor to estimate the ideal window size. Rather, their objective is to assess the robustness of conclusions of predictive ability tests to the choice of the estimation window size. The choice of the estimation window size has always been a concern for practitioners, since the use of different window sizes may lead to different empirical results in practice. In addition, arbitrary choices of window sizes have consequences about how the sample is split into in-sample and out-of-sample portions. Notwithstanding the choice of the window size is crucial, in the forecasting literature it is common to only report empirical results for one window size.

Inoue and Rossi (2010) note that reporting results based on one ad-hoc window size raises several concerns. One concern is that it might be possible that satisfactory results (or lack thereof) were obtained simply by chance, and are not robust to other window sizes. For example, this may happen because the predictive ability appears only in a sub-sample of the data, and whether the test can detect predictive ability depends on the estimation window size. A second concern is that it might be possible that the data were used more than once for the purposes of selecting the best forecasting model and thus the empirical results were the result of data snooping over many different window sizes and the search process was not ultimately taken into account when reporting the empirical results.[30] Ultimately, however, the estimation window is not a parameter of interest for the researcher: the objective is rather to test for equal predictive ability and, ideally, researchers would like to reach conclusions that are robust to the choice of the estimation window size.

Inoue and Rossi (2010) propose methodologies for comparing the out-of-sample forecasting performance of competing models that are robust to the choice of the estimation and evaluation window size by evaluating the models' relative forecasting performance for a variety of estimation window sizes, and then taking summary statistics. Their methodology can be applied to most of the tests of predictive ability that have been proposed in the literature, including tests for relative forecast comparisons as well as tests of forecast optimality.

Let $\Delta L_T(R)$ denote the test of equal predictive ability for non-nested model comparison proposed by either Diebold and Mariano (1995) or West (1996), and implemented using forecasts based either on a rolling window of size R or recursive/split estimation starting at observation R. Similarly, let $\Delta L_T^{\varepsilon}(R)$ denote Clark and McCracken's (2001) ENCNEW test

[30] Only rarely do researchers check the robustness of the empirical results to the choice of the window size by reporting results for a selected choice of window sizes.

for nested models comparison based either on rolling window estimation with window size R or recursive/split window estimation starting at observation R. Finally, let $\mathcal{W}_T(R)$ denote tests for forecast optimality analyzed by West and McCracken (1998), including tests of forecast encompassing (Clements and Hendry, 1993, Harvey, Leybourne and Newbold, 1998), tests for forecast rationality (Mincer and Zarnowitz, 1969) and tests of forecast uncorrelatedness (Granger and Newbold, 1986, and Diebold and Lopez, 1996) based on forecast errors obtained either by estimation on a rolling window of size R or recursive/split estimation starting at observation R.

They suggest the following statistics:

$$\mathcal{R}_T = \sup_{R\in\{\underline{R},\ldots\overline{R}\}} |\Delta L_T(R)|, \text{ and } \mathcal{A}_T = \frac{1}{\overline{R}-\underline{R}+1}\sum_{R=\underline{R}}^{\overline{R}} |\Delta L_T(R)|, \tag{26}$$

$$\mathcal{R}_T^{\varepsilon} = \sup_{R\in\{\underline{R},\ldots\overline{R}\}} \Delta L_T^{\varepsilon}(R) \text{ and } \mathcal{A}_T^{\varepsilon} = \frac{1}{\overline{R}-\underline{R}+1}\sum_{R=\underline{R}}^{\overline{R}} \Delta L_T^{\varepsilon}(R), \tag{27}$$

$$\mathcal{R}_T^{\mathcal{W}} = \sup_{R\in\{\underline{R},\ldots\overline{R}\}} \mathcal{W}_T(R), \text{ and } \mathcal{A}_T^{\mathcal{W}} = \frac{1}{\overline{R}-\underline{R}+1}\sum_{R=\underline{R}}^{\overline{R}} \mathcal{W}_T(R), \tag{28}$$

where $\underline{R}$ is the smallest window size considered by the researcher, $\overline{R}$ is the largest window size, and $\widehat{\Omega}_R$ is a consistent estimate of the long run variance matrix.[31] Inoue and Rossi's (2010) obtain asymptotic approximations to eqs. (26), (27) and (28) by letting the size of the window R be asymptotically a fixed fraction of the total sample size: $\zeta = \lim_{T\to\infty}(R/T) \in (0,1)$.[32]

The null hypothesis of equal predictive ability or forecast optimality at each window size for the $\mathcal{R}_T$ test is rejected at the significance level α when $\mathcal{R}_T > k_{\alpha,\zeta}^{\mathcal{R}}$ whereas the null hypothesis for the $\mathcal{A}_T$ test is rejected when $\mathcal{A}_T > k_{\alpha,\zeta}^{\mathcal{A}}$, where the critical values $\left(\alpha, k_{\alpha,\zeta}^{\mathcal{R}}\right)$ and $\left(\alpha, k_{\alpha,\zeta}^{\mathcal{A}}\right)$ for various values of $\underline{\zeta} \equiv \lim_{T\to\infty}(\underline{R}/T)$ and $\overline{\zeta} = 1-\underline{\zeta}$ are reported in the tables in Inoue and Rossi (2010). In practice, Inoue and Rossi (2010) recommend $\overline{\zeta} = 1-\underline{\zeta}$ and $\underline{\zeta} = 0.15$. For such values, Table A.4 in Appendix 1 reports the critical value for the statistics at the 5% significance level.

[31] See West (1996) for consistent variance estimates in eq. (26), Clark and McCracken (2001) for eq. (27) and West and McCracken (1998) for eq. (28). Inoue and Rossi's (2010) obtain asymptotic approximations to eqs. (26), (27) and (28) by letting the size of the window R be asymptotically a fixed fraction of the total sample size: $\zeta = \lim_{T\to\infty}(R/T) \in (0,1)$.

[32] Inoue and Rossi (2010) also consider cases where the window size is fixed – we refer interested readers to their paper for more details.

Interestingly, Inoue and Rossi (2010) show that in the presence of instabilities the power of rolling out-of-sample forecast tests depends crucially on the rolling window size, and that, similarly, the power of the recursive out-of-sample forecast tests does depend on the size of the first estimation window size. The intuition is as follows. Imagine that we are comparing the forecasting performance of two models, one of which (the large model) contains additional regressors relative to the competitor model (the small model). Suppose that the additional regressors are relevant only in a first part of the sample, and that they become insignificant in the later part of the sample. The finding of a superior performance of the larger model relative to the small model will clearly depend on when the predictive ability of the additional regressors disappears relative to the size of the estimation window. In fact, if the predictive ability disappears very early in the sample and the researcher uses a small window, he might have a chance to pick up the superior predictive ability of the large model; however, if the researcher uses a large window, he might miss the predictive ability since a large window will "wash out" the better performance of the large model. On the other hand, a large window would help finding evidence of superior predictive ability if there are no instabilities in the data because it provides more precise estimates.

Hansen and Timmermann's (2011) analysis is based on a similar concern about data mining over the split sample point in forecasts based on recursive estimation. They focus on nested models estimated via a recursive estimation scheme. They consider a different test statistic for nested models, namely the following MSFE-t-type statistic:

$$T_P(\rho) \equiv \frac{\sum_{t=R}^{P} \Delta L_{t+h}}{\widehat{\sigma}^2}, \tag{29}$$

where ΔL_{t+h} is the forecast error squared of the small model minus the forecast error squared of the large model, $\rho = \lim_{T\to\infty} (R/T)$ and $\widehat{\sigma}^2$ is a consistent estimate of the variance of ΔL_{t+h}. Following McCracken (2007) and generalizing his results, Hansen and Timmermann (2012) show that, under the null hypothesis that the parameter on the additional regressors in the large model are zero, the test statistic has the following limit distribution:

$$T_P(\rho) \Rightarrow 2\int u^{-1} B(u)' \Lambda dB(u) - \int u^{-2} B(u)' \Lambda B(u)\, du, \tag{30}$$

where Λ is a diagonal matrix with the eigenvalues of $\Sigma^{-1}\Omega$ on its main diagonal, and $B_j(u)$, $j = 1, ..., q$ are independent standard Brownian motions. Let the cumulative distribution function of $T_P(\rho)$ be denoted by $F_{\rho,\Lambda}$ and its p-value by $p(\rho)$, whose limiting distribution is a Uniform, $U(0,1)$.

Hansen and Timmermann (2011) make several contributions. The first is to show that the limiting distribution in (30) can be simplified to:

$$B(1)^2 - \rho^{-1} B(\rho)^2 + \ln \rho, \tag{31}$$

and can be simulated by $\sqrt{1-\rho}\,(Z_1^2 - Z_2^2) + \ln(\rho)$, where Z_1 and Z_2 are independent standard normal random variables. This limiting distribution is much simpler than the one derived in Clark and McCracken (2005), which is advantageous when deriving its p-values, especially when the number of extra regressors in the model is large. Hansen and Timmermann (2007) also show, via Monte Carlo simulations, that a researcher that data mines over several values of the window size, $\rho \in [\underline{\rho}, \overline{\rho}]$, that is a researcher that reports $p_{\min} = \min_{\rho \in [\underline{\rho}, \overline{\rho}]} p(\rho)$, would typically over-reject for large values of the split point ρ. That is, a spurious rejection of the null hypothesis of equal predictive ability is most likely to be found with large values of ρ whereas true rejections of a false null hypothesis are more likely to be found for small values of ρ.

If data were homoskedastic, Hansen and Timmermann (2007) recommend to first transform the test statistic as follows: $S_P(\rho) = (1-\rho)^{-1/2}[T_P(\rho) - q \ln \rho]$. In fact, the transformed statistic has a limiting distribution that does not depend on ρ in the homoskedastic case. However, in the heteroskedastic case the limiting distribution of $S_P(\rho)$ still depends on ρ and therefore does not have any advantages relative to using $T_P(\rho)$.

Hansen and Timmermann (2012) calculate the power of their proposed test $T_P(\rho)$ under local alternatives and show that the power of the test is highest when ρ is small. Thus, there is a trade-off between size and power in the presence of data mining over the sample split: the risk of rejecting the null hypothesis when it is true is highest when ρ is large; conversely, the power of the test is highest when ρ is small.

To resolve the data mining problem, Hansen and Timmermann (2012) recommend the following test statistic:

$$p_{\min} = \min_{\rho \in [\underline{\rho}, \overline{\rho}]} p(\rho).$$

There are several differences between this test statistic and the one proposed by Inoue and Rossi (2010). The first is that Hansen and Timmermann (2012) propose to minimize the p-value over the split-sample whereas Inoue and Rossi (2010) propose to maximize the test statistic over the estimation window size: the two would be equivalent if the test statistic were the same; however, note that for the case of nested models' forecast comparison (the case considered by Hansen and Timmermann, 2011), the latter focus on the MSFE-t test statistic (eq. 29) whereas Inoue and Rossi (2010) focus on the ENCNEW test. Another difference is that Inoue and Rossi (2010) consider the power of the test against parameter instabilities, whereas Hansen and Timmermann (2012) consider the power of the test in stationary environments. The advantage of the latter is that they can obtain detailed analytical power results and theoretically derive for which split-point the test has the largest rejection probability; the advantage of the former is that they consider the power of their test against predictive ability that appears only in a sub-sample of the data via Monte Carlo simulations, and can cover several test statistics for predictive ability. Finally, Hansen and Timmermann (2012) focus on recursive window estimation schemes, whereas Inoue and Rossi (2010) consider also rolling windows.

Hansen and Timmermann (2011) consider two interesting empirical analyses. The first is the predictability of stock returns, in particular the work by Welch and Goyal (2008), who found that the constant equity premium model produced better forecasts than models with predictors such as the default spread or the dividend yield. They find that the predictive ability is the strongest either for very small or very large values of ρ. A second empirical analysis focuses on inflation forecasts in a factor model. Their test does not find empirical evidence of superior predictive ability for the factor model over the simple autoregressive benchmark.

2.3.2 If There Are Instabilities, Does Estimation of Historic Breaks Help in Forecasting?

The presence of widespread instabilities in forecasting have inspired researchers to estimate models that allow for structural breaks. Several ways to incorporate time variation in the estimation of forecasting models have been proposed: (i) estimate models with multiple, discrete breaks at unknown points in time; or (ii) estimate time-varying parameter models where the parameters are allowed to change with each new observation, either according to a random walk or some other parametric process.

The detection of breaks is clearly an important issue in the literature: numerous in-sample testing procedures have been developed for detecting instabilities, each one of which depends on the assumptions made on the process underlying the instabilities. In particular, one-time, discrete breaks are typically detected by using Andrews' (1993) or Andrews and Ploberger's (1994) tests.[33] Examples of full sample estimation of models with a one-time

[33] Andrews (1993) proposed procedures to test for the presence of a one-time break at an unknown point in time. Bai (1997) demonstrated how to use Andrews' (1993) test to estimate the time of the break. Andrews and Ploberger (1994) developed optimal tests for structural breaks.

break include, among others, McConnell and Perez-Quiros (2000) for modeling the sharp decrease in U.S. GDP growth volatility, Stock and Watson (2002) and Inoue and Rossi (2011) for estimation of structural macroeconomic models that attempt to explain that decrease.[34] The presence as well as the timing of multiple, discrete breaks at unknown times can be detected by Bai and Perron's (1998) or Qu and Perron's (2007) procedure. Examples of full sample estimation of models with multiple discrete breaks include Rapach and Wohar's (2005) estimation of both inflation and real interest rates for several industrialized countries. The presence of small and persistent time variation in the parameters can be detected by Nyblom's (1989) or Elliott and Muller's (2006) test. Examples of full sample estimation of models with time-varying parameters include Cogley and Sargent (2001, 2005) and Cogley and Sbordone (2008), who model the parameters driving inflation and/or unemployment dynamics in the U.S. as a random walk. See Stock (1994) for an overview and discussion of in-sample tests for structural breaks.

While the literature discussed above has focused on the "in sample" detection and estimation of models with time-varying parameter, a more recent literature has attempted to utilize time-varying parameter models for forecasting. The latter is the objective of this Section. One major difference between in-sample detection of breaks and out-of-sample forecasting in the presence of breaks is that the particular type of instabilities does not matter in the former but may play an important role in the latter. In fact, as shown by Elliott and Muller (2006), conditional on the average magnitude of breaks being the same, the power of several, widely used tests for structural breaks is close over a wide range of breaking processes; thus, ignorance of which particular type of instability affects the data in practice does not matter for the goal of conducting an in-sample powerful test to detect whether there was a break in the data. Matters are very different when forecasting: the ability to forecast well may depend on the ability of successfully capturing and exploiting the form of instability affecting the data.

In what follows, we will review several papers that have successfully forecasted time

[34] McConnell and Perez-Quiros (2000) use structural break tests to identify a sharp decline in the volatility of output (as well as consumption and investment), labeled "the Great Moderation." Stock and Watson (2002, 2003) perform counterfactual VAR and New Keynesian model analyses and conclude that the Great Moderation was mainly caused by a decrease in the volatility of the shocks. Inoue and Rossi (2011) investigate the sources of the substantial decrease in output growth volatility in the mid-1980s by identifying which of the structural parameters in a representative New Keynesian and structural VAR models changed. They show that the Great Moderation was due not only to changes in shock volatilities but also to changes in monetary policy parameters, as well as in the private sector's parameters.

series out-of-sample using time-varying parameter models. We will focus on the following forecasting model (1), where, for simplicity, there are no control variables z_t:

$$y_{t+h} = \beta_t x_t + \varepsilon_{t+h}, \text{ for } t = 1, 2, ...T, \tag{32}$$

where different choices of how β_t evolves over time leads to different time-varying parameter models:

(i) Models with multiple, discrete breaks. Models with multiple, discrete structural breaks are such that:

$$\beta_t = \beta_1 \cdot 1(t < \tau_1) + \beta_2 \cdot 1(\tau_1 \leq t < \tau_2) + ... + \beta_K \cdot 1(\tau_{K-1} \leq t < \tau_K) + \beta_{K+1} \cdot 1(\tau_K \leq t),$$

where $\beta_1 \neq \beta_2 \neq ... \neq \beta_{K+1}$; K is the number of breaks, which give rise to $K+1$ regimes. Typically, except in very special circumstances, the time of the breaks $(\tau_1, ..., \tau_K)$ are unknown. One could assume, for example, that each regime is completely unpredictable based on the information in the previous regimes and, in the attempt of forecasting based only on the information available in the most recent regime, discard all data prior to time τ_K. Pesaran and Timmermann (2002) propose a Reversed Ordered Cusum (ROC) test, among other procedures. Although the ROC test estimates one break (the most recent one), nevertheless it is robust to the existence of multiple breaks since, in that case, it would focus on the most relevant break for forecasting purposes. Their procedure works as follows. Consider the linear model described by eq. (32), and let $\widehat{\beta}_{h,\tau} = \left(\frac{1}{T-\tau+1}\sum_{t=\tau}^{T} x_{t-h}x'_{t-h}\right)^{-1}\left(\frac{1}{T-\tau+1}\sum_{t=\tau}^{T} x_{t-h}y_t\right)$ be the OLS estimate of β_h using only observations from τ onwards, where $\tau = \widetilde{\tau}, \widetilde{\tau}-1, ..., 1$. $\widetilde{\tau}$ is a parameter chosen to guarantee that the estimate $\widehat{\beta}_{h,\tau}$ is meaningful; for example, Pesaran and Timmermann (2002) recommend $T - \widetilde{\tau} + 1$ to be set around two to three times the number of parameters in β_h. Also, let $\widehat{v}_\tau = \left(y_{\tau+h} - \widehat{\beta}'_{h,\tau}x_\tau\right)\left(1 + x'_\tau\left(\frac{1}{T-\tau+1}\sum_{t=\tau}^{T} x_t x'_t\right)^{-1} x_\tau\right)^{-1}$, $\tau = \widetilde{\tau}, \widetilde{\tau}-1, ..., 1$. The ROC squared test statistic is:

$$ROC_{s,T} = \left(\sum_{j=s}^{T} \widehat{v}_j^2\right)\left(\sum_{j=1}^{T} \widehat{v}_j^2\right)^{-1}, \ s = \widetilde{\tau}, \widetilde{\tau}-1, ..., 1. \tag{33}$$

The null hypothesis of the ROC squared test is the stability of the Mean Squared Error of the forecasting model and the test rejects when $ROC_{s,T}$ is outside the critical values provided in Brown et al. (1975).[35] As mentioned before, there are two issues with such

[35] Critical values depend on both the number of observations T as well as $\widetilde{\tau}$. Interested readers are referred to Brown et al. (1975).

procedure: not only the date of the latest break might be unknown and difficult to estimate precisely in finite samples, but, also, the parameter estimate might be imprecisely estimated if based only on data from $\tau_K \leq t \leq T$. In fact, measures of forecast accuracy such as the MSFE, which is the sum of the bias squared and the variance, would penalize a forecast depending on both its bias and its precision. Thus, by including data prior to τ_K it might be possible to improve the precision of the estimate at the cost of a higher bias. The choice of how many recent observations to use in estimating the parameters of a successful forecasting model clearly depends on this trade-off between bias and variance. Under special assumptions, it is possible to determine the optimal number of observations theoretically. For example, Pesaran and Timmermann (2007) focus on a linear model with exogenous, normal regressors and normal errors and forecast evaluation based on MSFEs. They show that the optimal number of observations (optimal in terms of unconditional MSFE) dated time τ_K (or earlier) to be used to estimate β_{K+1} is larger when: (i) the size of the break is smaller, (ii) $T - \tau_K$ is small; and (iii) the signal to noise ratio is small. Pesaran and Timmermann (2007) propose methodologies to determine optimally determine how many most recent observations to include in estimation. Among these, they propose: (i) an optimal number of observations based on the trade-off discussed above; (ii) cross-validation; and (iii) weighted forecast combinations. We will overview other methodologies proposed by Pesaran and Timmermann (2007) in Section 2.3.1.

Pesaran Pettenuzzo and Timmermann (2006) take a completely different approach. The novelty of their approach is to allow for the possibility of new breaks occurring in the forecasting period, whose properties depend on the size and duration of past breaks: if a break has happened in the past, they argue, it is also likely to happen in the future. Thus, it is important, for forecasting purposes, not only to identify past breaks, but be able to model the stochastic process that underlies the breaks so that the breaks themselves can be forecasted. To be concrete, their model is as follows: the data are drawn from several regimes, indexed by a state variable $s_t = 1, 2, ..., K + 1$, so that the sample of data, $\{y_t\}_{t=1}^{T}$ is drawn from the distribution $f(y_t | y_{t-1}, ..., y_1; \beta_s)$, where β_s is the parameter vector in regime s. The probability of moving from regime $s - 1$ to regime s is governed by a discrete first order Markov process with probability $p_{s-1,s}$, which is drawn from a known distribution with unknown parameters, for example a Beta distribution. The prior on the parameters of the Beta distribution are chosen to reflect prior beliefs about the mean duration of each regime. Finally, the parameters in each state, β_s, are drawn from a common distribution, for example a Normal distribution. This assumption allows Pesaran Pettenuzzo and Timmermann

(2006) to forecast the time series outside the estimation sample even if there are possible breaks in the out-of-sample period.

Pesaran Pettenuzzo and Timmermann (2006) assume a constant transition probability and a fixed number of regimes. Koop and Potter (2007) extend their framework to allow for regime changes where the number of regimes and their duration is unknown and unrestricted, and both the duration and the parameters in a future regime are allowed to depend on durations and parameters in a previous regime. They argue that these features are especially useful for forecasting, since breaks may occur out-of-sample: in their model, a new break can be forecast after the end of the sample and the size of the break depends on the properties of the previous regime, the history of previous breaks as well as a random element.

Another possibility is to estimate the parameters by using regime-switching models (Hamilton, 1988). Note how Pesaran, Pettenuzzo and Timmermann (2006) and Koop and Potter (2007) are different from regime-switching models: the latter are a special case when the parameters after a break are drawn from a discrete distribution with a finite number of states. If the states are not recurring, a standard regime-switching model will be misspecified and its parameter estimates will be inconsistent. In other words, regime-switching models assume that there is a finite number of states, and in the presence of regime changes the time series will always take value in each of these regimes (stationarity assumption). This is a very restrictive assumption for forecasting, and in fact regime-switching model do not seem very successful at forecasting: see Clements et al. (2004) for a review of the literature. Pesaran and Timmermann (2006) and Koop and Potter (2007) are also very different from in-sample models with multiple breaks (e.g. Bai and Perron, 1998), whose approach allows for multiple breaks but only for in-sample estimation and does not consider forecasting out-of-sample.

(ii) Models with time-varying parameters. There are several parametric specifications for models with time-varying parameters. For example, specifications may involve random walk parameters, such as:

$$\beta_t = \beta_{t-1} + \varepsilon_t^{rw},$$

or parameters that follow autoregressive specifications, such as time-varying autoregressive models:[36]

$$\beta_t = \sum_{j=1}^{p_\beta} \rho_j \beta_{t-j} + \varepsilon_t^{ar}.$$

All these approaches attempt to strike a balance between the desire of having parameters with a break at each point in time and the necessity of describing the time evolution of the parameters parsimoniously, which is clearly crucial for forecasting since parameter proliferation and the resulting imprecision of the parameter estimates penalizes forecasts, at least according to the typical MSFE loss function. Thus, these approaches describe breaks at each point in time using a smooth, parametric function that depends on a small number of parameters, for example the variance of ε_t^{rw} in the former, and the ρ_j's as well as the variance of ε_t^{ar} in the latter. Clearly, there are many choices of parametric functions for the evolution of the parameters.

One method that has been quite successful at forecasting in practice is the Unobserved Components Stochastic Volatility model proposed by Stock and Watson (2007). Their (univariate) model is as follows:

$$\begin{aligned} y_t &= \xi_t + \varepsilon_t^y, \\ \xi_t &= \xi_{t-1} + \varepsilon_t^\xi, \end{aligned} \tag{34}$$

where $\varepsilon_t^\xi \sim iidN\left(0, \sigma_{\xi,t}^2\right)$, $\varepsilon_t^y \sim iidN\left(0, \sigma_{y,t}^2\right)$, $ln\left(\sigma_{y,t}^2\right) = ln\left(\sigma_{y,t-1}^2\right) + \psi_{y,t}$, $ln\left(\sigma_{\xi,t}^2\right) = ln\left(\sigma_{\xi,t-1}^2\right) + \psi_{\xi,t}$ and $\left(\psi_{\eta,t}, \psi_{\xi,t}\right) \sim iidN\left(0, I\right)$. The model is estimated by Markov chain Monte Carlo, and the forecast of $y_{t+h|t}$ is the filtered estimate of ξ_t obtained by using only information available up to time t. Stock and Watson (2007) show that this model provides quite accurate inflation forecasts in the U.S.[37]

An alternative approach to model breaks due to level shifts which avoids imposing discrete regime changes is the nonlinear stochastic permanent break (stop-break) model considered by Engle and Smith (1999). Assuming $h = 1$, the model is such that:

$$\begin{aligned} y_{t+1} &= \beta_t + \varepsilon_{t+1} \\ \beta_t &= \beta_{t-1} + q_t \varepsilon_t \end{aligned} \tag{35}$$

[36] The latter may be generalized to the joint estimation of several variables in Vector Autoregressive models. The latter are typically estimated by Bayesian methods due to the computational difficulties in small samples arising in the estimation from imposing the structure of the time variation.

[37] Alternative approaches to proxy a slowly evolving inflation rate include the simple exponential smoothing method by Cogley (2002) and the autoregressive model with a shifting mean which evolves smoothly over time according to an exponential function, proposed by Gonzales, Hubrich and Terasvirta (2011), which can also be adapted to include exogenous information.

where ε_t is a martingale and q_t is a random variable bounded between zero and one. When the realized value of q_t is one, the realized shock at time t is permanent and y_t behaves like a random walk; when it is zero the shock is transitory and the conditional mean forecast is constant. By allowing q_t to vary between zero and one, the model builds a bridge between the constant mean forecast and the random walk.

(iii) Automatic model selection, impulse-indicator saturation and intercept corrections. An alternative set of methodologies for forecasting structural breaks is reviewed in Castle, Fawcett and Hendry (2011). They note that structural breaks resulting in location (mean) shifts are one of the major causes of forecast failure, as discussed in Clements and Hendry (1998, 2002 and 2006), whereas shifts on variables that have mean zero have smaller impact on forecasts (Hendry, 2000). Thus, their chapter focuses on forecasting breaks. Castle, Fawcett and Hendry (2011) note that predicting a break depends on whether it is possible to identify in advance the causes of such break; they argue that typically breaks are predictable although the lead time might be too short to be exploited in practice. For example, the financial crisis in 2007-2009 was not completely unpredictable: data on sub-prime loans and banks' leverage were signalling relevant information and *The Economist* had foreseen the possibility of a crisis well in advance; however, the extent of the off-balance-sheet loans and the policy responses became known only as the crisis unfolded, and were much more difficult to predict. They distinguish between breaks coming from two different sources: "regular" sources (i.e. economics) and other sources (i.e. politics, financial innovation). Their practical recommendation is then to monitor a wide variety of sources of information, including leading indicators,[38] disaggregated data (including news variables that are available at higher frequency and sectorial data),[39] prediction markets data and improved data at the forecast origin.

While monitoring a wide variety of data sources may provide useful information for forecasting, it necessitates methodologies for summarizing that information in practice. Castle, Fawcett and Hendry (2011, sec. 6) suggest using automatic model selection (Hendry and Krolzig, 2005, and Doornik, 2008) and impulse-indicator saturation. Other options include forecast combinations, model averaging and factor models. Automatic model selection se-

[38] See Marcellino (2009) for a review of the empirical performance of leading indicators in practice.

[39] See e.g. Hendry and Hubrich (2011) for forecasting aggregate variables via disaggregate components; Banbura et al. (2010) for incorporating higher frequency news indicators in forecasting;

Andersen et al. (2001) and Ferraro et al. (2012) provide examples of how using high frequency data (either news variables or oil price shocks) helps predict exchange rates.

quentially tests multiple variables using ad-hoc corrections to the critical values to take into account multiple model selection. Impulse-indicator saturation methods include a dummy variable for each observation to model possible breaks at each point in time, and then uses automatic model selection techniques to select the model. See Castle, Fawcett and Hendry (2011) for an extensive discussion. A further approach to estimation in the presence of instabilities is the intercept correction methodology proposed by Clements and Hendry (1996). They also discuss why parsimonious models often work better than larger models when there are breaks/instabilities; why double-differenced type models work well in the presence of breaks in the mean; when and why it helps to impose long-run (cointegrating) restrictions. We will not provide a detailed overview of such and related approaches due to space constraints and since they have already been covered in the previous volume of the Handbook series: see Clements and Hendry (2009, Section 7.2) for a thorough discussion of several of these methodologies.

We conclude this sub-section by reviewing the empirical evidence on the performance of models with breaks. Several researchers have evaluated the forecasting success of time-varying parameter models in practice. For example, Canova (2007) studies forecasting inflation in the G7 countries using real-time data. He compares the forecasting ability of univariate and multivariate time-varying autoregressive parameter models, and finds that time variations in the coefficients helps, but time varying univariate models perform better than multivariate ones. D'Agostino et al. (2009) use a multivariate time-varying coefficients VAR model with stochastic volatility, allowing for both changes in the coefficients and in the volatility, in an attempt to improve inflation forecasts. D'Agostino and Surico (2011) estimate time-varying VARs for the U.S. and evaluate their predictive ability relative to a time-varying univariate autoregression benchmark in forecasting inflation using two predictors: money growth, according to the quantity theory, or output growth, according to a Phillips curve. They also study whether inflation has become harder to forecast across different monetary policy regimes. They find that inflation predictability is the exception rather than the rule. Also, the forecasts produced by the bivariate model in inflation and money growth are significantly more accurate than the autoregressive forecasts only between WWII in 1939 and the Treasury-Federal Reserve accord in 1951. Output growth had predictive power for inflation in only two periods: between the great inflation of the 1970s to the early 1980s and between 1997 and 2000. Otherwise, under the gold standard, the Bretton Woods system and most of the great moderation sample, money growth and output growth had no marginal predictive power for inflation. Smith (2005) shows that the stop-break model (eq.

35) outperforms other nonlinear models in forecasting inflation out-of-sample. Bauwens, Korobilis and Rombouts (2011) compare the forecasting performance of several of the models we discussed in an extensive empirical analysis. In particular, the models they consider are Pesaran, Pettenuzzo, and Timmermann (2006), Koop and Potter (2007), D'Agostino et al. (2009), Stock and Watson's (2007) UCSV model as well as recursive and rolling OLS. Forecasting ability is judged by MSFEs as well as average predictive likelihood[40] in forecasting 23 univariate, quarterly U.S. macroeconomic time series from 1959 to 2010 following Stock and Watson (1996). Their empirical analysis finds extensive presence of structural breaks: at least three quarters of their series do have at least one structural break. They find that no single forecasting model stands out: in several instances, modeling the break process performs the best (in 83 percent of all series according to the RMSE criterion, and in 22 percent of all series according to the average predictive likelihood criterion), whereas in others rolling OLS forecasts performs the best, although the gains in terms of MSFEs are small. When the forecasting exercise starts at the beginning of the great recession (dated 2007), Pesaran, Pettenuzzo and Timmermann's (2006) method seem to perform very well. Finally, Guidolin and Timmermann (2007) use Markov-switching models to account for the presence of regimes in asset returns and show that they forecast well out-of-sample.

2.3.3 If There Are Instabilities, Do Forecast Combinations Help?

Since the seminal papers of Bates and Granger (1969), Granger and Newbold (1973), Diebold and Pauly (1987) and Hendry and Clements (2004), researchers have recognized the usefulness of forecast combinations in the presence of instabilities, and structural breaks are often cited as motivation for combining forecasts from different models. As noted in Timmermann (2006), the underlying idea is that models may differ in how they adapt to breaks: some models may adapt quickly, while others may only adjust very slowly. Thus, when breaks are small and recent, models with constant parameters may forecast more accurately than models that allow for time variation, and the converse is true in the presence of large breaks well in the past. Since detecting breaks is difficult in real time, it is possible that, across periods with varying degrees of instability, combining forecasts from models with different degrees of adaptability outperforms forecasts from each of the individual models. A similar reason why forecast combinations may work so well in practice is provided by Hendry and Clements (2002). In Hendry and Clements (2002), forecast breakdowns arise from shifts in

[40] The predictive likelihood is the predictive density evaluated at the actual (observed) value.

the mean of omitted variables, which result in unpredictable breaks in the intercept. However, by averaging forecasts over several regressions, breaks in the intercepts average out and the forecast combination is more robust to structural shifts than any of the individual regressions, provided that the intercept shifts are sufficiently uncorrelated across the different regressions. Stock and Watson (2008) argue that, in factor models, it is plausible that a similar argument could hold. In particular, even though factor loadings may be unstable, using many series to estimate the factors could average out instabilities as long as they are sufficiently independent across series. Then, factors might be precisely estimated even in the presence of instabilities in the individual relationships between the observable and the factors.

(i) Simple forecast combinations. Forecast combinations are obtained as follows: let $y^f_{t+h|t;i}$ be the forecast made at time t for horizon h using model "i", where $i = 1, ..., N$. The equal weight forecast combination is:[41]

$$y^{COMB,f}_{t+h|t} = \sum_{i=1}^{N} \omega_{t;i} y^f_{t+h|t;i}, \tag{36}$$

where $\omega_{t;i} = 1/N$. More generally, forecasts can be combined with unequal and possibly time-varying weights, $\omega_{t;i}$, which typically sum to unity.[42] In particular, Diebold and Pauly (1987) argued that forecast combination can greatly reduce forecast errors of models in the presence of a structural change. They considered rolling weighted least squares as well as time-varying parameter models as generalizations of equal weight forecast combinations: time-varying weights (which might, for example, be a function of time) might help in improving forecasts in the presence of instabilities. They showed, via numerical examples, that the improvement in forecasting ability can be substantial.

Several papers conjectured that the existence of instabilities could be a possible explanation behind the empirical success of forecast combinations in practice. Min and Zellner (1993) use forecast combination as a way to deal with heterogeneity arising from structural change. They propose a Bayesian approach to combine a constant linear regression model with a model with random walk time variation in the parameters. Hendry and Clements (2002) have shown, via Monte Carlo simulation exercises, that forecast combinations may

[41] When researchers are concerned about making equal weight forecast combinations robust to outliers, they implement a trimming. For example, in a 10 percent trimming, all forecasts generated at time t are ordered; then the 5 percent highest and the lowest 5 percent forecasts are discarded, and the remaining forecasts are combined with equal weights. See Stock and Watson (1999).

[42] That is, $\frac{1}{N} \sum_{i=1}^{N} \omega_{t;i} = 1$ for every t.

work well if there are intercept shifts in the data generating process. Aiolfi, Capistran and Timmermann (2010) also derive conditions under which, in a model with time-varying factor loadings, forecast combinations would provide more accurate forecasts than a model that uses either one of the two factors alone.

(ii) Ridge regression and inverse MSFE weights. A series of papers have proposed modifications of traditional forecast combination methods to improve forecasting ability in the presence of instabilities. Bates and Granger (1969) propose ridge regression method. That is, at each point in time forecasts are combined based on their historical performance in real time, that is, by comparing forecasts to the actual real time realizations in a previous sub-sample, and shrinking it towards equal weights. For example, let $Y^f_{t+h|t} = \left[y^f_{t+h|t;1}, ..., y^f_{t+h|t;i}, ..., y^f_{t+h|t;N}\right]'$. The weight vector, $W_t = [\omega_{t;1}, ..., \omega_{t;i}, ..., \omega_{t;N}]'$, is obtained as follows:

$$W^{BG}_{\tau} = \left(cI_N + \sum_t Y^f_{t+h|t} Y^{f\prime}_{t+h|t}\right)^{-1} \left(c \cdot \iota_N + \sum_t Y^f_{t+h|t} y_{t+h}\right) \tag{37}$$

where I_N is a $(N \times N)$ identity matrix, $c = k \cdot tr\left(N^{-1} \sum_{t=\tau-m}^{\tau} Y^f_{t+h|t} Y^{f\prime}_{t+h|t}\right)$, where k is the shrinkage coefficient (typical values of k are .001, .25 or 1), and ι_N is a $(N \times 1)$ vector of ones. Note that $\sum_t$ can be either $\sum_{t=1}^{\tau}$ or $\sum_{t=\tau-m}^{\tau}$, depending on whether researchers prefer a recursive or a rolling estimate of the combination regression. A special case is $k = 0$ in eq. (37), which leads to weighting each of the models by the inverse MSFE relative to the sum of the inverse MSFEs of the other models.[43] Alternative weight choices include predictive least squares (also known as the lowest historical MSFE method), which involves setting a weight equal to one to the model with the lowest historical MSFE and zero weight to the other models. Aiolfi and Timmermann (2006) propose to equally weighting only the forecasts with historical MSFEs in the lowest quartile of the MSFE distribution or incorporating a measure of the forecast performance by sorting forecasts into clusters based on their previous performance. The latter allows researchers to take into account the possibility that some models may be consistently better than others, and therefore that the good predictive ability of some models might be persistent over time.

(iii) Discounted MSFE. Another popular weighting scheme is the discounted MSFE method (see Diebold and Pauly, 1987); this method involves weighting forecasts by:

[43] This would correspond to the optimal weight when the forecast errors are uncorrelated across models.

$$\omega_{t;i} = \frac{\left(\sum_{s=s_0}^{t-h} \delta^{t-h-s} L^{(i)}(y_{t+h}, \widehat{\theta}_{i,t,R})\right)^{-1}}{\sum_{j=1}^{N} \left(\sum_{s=s_0}^{t-h} \delta^{t-h-s} L^{(j)}(y_{t+h}, \widehat{\theta}_{j,t,R})\right)^{-1}}$$

where $L^{(j)}(y_{t+h}, \widehat{\theta}_{j,t,R})$ was defined above eq. (8) for $j = 1, 2$, and here $j = 1, .., N$; s_0 is the initial time used to calculate the weights; and δ is the discount factor, e.g. $\delta = 1$ corresponds to Bates and Granger (1969) optimal weighting scheme when forecasts are uncorrelated across models. Other values of δ used by e.g. Stock and Watson (2004) are $\delta = 0.95$ and 0.9. See Stock and Watson (2004) and Clark and McCracken (2006) for other forecast combination weighting schemes.

(iv) Regime switching weights. Elliott and Timmermann (2005) propose forecast combinations where the combination weights are driven by regime switching in a latent state variable. The idea is that in relatively turbulent times one might want to put more weight on highly adaptive forecasts, whereas one may want to put more weight on stable forecasting models in relatively tranquil times. More in detail, Elliott and Timmermann (2005) consider a model where the joint distribution of the target variable and the vector of forecasts is conditionally Gaussian and driven by a latent state variable $S_{t+h} \in (1, 2, ..., k)$:

$$\begin{pmatrix} y_{t+h} \\ Y^f_{t+h|t} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{y,s_{t+h}} \\ \mu_{Yf,s_{t+h}} \end{pmatrix}, \begin{pmatrix} \sigma^2_{y,s_{t+h}} & \sigma'_{y,Yf,s_{t+h}} \\ \sigma_{y,Yf,s_{t+h}} & \sigma^2_{Yf,s_{t+h}} \end{pmatrix} \right),$$

and the unobservable state vector is generated by a first-order Markov chain with a transition probability matrix. They show that the proposed regime switching combination approach works well for a variety of macroeconomic variables in combining forecasts from survey data and time series models. Their Monte Carlo simulations show that time variation in the combination weights arises when the predictors and the target variable share a common factor structure driven by a hidden Markov process. See the detailed review by Timmermann (2006) for details on these and other methods for forecast combination.

The empirical evidence suggests that forecast combinations with equal weights perform the best in practice. Stock and Watson (2001) find that forecasts based on individual predictors tend to be very unstable over time whereas combinations tend to have better and more stable performance than the forecasts of the individual models that enter the combinations. They note that their finding is difficult to reconcile with the theory of forecast combinations in stationary environments. Stock and Watson (2003, 2004) note that forecast combinations with time-varying weights do not perform well in practice. On the other hand, Timmermann (2006) and Pesaran and Timmermann (2007) find that forecast combinations in models with varying degrees of adaptability to structural breaks at unknown times are better than forecasts from individual models. Clark and McCracken (2008) focus on forecasting with VARs in the presence of structural breaks. They show that simple equal weighted forecast combinations are consistently the best performers. It is also clear that forecast combinations are capable of predicting the equity premium better than the historical average, as shown by Rapach, Strauss and Zhou (2010), who argue that the success of forecast combinations is due to the presence of both instabilities and model uncertainty. Typically, forecast combinations are useful when researchers have available a large number of possible regressors, and estimating a joint model with all the regressors would lead to a very high parameter estimation error, which would penalize out-of-sample forecasts: in that case, researchers may combine forecasts obtained using each of the regressors, one at a time (e.g. Stock and Watson, 2003); note however that an alternative way of combining information based on a large number of different regressors is to use factor models (for brevity, we refer the reader to Stock and Watson, 2006, for a treatment of factor models). Aiolfi, Capistran and Timmermann (2010) also show that equal weighted forecast combinations of survey data outperform model-based forecasts from linear and non-linear univariate specifications as well as multivariate factor-augmented models for many macroeconomic variables and forecast horizons. They show that model instabilities are really important for explaining the gains due to forecast combinations. Occasionally, equal weighted forecast combinations of survey and model-based forecasts result in additional forecast improvements. Stock and Watson (2004) find that, in a seven countries database with a large number of predictors, the forecast combinations that perform the best are the ones with the least data adaptivity in their weighting schemes, such as equal weights. Note that the efficacy of equal weighted forecast combinations may depends on how the set of models is selected. Including several models that forecast very poorly might negatively affect the performance of forecast combinations. As shown in Mazzi et al. (2010) among others, if one uses some trimming to exclude models that forecast very poorly prior to taking the combination, equal weighted combinations are again effective. The recent and very detailed survey in Timmermann (2006, Section 4) discusses the usefulness of forecast combination as a hedge against model instability; in general, the main findings in Timmermann (2006) confirm that equal weighted forecast combinations outperform forecast combinations with time-varying weights.[44]

[44] There are exceptions, though. Ravazzolo, Verbeek and van Dijk (2007) provide one of the very few examples where models with time-varying weight schemes may forecast well when the data generating process has structural breaks. Their empirical application to forecasting returns of the S&P 500 index shows that

(v) Bayesian Model Averaging (BMA). BMA is an alternative method to implement forecast combinations with time-varying weights, implemented by estimating the weights by Bayesian methods. BMA techniques work as follows. They consider many possible models together with prior beliefs on the probability that each model is true. Then they compute the posterior probability that each model is the true one. Finally, they average the forecasts of the various models by using these posterior probabilities as weights. Thus, BMA are effectively forecast combinations, the only difference being that the weights are estimated by posterior probabilities. More formally, following Wright (2009), let the researcher prior belief about the probability that the true model is the i-th model be $P(\mathcal{M}_i)$, $i = 1, ..., N$. Also, let the posterior probability that the i-th model is the true model given the data $\mathcal{D}$ be:

$$P(\mathcal{M}_i|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{M}_i)\, P(\mathcal{M}_i)}{\sum_{j=1}^{N} P\left(\mathcal{D}|\mathcal{M}_j\right) P(\mathcal{M}_j)},$$

where $P\left(\mathcal{D}|\mathcal{M}_j\right)$ is the marginal likelihood of the j-th model. The marginal likelihood could be obtained by AIC or BIC, see e.g. Koop, Potter and Strachan (2008), Garratt, Koop and Vahey (2008) and Clark and McCracken (2006); the latter, for example, set $P\left(\mathcal{D}|\mathcal{M}_j\right)$ to be the information criterion times -0.5 times the estimation sample size. Typically, $P(\mathcal{M}_i) = 1/N$. The BMA forecast then weights each models' forecast by the posterior probability of the model:

$$y_{t+h|t}^{BMA,f} = \sum_{i=1}^{N} P(\mathcal{M}_i|\mathcal{D})\, y_{t+h|t;i}^{f} \tag{38}$$

Several papers suggest that BMA forecasts are very competitive in practice. Wright (2008) finds that BMA is quite useful for predicting exchange rates out-of-sample. In particular, BMA forecasts perform quite well relative to a driftless random walk, which is the toughest benchmark to beat in the exchange rate literature. Wright (2008) finds that, in most cases, BMA forecasts with a high degree of shrinkage have lower MSFEs than the random walk benchmark, although BMA forecasts are very close to those from the random walk forecast in magnitude. Wright (2009) finds that BMA provides better out-of-sample forecasts of U.S. inflation than equal weight forecast averaging. This superior performance is

---

time-varying weights might improve gains from investment strategies in the presence of transaction costs. Altavilla and Ciccarelli (2007) use the information contained in the revision history of inflation and GDP growth to improve the forecast accuracy of the models. They propose forecast combinations using weights that reflect both the relative ability that each model has at different points in time as well as different vintages to capture information on the revision process and improve forecasting performance, both in terms of precision and stability.

robust across sub-samples (before and after 1987), thus showing robustness to the possibility of forecast instabilities. Clark and McCracken (2010) provide empirical evidence on whether various forms of forecast averaging can improve real-time forecasts of small-scale VARs in the presence of instabilities (see Kozicki and Tinsley, 2001, Cogley and Sargent, 2005, Boivin and Giannoni, 2006, and Inoue and Rossi, 2011, among others, for empirical evidence of instabilities in VARs). The VARs that they consider include inflation, output and the interest rate. They consider BMA as well as alternative approaches to forecast averaging, such as equally weighted averages and MSFE-weighted averages as well as a large variety of methods robust to model instability, including different estimation window sizes, intercept corrections, allowing discrete breaks in parameters identified with break tests, and discounted least squares, BMA, among others. They show that the simplest forms of model averaging (such as equal weighted forecasts) consistently perform among the very best, whereas MSFE-weighted averages and factor models perform the worst. BMA's forecasts with high shrinkage perform well relative to VARs' and BVAR's forecasts, although not as well nor as consistently as simple equal weight forecast combinations.

A Monte Carlo analysis of the effects of parameter breaks on out-of-sample forecasting performance in BMAs is considered by Eklund and Karlsson (2005). They consider a Monte Carlo experiment where the parameter of one of the predictors is either constant or changes sign either at the beginning, in the middle, or towards the end of the data. When the parameters are constant, the true model is among the set of models to be estimated, whereas in the latter case, the true model is not. They compare the out-of-sample forecasting performance of typical BMA (whose weights depend on the posterior probabilities based on the marginal likelihood) with the performance of BMA models where the weights depend the posterior predictive density ("BMA with predictive likelihood"). The posterior predictive density is the density calculated in the out-of-sample portion of the data, that is observations $R+1, ..., T$ (the"hold-out sample", based on $P = T - R$ observations), using parameters estimated on data from 1 to, say, R ("training sample"). Differences between the performance of the typical BMA and the BMA with predictive likelihood suggest that the typical BMA may not be informative about the out-of-sample behavior. They show that results based on the typical BMA are very similar to those based on the BMA with predictive likelihood in the absence of a break, as long as P is large enough. However, in the presence of a break, the typical BMA fails to approximate the BMA with predictive likelihood: when the break is in the middle of the sample, the predictive likelihood performs significantly better provided the out-of-sample period is large enough. When the break is towards the end of the sample, the typical BMA always performs worse than the BMA with predictive likelihood. These results mirror the discussion in Section 2.3.1. In fact, when the true model is among the choice set, the predictive likelihood will select the true model asymptotically, although at a slower rate than the marginal likelihood since it relies on fewer observations; the two will perform similarly only when the sample size is large enough. When the true model is not among the choice set, the predictive likelihood will guard against over-fitting whereas the marginal likelihood will overfit.

Several new papers attempt to simultaneously address structural change and model certainty using a BMA approach. In particular, Ravazzolo, Paap, van Dijk and Franses (2007) allow for breaks of random magnitude in the parameters of forecasting regressions as well as uncertainty about the inclusion of models' predictors in their BMA framework. They attempt to predict U.S. excess stock returns using both macroeconomic and financial predictors. They find several breaks, which they relate to events such as oil crises, monetary policy changes, the 1987 stock market crash and the internet bubble. On the one hand, incorporating uncertainty on breaks and on the predictors does not lead to significant forecast improvements relative to linear models or traditional BMA; on the other hand, typical investors would be willing to pay several hundred basis points to switch to a strategy based on their forecasting model. Similarly, Groen, Paap and Ravazzolo (2009) propose a Phillips curve model for forecasting inflation by averaging across different model specifications selected from a set of potential predictors (lagged inflation, real activity data, term structure data, nominal data and surveys), where each of the models' specifications allow for stochastic breaks in regression parameters. The breaks are occasional random shocks. Like Ravazzolo, Paap, van Dijk, and Franses (2007), they find breaks that coincide with monetary policy regime changes or oil crises, and only little evidence of breaks in the variances or persistence. Koop and Korobilis (2009) propose a BMA where both the coefficient values as well as the entire forecast model can change over time (for example, a predictor might be useful during recessions but not in expansions). The advantage relative to Groen, Paap and Ravazzolo (2009) is that it can handle many more predictors.

In a more recent contribution, Pesaran, Schuermann and Smith (2009) propose to average forecasts not only across window sizes, as in Pesaran and Timmermann (2007), but also across models. They propose an "AveAve" approach where several models' forecasts are first averaged according to a Bayesian model averaging technique for a given window size, and then the procedure is repeated over several window sizes and their forecasts are averaged further. They show that the "AveAve" technique performs favorably in forecasting output growth and inflation across several countries relative to a simple equal weight forecast combination across window sizes and relative to an equal weight forecast combination across predictors (i.e. models).

2.3.4 Instabilities and Density Forecasts

So far, the discussion focused on conditional mean forecasting. To conclude, we discuss a few additional, related empirical results regarding density forecasts, including a brief overview of recent contributions in time-varying volatility forecasting for macroeconomics data.[45]

Regarding estimation of density forecasts in the presence of instabilities, researchers have proposed to use either forecast density combinations or to model the instabilities parametrically. Bayesian Model Averaging can be used to obtain forecast density combinations. For example, letting $f_{t+h|t;i}$ denote the forecast density of model i, $i = 1, .., N$, the BMA forecast density combination is:

$$f_{t+h|t}^{BMA} = \sum_{i=1}^{N} P\left(\mathcal{M}_i|\mathcal{D}\right) f_{t+h|t;i} \tag{39}$$

where $P\left(\mathcal{M}_i|\mathcal{D}\right)$ has been defined above eq. (38).

Hall and Mitchell (2007) discuss techniques to combine density forecasts. Their application to U.K. inflation density forecasts suggests that combining information across density forecasts can generate forecast improvements, a result similar to the forecast combination literature on point forecasts. They also discuss the estimation of the combination weights, although not in the presence of instabilities; see also Geweke and Amisano (2011). Jore, Mitchell and Vahey (2010) study the usefulness of combining forecast densities using many VARs and autoregressive models of output growth, inflation and interest rates. They propose a recursive-weight density combination strategy, based on the recursive logarithmic score of the forecast densities. They show that neither full-sample univariate combinations nor equal-weight combinations produce accurate real-time forecast densities for the Great Moderation period due to the existence of a structural break at the time of the Great Moderation. Their proposed recursive-weight density combination strategy gives competitive forecast densities by assigning a higher weight on rolling and break components that allow for the shifts in volatilities. Mazzi, Mitchell and Montana (2010) nowcast Euro-area output growth over the 2008-9 recession using density forecast combinations and economic indicators available at higher frequencies. They note that during the recent recession the relative forecasting

[45] For an extensive overview of volatility forecasting, in particular for financial variables, see Andersen et al. (2009).

performance of the models they consider changed abruptly. Billio et al. (2012) combine predictive densities using multivariate time-varying weights, where the weight dynamics is driven by the past performance of the predictive densities using learning mechanisms. The latter helps in identifying structural changes like the Great Moderation. An alternative to forecast density combinations is the estimation of models with time-varying parameters. For example, Clark (2011) focuses on density forecasts of U.S. GDP growth, unemployment, inflation and the interest rate in a Bayesian VAR with stochastic volatility, to better capture the decrease in volatility during the Great Moderation period. He demonstrates that adding stocastic volatility helps improving the real-time accuracy of density forecasts. Carriero, Clark and Marcellino (2012) extend the analysis to large VARs where the volatilities are driven by a single common factor and Koop and Korobilis (2012) propose new methods to estimate large dimensional VARs with time-varying parameters (including time-varying volatilities), where the models's dimension can change over time. Bache et al. (2011) consider how the density forecasting performance of a DSGE model with time-invariant parameters can be improved via combination with many VAR-based densities. They find that, although the DSGE models produce competitive point forecasts, their predictive densities are poorly calibrated. Densities become well calibrated only after merging the DSGE model with VARs allowing for breaks, although in this case the DSGE component receives little weight. Again, these results point to the importance of instabilities in practice. When combining density forecasts of the DSGE and the VARs with constant parameters, instead, the DSGE receives a larger weight, but only at horizons in which the predictive densities are mis-specified.[46] Potentially interesting alternative avenues for future research may include non-linear (logarithmic) combinations (e.g Kascha and Ravazzolo, 2010, although they don't focus on density forecasting and instabilities) and maximizing the forecasting performance not of the whole density, but on some regions of economic interest which might be more robust to instabilities. Regarding forecast density evaluation in the presence of instabilities, researchers might be interested in evaluating either the relative performance of density forecasts or the correct

[46] An interesting question is whether structural/economic restrictions and/or statistically motivated restrictions on the forecasting model might improve forecasts in the presence of instabilities. On the one hand, it might be possible that economic restrictions may render the forecasting model robust to the Lucas' critique since, if the parameters are "deep", they might be less subject to instabilities than reduced-form models. On the other hand, it might be possible that such restrictions may be invalid in the data, thus generating a mis-specified model whose forecasts may be less robust to instabilities. It might be possible that, by restricting the parameter space, there is less estimation error when parameters do shift, provided they remain within the parameter space (e.g. priors might improve forecasting performance, as in the Bayesian VAR literature).

specification of the density forecast of a selected model. Regarding the former, Manzana and Zerom (2009) focus on forecasting the distribution of inflation rather than its mean. They consider commonly used macroeconomic indicators and find that some indicators, such as the unemployment rate and housing starts, significantly improve forecasts of the distribution of core CPI inflation.[47] Regarding the latter, Rossi and Sekhposyan (2012b) empirically evaluate the correct specification of density forecasts of output growth and inflation based on a normal approximation in a large database of predictors similar to that considered in the empirical application in this Chapter.[48]

2.3.5 Summary of Findings

Overall, instabilities are a practical and serious concern for forecasters interested in evaluating predictive ability. Traditional forecast evaluation methods are inconsistent in the presence of instabilities. However, several alternative, robust procedures have been proposed. To determine Granger-causality, researchers might use Granger-causality tests robust to instabilities (Rossi, 2005); to assess which model forecasts the best, researchers can use Giacomini and Rossi's (2010a) Fluctuation and One-time reversal tests; to determine whether forecasts are rational, unbiased and/or optimal, researchers can rely on Rossi and Sekhposyan's (2011b) Fluctuation optimality tests. It is also possible to improve models' estimation in the presence of instabilities by either estimating historic breaks or by combining forecasts. The empirical evidence in the literature suggests that forecast combinations with equal weights provide the largest improvements in forecasting. Possible explanations of why forecast combinations may work well include finite sample error in the weights estimates (see Smith and Wallis, 2009) and different degrees of mis-specifications in the forecasting models, as determined by instabilities (see Hendry and Clements, 2004). BMA also performs quite well, whereas forecast combinations with time-varying weights do not perform as well. In addition, either averaging across window sizes or evaluating forecasting ability in a way robust to the choice of the window size usually improves the empirical evidence in favor of models' predictive ability.

[47] Amisano and Giacomini (2007) and Diks, Panchenkob and van Dijk (2011) are recent works that propose methodologies to evaluate the relative performance of density forecasts in stable environments.

[48] Diebold, Gunther and Tay (1998) and Corradi and Swanson (2006a) propose methodologies for evaluating the correct specification of density forecasts in stable environments – see Corradi and Swanson (2006b) for an excellent review. Rossi and Sekhposyan (2012a) propose tests to evaluate the correct specification of density forecasts in the presence of instabilities.

3 What is the Relationship Between In-sample and Out-of-sample Forecasting Ability in the Presence of Instabilities?

This section analyzes the relationship between models' in-sample fit and their out-of-sample forecasting performance in the presence of instabilities. First, we discuss the empirical evidence. Overall, the main message from the literature is that in-sample tests do not provide reliable guidance to out-of-sample forecasting ability. Then, we analyze the relationship between in-sample fit and out-of-sample forecasting ability. The difference between the two may be explained by structural breaks, overfitting, and different small sample properties of the estimates. We provide an overview of techniques that allow researchers to formally test whether in-sample fit provides enough guidance to out-of-sample forecasting performance via forecast breakdown tests (Clements and Hendry, 1998, 1999, and Giacomini and Rossi, 2009). When such tests reject, it is important to know why the in-sample fit is different from the out-of-sample forecasting performance, and we provide methods to empirically answer this question (Rossi and Sekhposyan, 2011a). Finally, Section 4 provides an empirical analysis of the presence of forecast breakdowns and their explanations in an empirical application to forecasting inflation and output growth using a large database of time series predictors.

3.1 Does In-sample Fit Provide Good Guidance to Out-of-Sample Forecasting Ability? The Empirical Evidence.

One area where researchers have explored whether in-sample fit provides guidance for out-of-sample forecasting ability is in predicting stock returns. Campbell (1987), Campbell and Shiller (1988), Bekaert and Hodrick (1992), Fama and French (1988), Perez-Quiros and Timmermann (2000) and Pesaran and Timmermann (1995) have found in-sample predictability in stock returns. However, more recent studies have documented that, although there is predictability in-sample, the true out-of-sample forecasting ability is much weaker: Bossaerts and Hillion (1999) find that stock returns on a variety of U.S. and international portfolios were unpredictable out-of-sample during the 1990s; Cooper, Gutierrez and Marcum (2005) find that relative returns on portfolios of stocks sorted on firm size, book-to-market value and past returns were not predictable out-of-sample during the period 1974-1997. Marquering and Veerbeek (2004) found that the trading strategies they study had predictive power only in the first half of the sample period they consider. Similarly, Sullivan, Timmermann and White (1999) find that the trading strategies they study were profitable before 1986 but not afterwards. Paye and Timmermann (2006) formally test for instabilities in return prediction models and find widespread instabilities. See also Goyal and Welch (2003) and Ang and Bekaert (2004).

A second area where in-sample fit does not provide reliable guidance for out-of-sample forecasting ability is when predicting exchange rates. Meese and Rogoff (1983a,b) have shown that, although models of exchange rate determination based on traditional fundamentals fit well in sample, their forecasting performance is much worse than a simple, a-theoretical random walk model. More recently, Sarno and Valente (2009) argued that the poor out-of-sample forecasting ability of exchange rate models may be caused by the poor performance of in-sample model-selection criteria, rather than by the lack of predictive content of the fundamentals.

Finally, a third area of interest is predicting output growth. Swanson (1998) shows that models with statistically significant in-sample monetary aggregates are not guaranteed to outperform simpler models out-of-sample. Furthermore, Swanson and White (1997) show that model selection based on the BIC fails to result in improved out-of-sample performance for several linear and non-linear models when predicting nine key macroeconomic variables. Giacomini and Rossi (2006) focus on predicting U.S. GDP using the U.S. yield curve. They also found significant failure of measures of in-sample fit for predicting GDP growth out-of-sample, and relate this failure to changes in monetary policy regimes.

3.2 The Theoretical Relationship Between Out-of-Sample Forecasts and In-sample Fit

The presence of model instability and/or overfitting might explain some of the differences between models' in-sample fit and out-of-sample forecasting ability. In fact, one important advantage of evaluating models on the basis of their out-of-sample forecasting ability is that out-of-sample procedures have power against structural breaks because they re-estimate their parameters over time by either rolling or recursive window estimation schemes. Clark and McCracken (2005) undertake an analytic investigation of the effects of structural breaks on parameters in tests of equal out-of-sample predictive ability and encompassing, as well as in-sample tests of predictive ability.

In what follows, we present a simplified example based on their results. Let $y_t = \beta_t + \varepsilon_t$, where $\varepsilon_t \sim iid(0, \sigma_t^2)$ and $t = 1, 2, ..., T$. Let $\beta_t = 0$ for $t \leq t^*$ and $\beta_t = \beta_2$ for $t > t^*$, and let $t^* = [\tau^* T]$, so that the breaks happens at a fixed fraction of the total sample size, $\tau^* \in [0, 1]$. Imagine a researcher interested in evaluating Granger-causality, which, in this example, simply means testing whether the constant is significant or not. The unrestricted model is a model with a constant, for which the in-sample fitted errors are $\widehat{\varepsilon}_{2,t} = y_t - T^{-1} \sum_{s=1}^{T} y_s$; the restricted model is a model with a zero mean, for which the in-sample fitted errors are $\widehat{\varepsilon}_{1,t} = y_t$. The Granger-causality test, GC_T, can be written as:

$$GC_T = T \left[T^{-1} \sum_{t=1}^{T} \left(\widehat{\varepsilon}_{1,t}^2 - \widehat{\varepsilon}_{2,t}^2 \right) \right] \left[T^{-1} \sum_{t=1}^{T} \widehat{\varepsilon}_{2,t}^2 \right]^{-1} \simeq T \left[\beta_2^2 \left(1 - \tau^*\right)^2 \right] + O_p\left(T^{1/2}\right). \qquad (40)$$

Thus, GC_T diverges to positive infinity as long as $\beta_2 \neq 0$, and will do so at rate T. However, it will diverge faster the larger is β_2 and the larger is $(1 - \tau^*)^2$. That is, since τ^* is bounded between 0 and 1, for a given value of β_2, the Granger-causality test statistic will be larger the smaller the value of τ^*, that is the earliest in the sample the parameter becomes different from zero. On the other hand, for a given value of τ^*, GC_T will be largest the bigger β_2 is, that is the more different from zero the constant is (zero is the restricted value of the parameter).

Now consider Diebold and Mariano's (1995) and West's (1996) tests. These tests are based on the one-step ahead out-of-sample forecast errors of the two models. The value of their test statistic will depend on when the break happens: whether it happens (a) after the sample split, or (b) before the sample split. It will also depend on the fraction of the sample used for forecast evaluation ($T - R = [(1 - \zeta) T]$, using the approximation $\zeta = \lim_{T,R \to \infty} (R/T)$) and on whether the parameters are re-estimated on rolling or expanding window estimation schemes. Let the out-of-sample forecast errors of the two models considered above be denoted by $\widehat{u}_{1,t+1|t} = y_{t+1}$ and $\widehat{u}_{2,t+1|t} = y_{t+1} - t^{-1} \sum_{s=1}^{t} y_s$ and let the loss function be quadratic. Clark and McCracken (2005) show that, in the recursive window case and for $h = 1$, the DMW_P statistic defined in eq. (9) is such that:

$$DMW_P = P \frac{\left[P^{-1} \sum_{t=R}^{T} \left(\widehat{u}_{1,t+1|t}^2 - \widehat{u}_{2,t+1|t}^2 \right) \right]}{\left[P^{-1} \sum_{t=R}^{T} \left(\widehat{u}_{1,t+1|t}^2 - \widehat{u}_{2,t+1|t}^2 \right)^2 \right]^{-1/2}}, \text{ where} \qquad (41)$$

$$\sum_{s=R}^{T} \left(\widehat{u}_{1,t+1|t}^2 - \widehat{u}_{2,t+1|t}^2 \right) \simeq \begin{cases} 2T\beta_2^2 \int_{\tau^*}^{1} \left[s^{-1} \left(s - \tau^*\right) - s^{-2} \left(s - \tau^*\right)^2 \right] ds = T\beta_2^2 \left(1 - \tau^*\right)^2, \text{ for (a)} \\ T\beta_2^2 \int_{\zeta}^{1} \left[\frac{2}{s} \left(s - \tau^*\right) - \frac{\left(s - \tau^*\right)}{s^2} \right] ds = T\beta_2^2 \left[1 - \zeta + \tau^{*2} - \frac{\tau^{*2}}{\zeta} \right], \text{ for (b),} \end{cases}$$

and the denominator is $O_p\left(T^{1/2}\right)$. In both cases, the dominating term in the DMW_P test statistic diverges to positive infinity. However, now the speed depends on P. As in the case of the GC_T test, the value of DMW_P is larger the larger is β_2, that is the bigger the predictive ability in the constant; and it is also larger the smaller τ^* in case (a), that is, the earlier the predictability shows up in the data.

Since both statistics diverge to infinity as the sample size diverges, in large samples both tests are likely to reject the null hypothesis provided $\beta_2 \neq 0$. Comparing (40) with (41), it is clear that the relative power of the two tests depends on the location of the break, τ^* and the fraction of the sample used for estimation purposes, ζ. We have shown in Section 2.2.1 that there exist situations in which the GC_T test has no power; thus, in such situations, out-of-sample forecast tests may have better power to select the correct model than in-sample Granger-causality tests. This argument prompted Rossi (2005) to design in-sample tests that have power against structural breaks in the parameters, reviewed in Section 2.2.1. Clark and McCracken (2005) have compared the performance of the GC_T and DMW_P tests with Rossi's (2005) $Exp - W_T^*$ test and shown that the latter is always more powerful when instabilities take the form of a one-time break. This suggests that, once one has determined the source of the possible advantage of out-of-sample predictive ability tests relative to in-sample tests, it may be possible to find an in-sample test that has better power properties. The latter point was also suggested by Inoue and Kilian (2006). However, out-of-sample forecast tests have power against a variety of alternatives, as Giacomini and Rossi (2009) have shown.

Giacomini and Rossi (2009) present a decomposition of the out-of-sample losses into a series of components that identify possible sources of differences between the out-of-sample predictive ability of a model relative to what was expected based on its in-sample fit. Their ultimate goal is to propose a theoretical framework for assessing whether a forecast model estimated over one period can provide good forecasts over a subsequent period. They formalize this idea by defining a forecast breakdown as a situation in which the out-of-sample performance of the model, judged by some loss function, is significantly worse than its in-sample performance. They show that one of the main causes of forecast breakdowns are instabilities in the data generating process and relate the properties of their forecast breakdown test to those of traditional structural break tests.

To gain some insight into the causes of forecast breakdowns, Giacomini and Rossi (2009) analyze the expectation of the difference between the out-of-sample forecast error relative to the average loss computed over the in-sample period. That is, for a given loss function $L(.)$

(for simplicity, we assume that the same loss is used for both estimation and out-of-sample forecast evaluation) and forecast horizon h, Giacomini and Rossi (2009) propose analyzing the sequence of P out-of-sample "surprise losses":

$$SL_{t+h} = L_{t+h} - \overline{L}_t, \text{ for } t = R, R+1, ..., T, \tag{42}$$

where L_{t+h} is the out-of-sample forecast error loss and $\overline{L}_t$ is the in-sample average loss. The latter depend on the forecasting scheme. Let $\overline{\Sigma}_j$ denote the relevant sample average depending on the forecasting scheme: $\overline{\Sigma}_j = t^{-1}\sum_{j=1}^{t}$ for the recursive scheme, $\overline{\Sigma}_j = R^{-1}\sum_{j=t-R+1}^{t}$ for the rolling scheme with window size R, and $R^{-1}\sum_{j=1}^{R}$ for the fixed scheme; thus, $\overline{L}_t = \overline{\sum}_j L_j$. For example, in the case of a quadratic loss, L_{t+h} is the squared out-of-sample forecast error of a model, and $\overline{L}_t$ is the in-sample mean squared (fitted) error. They further define β_t^* to be such that $E\left(\partial L_t\left(\beta_t^*\right)/\partial\beta\right) = 0$ and $\widehat{\beta}_t$ to be the in-sample parameter estimate at time t estimated via either fixed, recursive or rolling estimation scheme, $t = 1, 2, \ldots, T$. Also, let $\overline{\widehat{\beta}}_t$, $\overline{\beta}_j^*$ denote intermediate points between $\left(\widehat{\beta}_t, \beta_t^*\right)$, $\left(\beta_t^*, \beta_j^*\right)$, respectively.

Giacomini and Rossi (2009) decompose the expectation of the average surprise losses over the out-of-sample portion of the data, eq. (42), into components grouped into parameter instabilities, other instabilities and estimation uncertainty. They define "Forecast breakdowns" (see Clements and Hendry (1998, 1999) as situations where:

$$E\left(P^{-1/2}\sum_{t=R}^{T} SL_{t+h}(\widehat{\beta}_t)\right) \neq 0.$$

Their decomposition in shows that forecast breakdowns can be caused by several factors. To be concrete, let's derive the decomposition when there are both breaks in parameters and breaks in the variance of the errors, for the special case of a linear regression model, a fixed forecasting scheme and a quadratic loss. Consider the following simplified example, where $L(e) = e^2$, the forecasting scheme is fixed, and the model is: $y_{t+1} = x_t'\beta_t + \varepsilon_{t+1}$, where: $\varepsilon_t = \sigma_t u_t$; the $(p \times 1)$ vector x_t is *i.i.d.* with $E\left(x_t x_t'\right) \equiv J; \beta_t = \beta + P^{-1/4}\Delta\beta \cdot 1\left(t \geq R\right)$; $\sigma_t^2 = \sigma^2 + P^{-1/2}\Delta\sigma^2 \cdot 1\left(t \geq R\right) + \rho\varepsilon_{t-1}^2$ ($\Delta\sigma^2$ can be negative) and u_t is i.i.d.(0,1). This specification allows for ARCH and two types of structural breaks: a break in the conditional mean parameters at time R (from β to $\beta + \Delta\beta$), and a break in the unconditional variance of the errors at time R (from $\sigma^2/\left(1-\rho\right)$ to $\left(\sigma^2 + \Delta\sigma^2\right)/\left(1-\rho\right)$). Giacomini and Rossi (2009) show that:

$$E\left(P^{-1/2}\sum_{t=R}^{T} SL_{t+h}(\widehat{\beta}_t)\right) = \underbrace{\frac{\Delta\sigma^2}{1-\rho}}_{\text{“other instabilities”}} + \underbrace{\frac{1}{2}\Delta\beta' J\Delta\beta}_{\text{“parameter instabilities II”}} + \underbrace{2\frac{P^{1/2}}{R}\frac{\sigma^2}{1-\rho}p}_{\text{“overfitting"}}. \tag{43}$$

First, note from (43) that a forecast breakdown can be caused by a "small" positive break in the variance of the disturbances and/or a "large" break (positive or negative) in the conditional mean parameters. However, the presence of ARCH does not cause a forecast breakdown. Second, expression (43) implies that breaks in parameters and in the variance of the errors could have opposite effects on the forecast performance, and thus not necessarily cause a forecast breakdown (e.g., if $\Delta\sigma^2 \leq -.5\Delta\beta' J \Delta\beta$). In other words, there could be a bias-variance trade-off between breaks in the model's parameters (which result in biased forecasts) and breaks in the variance of the errors which does not necessarily result in a discrepancy between in-sample fit and out-of-sample forecasting performance. Indirect approaches that jointly test for breaks in conditional mean and variance parameters may instead detect both breaks and thus incorrectly conclude that the forecast performance of the model necessarily deteriorates. Finally, under their assumptions, the overfitting component is present only in finite samples and is proportional to the number of parameters, the variance of the disturbances and the factor $P^{1/2}/R$. Giacomini and Rossi (2009) further discuss the effects of overfitting on the properties of the forecast breakdown test in greater detail and propose an overfitting-corrected version of their test based on a small sample approximation where the number of regressors is large relative to the total sample size.

Other additional, important points on the relationship between in-sample fit and out-of-sample forecasting ability were made by Inoue and Kilian (2004). Inoue and Kilian (2004) note that there are important cases where strong in-sample evidence and weak out-of-sample evidence are not necessarily an indication that in-sample tests are not reliable. For example, in-sample tests rely on a larger sample size than out-of-sample tests (which have to reserve a portion of the data for out-of-sample forecast validation), so that they may have higher power. If the data are stationary, Inoue and Kilian's (2004) explanation implies that we should discount the results out-of-sample tests when the latter fail to confirm the findings of predictability using in-sample tests.[49] Another interesting point that Inoue and Kilian (2004) make is that it is not necessarily true that out-of-sample tests are more robust to data mining than in-sample tests: the problem is that out-of-sample tests are not truly "out-of-sample", since the researcher is free to experiment with alternative predictors in the

[49] Inoue and Kilian (2004) also consider the possibility of breaks.

out-of-sample portion of the data until he finds a significant predictor.[50]

An interesting question is why there are instabilities in the forecasting performance and why they might explain the gap between in-sample fit and out-of-sample forecasting ability, such as that described in eq. (43). Timmermann (2008) provides an intriguing explanation based on the economic analysis of the stock market. In particular, he argues that forecasters of stock returns face a moving target that changes over time: "just as the forecaster may think that he has figured out how to predict returns, the dynamics of market prices will, in all likelihood, have moved on – possibly as a consequence of the forecaster's own efforts" (Timmermann, 2008, p. 1). That is, forecasters constantly search across competing approaches and investment strategies and make use of all available in-sample information. Once a successful forecast strategy is found, more and more forecasters and investors will try to exploit it, and it will start to have an impact on prices so that the predictability effectively gets incorporated in the current price and it disappears. Timmermann (2008) conjectures that such competition across forecasters and investors generates instabilities in the models' out-of-sample forecasting performance. Interestingly, it might then be that the lack of predictability is not due to the inexistence of predictability, or worse to the lack of skills of forecasters, but to the fact that predictive opportunities are exploited efficiently: an example of "post hoc ergo propter hoc". Note that, as a consequence of Timmermann's (2008) argument, if the predictability of successful models were based on actual observed variables whose information was effectively exploited by forecasters, econometrician's regressions should be able to uncover such relationships; however, successful models might be too complicated to be captured by econometrician's simple time series regressions, in part also due to their instabilities over time.[51]

Finally, note that the main focus of this Section is on the relationship between in-sample model's fit and out-of-sample forecasting ability in the presence of instabilities. For com-

[50] Inoue and Kilian (2006) focus instead on the consistent selection of forecasting models based on the MSFEs, rather than on testing, and show that selecting models based on MSFEs may lead to choosing over-parameterized models under the assumption that the window size used for estimation is a fixed fraction of the total sample size.

[51] See also Schwert (2003) for a similar argument. He argues that it has been observed that anomalies in financial markets may disappear after being documented in the literature. This raises the question whether the disappearance is due to sample selection bias or to the practitioners' focus on anomalies. In the former case, there was no anomaly to start with; in the second case, it is possible that the anomaly was identified by practitioners and then disappeared because practitioners take anomalies into account in their trading pattern so that profitable transactions vanish.

pleteness, let us mention two recent papers that focus on the relationship between in-sample fit and out-of-sample forecasting ability, although they focus on stationary environments. The first is Hansen (2009). Hansen (2009) derives the joint limiting distribution of in-sample fit and out-of-sample forecasts at the true, or pseudo-true, parameter values. His results indicate that for a broad class of loss functions the two are strongly negatively correlated. The consequence of this result is that good in-sample fit leads into poor out-of-sample fit. In particular, an example in Hansen (2009) shows that, under some simplifying assumptions (e.g. the data are iid Normal and the loss is quadratic), then the in-sample fitted error ($\widehat{u}_{1,t}^2$) and the out of sample forecast error ($\widehat{u}_{1,t+1|t}^2$) are jointly distributed as $\left(\widehat{u}_{1,t}^2; \widehat{u}_{1,t+1|t}^2\right) \underset{d}{\rightarrow} \left(Z_1^2, -Z_1^2 + 2Z_2Z_2\right)$, where Z_1, Z_2 are iid Normals, independent of each other. This shows that the source of advantage of models' in-sample fit (Z_1^2) is exactly the same component that penalizes models' out-of-sample fit.

The second paper is the work by Calhoun (2011). Calhoun (2011) focuses on the asymptotic distribution of tests of forecast comparisons in models where the number of predictors used by the larger model increases with the sample size. Under these assumptions, he shows that out-of-sample tests can test hypotheses about measures of models' forecasting performance if the fraction of the sample used for out-of-sample evaluation is small. Furthermore, in-sample tests as well as Clark and McCracken's (2001, 2005a), McCracken's (2007) and Clark and West's (2006, 2007) tests will choose the larger model too often even if the smaller model is more accurate.

3.3 How Can Researchers Formally Establish Whether In-sample Fit is Indicative of Out-of-Sample Forecasting Ability?

Giacomini and Rossi (2009) propose a test to detect and predict forecast breakdowns in a model. Their notion of a forecast breakdown is a formalization and generalization of what Clements and Hendry (1998, 1999) called a "forecast failure", described as a "deterioration in forecast performance relative to the anticipated outcome" (Clements and Hendry, 1999, p. 1). Giacomini and Rossi (2009) formalize the definition of a forecast breakdown by comparing the model's out-of-sample performance to its in-sample performance using the notion of surprise losses, SL_{t+h}, defined in eq. (42). Their test for predicting forecast breakdowns is obtained as follows. Consider the sequence of P out-of-sample surprise losses SL_{t+h} and select a $p-$dimensional vector of forecast breakdown predictors X_t (which can include a constant, lagged surprise losses, and various predictors such as business cycle leading indicators as well as economically meaningful variables). Then, estimate the following model:

$$SL_{t+h} = a_0 + a_1' X_t + \varepsilon_{t+h}. \quad (44)$$

and testing whether $a_0 = a_1 = 0$. When the null hypothesis is rejected, the model experienced a forecast breakdown, which implies that the model (44) could be used to predict future forecast breakdowns.[52]

A special case is the test to detect past forecast breakdown. For simplicity of exposition, let's focus on this simple case. In this case, additional regressors X_t are not included, so that the researcher tests whether the surprise losses are zero in expectation. The "forecast breakdown" test statistic is then:

$$t_{R,P,h} = \frac{\overline{SL}_P}{\widehat{\sigma}_{SL}}, \quad (45)$$

where $\overline{SL}_P = P^{-1/2} \sum_{t=R}^{T} SL_{t+h}$ and $\widehat{\sigma}_{SL}^2$ is the appropriate, consistent estimate of the variance of the average surprise losses provided by Giacomini and Rossi (2009); for example, in the recursive estimation case, $\widehat{\sigma}_{SL}^2$ is simply the HAC variance estimate of the surprise losses. The test rejects the null hypothesis at the $\alpha\%$ confidence level whenever $t_{R,P,h}$ is greater than the $(1-\alpha) - th$ quantile of a standard Normal distribution.[53]

3.4 How to Empirically Determine Why In-sample Fit Differs From Out-of-Sample Forecasting Ability?

While the test proposed by Giacomini and Rossi (2009) has power to detect forecast breakdowns, it is not possible to use it to determine what is the source of the forecast breakdown. Rossi and Sekhposyan (2011a) take Giacomini and Rossi's (2009) decomposition a step further by developing a new methodology to identify the sources of models' forecasting performance. The methodology decomposes the models' forecasting performance into asymptotically uncorrelated components that measure instabilities in the forecasting performance, predictive content and over-fitting.

Rossi and Sekphosyan (2011a) define predictive content as the correlation between in-sample and out-of-sample measures of fit, similarly to Giacomini and Rossi (2009). When

[52] Note that the estimate of the variance to be used to implement the test $a_0 = a_1 = 0$ is complicated by parameter estimation uncertainty, and it is provided in Giacomini and Rossi (2009).

[53] The overfitting component is always positive and will be a cause of forecast breakdown in finite samples. Under special assumptions, Giacomini and Rossi (2009) also provide an overfitted-corrected test for forecast breakdown.

the correlation is small, the in-sample measures of fit have no predictive content for the out-of-sample and vice versa. An interesting case occurs when the correlation is strong, but negative: in this case, the in-sample predictive content is strong yet misleading for the out-of-sample. Rossi and Sekhposyan (2011a) define over-fitting as a situation in which a model fits well in-sample but loses predictive ability out-of-sample; that is, where in-sample measures of fit fail to be informative regarding the out-of-sample predictive content.

To capture predictive content and over-fitting, they consider the following regression:

$$\Delta L_{t+h} = a \cdot \Delta \mathcal{L}_t + u_{t+h} \quad \text{for} \quad t = R, R+1, ..., T, \tag{46}$$

where ΔL_{t+h} is the sequence of estimated out-of-sample loss differences of two models evaluated at the estimated parameter values defined in eq. (8) and $\Delta \mathcal{L}_t$ denotes the in-sample loss difference of the two models.

Let $\widehat{a} \equiv \left(\frac{1}{P} \sum_{t=R}^{T} \Delta \mathcal{L}_t^2 \right)^{-1} \left(\frac{1}{P} \sum_{t=R}^{T} \Delta \mathcal{L}_t \Delta L_{t+h} \right)$ denote the OLS estimate of a in regression (46), $\widehat{a}\Delta \mathcal{L}_t$ and $\widehat{u}_{t+h}$ denote the corresponding fitted values and regression errors: $\Delta L_{t+h} = \widehat{a}\Delta \mathcal{L}_t + \widehat{u}_{t+h}$. Note that regression (46) does *not* include a constant, so that the error term measures the average out-of-sample loss not explained by in-sample performance. Then, the average MSFE can be decomposed as:

$$\frac{1}{P} \sum_{t=R}^{T} \Delta L_{t+h} = B_P + U_P, \tag{47}$$

where $B_P \equiv \widehat{a} \left(\frac{1}{P} \sum_{t=R}^{T} \Delta \mathcal{L}_t \right)$ and $U_P \equiv \frac{1}{P} \sum_{t=R}^{T} \widehat{u}_{t+h}$. B_P can be interpreted as the component that was predictable on the basis of the in-sample relative fit of the models (predictive content), whereas U_P is the component that was unexpected (over-fitting).

Let $A_{\tau,P} = m^{-1} \sum_{t=R+\tau-m}^{R+\tau-1} \Delta L_{t+h} - \frac{1}{P} \sum_{t=R}^{T} \Delta L_{t+h}$, and $\overline{A}_{\tau,P} \equiv E\left(A_{\tau,P}\right), \overline{B}_P \equiv \beta E\left(\Delta \mathcal{L}_t\right)$, $\overline{U}_P \equiv E\left(\Delta L_{t+h}\right) - \beta E\left(\Delta \mathcal{L}_t\right)$. Rossi and Sekhposyan (2011a) propose the following decomposition:

$$\frac{1}{m} \sum_{t=R+\tau-m}^{R+\tau-1} \left[\Delta L_{t+h} - E\left(\Delta L_{t+h}\right)\right] = \left(A_{\tau,P} - \overline{A}_{\tau,P}\right) + \left(B_P - \overline{B}_P\right) + \left(U_P - \overline{U}_P\right). \tag{48}$$

They consider three null hypotheses: (i) Constant predictive ability: $H_{0,A} : \overline{A}_{\tau,P} = 0$ for all $\tau = m, m+1, ..., P$; (ii) No predictive content: $H_{0,B} : \overline{B}_P = 0$; and (iii) No overfitting: $H_{0,U} : \overline{U}_P = 0$. Under the null hypotheses, they show that the three components, $A_{\tau,P}$, B_P and U_P, are asymptotically independent. Thus, the components in decomposition (48) can be used to construct three test statistics to test each of the null hypotheses: constant predictive ability, predictive content, and overfitting:

$$\begin{aligned} \Gamma_P^{(A)} &\equiv \sup_{\tau=m,\ldots,P} |\sqrt{P}\widehat{\sigma}_A^{-1}A_{\tau,P}|, \\ \Gamma_P^{(B)} &\equiv \sqrt{P}\widehat{\sigma}_B^{-1}B_P, \\ \Gamma_P^{(U)} &\equiv \sqrt{P}\widehat{\sigma}_U^{-1}U_P. \end{aligned} \tag{49}$$

The $\Gamma_P^{(A)}$ test rejects "constant predictive ability" when $\Gamma_P^{(A)} > k_{\alpha,\delta}^{RS}$, where $k_{\alpha,\delta}^{RS}$, the critical values for the $\Gamma_P^{(A)}$ test, are reported in Rossi and Sekhposyan's (2011a) Table 1 and depend on $\delta = \lim_{T\to\infty} (m/P)$. The $\Gamma_P^{(B)}$ test rejects "no predictive content" when $\left|\Gamma_P^{(B)}\right| > z_{\alpha/2}$, where $z_{\alpha/2}$ is the $\alpha/2 - th$ percentile of a standard Normal distribution. Similarly, $\Gamma_P^{(U)}$ test rejects "no overfitting" when $\left|\Gamma_P^{(U)}\right| > z_{\alpha/2}$. For convenience, we report Rossi and Sekhposyan's (2011a) critical values for tests with significance level 5% in Table A.5 in Appendix 1. For the same significance level, $z_{\alpha/2} = 1.645$.

To gain intuition, consider a simple example where the true data generating process (DGP) is $y_{t+h} = \beta + \varepsilon_{t+h}$, where $\varepsilon_{t+h} \sim iidN(0, \sigma^2)$. Rossi and Sekhposyan (2011a) compare the forecasts of two nested models for y_{t+h} made at time t, based on parameter estimates obtained via a rolling estimation scheme with a fixed window size. The first (unrestricted) model includes a constant only, so that its forecasts are $\widehat{\beta}_{t,R} = \frac{1}{R}\sum_{j=t-h-R+1}^{t-h} y_{j+h}$, $t = R$, $R+1$, ..., T, and the second (restricted) model sets the constant to be zero, so that its forecast is zero. Consider the (quadratic) forecast error loss difference between the first and the second model, $\Delta L_{t+h} = \left(y_{t+h} - \widehat{\beta}_{t,R}\right)^2 - y_{t+h}^2$, and the (quadratic) in-sample loss difference $\Delta\mathcal{L}_t = \left(y_t - \widehat{\beta}_{t,R}\right)^2 - y_t^2$. Let $a \equiv E(\Delta L_{t+h}\Delta\mathcal{L}_t)/E(\Delta\mathcal{L}_t^2)$. Rossi and Sekhposyan (2011a) show that $a = (\beta^4 + 4\sigma^2\beta^2 + (4\sigma^2 + 2\sigma^2\beta^2)/R)^{-1}(\beta^4 - 3\sigma^2/R^2)$.When the models are nested, in small samples $E(\Delta\mathcal{L}_t) = -(\beta^2 + \sigma^2/R) < 0$, as the in-sample fit of the larger model is always better than that of the small one. Consequently, $E(B_P) = aE(\Delta\mathcal{L}_t) = 0$ only when $a = 0$. The calculations show that the numerator for a has two distinct components: the first, β^4, is an outcome of the mis-specification in the second model; the other, $3\sigma^2/R^2$, changes with the sample size and "captures" estimation uncertainty in the first model. When the two components are equal, the in-sample loss differences have no predictive content for the out-of-sample. When the mis-specification component dominates, in-sample loss differences provide information content for the out-of-sample. On the other hand, when a is negative, though the in-sample fit has predictive content for the out-of-sample, it is misleading in that it is driven primarily by the estimation uncertainty. For any given value of a, $E(B_P) = aE(\Delta\mathcal{L}_t) =$

$-a(\beta^2 + \sigma^2/R)$. By construction, $E(U_P) = E(\Delta L_{t+h}) - E(B_P) = (\sigma^2/R - \beta^2) - E(B_P)$. Similar to the case of B_P, the component designed to measure over-fitting is affected by both mis-specification and estimation uncertainty. One should note that for $a > 0$, the mis-specification component affects both $E(B_P)$ and $E(U_P)$ in a similar direction, while the estimation uncertainty moves them in opposite directions. Estimation uncertainty penalizes the predictive content B_P and makes the unexplained component U_P larger.

Rossi and Sekhposyan (2011a) use their proposed method to understand why exchange rate forecasts based on the random walk are superior to those of economic models on average over the out-of-sample period. They find that lack of predictive content is the major explanation for the lack of short-term forecasting ability of the economic models, whereas instabilities play a role especially for medium term (one-year ahead) forecasts.

3.5 Summary of Findings

The finding that in-sample fit is not indicative of out-of-sample forecasting performance is widespread in economics and finance. However, recent developments allow researchers to test and predict forecast breakdowns, that is situations where the in-sample fit does not provide enough guidance to out-of-sample forecasting performance, as well as methodologies to decompose models' relative out-of-sample forecast error losses into separate components to identify the contributions of instabilities, actual predictive content and overfit in explaining the models' performance. The next section sheds some light on the empirical importance of forecast breakdowns in practice and the reasons behind the breakdowns.

4 Empirical Evidence

This section revisits the empirical evidence on forecasting in the presence of instability since the seminal work by Stock and Watson (2003). Our main goal is to establish whether the empirical conclusions they reached are still valid, and whether the recent estimation and forecast evaluation techniques reviewed in this chapter change our perspectives on the empirical evidence of forecastability of output growth and inflation. We focus on the same database in Stock and Watson (2003), with the main difference that our database is updated to the latest available sample, and we perform a series of estimation techniques and tests that are substantially more extended than theirs.

We consider forecasting quarterly output growth and inflation h-periods into the future. Let the regression model be:

$$Y_{t+h}^{h} = \beta_0 + \beta_1 (L) X_t + \beta_2 (L) Y_t + u_{t+h}^{h}, \ t = 1, ..., T \tag{50}$$

where the dependent variable is either $Y_{t+h}^{h} = (400/h) \ln(RGDP_{t+h}/RGDP_t)$ when forecasting real GDP growth ($RGDP_t$ is real GDP at time t) or $Y_{t+h}^{h} = (400/h) \ln(P_{t+h}/P_t) - 400 \ln (P_t/P_{t-1})$ when forecasting inflation growth (P_t is the price level at time t), h is the forecast horizon and equals four, so that the forecasts involve annual percent growth rates of GDP and inflation. $\beta_1 (L) = \sum_{j=0}^{p} \beta_{1j} L^j$ and $\beta_2 (L) = \sum_{j=0}^{q} \beta_{2j} L^j$, where L is the lag operator. We consider several explanatory variables, X_t, one at a time. The explanatory variable, X_t, is either an interest rate or a measure of real output or unemployment, price, money or earnings. We consider data for five countries: Canada (labeled "CN"), France (labeled "FR"), Germany (labeled "GY"), Italy (labeled "IT"), Japan (labeled "JP"), the U.K. (labeled "UK") or the U.S. (labeled "US"). Following Stock and Watson (2003), the data are transformed to eliminate stochastic or deterministic trends. For a detailed description of the variables that we consider (and their transformations), see the Not-for-Publication Appendix available at: http://www.econ.upf.edu/~rossi/. In this empirical analysis, we focus in particular on predicting CPI inflation and output (real GDP) growth using econometric models and techniques that allow for instabilities. We utilize quarterly, finally revised data available in January 2011. The earliest starting point of the sample that we consider is January 1959, although several series have a later starting date due to data availability constraints. For the out-of-sample forecasting exercise, we estimate the number of lags (p and q) recursively by BIC unless otherwise noted; the estimation scheme is rolling with a window size of 40 observations.[54] Tests are implemented using HAC-robust variance estimates,

[54] We consider only rolling forecasts due to space constraints.

where the truncation parameter is $T^{1/5}$.

Faust and Wright (2012, in this Handbook) consider several other models that are useful for forecasting inflation, in particular judgemental forecasts as well as a fixed coefficient autoregressive benchmark with a judgemental starting point and a judgemental long run value, which, they show, provides very competitive forecasts. There are two main differences between the empirical results in this chapter and Faust and Wright (2012). The latter focus on real-time data and their sample, which is constrained by the availability of judgemental forecasts, starts in 1985. We focus on fully revised data that were available in January 2011 since our objective is to study the behavior of inflation over a longer sample period, which is important in order to uncover potential instabilities in the forecasting performance of the models.

Unless otherwise noted, in all the tables and figures, Panel A reports results for forecasting inflation and Panel B for output growth.

4.1 "Is the Predictive Content Stable Over Time?"

In this section, we test whether the predictive content is stable over time. We focus on testing the stability of the predictive content by using both traditional Granger-causality tests, out-of-sample forecast comparison tests, and forecast rationality tests, as well as their versions robust to instabilities. Then, we evaluate the forecasting ability of time-varying coefficient models and forecast combinations.

4.1.1 Do Traditional Macroeconomic Time Series Granger-cause Inflation and Output Growth?

Table 1 reports results of Granger-causality tests as well as Rossi's (2005) Granger-causality tests robust to instabilities. For each of the predictors that we consider (reported in the first column), transformed in several possible ways (described in the second column), and for each of the countries that we consider (described in the remaining columns), the table reports p-values of traditional Granger-causality tests (upper row) and p-values of Rossi's (2005) Granger-causality test robust to instabilities (lower row, in parentheses), QLR_T^*, defined in eq. (3).[55] The table shows two interesting empirical results. First, the traditional Granger-

[55] The Granger-causality tests focus on jointly testing whether $\beta_{10} = ... = \beta_{1p} = 0$ in regression (50). Note that in the table several predictability tests are reported, one for each predictor, although the multiple testing aspect is not taken into account in the calculation of the p-values. There are currently no available

causality tests show that many of the predictors that we consider do help predicting both inflation and output growth since, in most cases, the p-values are close to zero. The tables show which predictors are most useful. For example, inflation does not Granger-cause output growth in most countries, but some measures of unemployment do. Second, in several cases traditional Granger-causality tests do not find predictive ability whereas Rossi's (2005) test does, thus indicating that there is Granger-causality once instability has been taken into account. For example, only selected interest rates Granger-cause inflation, although almost all interest rates do Granger-cause inflation if we take instabilities into account.

INSERT TABLE 1 HERE

To get a sense of how important instabilities are, Figure 1 reports scatterplots of the p-values of the traditional Granger-causality tests (on the horizontal axis) and of Rossi's (2005) Granger-causality test robust to instabilities (on the vertical axis). Panel A in Figure 1 reports results for forecasting inflation and Panel B for output growth. Each dot in the figure corresponds to one of the series that we consider. The dotted lines represent p-values of 5%, and divide the picture in four quadrants. Dots in the upper right quadrant correspond to series where no Granger-causality is found by either traditional tests or by Granger-causality tests robust to instabilities. Dots in the lower left quadrant (close to the origin) correspond to series where Granger-causality is found by both traditional and robust tests. The upper left and in the lower right quadrants focus on cases in which the two tests disagree. Dots in the lower right quadrant correspond to series where traditional Granger-causality tests do not find evidence of predictive ability whereas Rossi's (2005) robust Granger-causality test does find predictive ability. Similarly, dots in the upper left quadrant correspond to series where traditional Granger-causality tests do find evidence of predictive ability whereas Rossi's (2005) robust Granger-causality test does not.

Panel A in Figure 1 shows that there are many dots concentrated in the lower left panel, indicating that both tests do find Granger-causality for several inflation predictors. However, there are many more dots in the lower right quadrant than in the upper left one, thus indicating that there are several cases where Granger-causality is uncovered only by using tests that are robust to instabilities. Similar results hold for forecasting output growth, reported in Panel B. We conclude that properly taking into account instabilities is very important when evaluating whether traditional macroeconomic time series Granger-

tests for multiple forecast comparisons robust to instabilities.

cause either inflation or output growth and in several cases overturns the empirical results based on traditional Granger-causality tests.

INSERT FIGURE 1 HERE

4.1.2 Do Traditional Macroeconomic Time Series Beat an Autoregressive Benchmark Model in Out-of-Sample Forecast Comparisons Tests?

We next consider the predictive ability of the same macroeconomic variables for forecasting inflation and output growth out-of-sample. The benchmark is the autoregressive model, and the forecast horizon is four quarters. Results are broadly similar for the random walk without drift benchmark and for other forecast horizons. We consider both traditional out-of-sample forecast comparison tests as well as Giacomini and Rossi's (2010a) forecast comparisons tests robust to instabilities.[56]

Tables 2 and 3 report results of traditional out-of-sample forecast comparison tests. The first line in Table 2 reports the RMSFE of the benchmark autoregressive (AR) model (labeled "ARrmse"). In subsequent rows, for every explanatory variable, the first line in Table 2 reports the ratio of the MSFE of the model relative to the MSFE of the autoregressive benchmark, so that values less than unity indicate that the model forecasts better than the autoregressive benchmark; the second line (in parentheses) reports the p-value of the one-sided DMW_P test statistic, eq. (9). The p-values of the DMW_P test statistic used in this empirical application are obtained using the critical values in Giacomini and White (2006). The table shows little empirical evidence in favor of predictive ability for the models. However, there are some exceptions: for predicting inflation one year ahead, some measures of interest rates are useful in some countries, and some measures of output and unemployment gap are useful for France and Italy; when predicting output growth, several interest rates are useful for various countries, as well as industrial production and the employment gap for Canada, Italy and the U.S.

INSERT TABLE 2 HERE

We now turn to out-of-sample forecast comparison tests that are robust to instabilities. Table 3 reports results for the Giacomini and Rossi's (2010a) Fluctuation test, eq. (13). The

[56] A similar exercise was undertaken by Rossi and Sekhposyan (2010) for the US only. There are two differences relative to Rossi and Sekhposyan (2010): their sample ended in 2005 whereas ours is updated to 2010, and they also considered real-time forecasts, which we do not.

test is implemented by choosing $\delta = 0.375$, which, for example, gives a window size of 60 out-of-sample observations when the total number of out-of-sample forecasts is 160. Asterisks denote significance at the 5% level. In many cases we find empirical evidence that the model with macroeconomic predictors forecasts better. In particular, there is evidence that some interest rates (e.g. real overnight and T-bill rates), output measures (e.g. real GDP, unemployment, etc.), stock prices and some measures of money supply were useful predictors for inflation at some point in time. Similarly, the spread, stock prices, unemployment, capital utilization and several measures of money supply were useful predictors for output growth at some point in time. Figure 2 reports a scatterplot of the p-values of the traditional DMW_P "average-out-of-sample" traditional test statistic (labeled $MSE-t$, on the horizontal axis)[57] and of the Giacomini and Rossi's (2010a) Fluctuation test (on the vertical axis). Figure 2 is interpreted as follows: dots on the right of the vertical critical value line represent successful predictors according to the traditional test, whereas dots above the horizontal critical value line represent successful predictors according to the Fluctuation test. Clearly, both Panels A and B show that several of the dots are in the upper, left quadrant. Thus, even though in many cases traditional tests would not find evidence that any of the predictors are useful for forecasting inflation or output growth, the Fluctuation test uncovers that they were indeed useful predictors at some point in time. The problem is that their predictive ability was masked by instabilities.

INSERT TABLE 3 AND FIGURE 2 HERE

A scatterplot of the in-sample versus the out-of-sample tests suggests that in-sample tests typically find more predictive ability than out-of-sample tests. Figure 3 plots results for traditional tests, whereas Figure 4 focuses on the robust tests. The main conclusion is that out-of-sample tests are a tougher benchmark to beat, due to the reasons discussed in Section 3, and confirms one of the main themes in this Chapter, namely that in-sample tests do not provide reliable guidance to out-of-sample forecasting ability.

INSERT FIGURES 3 AND 4 HERE

It would also be interesting to investigate the behavior of the relative predictive ability over time by plotting the Fluctuation tests for each predictor. However, this is infeasible due to space constraints. Instead, we report the percentage of predictors whose Fluctuation

[57] P-values for the DMW_P test are calculated using Giacomini and White (2006).

test is outside the critical value at each point in time. Figure 5 reports the results. Panel A in Figure 5 shows that the largest percentages of rejections for inflation forecasts happened around the mid- to late 1980s, whereas there is much less empirical evidence in favor of the predictors in the late 2000s. Results are similar for output (Panel B), except that there seems to be more predictive ability in forecasting output growth in the late 1990s and early 2000s relative to inflation.

INSERT FIGURE 5 HERE

Table 4 reports results for the Clark and McCracken's (2001) ENCNEW test statistic. See Clark and McCracken (this Handbook) and Busetti, Marcucci and Veronese (2011) for an analysis of the relative properties of the ENCNEW test relative to other tests proposed in the literature in stationary environments. The latter test finds much more evidence in favor of predictive ability than the test reported in Table 2. Several measures of interest rates significantly help predicting inflation for most countries, as well as several measures of output and money. Predicting output growth is instead much harder, and only selected measures of interest rates seem to work well across countries. The reason Tables 2 and 4 reach different conclusions is because of the different null hypothesis of the two tests. Table 2 tests for equal predictive ability at the estimated parameter values, whereas Table 4 tests for equal predictive ability under the assumption that the autoregressive benchmark model is the truth.

INSERT TABLE 4 HERE

4.1.3 Are Forecasts Rational?

Table 5 reports the results of Mincer and Zarnowitz' (1969) tests for forecast rationality. For every explanatory variable, the table reports the p-value of the traditional Mincer and Zarnowitz' (1969) test statistic, eq. (18). The table shows that rationality is almost never rejected. However, results are very different when considering robust forecast rationality tests. Rejections at 5% significance level for the Rossi and Sekhposyan's (2011b) Fluctuation rationality test, eq. (19), are reported by asterisks. There are several instances where rationality is rejected, in particular when using interest rates and monetary aggregates for predicting inflation in several countries, as well as for almost all predictors of output growth. Figure 6 reports a scatterplot of the traditional Mincer and Zarnowitz's (1969) test statistic (on the horizontal axis) and of Rossi and Sekhposyan's (2011b) Fluctuation rationality test (on the vertical axis).[58] The figure shows that in several cases one would not find evidence against rationality by using the traditional tests, but would reject rationality using the Fluctuation rationality test. That is, there is empirical evidence that forecasts were not rational at least at some point in time.

INSERT TABLE 5 AND FIGURE 6 HERE

Results are very similar for forecast unbiasedness tests – see Panel C and D in Table 5, which report results for tests for traditional forecast unbiasedness and for robust unbiasedness tests (Rossi and Sekhposyan, 2011b), and Figure 7, which reports scatterplots of p-values for the same tests.

INSERT FIGURE 7 HERE

4.1.4 Are the Empirical Conclusions Robust to the Choice of the Window Size?

Table 6 reports results for Pesaran and Timmermann's (2007) "Ave" procedure for combining forecasts across window sizes, eq. (24), relative to the autoregressive benchmark. For each regressor, the first row reports the ratio of the MSFE of the "Ave" forecast relative to the MSFE of the autoregressive benchmark, and the second line reports p-values of the Diebold and Mariano's (1995) and Giacomini and White's (2006) test. In the case of inflation, reported in Panel A, the procedure is capable of improving the forecasting performance of several predictors; in particular, for the U.S., the successful predictors include several interest rates (Treasury bills, bonds, overnight rates, both nominal and real), stock prices, several output measures (including GDP, capital utilization, unemployment) and producer price indices. The last row of the table reports similar results for the Pesaran, Schuermann and Smith's (2009) "Ave-Ave" procedure, which combines all predictors across all windows. Interestingly, the "Ave-Ave" procedure does perform significantly better than the autoregressive benchmark for all countries.

Turning to forecasting output growth, Panel B shows that Pesaran and Timmermann's (2007) "Ave" procedure is also useful for predicting output growth, although to a smaller extent. A few predictors, among which the first difference of the real overnight interest rate, become statistically significant for almost all countries, as well as exchange rates, stock prices and money measures. Again, the last row shows that the "Ave-Ave" procedure does perform significantly better than the autoregressive benchmark for all countries.

[58] Note that, in this case, for simplicity, unlike in the previous tables, we report the test statistic value rather than its p-value.

INSERT TABLE 6 HERE

Results are even more striking when considering Inoue and Rossi's (2011) forecast comparison test procedure robust to the choice of the window size, eq. (27). Rejections of the test at the 5% significance level are marked by asterisks in Table 6. The table shows that it is possible to reject the benchmark model for almost every predictor for some choice of the window size. Overall, we conclude that the choice of the window size significantly affects the empirical evidence on predictive ability, and that methodologies that average information across window sizes are typically quite successful.

4.1.5 Do Time-Varying Estimation Models and Forecast Combinations Improve Forecasts?

We consider four techniques that have been used in the literature to estimate models in the presence of instabilities and which we reviewed in Section 2.3: forecast combinations with equal weights (labeled "EWA"), Bayesian model averaging (labeled "BMA"), factor-augmented Autoregressive models (labeled "FAAR"), and, for predicting inflation, Stock and Watson's (2007) UCSV model (labeled "UCSV"). Unreported results show that intercept corrections never improve over the autoregressive benchmark for any of the predictors.

We follow Faust and Wright (2009) and Wright (2009) in the estimation. In particular, for the BMA model, eq. (34), we assign the same prior used in Faust and Wright (2009): the prior over the parameters of the n models is such that, if each model is $y_{t+h} = \beta_i' x_{it} + \varepsilon_{i,t+h}$, where $\varepsilon_{i,t+h} \sim N(0, \sigma^2)$, then the prior for β_i conditional on σ is $N\left(\overline{\beta}, \phi\left(\sigma^2 \sum_{t=1}^{T} x_{it} x_{it}'\right)^{-1}\right)$, $\phi = 2$, the marginal prior for σ is proportional to $1/\sigma$. The models' forecasts are produced based on the posterior mean of the parameters. The n-forecasts are then combined by a weighted average; the weights are determined by the posterior probability that each model is correct. The FAAR model is estimated as follows: $y_{t+h} = \beta_0 + \sum_{i=1}^{p} \beta_i z_{it} + \sum_{j=0}^{q} \gamma_j y_{t-j} + \varepsilon_t$ where z_{it} are the first m principal components; p and q are simultaneously chosen by BIC. The max number of lags for y that we consider is 4, and the maximum number of principal components is 6.

Results for traditional out-of-sample forecast comparison tests relative to the autoregressive benchmark are reported in Table 7. The table reports the ratio of the MSFE of each of the models relative to the autoregressive benchmark as well as the p-value of the DMW_P test, eq. (41), using Giacomini and White's (2006) critical values in parenthesis. The table shows that equal weighted forecast combinations perform significantly better than the benchmark for forecasting inflation in most countries except Italy and France, in which cases the MSFE is nevertheless not much worse than that of the benchmark. The UCSV model also performs quite well especially for Germany, Japan and the U.S., although its forecasts are not better than the equal weighting average. BMA works quite well too: in most countries, it has a lower MSFE than the autoregressive model, although the difference is not significant except for France. FAAR models do not perform particularly well.

When forecasting output growth, forecast combinations are still the preferred choice for all countries except in the case of Japan, where the FAAR model performs better (although not significantly so) than the autoregressive benchmark. Again, BMA's forecasts are better than the autoregressive benchmark for several countries, although not significantly so except in the case of Germany.

Finally, we consider forecast comparisons tests robust to instabilities. According to Giacomini and Rossi's (2010a) Fluctuation test, reported in Table 8, when forecasting inflation both EWA and UCSV models beat the benchmark for all countries; similar results hold for the BMA in all but two countries. The Fluctuation test instead does not find any predictability in FAAR models except for Canada and Germany. Results are overall very similar for predicting output growth except that FAAR models do better.

INSERT TABLES 7 AND 8 HERE

Figure 8 reports plots the Fluctuation test over time for each of the models that we consider.[59] Panels A-D report results for forecasting inflation. Panel A shows the forecasting ability of EWA models is very strong, and suggests it is strong especially in the early 1980s; results are similar for BMA (Panel B). Panel C shows that FAAR models were never better than the benchmark, whereas Panel D shows that the UCSV model is better than the benchmark, both at the beginning of the sample but especially in the late 2000s. Panels E-G in Figure 8 show similar results for forecasting output.

INSERT FIGURE 8 HERE

4.2 "In-sample Versus Out-of-Sample"

We conclude the empirical analysis by considering two additional empirical questions. The first is whether there are forecast breakdowns. The second is what are the sources of the difference between in-sample fit and out-of-sample forecasting ability.

[59] The Fluctuation test is implemented using a centered moving window.

Table 9 considers Giacomini and Rossi's (2009) forecast breakdown test, eq. (45). The table shows that most predictors, with rare exceptions, have been experiencing forecast breakdowns. This is true both when forecasting inflation as well as output growth. For most of the series, the p-value of the forecast breakdown test is zero, which implies that the empirical evidence in favor of forecast breakdowns is very strong. Thus, the in-sample fit is not indicative of the out-of-sample performance for most predictors.

Finally, Table 10 investigates the causes of the differences between the in-sample fit and the forecasting ability of the candidate predictors relative to the autoregressive model by using Rossi and Sekhposyan's (2011b) test, eq. (49). Rossi and Sekhposyan's (2011b) decomposition, eq. (48), applies to the relative (de-meaned) MSFE differences in the numerator of the Diebold and Mariano's (1995) test statistic. The decomposition investigates the contributions of time-variation, over-fitting and marginal predictive content to explain the difference between in-sample fit and out-of-sample forecasting ability of the models. From Table 2, which reported the ratio of the MSFE differences, we know that the MSFE of the autoregressive model is lower than that of the predictors's model for most predictors. Thus, Rossi and Sekhposyan's (2011b) decomposition helps understand why the predictors' model does not significantly improve over the autoregressive model in forecasting out-of-sample.

In the case of forecasting inflation (Panel A), the $\Gamma_P^{(A)}$ test points to the existence of instabilities in most series and for most countries. In several cases, in particular in the case of nominal interest rates, the $\Gamma_P^{(B)}$ statistic is positive and significant and the $\Gamma_P^{(U)}$ component is significant only rarely, suggesting that nominal interest rates may have some predictive content for inflation and the main reason for their poor performance is instability. In several other cases, in particular when considering real interest rates as well as employment/unemployment and capital utilization, the B_P component is instead significantly negative, thus suggesting that not only there is instability but also that in-sample fit is misleading. In the case of stock prices and some measures of real activity, overfitting (U_P) is also important. For U.S. data, in particular, Table 10 shows that money does have predictive content for inflation, although it is highly unstable in most cases; the in-sample predictive content of measures of real activity and some nominal interest rates (e.g. the 3-month T-bill, and the 5 and 10 years maturity bonds), instead, is negatively correlated with out-of-sample predictive content. The least empirical evidence of overfitting and the most empirical evidence of predictive content seem to be related to inflation predictors such as the monetary base and M1.

INSERT TABLES 9 AND 10 HERE

Predicting output shares similar features but also interesting differences. As in the case of predicting inflation, instabilities are really important. Overall, however, notwithstanding instabilities, most interest rates demonstrate significant predictive ability on average over the sample, as well as real output measures such as employment, capital utilization, several measures of money growth and inflation. The reason for their poor performance is attributed to the fact that, for most series, overfitting is also significantly present, and that undermines the positive effects of the predictive content. The in-sample fit of exchange rates and stock prices, instead, is significantly misleading for predicting output growth out-of-sample. For the U.S., in particular, interest rates and money measures seem to have potential explanatory power, although undermined by instabilities; exchange rates and stock prices instead, mostly overfit.

5 Conclusions

This chapter shows that there are two important stylized facts regarding the forecasting ability of economic models. The first is that the predictive content is unstable over time. The second is that in-sample predictive content does not necessarily guarantee out-of-sample predictive ability, nor the stability of the predictive relation over time. These issues were discussed, among others, in an influential paper by Stock and Watson (2003), who also provided empirical evidence using a large database of macroeconomic predictors for both inflation and output growth. As we show, these issues are important not only in Stock and Watson's (2003) database, but also in several models and databases commonly considered in macroeconomics, finance, as well as international finance.

However, several new methods for estimation and inference have been developed in the recent literature to help researchers and practitioners to deal with these issues. In particular, researchers who are interested in evaluating predictive ability, but worry about the predictive content being unstable over time, can rely on Granger-causality tests robust to instabilities (Rossi, 2005), out-of-sample forecast comparison tests robust to instabilities (Giacomini and Rossi, 2010), and forecast rationality tests robust to instabilities (Rossi and Sekhposyan, 2011b). Instabilities can be exploited to improve the estimation of the forecasting models, for example by estimating historic breaks via structural breaks or time-varying parameter models (Pesaran and Timmermann's (2007) "ROC" procedures and Stock and Watson's (2007) UCSV model) or models with multiple discrete breaks (Pesaran, Pettenuzzo and Timmermann, 2006, and Koop and Potter, 2007), or by combining models' forecasts either via equal weights, Bayesian model averaging or across window sizes (Pesaran and Timmermann's (2002) "Ave" procedure) or across recursive and rolling schemes (Clark and McCracken, 2009) or by intercept corrections (Clemens and Hendry, 1996). Other tools involve inference robust to the choice of the window size (Inoue and Rossi, 2011, and Hansen and Timmermann, 2011).

Researchers should also worry about the fact that in-sample fit does not guarantee good out-of-sample forecasting performance. Forecast breakdown tests (Clemens and Hendry, 1998, and Giacomini and Rossi, 2009) can be used to establish when that is the case, and Rossi and Sekhposyan's (2011b) decomposition can be used to determine the reasons behind the difference between in-sample fit and out-of-sample forecasting performance.

An empirical application to the updated Stock and Watson's (2003) large database of macroeconomic predictors for inflation growth and real GDP growth highlights the following, general conclusions:

(i) there is substantially more empirical evidence in favor of Granger-causality of typical macroeconomic predictors when using Granger-causality tests robust to instabilities;

(ii) there is also substantially more empirical evidence in favor of out-of-sample forecasting ability when using out-of-sample forecast tests robust to instabilities;

(iii) there is more empirical evidence against forecast rationality when one allows for instabilities;

(iv) given the widespread empirical importance of instabilities, it comes at no surprise that the choice of the window size is crucial; forecast combinations across window sizes tend to perform well out-of-sample, and the empirical evidence in favor of predictive ability is clearly stronger across predictors when using methods that are robust to the choice of the window size;

(v) equal weighted averaging is among the time-varying estimation models that perform the best out-of-sample; Bayesian model averaging and the UCSV model by Stock and Watson (2007) also do very well (the latter in the special case of forecasting inflation) although not as well as equal weighted forecast combination. Factor autoregressive models tend to perform worse than an autoregressive benchmark;

(vi) there is substantial evidence of forecast breakdowns, which is related not only to instabilities, but also poor predictive ability of the regressors; in several cases, even if the regressors have predictive power, it appears to be undermined by overfitting.

The results in this chapter suggest several avenues for future research. First, equal weight forecast averaging is one of the most successful and stable forecast methodologies in the presence of instabilities. Understanding why that is the case might provide useful guidelines for improving the estimation of time-varying parameter models (see Hendry and Clements, 2002, and Timmermann, 2006). Second, the widespread presence of forecast breakdowns suggests the need of improving ways to select good forecasting models in-sample. In addition, it is also very important to understanding the economic causes of such breakdowns in forecasting accuracy. Developing such procedures is an important area for future research.

References

[1] Agnew, R.A. (1982), "Econometric Forecasting Via Discounted Least Squares," *Naval Research Logistics Quarterly* 29(2), 291-302.

[2] Aiolfi, M. and A. Timmermann (2004), "Structural Breaks and the Performance of Forecast Combinations," *mimeo* UCSD.

[3] Aiolfi, M. and A. Timmermann (2006), "Persistence of Forecasting Performance and Conditional Combination Strategies," *Journal of Econometrics* 135(1-2), 31-53.

[4] Aiolfi, M., C. Capistran and A. Timmermann (2010), "Forecast Combinations," mimeo, UCSD.

[5] Altavilla, C. and M. Ciccarelli (2007), "Information Combination and Forecastability. Evidence From Vintages of Time-series Data," *European Central Bank Working Paper Series* 846.

[6] Amato, R. and N.R. Swanson (1998), "Money and Output Viewed Through a Rolling Window," *Journal of Monetary Economics* 41, 455-474.

[7] Amisano, G. and R. Giacomini (2007), "Comparing Density Forecasts via Weighted Likelihood Ratio Tests," *Journal of Business and Economic Statistics* 25, 177-190.

[8] Andersen, T., T. Bollerslev, P. Christoffersen and F. Diebold (2006), "Volatility and Correlation Forecating" in: G. Elliott, C. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting Vol. 1*, North Holland: Elsevier.

[9] Andersen, T.G., T. Bollerslev, F.X. Diebold and C. Vega (2003), "Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange," *American Economic Review* 93(1), 38-62.

[10] Andreou, E., E. Ghysels and A. Kourtellos (2010), "Regression Models with Mixed Sampling Frequencies," *Journal of Econometrics* 158, 246-261.

[11] Andrews, D.W. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 817-858.

[12] Andrews, D.W. (1993), "Tests for Parameter Instability and Structural Change With Unknown Change Point," *Econometrica* 61(4), 821-856.

[13] Andrews, D.W. and W. Ploberger (1994), "Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative," *Econometrica* 62, 1383-1414.

[14] Ang, A. and G. Bekaert (2004), "Stock Return Predictability: Is it There?," *mimeo*, Columbia University and NBER.

[15] Bacchetta, P., E. van Wincoop and T. Beutler (2009), "Can Parameter Instability Explain the Meese-Rogoff Puzzle?," in: *NBER International Seminar on Macroeconomics 2009*, 125-173 National Bureau of Economic Research.

[16] Bacchetta, P. and E. van Wincoop (2009) "On the Unstable Relationship between Exchange Rates and Macroeconomic Fundamentals," *NBER Working Papers* 15008.

[17] Bache, I.W., A.S. Jore, J. Mitchell and S.P. Vahey (2011), Combining VAR and DSGE Forecast Densities, *Journal of Economic Dynamics and Control* 35, 1659-1670.

[18] Bai, J. (1997), "Estimation of a Change Point in Multiple Regression Models," *Review of Economics and Statistics* 79(4), 551-563.

[19] Bai, J. and P. Perron (1998), "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica* 66(1), 47-78.

[20] Banbura, M., D. Giannone and L. Reichlin (2010), "Nowcasting", in: M.P. Clements and D.F. Hendry (eds.), *Oxford Handbook on Economic Forecasting*, forthcoming.

[21] Bates, J.M. and C. W.J. Granger (1969), "The Combination of Forecasts," *Operations Research Quarterly* 20, 451-468.

[22] Bauwens, L., D. Korobilis and J.V.K. Rombouts (2011), "A Comparison of Forecasting Procedures for Macroeconomic Series: The Contribution of Structural Break Models," *mimeo*, available at: http://neumann.hec.ca/pages/jeroen.rombouts/research/bkkr4.pdf

[23] Bekaert, G. and R.J. Hodrick (1992), "Characterizing Predictable Components in Excess Returns in Equity and Foreign Exchange Markets," *Journal of Finance* 47, 467–509.

[24] Beckmann J., A. Belke and M. Kuhl (2011), "The Dollar-Euro Exchange Rate and Macroeconomic Fundamentals: a Time-Varying Coefficient Approach," *Review of World Economics* 147 (1), 11-40.

[25] Billio, M., R. Casarin, F. Ravazzolo F. and H.K. van Dijk (2011), "Combining Predictive Densities Using Nonlinear Filtering with Applications to US Economics Data", *mimeo*, Norges Bank.

[26] Blume, M. E. (1975), "Betas and Their Regression Tendencies," *Journal of Finance* 30, 785-795.

[27] Boivin, J. and M. Giannoni (2006), "Has Monetary Policy Become More Effective?," *Review of Economics and Statistics* 88(3), 445-462.

[28] Bordo, M.D. and J.G. Haubrich (2008), "The Yield Curve as a Predictor of Growth: Long-Run Evidence, 1875-1997," *Review of Economics and Statistics* 90(1), 182-185.

[29] Bossaerts, P. and P. Hillion, P. (1999), "Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?," *Review of Financial Studies* 12, 405–428.

[30] Brown, R.G. (1963), *Smoothing, Forecasting and Prediction*, Englewood Cliffs: Prentice Hall.

[31] Brown, R.L., J. Durbin and J.M. Evans (1975), "Techniques for Testing the Constancy of Regression Relationships Over Time," *Journal of the Royal Statistical Society* Series B 37, 149-192.

[32] Busetti, F. J. Marcucci and G. Veronese (2011), "Comparing Forecast Accuracy: A Monte Carlo Investigation", *mimeo*, Bank of Italy

[33] Calhoun, G. (2011), "Out-of-Sample Comparison of Overfit Models," *mimeo*, Iowa State University.

[34] Calhoun, G. and G. Elliott (2012), "Why Do Nonlinear Models Provide Poor Macroeconomic Forecasts?," *mimeo*, UCSD.

[35] Campbell, J.Y. (1987), "Stock Returns and the Term Structure," *Journal of Financial Economics* 18, 373–399.

[36] Campbell, J.Y. and R.J. Shiller (1988), "Stock Prices, Earnings, and Expected Dividends," *Journal of Finance* 43, 661–676.

[37] Canova, F. (1993), "Modelling and Forecasting Exchange Rates with a Bayesian Time-varying Coefficient Model," *Journal of Economic Dynamics and Control* 17(1-2), 233-261.

[38] Canova, F. (2007), "G-7 Inflation Forecasts: Random Walk, Phillips Curve, or What Else?," *Macroeconomic Dynamics* 11(01), 1-30.

[39] Carriero, A., T. Clark and M. Marcellino (2012), "Common Drifting Volatility in Large Bayesian VARs", *CEPR Discussion Paper* 8894.

[40] Carstensen, K., K. Wohlrabe and C. Ziegler (2010), "Predictive Ability of Business Cycle Indicators under Test: A Case Study for the Euro Area Industrial Production," *CESIFO* W.P. n. 3158.

[41] Castelnuovo, E., L. Greco and D. Raggi (2008), "Estimating Regime-Switching Taylor Rules with Trend Inflation," *Bank of Finland Research Discussion Papers* 20/2008.

[42] Castle, J., N.W.P. Fawcett and D.F. Hendry (2012), "Forecasting Breaks and Forecasting during Breaks," in Clements, M.P. and D.F. Hendry (eds.), *Oxford Handbook of Economic Forecasting,* Oxford University Press, 315-354.

[43] Chen, Y., K. Rogoff and B. Rossi (2010), "Can Exchange Rates Forecast Commodity Prices?," *Quarterly Journal of Economics* 125(3), 1145-1194.

[44] Clark, T.E. (2011), "Real-Time Density Forecasts from Bayesian Vector Autoregressions with Stochastic Volatility", *Journal of Business and Economic Statistics* 29(3), 327-341.

[45] Clark, T.E. and M.W. McCracken (2001), "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics* 105(1), 85-110.

[46] Clark, T.E. and M.W. McCracken (2005), "The Power of Tests of Predictive Ability in the Presence of Structural Breaks," *Journal of Econometrics* 124, 1-31.

[47] Clark T. E. and M. W. McCracken (2008), "Forecasting with Small Macroeconomic VARs in the Presence of Instability". In: *Forecasting in the Presence of Structural Breaks and Model Uncertainty*, Rapach D. E., Wohar M. E. (eds). Emerald Group Publishing: Bingley, U.K., 93–147.

[48] Clark, T.E. and M.W. McCracken (2009), "Improving Forecast Accuracy by Combining Recursive and Rolling Forecasts," *International Economic Review* 50(2), 363-395.

[49] Clark, T.E. and M.W. McCracken (2010), "Averaging Forecasts from VARs with Uncertain Instabilities," *Journal of Applied Econometrics* 25(1), 5-29.

[50] Clark, T.E. and K.D. West (2006), "Using Out-of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis," *Journal of Econometrics* 135, 155-186.

[51] Clark, T.E. and K.D. West (2007), "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models," *Journal of Econometrics* 138, 291-311.

[52] Clemen, R.T. and R.L. Winkler (1986), "Combining Economic Forecasts," *Journal of Business and Economic Statistics* 4, 39-46.

[53] Clements, M.P. and D.F. Hendry (1993), "On the Limitations of Comparing Mean Square Forecast Errors," *Journal of Forecasting* 12, 617-637.

[54] Clements, M.P. and D.F. Hendry (1996), "Intercept Corrections and Structural Change," *Journal of Applied Econometrics* 11, 475-494.

[55] Clements, M.P. and D.F. Hendry (1998), *Forecasting Economic Time Series*, Cambridge: Cambridge University Press.

[56] Clements, M.P. and D.F. Hendry (1999), "Some Methodological Implications of Forecast Failure," *mimeo*, Warwick University and Nuffield College.

[57] Clements, M. P. and D. F. Hendry (2002). Explaining forecast failure in macroeconomics. In: Clements, M. P. and D.F. Hendry (eds.), *A Companion to Economic Forecasting*. Oxford: Blackwells, 539–571.

[58] Hendry, D.F. and M.P. Clements (2004), "Pooling of Forecasts," *Econometrics Journal* 7(1), 1-31.

[59] Clements, M. P. and D. F. Hendry (2006), "Forecasting with breaks in data processes", in: Elliott, G., C.W.J. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting Vol. 1*, Elsevier, 605–657.

[60] Clements, M.P., P.H. Franses and N.R. Swanson (2004), "Forecasting Economic and Financial Time-Series with Non-linear models," *International Journal of Forecasting* 20, 169-183.

[61] Cogley, T. (2002), "A Simple Adaptive Measure of Core Inflation", *Journal of Money, Credit and Banking* 34, 94-113.

[62] Cogley, T. and T. Sargent (2001), "Evolving Post–World War II U.S. Inflation Dynamics," in B. Bernanke and K. Rogoff (eds.), *NBER Macroeconomics Annual,* Cambridge, MA: MIT Press.

[63] Cogley, T. and T. Sargent (2005), "Drifts and Volatilities: Monetary Policies and Outcomes in the post World War II U.S.," *Review of Economic Dynamics* 8, 262-302.

[64] Cogley, T. and A. Sbordone (2008), "Trend Inflation, Indexation, and Inflation Persistence in the New Keynesian Phillips Curve," *American Economic Review* 98, 2101–2126.

[65] Cooper, M., R.C. Gutierrez and W. Marcum (2005), "On the Predictability of Stock Returns in Real Time," *Journal of Business* 78 (2), 469-499.

[66] Corradi, V. and N. R. Swanson (2006a), "Bootstrap Conditional Distribution Tests in the Presence of Dynamic Mis-specification", *Journal of Econometrics* 133, 779-806.

[67] Corradi, V. and N. R. Swanson (2006b), "Predictive Density Evaluation", in: G. Elliott, C. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Elsevier, 197-284.

[68] Croushore, D. and T. Stark (2001), "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics* 105(1), 111-130.

[69] Croushore, D. (2011), "Two Dimensions of Forecast Evaluation: Vintages and Subsamples," *mimeo*, U. of Richmond.

[70] D'Agostino, A., L. Gambetti and D. Giannone (2009), "Macroeconomic Forecasting and Structural Change," *mimeo*, European Central Bank.

[71] D'Agostino, A., D. Giannone and P. Surico (2008), "(Un)Predictability and Macroeconomic Stability," *CEPR Discussion Paper* No. DP6594.

[72] D'Agostino, A. and P. Surico (2011), "A Century of Inflation Forecasts," *Review of Economics and Statistics*, forthcoming.

[73] Díaz, J. and G. Leyva (2008), "Forecasting Chilean Inflation in Difficult Times," *Central Bank of Chile* Working Papers 511.

[74] Diks, C., V. Panchenkob and D. van Dijk (2011), "Likelihood-based Scoring Rules for Comparing Density Forecasts in Tails", *Journal of Econometrics* 163, 215–230.

[75] Diebold, F. X., T.A. Gunther, and A.S. Tay (1998), "Evaluating Density Forecasts with Applications to Financial Risk Management", *International Economic Review* 39(4), 863-883.

[76] Diebold, F. X. and J. Lopez (1996), "Forecast Evaluation and Combination," in: G.S. Maddala and C.R. Rao (eds.), *Handbook of Statistics*, North-Holland, 241–268.

[77] Diebold, F.X. and R. S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics,* 13, 253-263.

[78] Diebold, F.X. and P. Pauly (1987), "Structural Change and the Combination of Forecasts," *Journal of Forecasting* 6, 21-40.

[79] Doornik, J.A. (2008), "Encompassing and Automatic Model Selection", *Oxford Bulletin of Economics and Statistics* 70, 915–925.

[80] Edge, R.M, and R.S. Gürkaynak (2010), "How Useful Are Estimated DSGE Model Forecasts for Central Bankers?," CEPR Discussion Papers 8158, Brooking Papers on Economic Activity, *forthcoming.*

[81] Eklund, J. and S. Karlsson (2005), "Forecast Combination and Model Averaging using Predictive Measures," *Econometric Reviews* 26(2-4), 329-363.

[82] Elliott, G. (2009), "Forecasting with Trending Data" in: G. Elliott, C. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting Vol. 1*, North Holland: Elsevier.

[83] Elliott, G. and U. Muller (2006), "Efficient Tests for General Persistent Time Variation in Regression Coefficients," *Review of Economic Studies* 73, 907-940.

[84] Elliott, G. and U. Muller (2007), "Confidence Sets for the Date of a Single Break in Linear Time Series Regressions," *Journal of Econometrics* 141, 1196-1218.

[85] Elliott, G. and A. Timmermann (2004), "Optimal Forecast Combination Under General Loss Function and Forecast Error Distributions," *Journal of Econometrics* 122, 47-79.

[86] Elliott, G. and A. Timmermann (2005), "Optimal Forecast Combination Under Regime Switching," *International Economic Review*, 1081-1102.

[87] Elliott, G. and A. Timmermann (2008), "Economic Forecasting," *Journal of Economic Literature* 46(1), 3–56.

[88] Engel, C., N.C. Mark, and K.D. West (2008), "Exchange Rate Models Are Not as Bad as You Think," *NBER Macroeconomics Annual*.

[89] Engle, R.F. and A.D. Smith (1999), "Stochastic Permanent Breaks", *Review of Economics and Statistics* 81, 553-574.

[90] Fama, E.F. and K.R. French (1988), "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics* 22, 3–25.

[91] Fama, E. and K.R. French (1997), "Industry Cost of Equity," *Journal of Financial Economics* 43, 153-193.

[92] Faroque, A., W. Veloce and J. Lamarche (2009), "Have Structural Changes Eliminated the Out-of-Sample Ability of Financial Variables To Forecast Real Activity After the Mid-1980s? Evidence From the Canadian Economy," *Brock University Working Papers* 0910.

[93] Faust, J. and J. Wright (2009), "Comparing Greenbook and Reduced Form Forecasts using a Large Realtime Dataset," *Journal of Business and Economic Statistics* 27, 468-479.

[94] Faust, J. and J. Wright (2012), "Forecasting Inflation", in: in: G. Elliott and A. Timmermann (eds.), *Handbook of Economic Forecasting Vol. 2*, North Holland: Elsevier.

[95] Fernandez, Koenig and Nikolosko-Rzhevskyy (2008), "The Relative Performance of Alternative Taylor Rule Specifications," *Dallas Fed Staff Papers* 6.

[96] Ferraro, D., K. Rogoff and B. Rossi (2011), "Can Oil Prices Forecast Exchange Rates?," *Federal Reserve Bank of Philadelphia* WP 11-34.

[97] Figlewski, S. and T. Urich (1983), "Optimal Aggregation of Money Supply Forecasts: Accuracy, Profitability and Market Efficiency," *Journal of Finance* 28, 695-210.

[98] Galvão, A. B. (2011), "Changes in Predictive Ability with Mixed Frequency Data," *mimeo*, Queen Mary University of London.

[99] Galvão, A. B. (2006). "Structural Break Threshold VARs for Predicting the Probability of U.S. Recessions Using the Spread," *Journal of Applied Econometrics* 21: 463–87.

[100] Garratt, A., G. Koop and S.P. Vahey (2008), "Forecasting Substantial Data Revisions in the Presence of Model Uncertainty," *Economic Journal* 118(530), 1128-1144.

[101] Geweke, J. and G. Amisano (2007), "Optimal Prediction Pools," *Journal of Econometrics* 164(1), 130-141.

[102] Giacomini, R. and B. Rossi (2006), "How Stable is the Forecasting Performance of the Yield Curve for Output Growth?," *Oxford Bulletin of Economics and Statistics* 68(s1), 783-795.

[103] Giacomini, R. and B. Rossi (2009), "Detecting and Predicting Forecast Breakdowns," *Review of Economic Studies* 76(2), 2009.

[104] Giacomini, R. and B. Rossi (2010a), "Forecast Comparisons in Unstable Environments", *Journal of Applied Econometrics* 25(4), 595-620.

[105] Giacomini, R. and B. Rossi (2010b), "Model Comparisons in Unstable Environments," *mimeo.*

[106] Giacomini, R. and H. White (2006), "Tests of Conditional Predictive Ability," *Econometrica,* 74, 1545-1578.

[107] Gonzales, A., K. Hubrich and T. Terasvirta (2011), "Forecasting Inflation with Gradual Regime Shifts and Exogenous Information", *European Central Bank Working Paper* No. 1363.

[108] Gonzalo, J. and J. Pitarakis (2010), "Regime Specific Predictability in Predictive Regressions," Universidad Carlos III Working Papers we097844.

[109] Goyal, A. and I. Welch (2003), "Predicting the Equity Premium with Dividend Ratios," *Management Science* 49, 639–654.

[110] Goyal, A. and I. Welch (2008), "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction", *Review of Financial Studies* 21(4), 1455–1508.

[111] Granger, C.W.J. and P. Newbold (1973), "Some Comments on the Evaluation of Economic Forecasts," *Applied Economics* 5, 35-47.

[112] Granger, C.W.J. and P. Newbold (1986), *Forecasting Economic Time Series* (2nd ed.), New York: Academic Press.

[113] Groen, J.J., R. Paap and F. Ravazzolo (2009), "Real-Time Inflation Forecasting in a Changing World," *Norges Bank Working Paper* 2009/16.

[114] Guidolin, M. and A. Timmermann (2007), "Asset Allocation under Multivariate Regime Switching," *Journal of Economic Dynamics and Control*, 3503-3544.

[115] Hall, S.G. and J. Mitchell (2007), "Combining Density Forecasts," *International Journal of Forecasting* 23(1), 1-13.

[116] Hamilton, J. (1988) "Rational Expectations Econometric Analysis of Changes in Regime," *Journal of Economic Dynamics and Control* 12(2-3), 385-413.

[117] Hansen, B. (2000), "Testing for Structural Change in Conditional Models," *Journal of Econometrics* 97, 93-115.

[118] Hansen, P.R. (2009), "In-Sample Fit and Out-of-Sample Fit: Their Joint Distribution and Its Implications for Model Selection", *mimeo.*

[119] Hansen, P.R. and A. Timmermann (2011), "Choice of Sample Split in Out-of-Sample Forecast Evaluation," *mimeo.*

[120] Harvey, A. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.

[121] Harvey, D.I., S.J. Leybourne and P. Newbold (1998), "Tests for Forecast Encompassing," *Journal of Business and Economic Statistics* 16 (2), 254-259.

[122] Hayashi, F. (2000), *Econometrics*, Princeton: Princeton University Press.

[123] Hendry, D.F. (2000), "On Detectable and Non-detectable Structural Change," *Structural Change and Economic Dynamics* 11, 45–65.

[124] Hendry, D.F. and M.P. Clements (2002), "Pooling of Forecasts," *Econometrics Journal* 5, 1-26.

[125] Hendry, D.F. and K. Hubrich (2011), "Combining Disaggregate Forecasts or Combining Disaggregate Information to Forecast an Aggregate," *Journal of Business and Economic Statistics* 29(2), 216-227.

[126] Hendry, D. F. and H.M. Krolzig (2005), "The properties of automatic Gets modelling," *Economic Journal*, 115, C32–C61.

[127] Henkel, S., J. S. Martin, and F. Nardari (2012), "Time-Varying Short-Horizon Predictability," *Journal of Financial Economics*, forthcoming.

[128] Holt, C.C. (1957), "Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages," *ONR Research Memorandum* 52, Carnegie Institute of Technology.

[129] Inoue, A. and L. Kilian (2004), "In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?," *Econometric Reviews* 23(4), 371-402.

[130] Inoue, A. and L. Kilian (2006), "On the Selection of Forecasting Models," *Journal of Econometrics* 130(2), 273-306.

[131] Inoue, A. and B. Rossi (2010), "Out of Sample Forecast Tests Robust to the Window Size Choice," Duke University ERID Working Paper 94.

[132] Inoue, A. and B. Rossi (2011), "Identifying the Sources of Instabilities in Macroeconomic Fluctuations," *Review of Economics and Statistics*, forthcoming.

[133] Jore, A.S., J. Mitchell and S. P. Vahey (2010), "Combining Forecast Densities from VARs with Uncertain Instabilities," *Journal of Applied Econometrics* 25(4), 621-634.

[134] Kang, H. (1986), "Unstable Weights in the Combination of Forecasts," *Management Science* 32:683-695.

[135] Kascha, C. and F. Ravazzolo (2010), "Combining Inflation Density Forecasts," *Journal of Forecasting* 29(1-2), 231-250.

[136] Koop, G. and D. Korobilis (2009), "Forecasting Inflation Using Dynamic Model Averaging", *Rimini Centre for Economic Analysis Working Paper* 34_09.

[137] Koop. G. and Korobilis (2012), "Large Time-Varying Parameter VARs", *mimeo,* University of Glasgow.

[138] Koop, G. and S.M. Potter (2004), "Forecasting in Large Macroeconomic Panels Using Bayesian Model Averaging," *Econometrics Journal* 7, 550-565.

[139] Koop, G. and S.M. Potter (2007), "Estimation and Forecasting in Models with Multiple Breaks," *Review of Economic Studies* 74, 763-789.

[140] Koop, G., S.M. Potter and R.W. Strachan (2008), "Re-examining the Consumption-Wealth Relationship: the Role of Model Uncertainty," *Journal of Money, Credit and Banking* 40(2-3), 341-367.

[141] Kozicki, S. and P.A. Tinsley (2001), "Shifting Endpoints in the Term Structure of Interest Rates," *Journal of Monetary Economics* 47, 613-652.

[142] Maheu, J. M. and S. Gordon (2008), "Learning, Forecasting and Structural Breaks," *Journal of Applied Econometrics* 23, 553-583.

[143] Manzana, S. and D. Zerom (2009), "Are Macroeconomic Variables Useful for Forecasting the Distribution of U.S. Inflation?," *mimeo*, Baruch College.

[144] Marcellino, M. (2006), "Leading Indicators", in: Elliot, G., C.W.J. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Ch. 16. Amsterdam: Elsevier.

[145] Marquering, W. and M. Verbeek (2004), "The Economic Value of Predicting Stock Index Returns and Volatility," *Journal of Financial and Quantitative Analysis* 39, 407–429.

[146] Mazzi, G.L., J. Mitchell and G. Montana (2010), "Density Nowcasts and Model Combination: Nowcasting Euro-area GDP Growth Over the 2008-9 Recession", *mimeo.*

[147] McConnell, M. and G. Perez-Quiros (2000), "Output Fluctuations in the United States: What Has Changed since the Early 1980's?" *American Economic Review* 90, 1464–1476.

[148] Meese, R. and K. S. Rogoff (1983a), "Exchange Rate Models of the Seventies. Do They Fit Out of Sample?," *Journal of International Economics* 14, 3-24.

[149] Meese, R. and K.S. Rogoff (1983b), "The Out of Sample Failure of Empirical Exchange Rate Models," in Jacob Frankel (ed.), *Exchange Rates and International Macroeconomics*, Chicago: University of Chicago Press for NBER.

[150] Meese, R. and K.S. Rogoff (1988), "Was It Real? The Exchange Rate–Interest Differential Relation over the Modern Floating Rate Period," *Journal of Finance* 43 (1988), 923–948.

[151] Min, C. and A. Zellner (1993), "Bayesian and Non-Bayesian Methods for Combining Models and Forecasts with Applications to Forecasting International Growth Rates," *Journal of Econometrics* 56, 89-118.

[152] Mincer, J. and V. Zarnowitz (1969), "The Evaluation of Economic Forecasts," in: J. Mincer (ed.), *Economic Forecasts and Expectations*, New York: National Bureau of Economic Research, 81–111.

[153] Newey, W. and K. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-708.

[154] Nyblom, J. (1989), "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association* 84, 223-230.

[155] Palm, F.C. and A. Zellner (1992), "To Combine or Not to Combine? Issues of Combining Forecasts," *Journal of Forecasting* 11:687-701.

[156] Patton, A. and A. Timmermann (2007), "Properties of Optimal Forecasts Under Asymmetric Loss and Nonlinearity," *Journal of Econometrics* 140, 884-918.

[157] Patton, A. and A. Timmermann (2011), "Forecast Rationality Tests Based on Multi-Horizon Bounds," *Journal of Business and Economic Statistics*, forthcoming.

[158] Paye, B. and A. Timmermann (2006), "Instability of Return Prediction Models". *Journal of Empirical Finance* 13(3), 274-315.

[159] Paye, B. S., D. Vol (2012), "Predictive Regressions for Aggregate Stock Market Volatility Using Macroeconomic Variables," *Journal of Financial Economics*, forthcoming.

[160] Perez-Quiros, G. and A. Timmermann (2000), "Firm Size and Cyclical Variations in Stock Returns," *Journal of Finance* 55, 1229–1262.

[161] Pesaran, M.H., and A. Pick (2011), Forecast Combination Across Estimation Windows, *Journal of Business and Economic Statistics* 29 (2), 307-318.

[162] Pesaran, M.H., T. Schuermann and L. V. Smith (2009), "Forecasting economic and financial variables with global VARs," *International Journal of Forecasting* 25(4), 642-675.

[163] Pesaran, M.H. and A. Timmermann (1995), "Predictability of Stock Returns: Robustness and Economic Significance," *Journal of Finance* 50, 1201–1228.

[164] Pesaran, M.H. and A. Timmermann (2002), "Market Timing and Return Prediction Under Model Instability," *Journal of Empirical Finance* 9(5), 495–510.

[165] Pesaran, M.H. and A. Timmermann (2005), "Small Sample Properties of Forecasts from Autoregressive Models under Structural Breaks," *Journal of Econometrics* 129, 183-217.

[166] Pesaran, M. H., D. Pettenuzzo and A. Timmermann (2006), "Forecasting Time Series Subject to Multiple Structural Breaks," *Review of Economic Studies* 73, 1057-1084.

[167] Pesaran, M.H. and A. Timmermann (2007), "Selection of Estimation Window in the Presence of Breaks," *Journal of Econometrics* 137(1), 134-161.

[168] Qu, Z. and P. Perron (2007), "Estimating and Testing Multiple Structural Changes in Multivariate Regressions", *Econometrica* 75, 459-502.

[169] Rapach, D.E., J. K. Strauss and G. Zhou (2010), "Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy," *Review of Financial Studies* 23(2), 821-862.

[170] Rapach, D.E. and M.E. Wohar (2005), "Regime Changes in International Real Interest Rates: Are They a Monetary Phenomenon?," *Journal of Money, Credit, and Banking* 37(5), 887-906.

[171] Rapach, D.E. and M.E. Wohar (2006), "Structural Breaks and Predictive Regression Models of Aggregate U.S. Stock Returns," *Journal of Financial Econometrics* 4(20), 238-274.

[172] Ravazzolo, F., R. Paap, D. van Dijk and P.H. Franses (2007), "Bayesian Model Averaging in the Presence of Structural Breaks," *mimeo*, Erasmus University.

[173] Ravazzolo, F., M. Verbeek and H.K. Van Dijk (2007), "Predictive Gains from Forecast Combinations Using Time Varying Model Weights," *mimeo*, available at SSRN: http://ssrn.com/abstract=1012574.

[174] Rime, D., L. Sarno and E. Sojli (2010), Exchange Rate Forecasting, Order Flow and Macroeconomic Information", *Journal of International Economics* 80(1), 72-88.

[175] Rogoff, K.S. and V. Stavrakeva (2008), "The Continuing Puzzle of Short Horizon Exchange Rate Forecasting," *NBER Working Papers* 14071

[176] Rossi, B. (2005), "Optimal Tests for Nested Model Selections With Underlying Parameter Instabilities," *Econometric Theory* 21(5), 962-990.

[177] Rossi, B. (2006), "Are Exchange Rates Really Random Walks? Some Evidence Robust to Parameter Instability," *Macroeconomic Dynamics* 10(1), 20-38.

[178] Rossi, B. (2011), "Comment to: Forecast Rationality Tests Based on Multi-Horizon Bounds," *Journal of Business and Economic Statistics*, forthcoming.

[179] Rossi, B. and T. Sekhposyan (2010), "Have Models' Forecasting Performance Changed Over Time, and When?," *International Journal of Forecasting* 26(4).

[180] Rossi, B. and T. Sekhposyan (2011a), "Understanding Models' Forecasting Performance," *Journal of Econometrics* 164, 158-172.

[181] Rossi, B. and T. Sekhposyan (2011b), "Forecast Optimality Tests in the Presence of Instabilities," *mimeo*, Duke University.

[182] Rossi, B. and T. Sekhposyan (2012a), "Tests of Specification and Distributional Change for Predictive Densities", *mimeo*.

[183] Rossi, B. and T. Sekhposyan (2012b), "Evaluating Predictive Densities for U.S. Output Growth and Inflation in a Large Macroeconomic Data Set", *mimeo*.

[184] Sarno, L. and G. Valente (2009), "Exchange Rates and Fundamentals: Footloose or Evolving Relationship?," *Journal of the European Economic Association* 7(4), 786-830.

[185] Schinasi, G. and P. Swamy (1989), "The Out-of-Sample Forecasting Performance of Exchange Rate Models when Coefficients Are Allowed to Change," *Journal of International Money and Finance* 8, 375-390.

[186] Schrimpf, A. and Q.W. Wang (2010), "A Reappraisal of the Leading Indicator Properties of the Yield Curve Under Structural Instability," *International Journal of Forecasting* 26(4), 836-857.

[187] Schwert, G.W. (2003), "Anomalies and Market Efficiency", in: Constantinides, G.M., M. Harris and R. Stulz (eds.), *Handbook of the Economics of Finance*, Elsevier.

[188] Smith, A. (2005), "Forecasting in the Presence of Level Shifts," *Journal of Forecasting* 24(8), 557-574.

[189] Smith, J. and K. Wallis (2009), "A Simple Explanation of the Forecast Combination Puzzle," *Oxford Bulletin of Economics and Statistics* 71, 331-355.

[190] Stock, J.H. (1994), "Unit Roots, Structural Breaks, and Trends", in: Engle, R. and D. McFadden (eds.), *Handbook of Econometrics* Vol. IV, Amsterdam: Elsevier, 2740-2843.

[191] Stock, J.H. and M.W. Watson (1996), "Evidence on Structural Stability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics* 14, 11-30.

[192] Stock, J.H. and M.W. Watson (1999), "A Dynamic Factor Model Framework for Forecast Combination," *Spanish Economic Review* 1, 91-121.

[193] Stock, J.H. and M.W. Watson (2001), "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series," in R.F. Engle and H. White (eds.), *Festschrift in Honour of Clive Granger*, Cambridge University Press, 1-44.

[194] Stock, J. H. and M. W. Watson (2002), "Has the Business Cycle Changed and Why?" in M. Gertler and K. Rogoff (eds.), *NBER Macroeconomics Annual* (Cambridge, MA: MIT Press).

[195] Stock, J.H. and M.W. Watson (2003), "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature* XLI, 788-829.

[196] Stock, J.H. and M.W. Watson (2004), "Combination Forecasts of Output Growth in a Seven Country Data Set," *Journal of Forecasting* 23, 405-430.

[197] Stock, J.H. and M.W. Watson (2006), "Forecasting With Many Predictors," in: G. Elliott, C. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting Vol. 1*, North Holland: Elsevier.

[198] Stock, J.H. and M.W. Watson (2007), "Why Has Inflation Become Harder to Forecast?," *Journal of Money, Credit and Banking* 39 (1), 3–34.

[199] Stock, J.H. and M.W. Watson (2008), "Forecasting in Dynamic Factor Models Subject to Structural Instability," in: Castle, J. and N. Shephard (eds.), *The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry*, Oxford: Oxford University Press.

[200] Stock, J.H. and M.W. Watson (2010), "Modeling Inflation After the Crisis," *mimeo.*

[201] Sullivan, R., A. Timmermann and H. White, H. (1999), "Data-snooping, Technical Trading Rule Performance, and the Bootstrap," *Journal of Finance* 54, 1647–1691.

[202] Swanson, N.R. (1998), "Money and Output Viewed Through a Rolling Window," *Journal of Monetary Economics* 41, 455-473.

[203] Swanson, N.R. and H. White (1995), "A Model Selection Approach to Assessing the Information in the Term Structure Using Linear Models and Artifical Neural Networks," *Journal of Business and Economic Statistics* 13, 265-275.

[204] Swanson, N.R. and H. White (1997), "A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks," *Review of Economics and Statistics* 79, 540-550.

[205] Terasvirta, Timo (2009), "Forecasting Economic Variables with Nonlinear Models," in: G. Elliott, C. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting Vol. 1*, North Holland: Elsevier.

[206] Timmermann, A. (2006), "Forecast Combinations," in: G. Elliott, C. Granger and A. Timmermann, *Handbook of Economic Forecasting Vol. 1*, North Holland: Elsevier.

[207] Timmermann, A. (2008), "Elusive Return Predictability," *International Journal of Forecasting* 24(1), 1-18.

[208] Wheelock, D.C. and M.E. Wohar (2009), "Can the Term Spread Predict Output Growth and Recessions? A Survey of the Literature," *Federal Reserve Bank of St. Louis Review* 91(5), 419-440.

[209] West, K.D. (1996), "Asymptotic Inference about Predictive Ability," *Econometrica*, 64, 1067-1084.

[210] West, K.D. (2006), "Forecast Evaluation," in: G. Elliott, C. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting Vol. 1*, North Holland: Elsevier.

[211] West, K.D., and M.W. McCracken (1998), "Regression-Based Tests of Predictive Ability," *International Economic Review* 39(4), 817-840.

[212] Winters, P.R. (1960), "Forecasting Sales by Exponentially Weighted Moving Averages," *Management Science* 6, 324-342.

[213] Wright, J.H. (2008), "Bayesian Model Averaging and Exchange Rate Forecasts," *Journal of Econometrics* 146(2), 329-341.

[214] Wright, J.H. (2009), "Forecasting US Inflation by Bayesian Model Averaging," *Journal of Forecasting* 28(2), 131-144.

[215] Wolff, C. (1987), "Time-Varying Parameters and the Out-of-Sample Forecasting Performance of Structural Exchange Rate Models," *Journal of Business and Economic Statistics* 5, 87-97.

6 Appendix 1. Critical Value Tables

Table A.1. Critical values for Rossi's (2005) Robust Granger-causality Test

p	QLR_T^*	$Exp - Wald_T^*$	$Mean - Wald_T^*$
1	9.826	3.134	5.364
2	14.225	5.015	8.743
3	17.640	6.738	11.920
4	21.055	8.191	14.362
5	24.550	9.824	17.523
6	27.377	11.203	19.877
7	30.414	12.630	22.389
8	33.717	14.225	25.397
9	36.552	15.537	27.844
10	39.020	16.761	30.039

Notes. The table reports asymptotic critical values of Rossi's (2005) QLR_T^* , $Exp{-}Wald_T^*$ and $Mean - Wald_T^*$ test statistics for tests of nominal size equal to 5%. See Section 2.2.1 for details.

Table A.2. Critical values for Giacomini and Rossi's (2010a) Fluctuation Test (k_α^{GR})

δ	Two-sided Test	One-sided Test
.1	3.393	3.176
.2	3.179	2.938
.3	3.012	2.770
.4	2.890	2.624
.5	2.779	2.475
.6	2.634	2.352
.7	2.560	2.248
.8	2.433	2.080
.9	2.248	1.975

Note. The table reports the critical values (k_α^{GR}) of the Fluctuation test in Proposition 1 in Giacomini and Rossi (2010a). The nominal size of the test is 5%, $\delta = m/P$, where m is the size of the rolling window used for out-of-sample smoothing and P is the out-of-sample size. See Section 2.2.2 for details.

Table A.3. Critical Values for Rossi and Sekhposyan (2011b) Fluctuation Optimality Test

p	μ : 0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
1	12.08	10.59	9.65	8.75	7.75	6.96	6.49	6.12	5.37
2	23.93	21.01	18.81	16.90	16.45	14.51	13.29	11.95	10.65

Note. The table reports critical values for the Fluctuation optimality test in Rossi and Sekhposyan (2011b). The nominal size of the test is equal to 5%, $\mu = m/P$, where m is the size of the rolling window used for out-of-sample smoothing, P is the out-of-sample size, and p is the number of restrictions. See Section 2.2.3 for details.

Table A.4. Critical Values for Inoue and Rossi's (2010) Test Statistics

Test Statistics:	Critical Values:				
A. Forecast Comparison Tests					
Non-Nested Models					
$\mathcal{R}_T$	2.7231				
$\mathcal{A}_T$	1.7292				
p :	1	2	3	4	5
Nested Models					
$\mathcal{R}_T^{\mathcal{E}}$ (rolling window)	5.1436	7.1284	8.4892	9.7745	10.823
$\mathcal{A}_T^{\mathcal{E}}$ (rolling window)	1.7635	2.4879	2.9559	3.39	3.7427
$\mathcal{R}_T^{\mathcal{E}}$ (recursive window)	3.0078	4.2555	5.0577	6.1064	6.3340
$\mathcal{A}_T^{\mathcal{E}}$ (recursive window)	1.4955	2.1339	2.3919	2.9668	2.9717
B. Forecast Optimality					
$\mathcal{R}_T^{\mathcal{W}}$ (forecast optimality)	1.3342	2.4634	3.5569	4.6451	5.7182
$\mathcal{A}_T^{\mathcal{W}}$ (forecast optimality)	1.1424	2.2009	3.245	4.2848	5.3166

Note. The table reports critical values for the Inoue and Rossi's (2010) test statistics. The nominal size of the test is 5%, $\underline{\zeta} = 0.15$, p is either the number of regressors in the large model in excess of those in the small model (for the nested models' forecast comparison tests) or the number of regressors used to check forecast optimality (for the forecast optimality tests). See Section 2.3.1 for details.

Table A.5. Critical Values for Rossi and Sekhposyan's (2011a) $\Gamma_P^{(A)}$ Test

δ :	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
$k_{\alpha,\delta}^{RS}$:	10.496	6.609	4.842	3.738	2.984	2.412	1.900	1.446	0.952

Note. The table reports critical values $k_{\alpha,\delta}^{RS}$ for the test statistic $\Gamma_P^{(A)}$. The nominal significance level is equal to 5%. See Section 3.4 for details.

7 Tables

Table 1, Panel A (Inflation). Granger-causality and Rossi's (2005) p-values

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
rtbill	lev	0.01	0.00	0.00	0.01	0.31	0.00	0.00
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rbnds	lev	- -	- -	- -	- -	- -	0.00	0.00
		- -	- -	- -	- -	- -	(0.00)	(0.00)
rbndm	lev	0.01	- -	- -	0.00	- -	- -	0.02
		(0.00)	- -	- -	(0.00)	- -	- -	(0.00)
rbndl	lev	0.02	0.00	0.01	0.00	0.31	0.00	0.01
		(0.00)	(0.00)	(0.00)	(0.00)	(0.04)	(0.00)	(0.00)
rovnght	1d	0.12	0.00	0.00	0.16	0.01	0.82	0.02
		(0.10)	(0.06)	(0.00)	(0.46)	(0.00)	(0.27)	(0.04)
rtbill	1d	0.11	0.00	0.00	0.46	0.01	0.00	0.00
		(0.30)	(0.03)	(0.00)	(0.77)	(0.00)	(0.00	(0.00)
rbnds	1d	- -	- -	- -	- -	- -	0.00	0.01
		- -	- -	- -	- -	- -	(0.00)	(0.07)
rbndm	1d	0.10	- -	- -	0.80	- -	- -	0.07
		(0.00)	- -	- -	(0.01	- -	- -	(0.43)
rbndl	1d	0.10	0.01	0.01	0.65	0.06	0.00	0.16
		(0.06)	(0.14)	(0.05)	(0.00	(0.30	(0.00	(0.34)
rrovnght	lev	0.11	0.01	0.01	0.09	0.00	0.62	0.00
		(0.00)	(0.00)	(0.00)	(0.06)	(0.00)	(0.75)	(0.00)
rrtbill	lev	0.02	0.00	0.01	0.83	0.00	0.00	0.00
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rrbnds	lev	- -	- -	- -	- -	- -	0.00	0.00
		- -	- -	- -	- -	- -	(0.00)	(0.00)
rrbndm	lev	0.00	- -	- -	0.85	- -	- -	0.26
		(0.00)	- -	- -	(0.06)	- -	- -	(0.00)
rrbndl	lev	0.00	0.04	0.00	0.99	0.00	0.00	0.22
		(0.00)	(0.00)	(0.00	(0.06)	(0.00)	(0.00)	(0.00)
rrovnght	1d	0.00	0.00	0.00	0.16	0.00	0.82	0.00
		(0.00)	(0.06)	(0.02)	(0.46)	(0.00)	(0.69)	(0.00)
rrtbill	1d	0.00	0.00	0.00	0.06	0.00	0.00	0.00
		(0.00)	(0.04)	(0.00)	(0.62)	(0.00)	(0.00)	(0.00)
rrbnds	1d	- -	- -	- -	- -	- -	0.00	0.00
		- -	- -	- -	- -	- -	(0.00)	(0.00)
rrbndm	1d	0.00	- -	- -	0.80	- -	- -	0.07
		(0.00)	- -	- -	(0.24)	- -	- -	(0.15)
rrbndl	1d	0.00	0.01	0.00	0.65	0.00	0.00	0.16
		(0.00)	(0.26)	(0.00)	(0.01)	(0.00)	(0.00)	(0.10)
rspread	lev	0.93	0.03	0.59	0.00	0.00	0.24	0.00
		(0.05)	(0.00)	(0.81)	(0.00)	(0.00)	(0.00)	(0.00)
exrate	ln1d	0.96	0.02	0.27	0.79	0.00	0.23	0.07
		(0.68)	(0.00)	(0.61)	(0.63)	(0.00)	(0.38)	(0.02)
rexrate	ln1d	0.32	0.04	0.14	0.28	0.00	0.68	0.07
		(0.22)	(0.00)	(0.40)	(0.57)	(0.00)	(0.58)	(0.02)
stockp	ln1d	0.30	0.61	0.74	0.86	0.03	0.01	0.42
		(0.00)	(0.45)	(0.00)	(0.20)	(0.00)	(0.00)	(0.47)
rstockp	ln1d	0.14	0.84	0.58	0.89	0.02	0.06	0.22
		(0.00)	(0.48)	(0.00)	(0.56)	(0.00)	(0.00)	(0.11)
rgdp	ln1d	0.00	0.02	0.00	0.38	0.05	0.21	0.01
		(0.00)	(0.01)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)
rgdp	gap	0.00	0.00	0.00	0.21	0.25	0.02	0.01
		(0.00)	(0.00)	(0.00)	(0.14)	(0.32)	(0.00)	(0.00)

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
ip	ln1d	0.00	0.26	0.01	0.04	0.02	0.14	0.01
		(0.00)	(0.24)	(0.00)	(0.06)	(0.02)	(0.15)	(0.00)
ip	gap	0.00	0.09	0.04	0.02	0.07	0.03	0.01
		(0.00)	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
capu	lev	0.06	0.41	0.00	0.01	0.07	0.78	0.00
		(0.35)	(0.02)	(0.00)	(0.00)	(0.00)	(0.09)	(0.00)
emp	ln1d	0.00	0.07	0.00	0.17	0.18	0.00	0.08
		(0.00)	(0.52)	(0.00)	(0.58)	(0.66)	(0.00)	(0.00)
emp	gap	0.00	0.07	0.00	0.05	0.00	0.00	0.01
		(0.00)	(0.00)	(0.00)	(0.48)	(0.00)	(0.00)	(0.00)
unemp	lev	0.00	0.00	0.00	0.05	0.06	0.00	0.03
		(0.00)	(0.00)	(0.00)	(0.17)	(0.00)	(0.00)	(0.00)
unemp	1d	0.00	0.02	0.01	0.14	0.02	0.00	0.01
		(0.00)	(0.00)	(0.00)	(0.20)	(0.00)	(0.00)	(0.00)
unemp	gap	0.00	0.05	0.00	0.06	0.00	0.00	0.01
		(0.00)	(0.04)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
pgdp	ln1d	0.49	0.33	0.38	0.15	0.42	0.19	0.06
		(0.00)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)
pgdp	ln2d	0.06	0.96	0.62	0.43	0.85	0.50	0.96
		(0.37)	(1.00)	(1.00)	(0.42)	(0.65)	(0.81)	(0.00)
ppi	ln1d	0.14	0.00	0.57	0.00	0.00	0.00	0.00
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00	(0.00)	(0.00)
ppi	ln2d	0.00	0.01	0.03	0.00	0.00	0.00	0.01
		(0.00)	(0.00)	(0.33)	(0.00)	(0.00)	(0.00)	(0.07)
earn	ln1d	0.85	0.41	0.03	0.96	0.12	0.00	0.97
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)
earn	ln2d	0.19	0.06	0.31	0.08	0.01	0.00	0.12
		(0.40)	(0.00)	(0.82)	(0.02)	(0.00)	(0.00)	(0.31)
mon0	ln1d	0.08	- -	0.16	- -	- -	0.04	0.01
		(0.00)	- -	(0.31)	- -	- -	(0.00)	(0.00)
mon0	ln2d	0.75	- -	0.59	- -	- -	0.00	0.91
		(0.54)	- -	(0.89)	- -	- -	(0.11)	(0.16)
mon1	ln1d	0.30	0.93	0.02	0.47	0.10	0.16	0.03
		(0.00)	(0.00)	(0.00)	(0.00)	(0.42)	(0.05)	(0.00)
mon1	ln2d	0.65	0.61	0.15	0.66	0.69	0.75	0.72
		(0.76)	(0.00)	(0.47)	(0.73)	(0.86)	(0.10)	(0.01)
mon2	ln1d	0.00	0.86	0.00	0.04	0.14	0.93	0.13
		(0.00)	(0.00)	(0.01)	(0.00)	(0.03)	(0.82)	(0.00)
mon2	ln2d	0.00	0.91	0.74	0.53	0.46	0.18	0.16
		(0.00)	(0.32)	(1.00)	(0.73)	(0.81)	(1.00)	(0.05)
mon3	ln1d	0.00	0.62	0.74	0.00	0.21	0.06	0.29
		(0.02)	(0.00)	(0.32)	(0.00)	(0.00)	(0.00)	(0.04)
mon3	ln2d	0.38	0.73	0.94	0.62	0.13	0.12	0.54
		(0.47)	(0.00)	(1.00)	(0.90)	(0.00)	(0.06)	(0.01)
rmon0	ln1d	0.00	- -	0.07	- -	- -	0.00	0.00
		(0.00)	- -	(0.13)	- -	- -	(0.00)	(0.00)
rmon1	ln1d	0.01	0.27	0.00	0.01	0.01	0.05	0.00
		(0.00)	(0.68)	(0.00)	(0.00)	(0.08)	(0.03)	(0.00)
rmon2	ln1d	0.00	0.18	0.00	0.03	0.00	0.02	0.00
		(0.00)	(0.32)	(0.00)	(0.06)	(0.00)	(0.11)	(0.00)
rmon3	ln1d	0.00	0.03	0.03	0.21	0.01	0.00	0.00
		(0.00)	(0.00)	(0.00)	(1.00)	(0.00)	(0.00)	(0.00)

Table 1, Panel B. (Output) Granger Causality and Rossi's (2005) p-values

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
rtbill	lev	0.00	0.98	0.02	0.42	0.00	0.02	0.03
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rbnds	lev	- -	- -	- -	- -	- -	0.19	0.06
		- -	- -	- -	- -	- -	(0.00)	(0.00)
rbndm	lev	0.04	- -	- -	0.21	- -	- -	0.30
		(0.00)	- -	- -	(0.00)	- -	- -	(0.00)
rbndl	lev	0.10	0.90	0.76	0.15	0.00	0.12	0.38
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rovnght	1d	0.11	0.00	0.86	0.00	0.07	0.57	0.00
		(0.00)	(0.00)	(0.05)	(0.00)	(0.00)	(0.29)	(0.00)
rtbill	1d	0.00	0.12	0.46	0.74	0.12	0.00	0.01
		(0.00)	(0.00)	(0.13)	(1.00)	(0.00)	(0.00)	(0.00)
rbnds	1d	- -	- -	- -	- -	- -	0.00	0.02
		- -	- -	- -	- -	- -	(0.00)	(0.00)
rbndm	1d	0.01	- -	- -	0.91	- -	- -	0.03
		(0.00)	- -	- -	(0.00)	- -	- -	(0.00)
rbndl	1d	0.03	0.06	0.30	0.81	0.47	0.00	0.02
		(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
rrovnght	lev	0.11	0.13	0.03	0.12	0.26	0.22	0.62
		(0.00)	(0.02)	(0.00)	(0.00)	(0.04)	(0.48)	(0.61)
rrtbill	lev	0.05	0.96	0.04	0.38	0.44	0.14	0.09
		(0.04)	(0.00)	(0.00)	(0.00)	(0.10)	(0.16)	(0.13)
rrbnds	lev	- -	- -	- -	- -	- -	0.01	0.07
		- -	- -	- -	- -	- -	(0.02)	(0.21)
rrbndm	lev	0.54	- -	- -	0.26	- -	- -	0.00
		(0.07)	- -	- -	(0.00)	- -	- -	(0.00)
rrbndl	lev	0.84	0.98	0.01	0.30	0.24	0.02	0.00
		(0.05)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rrovnght	1d	0.94	0.48	0.00	0.42	0.81	0.32	0.80
		(0.84)	(1.00)	(0.08)	(0.27)	(1.00)	(0.80)	(0.34)
rrtbill	1d	0.72	0.98	0.00	0.26	0.55	0.29	0.45
		(0.74)	(1.00)	(0.00)	(0.77)	(0.83)	(0.25)	(1.00)
rrbnds	1d	- -	- -	- -	- -	- -	0.31	0.65
		- -	- -	- -	- -	- -	(0.22	(1.00)
rrbndm	1d	0.65	- -	- -	0.45	- -	- -	0.60
		(0.48)	- -	- -	(0.29)	- -	- -	(1.00)
rrbndl	1d	0.52	0.95	0.01	0.48	0.55	0.15	0.56
		(0.42)	(1.00)	(0.15)	(0.31)	(1.00)	(0.36)	(0.89)
rspread	lev	0.00	0.00	0.00	0.00	0.21	0.58	0.00
		(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)
exrate	ln1d	0.81	0.79	0.16	0.14	0.66	0.91	0.85
		(0.58)	(1.00)	(0.88)	(0.00)	(0.00)	(0.05)	(0.41)
rexrate	ln1d	0.94	0.78	0.15	0.14	0.74	0.82	0.85
		(0.79)	(1.00)	(0.87)	(0.00)	(0.08)	(0.08)	(0.41)
stockp	ln1d	0.00	0.04	0.03	0.01	0.00	0.00	0.00
		(0.00)	(0.00)	(0.04)	(0.02)	(0.00)	(0.00)	(0.00)
rstockp	ln1d	0.00	0.04	0.01	0.01	0.00	0.00	0.00
		(0.00)	(0.46)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
ip	ln1d	0.00	0.54	0.00	0.95	0.68	0.12	0.03
		(0.00)	(1.00)	(0.00)	(0.31)	(0.00)	(0.04)	(0.18)
ip	gap	0.20	0.03	0.06	0.29	0.04	0.08	0.92
		(0.05)	(0.00)	(0.12)	(0.27)	(0.00)	(0.02)	(0.71)
capu	lev	0.07	0.01	0.06	0.12	0.00	0.05	0.00
		(0.00)	(0.18)	(0.00)	(0.00)	(0.00)	(0.15)	(0.00)
emp	ln1d	0.17	0.90	0.51	0.00	0.60	0.00	0.52
		(0.00)	(0.00)	(0.76)	(0.00)	(0.00)	(0.00)	(0.64)
emp	gap	0.35	0.90	0.71	0.00	0.29	0.00	0.60
		(0.05)	(0.03)	(0.20)	(0.00)	(0.00)	(0.00)	(0.31)
unemp	lev	0.95	0.07	0.02	0.12	0.00	0.00	0.00
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
unemp	1d	0.11	0.03	0.17	0.53	0.04	0.00	0.45
		(0.55)	(0.00)	(0.38)	(0.12)	(0.00)	(0.00)	(0.68)
unemp	gap	0.05	0.00	0.01	0.63	0.01	0.00	0.00
		(0.26)	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)
pgdp	ln1d	0.02	0.36	0.30	0.00	0.53	0.01	0.02
		(0.00)	(0.00)	(0.02)	(0.00)	(0.37)	(0.00)	(0.00)
pgdp	ln2d	0.01	0.69	0.17	0.67	1.00	0.83	0.56
		(0.37)	(1.00)	(1.00)	(1.00)	(0.84)	(0.74)	(1.00)
cpi	ln1d	0.05	0.95	0.00	0.00	0.36	0.00	0.00
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
cpi	ln2d	0.83	0.06	0.00	0.47	0.46	0.29	0.02
		(0.56)	(0.00)	(0.10)	(0.38)	(0.73)	(0.52)	(0.00)
ppi	ln1d	0.18	0.00	0.00	0.25	0.11	0.00	0.00
		(0.07)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ppi	ln2d	0.31	0.42	0.56	0.02	0.70	0.01	0.77
		(0.89)	(0.63)	(0.86)	(0.39)	(0.00)	(0.00)	(1.00)
earn	ln1d	0.95	0.35	0.14	0.00	0.35	0.14	0.22
		(0.05)	(0.00)	(0.01)	(0.00)	(0.00)	(0.02)	(0.01)
earn	ln2d	0.88	0.76	0.93	0.26	0.47	0.45	0.88
		(0.20)	(1.00)	(0.70)	(1.00)	(0.00)	(0.79)	(1.00)
mon0	ln1d	0.02	- -	0.79	- -	- -	0.57	0.38
		(0.00)	- -	(0.31)	- -	- -	(0.24)	(0.00)
mon0	ln2d	0.52	- -	0.43	- -	- -	0.36	0.03
		(0.90)	- -	(1.00)	- -	- -	(1.00)	(0.18)
mon1	ln1d	0.01	0.92	0.01	0.14	0.01	0.13	0.44
		(0.00)	(1.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
mon1	ln2d	0.09	0.77	0.69	0.91	0.49	0.69	0.71
		(0.48)	(0.83)	(0.63)	(1.00)	(0.80)	(1.00)	(1.00)
mon2	ln1d	0.19	0.77	0.03	0.10	0.00	0.89	0.00
		(0.00)	(1.00)	(0.00)	(0.00)	(0.00)	(0.27)	(0.00)
mon2	ln2d	0.30	0.88	0.38	0.52	0.00	0.52	0.42
		(0.50)	(0.80)	(0.82)	(1.00)	(0.00)	(1.00)	(0.83)
mon3	ln1d	0.34	0.90	0.32	0.95	0.00	0.03	0.13
		(0.03)	(0.79)	(0.01)	(0.00)	(0.00)	(0.28)	(0.00)
mon3	ln2d	0.19	0.98	0.54	0.69	0.25	0.62	0.48
		(1.00)	(1.00)	(1.00)	(1.00)	(0.00)	(1.00)	(1.00)
rmon0	ln1d	0.01	- -	0.74	- -	- -	0.06	0.22
		(0.00)	- -	(0.67)	- -	- -	(0.04)	(0.00)
rmon1	ln1d	0.00	0.56	0.00	0.53	0.02	0.27	0.04
		(0.00)	(0.80)	(0.00)	(1.00)	(0.00)	(0.00)	(0.00)
rmon2	ln1d	0.87	0.23	0.03	0.60	0.00	0.02	0.00
		(0.29)	(0.00)	(0.00)	(1.00)	(0.00)	(0.06)	(0.00)
rmon3	ln1d	0.98	0.42	0.19	0.21	0.13	0.57	0.00
		(0.17)	(0.00)	(0.00)	(0.17)	(0.02)	(1.00)	(0.00)

Table 2, Panel A (Inflation). Relative MSFE and p-values

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
AR rmse	ln2d	1.79	1.68	1.47	3.05	3.15	3.61	2.04
rtbill	lev	1.18	1.43	0.99	1.04	1.27	1.49	0.95
		(0.88)	(0.93)	(0.45)	(0.60)	(0.72)	(0.93)	(0.29)
rbnds	lev	- -	- -	- -	- -	- -	1.25	1.06
		(- -)	(- -)	(- -)	(- -)	(- -)	(0.99)	(0.73)
rbndm	lev	1.18	- -	- -	1.58	- -	- -	1.12
		(0.90)	(- -)	(- -)	(0.96)	(- -)	(- -)	(0.87)
rbndl	lev	1.23	1.37	1.00	1.66	2.46	1.19	1.11
		(0.92)	(0.91)	(0.48)	(0.98)	(0.96)	(0.88)	(0.87)
rovnght	1d	1.12	1.10	0.96	1.73	1.08	1.05	1.02
		(0.88)	(0.97)	(0.25)	(0.99)	(0.78)	(0.72)	(0.55)
rtbill	1d	1.05	1.12	0.96	0.89	1.29	1.17	1.05
		(0.82)	(0.91)	(0.28)	(0.11)	(0.94)	(0.84)	(0.65)
rbnds	1d	- -	- -	- -	- -	- -	1.02	1.05
		(- -)	(- -)	(- -)	(- -)	(- -)	(0.55)	(0.75)
rbndm	1d	1.03	- -	- -	1.18	- -	- -	1.05
		(0.76)	(- -)	(- -)	(0.81)	(- -)	(- -)	(0.83)
rbndl	1d	1.07	1.18	1.01	1.26	3.30	0.97	1.05
		(0.81)	(0.98)	(0.62)	(0.87)	(0.97)	(0.40)	(0.80)
rrovnght	lev	1.19	1.82	1.08	1.80	1.44	1.48	1.33
		(0.92)	(1.00)	(0.80)	(0.99)	(0.98)	(0.99)	(0.92)
rrtbill	lev	1.46	1.70	1.13	1.61	1.62	1.27	1.49
		(0.96)	(1.00)	(0.90)	(0.92)	(0.89)	(0.88)	(0.95)
rrbnds	lev	- -	- -	- -	- -	- -	1.23	1.42
		(- -)	(- -)	(- -)	(- -)	(- -)	(0.76)	(0.91)
rrbndm	lev	1.48	- -	- -	1.57	- -	- -	1.49
		(0.98)	(- -)	(- -)	(0.99)	(- -)	(- -)	(0.92)
rrbndl	lev	1.41	1.70	0.90	1.77	2.05	1.03	1.36
		(0.97)	(1.00)	(0.21)	(0.98)	(0.97)	(0.56)	(0.89)
rrovnght	1d	1.07	1.15	0.98	1.03	0.88	1.12	1.07
		(0.74)	(0.97)	(0.39)	(0.64)	(0.09)	(0.95)	(0.71)
rrtbill	1d	1.00	1.00	0.98	1.01	1.02	1.00	1.13
		(0.51)	(0.51)	(0.39)	(0.54)	(0.55)	(0.51)	(0.83)
rrbnds	1d	- -	- -	- -	- -	- -	1.01	1.02
		(- -)	(- -)	(- -)	(- -)	(- -)	(0.54)	(0.60)
rrbndm	1d	0.84	- -	- -	1.20	- -	- -	0.95
		(0.00)	(- -)	(- -)	(0.84)	(- -)	(- -)	(0.19)
rrbndl	1d	0.94	0.99	0.95	1.25	2.20	0.92	0.95
		(0.20)	(0.40)	(0.19)	(0.90)	(0.87)	(0.26)	(0.17)
rspread	lev	1.10	1.32	1.10	1.52	0.93	1.27	1.01
		(0.98)	(0.97)	(0.99)	(0.98)	(0.37)	(0.88)	(0.54)
exrate	ln1d	1.09	- -	- -	- -	- -	- -	- -
		(0.98)	(- -)	(- -)	(- -)	(- -)	(- -)	(- -)
rexrate	ln1d	1.11	- -	- -	- -	- -	- -	- -
		(0.98)	(- -)	(- -)	(- -)	(- -)	(- -)	(- -)
stockp	ln1d	1.10	1.01	1.09	1.84	1.06	0.89	1.04
		(0.95)	(0.61)	(0.93)	(0.92)	(0.88)	(0.18)	(0.72)
rstockp	ln1d	1.10	1.01	1.08	1.78	1.00	0.94	1.04
		(0.93)	(0.65)	(0.94)	(0.91)	(0.50)	(0.28)	(0.69)
rgdp	ln1d	0.93	1.06	0.88	1.10	1.00	1.11	0.86
		(0.28)	(0.77)	(0.08)	(0.99)	(0.52)	(0.79)	(0.12)
rgdp	gap	0.91	1.34	0.87	1.27	0.98	0.96	0.90
		(0.28)	(0.98)	(0.10)	(0.97)	(0.17)	(0.38)	(0.26)

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
ip	ln1d	0.98	1.11	1.01	1.02	0.86	1.07	0.90
		(0.45)	(0.97)	(0.58)	(0.58)	(0.12)	(0.82)	(0.18)
ip	gap	1.02	1.11	1.10	1.08	1.05	1.11	0.91
		(0.55)	(0.83)	(0.89)	(0.80)	(0.66)	(0.81)	(0.23)
capu	lev	1.38	2.57	0.96	0.97	1.25	0.88	0.92
		(0.92)	(0.99)	(0.36)	(0.43)	(0.93)	(0.23)	(0.33)
emp	ln1d	1.01	1.76	1.16	1.11	1.01	1.08	0.83
		(0.53)	(0.98)	(0.79)	(0.96)	(0.54)	(0.64)	(0.07)
emp	gap	0.82	2.13	0.92	1.20	1.10	1.69	0.88
		(0.12)	(0.99)	(0.25)	(0.98)	(0.88)	(0.96)	(0.19)
unemp	lev	1.08	1.26	0.99	1.50	1.94	1.22	1.09
		(0.69)	(0.93)	(0.47)	(0.93)	(1.00)	(0.89)	(0.75)
unemp	1d	0.97	1.02	0.93	0.93	1.02	0.86	0.93
		(0.41)	(0.62)	(0.24)	(0.19)	(0.60)	(0.19)	(0.31)
unemp	gap	0.87	1.47	0.93	1.11	1.05	1.25	0.87
		(0.22)	(0.98)	(0.18)	(0.98)	(0.76)	(0.84)	(0.16)
pgdp	ln1d	1.08	1.32	1.13	1.58	0.99	1.09	1.16
		(0.96)	(0.96)	(0.92)	(0.99)	(0.31)	(0.81)	(0.93)
pgdp	ln2d	1.04	1.17	1.07	1.12	1.00	1.05	0.99
		(0.91)	(0.97)	(0.98)	(0.88)	(0.32)	(0.91)	(0.30)
cpi	ln1d	- -	- -	- -	- -	- -	- -	- -
		(- -)	(- -)	(- -)	(- -)	(- -)	(- -)	(- -)
cpi	ln2d	- -	- -	- -	- -	- -	- -	- -
		(- -)	(- -)	(- -)	(- -)	(- -)	(- -)	(- -)
ppi	ln1d	1.24	- -	1.12	0.83	1.28	1.12	1.22
		(0.99)	(- -)	(0.86)	(0.13)	(0.90)	(0.73)	(0.96)
ppi	ln2d	0.92	- -	1.03	0.98	0.85	0.86	0.99
		(0.12)	(- -)	(0.64)	(0.40)	(0.18)	(0.26)	(0.41)
earn	ln1d	1.18	1.32	1.03	1.38	0.96	1.15	1.15
		(0.98)	(0.97)	(0.87)	(1.00)	(0.21)	(0.85)	(1.00)
earn	ln2d	1.10	1.04	1.00	0.99	0.96	1.06	1.03
		(0.94)	(0.97)	(0.61)	(0.47)	(0.10)	(0.65)	(0.97)
mon0	ln1d	1.20	- -	2.08	- -	- -	1.02	1.08
		(0.98)	(- -)	(0.87)	(- -)	(- -)	(0.59)	(0.90)
mon0	ln2d	1.07	- -	1.08	- -	- -	1.15	1.02
		(0.96)	(- -)	(0.89)	(- -)	(- -)	(0.91)	(0.76)
mon1	ln1d	1.23	3.45	1.10	1.31	1.09	0.98	1.03
		(0.97)	(0.93)	(0.88)	(0.99)	(0.87)	(0.37)	(0.67)
mon1	ln2d	1.06	1.03	1.06	1.11	0.97	1.22	0.96
		(0.97)	(0.73)	(0.92)	(0.93)	(0.19)	(0.84)	(0.16)
mon2	ln1d	1.17	2.29	1.05	1.91	1.15	1.16	1.06
		(0.91)	(0.93)	(0.77)	(0.94)	(0.84)	(0.97)	(0.90)
mon2	ln2d	1.13	1.11	1.06	1.13	1.09	0.87	1.04
		(0.92)	(0.90)	(0.92)	(0.91)	(0.99)	(0.17)	(0.79)
mon3	ln1d	1.09	0.93	1.01	1.53	1.37	1.10	1.20
		(0.95)	(0.28)	(0.60)	(0.96)	(0.98)	(0.97)	(0.96)
mon3	ln2d	1.03	1.13	1.02	1.05	1.10	1.02	1.00
		(0.69)	(1.00)	(0.97)	(0.83)	(0.91)	(0.83)	(0.38)
rmon0	ln1d	0.99	- -	2.21	- -	- -	0.80	1.09
		(0.47)	(- -)	(0.88)	(- -)	(- -)	(0.11)	(0.70)
rmon1	ln1d	1.09	1.24	1.04	1.14	1.01	1.15	0.93
		(0.93)	(0.98)	(0.69)	(0.97)	(0.55)	(0.90)	(0.29)
rmon2	ln1d	0.85	1.48	1.05	1.17	0.95	1.17	0.99
		(0.23)	(0.99)	(0.62)	(0.87)	(0.36)	(0.89)	(0.46)
rmon3	ln1d	0.77	5.13	0.89	1.20	1.11	1.10	1.01
		(0.11)	(0.96)	(0.02)	(0.96)	(0.75)	(0.97)	(0.53)

Table 2, Panel B (Output). Relative MSFE and p-values

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
AR rmse	ln1d	2.34	1.68	3.38	5.01	3.35	2.51	2.46
rtbill	lev	1.05 (0.63)	1.25 (0.99)	0.92 (0.13)	1.48 (0.99)	1.61 (0.99)	1.03 (0.59)	1.07 (0.62)
rbnds	lev	- - (- -)	- - (- -)	- - (- -)	- - (- -)	- - (- -)	0.96 (0.38)	1.07 (0.66)
rbndm	lev	1.10 (0.72)	- - (- -)	- - (- -)	1.30 (0.98)	- - (- -)	- - (- -)	1.22 (0.91)
rbndl	lev	1.21 (0.87)	1.24 (0.99)	0.99 (0.48)	1.21 (1.00)	1.14 (0.92)	1.02 (0.61)	1.24 (0.95)
rovnght	1d	1.03 (0.76)	1.16 (0.99)	1.10 (0.99)	1.07 (0.96)	1.09 (0.90)	1.01 (0.75)	1.04 (0.61)
rtbill	1d	0.99 (0.47)	1.21 (0.98)	1.07 (0.98)	1.02 (0.78)	1.13 (0.92)	1.07 (0.76)	1.13 (0.75)
rbnds	1d	- - (- -)	- - (- -)	- - (- -)	- - (- -)	- - (- -)	0.98 (0.40)	1.05 (0.65)
rbndm	1d	1.07 (0.93)	- - (- -)	- - (- -)	1.14 (0.98)	- - (- -)	- - (- -)	1.06 (0.70)
rbndl	1d	1.05 (0.80)	1.20 (0.98)	1.11 (0.91)	1.10 (0.95)	1.04 (0.84)	0.99 (0.45)	1.03 (0.63)
rrovnght	lev	1.09 (0.79)	1.16 (0.95)	0.99 (0.43)	0.88 (0.17)	1.12 (0.95)	1.24 (0.97)	1.25 (0.99)
rrtbill	lev	1.13 (0.91)	1.63 (0.98)	1.00 (0.50)	1.41 (0.98)	0.96 (0.34)	1.23 (0.98)	1.42 (0.98)
rrbnds	lev	- - (- -)	- - (- -)	- - (- -)	- - (- -)	- - (- -)	1.06 (0.75)	1.49 (0.96)
rrbndm	lev	1.21 (0.99)	- - (- -)	- - (- -)	1.23 (0.83)	- - (- -)	- - (- -)	1.59 (0.96)
rrbndl	lev	1.18 (0.97)	1.72 (0.98)	1.04 (0.93)	1.22 (0.78)	1.11 (0.87)	1.14 (0.86)	1.55 (0.94)
rrovnght	1d	1.03 (0.78)	1.02 (0.97)	1.00 (0.51)	1.02 (0.98)	1.04 (0.88)	1.08 (0.92)	1.07 (0.97)
rrtbill	1d	1.06 (0.97)	1.05 (0.99)	0.97 (0.07)	1.05 (0.94)	1.25 (0.94)	1.16 (0.97)	1.36 (0.93)
rrbnds	1d	- - (- -)	- - (- -)	- - (- -)	- - (- -)	- - (- -)	1.04 (0.81)	1.38 (0.97)
rrbndm	1d	1.02 (0.87)	- - (- -)	- - (- -)	1.70 (0.89)	- - (- -)	- - (- -)	1.36 (0.97)
rrbndl	1d	1.02 (0.89)	1.15 (0.96)	0.96 (0.12)	1.81 (0.89)	1.02 (0.72)	1.03 (0.71)	1.34 (0.97)
rspread	lev	1.04 (0.57)	1.02 (0.60)	1.22 (0.84)	0.95 (0.31)	1.10 (0.98)	1.26 (0.99)	0.71 (0.02)
exrate	ln1d	1.07 (0.92)	- - (- -)	- - (- -)	- - (- -)	- - (- -)	- - (- -)	- - (- -)
rexrate	ln1d	1.08 (0.94)	- - (- -)	- - (- -)	- - (- -)	- - (- -)	- - (- -)	- - (- -)
stockp	ln1d	0.98 (0.37)	0.82 (0.02)	1.01 (0.55)	1.36 (0.97)	0.87 (0.01)	1.05 (0.68)	0.89 (0.06)
rstockp	ln1d	0.94 (0.24)	0.83 (0.04)	1.02 (0.64)	1.24 (0.98)	0.88 (0.03)	1.01 (0.53)	0.87 (0.06)
rgdp	ln1d	- - (- -)	- - (- -)	- - (- -)	- - (- -)	- - (- -)	- - (- -)	- - (- -)
rgdp	gap	- - (- -)	- - (- -)	- - (- -)	- - (- -)	- - (- -)	- - (- -)	- - (- -)

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
ip	ln1d	0.96	0.94	1.10	1.13	0.94	1.10	1.00
		(0.12)	(0.16)	(0.86)	(0.99)	(0.13)	(0.91)	(0.48)
ip	gap	1.12	0.97	1.45	1.26	1.10	1.10	1.22
		(0.98)	(0.31)	(0.94)	(1.00)	(0.92)	(0.91)	(1.00)
capu	lev	2.20	1.29	0.89	1.33	1.11	1.04	0.99
		(0.94)	(0.86)	(0.22)	(1.00)	(0.86)	(0.64)	(0.47)
emp	ln1d	1.03	1.01	1.03	1.03	1.03	1.12	1.03
		(0.78)	(0.54)	(0.61)	(0.93)	(0.87)	(0.92)	(0.82)
emp	gap	1.21	1.11	1.04	1.27	1.35	1.41	1.38
		(0.99)	(0.78)	(0.75)	(0.97)	(1.00)	(0.96)	(0.99)
unemp	lev	1.07	0.92	1.05	1.04	1.14	1.24	1.01
		(0.72)	(0.30)	(0.63)	(0.59)	(0.89)	(0.87)	(0.52)
unemp	1d	1.10	1.02	0.81	1.04	1.06	1.02	0.99
		(0.95)	(0.62)	(0.14)	(0.83)	(0.97)	(0.64)	(0.38)
unemp	gap	1.11	1.03	1.00	1.14	1.15	1.22	1.06
		(0.93)	(0.62)	(0.51)	(0.98)	(0.97)	(0.94)	(0.77)
pgdp	ln1d	1.14	1.35	1.02	1.03	1.00	1.10	1.13
		(0.91)	(0.96)	(0.70)	(0.60)	(0.49)	(0.92)	(0.75)
pgdp	ln2d	1.03	1.04	1.00	1.51	1.01	1.07	1.24
		(0.93)	(0.88)	(0.28)	(0.93)	(0.71)	(0.95)	(0.99)
cpi	ln1d	1.02	1.41	0.92	1.41	1.00	1.16	1.00
		(0.59)	(0.94)	(0.20)	(0.85)	(0.50)	(0.87)	(0.50)
cpi	ln2d	1.03	1.27	0.96	1.61	1.19	1.09	1.15
		(0.94)	(0.96)	(0.04)	(0.89)	(0.93)	(0.76)	(0.86)
ppi	ln1d	1.19	- -	0.97	1.19	1.38	1.15	1.20
		(0.97)	(- -)	(0.35)	(0.93)	(0.96)	(0.91)	(0.84)
ppi	ln2d	1.02	- -	1.02	1.12	1.34	1.12	1.02
		(0.80)	(- -)	(0.88)	(1.00)	(0.98)	(0.86)	(0.83)
earn	ln1d	1.10	1.25	1.01	0.96	1.12	0.99	1.21
		(0.91)	(0.99)	(0.61)	(0.33)	(0.97)	(0.46)	(0.99)
earn	ln2d	1.01	1.13	1.01	1.11	0.99	1.01	1.04
		(0.67)	(0.82)	(0.94)	(0.69)	(0.15)	(0.81)	(0.94)
mon0	ln1d	1.19	- -	1.06	- -	- -	1.15	1.22
		(0.91)	(- -)	(0.94)	(- -)	(- -)	(0.94)	(1.00)
mon0	ln2d	1.04	- -	1.02	- -	- -	0.98	0.99
		(0.84)	(- -)	(0.75)	(- -)	(- -)	(0.17)	(0.29)
mon1	ln1d	1.04	1.81	0.87	1.02	1.07	1.05	1.08
		(0.59)	(0.98)	(0.08)	(0.70)	(1.00)	(0.68)	(0.92)
mon1	ln2d	1.04	1.02	0.99	1.02	1.08	1.11	1.01
		(0.94)	(0.82)	(0.35)	(0.93)	(0.97)	(0.82)	(0.73)
mon2	ln1d	1.24	1.53	1.14	1.03	1.11	1.16	0.97
		(0.92)	(1.00)	(0.96)	(0.67)	(0.86)	(0.89)	(0.34)
mon2	ln2d	1.02	1.02	1.00	1.03	1.02	0.99	1.08
		(0.81)	(0.67)	(0.60)	(0.94)	(0.77)	(0.19)	(0.92)
mon3	ln1d	1.11	1.04	1.16	1.31	0.84	1.00	1.28
		(0.74)	(0.64)	(0.99)	(0.98)	(0.05)	(0.50)	(1.00)
mon3	ln2d	1.03	0.97	1.01	1.02	0.96	1.08	1.08
		(0.86)	(0.14)	(0.90)	(0.92)	(0.09)	(0.77)	(0.86)
rmon0	ln1d	1.10	- -	1.11	- -	- -	1.02	0.99
		(0.70)	(- -)	(0.95)	(- -)	(- -)	(0.59)	(0.47)
rmon1	ln1d	1.09	1.02	0.87	1.04	1.09	1.04	0.95
		(0.63)	(0.69)	(0.05)	(0.95)	(0.96)	(0.71)	(0.37)
rmon2	ln1d	1.19	1.06	1.04	0.99	1.47	1.01	0.78
		(0.93)	(0.73)	(0.62)	(0.40)	(0.92)	(0.80)	(0.05)
rmon3	ln1d	1.17	1.51	1.10	1.08	1.00	1.07	0.85
		(0.98)	(0.88)	(0.97)	(0.88)	(0.49)	(0.80)	(0.09)

Table 3, Panel A. (Inflation) Fluctuation Test
(Critical Value = 2.62)

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
rtbill	lev	4.23*	1.30	15.68*	1.74	0.12	3.47*	6.41*
rbnds	lev	- -	- -	- -	- -	- -	1.50	4.05*
rbndm	lev	1.77	- -	- -	0.36	- -	- -	2.96*
rbndl	lev	4.93*	4.35*	6.07*	0.05	0.02	-1.40	3.00*
rovnght	1d	1.76	-0.27	9.82*	1.32	0.49	2.98	6.55*
rtbill	1d	3.97*	0.26	22.71*	12.47*	0.05	1.98	4.69*
rbnds	1d	- -	- -	- -	- -	- -	10.18*	2.42
rbndm	1d	4.08*	- -	- -	0.22	- -	- -	-0.42
rbndl	1d	1.82	-0.05	5.44*	0.17	0.10	8.31*	3.27*
rrovnght	lev	-0.81	0.27	4.87*	0.47	0.45	-1.37	0.44
rrtbill	lev	1.06	-0.02	3.80*	0.64	1.75	0.09	0.69
rrbnds	lev	- -	- -	- -	- -	- -	1.71	0.73
rrbndm	lev	2.18	- -	- -	0.03	- -	- -	1.33
rrbndl	lev	2.63*	-0.04	7.31*	0.08	1.93	5.47*	1.88
rrovnght	1d	3.57*	-0.34	9.04*	8.09*	15.20*	3.29	3.39*
rrtbill	1d	7.62*	9.47*	17.52*	8.67*	18.85*	10.99*	1.49
rrbnds	1d	- -	- -	- -	- -	- -	17.08*	10.24*
rrbndm	1d	20.63*	- -	- -	0.16	- -	- -	14.02*
rrbndl	1d	9.29*	10.37*	7.97	0.23	0.25	8.63*	20.31*
rspread	lev	-4.29	0.96	1.13	0.03	17.73*	1.80	2.99*
exrate_a	ln1d	-2.28	- -	- -	- -	- -	- -	- -
rexrate_a	ln1d	-0.93	- -	- -	- -	- -	- -	- -
stockp	ln1d	-4.33	5.48*	2.74*	-0.02	0.17	9.31*	5.94*
rstockp	ln1d	-6.64	5.16*	1.77	-0.02	3.37*	8.98*	5.34*
rgdp	ln1d	11.69*	3.27*	11.22*	0.57	3.41*	9.19*	6.98*
rgdp	gap	12.21*	0.93	10.85*	-0.07	11.37*	12.76*	5.30*
ip	ln1d	4.90*	-0.46	9.54*	9.42*	13.84*	2.29	9.48*
ip	gap	5.31*	-0.81	1.46	-1.33	1.43	3.62*	8.49*
capu	lev	1.77	0.13	8.15*	2.72*	2.78*	10.50*	8.54*
emp	ln1d	12.20*	1.11	0.84	0.14	2.62*	8.72*	9.40*
emp	gap	14.26*	0.28	8.66*	0.23	-0.17	0.83	8.49*
unemp	lev	8.93*	-0.24	8.18*	-0.05	0.00	5.81*	6.07*
unemp	1d	12.43*	3.38*	9.63*	10.60*	1.31	8.48*	7.32*
unemp	gap	11.25*	-0.17	8.67*	2.21	1.05	6.63*	6.20*
pgdp	ln1d	3.16*	0.51	1.63	0.25	4.30*	0.55	0.44
pgdp	ln2d	3.37*	-2.99	1.52	0.39	5.54*	0.59	15.44*
cpi	ln1d	- -	- -	- -	- -	- -	- -	- -
cpi	ln2d	- -	- -	- -	- -	- -	- -	- -
ppi	ln1d	-0.67	- -	10.64*	5.95*	0.41	-0.04	-0.73
ppi	ln2d	10.22*	- -	16.30*	5.31*	8.46*	11.53*	12.78*
earn	ln1d	2.42	0.44	4.10	-0.30	10.88*	4.52*	0.29
earn	ln2d	-0.85	0.75	5.86	22.81*	11.48*	2.22	-0.07
mon0	ln1d	-0.77	- -	0.81	- -	- -	8.68*	4.92*
mon0	ln2d	0.94	- -	1.09	- -	- -	0.65	5.01*
mon1	ln1d	-1.39	0.21	4.03*	0.72	1.42	6.26*	5.47*
mon1	ln2d	0.62	7.39*	2.78*	0.39	11.60*	1.15	25.98*
mon2	ln1d	4.36*	0.61	3.60*	0.29	0.24	0.52	5.06*
mon2	ln2d	1.73	0.79	0.82	0.29	0.50	12.54*	3.73*
mon3	ln1d	0.42	7.16*	8.38*	0.28	3.09*	1.45	-0.33
mon3	ln2d	2.70*	-0.02	1.71	3.04*	1.61	4.13*	18.62*
rmon0	ln1d	9.91*	- -	0.90	- -	- -	9.68*	4.42*
rmon1	ln1d	-0.07	-3.66	9.05*	1.41*	5.21*	0.61	7.65*
rmon2	ln1d	12.59*	0.78	3.36*	1.40	19.58*	3.31*	7.77*
rmon3	ln1d	9.74*	0.11	12.30*	0.62	4.39*	-0.01	6.34*

Table 3, Panel B. (Output) Fluctuation Test
(Critical Value = 2.62)

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
rtbill	lev	6.10*	-2.30	9.06*	0.58	1.60	9.42*	0.95
rbnds	lev	- -	- -	- -	- -	- -	6.93*	1.14
rbndm	lev	4.10*	- -	- -	-0.64	- -	- -	-0.61
rbndl	lev	0.41	-2.00	6.84*	-7.02	6.82*	6.57*	-1.63
rovnght	1d	6.73*	-0.20	-0.24	0.54	0.97	8.82*	10.57*
rtbill	1d	8.86*	0.43	0.27	2.49	0.47	3.88*	3.45*
rbnds	1d	- -	- -	- -	- -	- -	17.61	6.99*
rbndm	1d	1.55	- -	- -	-0.23	- -	- -	1.93
rbndl	1d	3.14*	1.20	1.02	0.12	3.76*	2.77*	4.44*
rrovnght	lev	-1.98	0.45	6.83*	7.24*	2.74*	1.57	-0.56
rrtbill	lev	1.60	1.53	4.93*	0.44	2.99*	1.84	-0.21
rrbnds	lev	- -	- -	- -	- -	- -	6.63*	-0.06
rrbndm	lev	0.57	- -	- -	3.14*	- -	- -	1.16
rrbndl	lev	1.55	1.53	2.51	1.94	6.44*	7.70*	1.49
rrovnght	1d	3.80*	1.33	7.05*	0.55	4.05*	0.41	0.96
rrtbill	1d	-2.43	0.08	9.58*	0.54	0.00	-0.02	0.11
rrbnds	1d	- -	- -	- -	- -	- -	7.22*	-0.01
rrbndm	1d	2.31	- -	- -	0.25	- -	- -	0.04
rrbndl	1d	1.23	0.17	9.13*	0.19	5.06*	5.50*	0.03
rspread	lev	3.23*	11.61*	2.26	7.82*	4.36*	0.97	10.75*
exrate_a	ln1d	3.50*	- -	- -	- -	- -	- -	- -
rexrate_a	ln1d	-0.17	- -	- -	- -	- -	- -	- -
stockp	ln1d	8.56*	12.09*	2.76*	0.05	14.45*	3.58*	13.45*
rstockp	ln1d	8.93*	9.83*	0.78	0.06	9.18*	7.74*	14.53*
rgdp	ln1d	- -	- -	- -	- -	- -	- -	- -
rgdp	gap	- -	- -	- -	- -	- -	- -	- -
ip	ln1d	9.20*	9.41*	0.88	-1.04	13.15*	0.88	9.13*
ip	gap	-2.05	10.92*	-0.28	-3.94	4.53*	1.72	1.34
capu	lev	0.43	8.08*	7.05*	-0.41	10.05*	2.78*	12.19*
emp	ln1d	3.17*	5.38*	6.39*	2.16	3.66*	1.29	10.48*
emp	gap	-0.01	0.13	5.79*	0.85	1.09	1.06	0.62
unemp	lev	4.02*	7.71*	5.81*	4.74*	3.14*	5.25*	14.37*
unemp	1d	1.92	8.76*	7.60*	2.81*	3.24*	3.53*	11.53*
unemp	gap	0.75	8.92*	6.11*	0.25	4.43*	5.05*	9.21*
pgdp	ln1d	1.96	-3.61	6.38*	7.13*	8.27*	6.07*	2.17
pgdp	ln2d	1.72	2.90*	11.45*	0.20	4.93*	0.08	0.17
cpi	ln1d	5.81*	7.82*	9.06*	1.11	2.81*	8.96*	11.47*
cpi	ln2d	-0.76	0.04	9.10*	-0.00	0.79	2.14	0.50
ppi	ln1d	3.49*	- -	13.27*	3.63*	1.02	6.29*	-0.62
ppi	ln2d	6.86*	- -	4.38*	-1.57	0.51	2.88*	4.71*
earn	ln1d	3.49*	-0.33	5.35*	9.07*	0.29	6.83*	-1.59
earn	ln2d	3.07*	1.78	1.07	3.07*	11.28*	5.22*	2.14
mon0	ln1d	1.63	- -	2.66*	- -	- -	-0.10	-1.39
mon0	ln2d	2.06	- -	2.37	- -	- -	9.37*	17.27*
mon1	ln1d	3.76*	0.09	25.26*	2.81*	1.21	-0.00	1.83
mon1	ln2d	1.11	4.22*	13.52*	0.72	1.11	0.04	1.18
mon2	ln1d	3.54*	-0.04	2.88*	1.97	4.18*	1.16	9.33*
mon2	ln2d	7.62*	7.33*	8.54*	2.28	6.04*	17.58*	1.18
mon3	ln1d	7.64*	2.61	-0.32	0.84	10.30*	10.44*	-0.28
mon3	ln2d	3.61*	8.80*	-0.82	7.00*	8.23*	0.04	0.47
rmon0	ln1d	-0.12	- -	1.12	- -	- -	4.42*	9.02*
rmon1	ln1d	1.90	4.47*	20.01*	2.79*	3.13*	0.02	11.95*
rmon2	ln1d	4.54*	1.19	6.87*	3.99*	1.44	3.45*	11.74*
rmon3	ln1d	2.34*	0.75	-0.08	1.60	5.68*	3.67*	9.69*

Table 4, Panel A. (Inflation) ENCNEW Test

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
rtbill	lev	10.24*	14.83*	5.52*	10.49*	36.92*	8.04*	55.33*
rbnds	lev	- -	- -	- -	- -	- -	13.65*	26.38*
rbndm	lev	9.20*	- -	- -	16.43*	- -	- -	18.60*
rbndl	lev	11.73*	15.57*	12.78*	5.57*	36.93*	32.12*	12.87*
rovnght	1d	1.31	-1.41	7.62*	-11.74	9.57*	4.92*	21.93*
rtbill	1d	4.30*	-3.42	7.93*	6.22*	0.24	4.86*	21.81*
rbnds	1d	- -	- -	- -	- -	- -	14.30*	9.96*
rbndm	1d	3.33	- -	- -	26.86*	- -	- -	5.33*
rbndl	1d	6.72*	-6.59	2.76*	18.54*	16.67*	32.39*	13.67*
rrovnght	lev	5.55*	-6.18	- -	-12.40	- -	- -	- -
rrtbill	lev	-3.19			6.02*			
rrbnds	lev	- -						
rrbndm	lev				-9.64			
rrbndl	lev	- -						
rrovnght	1d	4.93*	-1.05		2.90*		1.66	16.45*
rrtbill	1d	8.70*		11.45*	2.32*		13.20*	
rrbnds	1d							
rrbndm	1d				20.35*			
rrbndl	1d				14.09*			
rspread	lev	-2.83	-5.44	-3.47	-0.25	34.78*		30.39*
exrate	ln1d	-1.20						
rexrate	ln1d	-1.03	- -					
stockp	ln1d	0.83	0.15	5.21*	-7.77	5.69*	24.58*	5.90*
rstockp	ln1d	1.49	0.04	2.57*	-7.12	13.72*	18.44*	9.25*
rgdp	ln1d	27.35*	4.19*	21.74*	-2.58	12.10*	6.50*	39.23*
rgdp	gap	43.30*	9.25*	33.56*	-4.35	4.58*	23.41*	46.93*
ip	ln1d	33.32*	-4.23	5.38*	10.04*	31.37*	1.99*	27.06*
ip	gap	34.10*	6.40*	3.40*	3.62*	13.52*	- -	36.12*
capu	lev	3.34	11.46*	22.60*	48.43*	31.53*	- -	43.82*
emp	ln1d	32.69*	1.80	10.73*	-3.17	6.66*	- -	49.76*
emp	gap	47.01*	0.75	32.24*	-1.36	2.91*	25.52*	54.13*
unemp	lev	16.46*	6.42*	24.97*	-2.80	-9.69	21.16*	23.96*
unemp	1d	27.06*	8.34*	23.34*	11.02*	9.13*	45.00*	37.62*
unemp	gap	43.13*	4.63*	16.43*	-3.65	7.90*	42.24*	46.61*
pgdp	ln1d	- -	-0.89	-2.17	- -	3.87*	4.55*	- -
pgdp	ln2d	0.97	-1.43	-0.47	- -	1.13	-1.54	1.45
cpi	ln1d	- -	- -	- -	- -	- -	- -	- -
cpi	ln2d	- -	- -	- -	- -	- -	- -	- -
ppi	ln1d	3.60*	- -	2.21	20.44*	16.54*	31.83*	- -
ppi	ln2d	22.31*	- -	- -	9.17*	97.33*	- -	11.18*
earn	ln1d	0.65	-5.51	-0.18	9.66*	7.98*	4.28*	- -
earn	ln2d	-3.86	-1.91	0.30	6.66*	5.20*	9.31*	-0.76
mon0	ln1d	-4.35	- -	18.24*	- -	- -	24.17*	2.18*
mon0	ln2d	-2.76	- -	0.10	- -	- -	10.82*	2.20*
mon1	ln1d	-8.00	- -	3.69*	2.19*	5.43*	- -	12.09
mon1	ln2d	-2.10	- -	-0.06	-2.12	3.20*	- -	- -
mon2	ln1d	8.33*	- -	7.78*	-4.51	15.27*	-2.38	2.14*
mon2	ln2d	4.41*	- -	-1.85	-2.09	-2.06	9.85*	1.54
mon3	ln1d	5.68*	- -	1.75	-1.89	-1.83	-0.86	- -
mon3	ln2d	2.21	- -	-0.68	0.09	-0.29	-0.35	- -
rmon0	ln1d	18.34*	- -	19.43*	- -	- -	- -	- -
rmon1	ln1d	0.82	- -	8.73*	0.16	36.51*	- -	- -
rmon2	ln1d	48.66*	- -	21.77*	-0.84	- -	- -	18.66*
rmon3	ln1d	50.24*	- -	12.30*	-2.90	- -	-1.03	- -

Table 4, Panel B. (Output) ENCNEW Test

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
rtbill	lev	44.28*	-0.88	24.16*	-7.88	-2.49	33.91*	- -
rbnds	lev	- -	- -	- -	- -	- -	28.128	- -
rbndm	lev	- -	- -	- -	4.948*	- -	- -	20.86*
rbndl	lev	- -	-0.62	36.77*	-4.13	1.61	20.83*	- -
rovnght	1d	2.00*	-5.10	-3.77	-4.57	0.57	-0.09	- -
rtbill	1d	23.14*	-5.93	-1.55	-0.46	2.72*	11.55*	- -
rbnds	1d	- -	- -	- -	- -	- -	10.02*	- -
rbndm	1d	12.07*	- -	- -	-2.67	- -	- -	- -
rbndl	1d	18.18*	-5.70	13.92*	-2.99	5.86*	29.50*	- -
rrovnght	lev	7.87*	7.14*	7.33*	36.20*	-4.36	-5.11	-10.31
rrtbill	lev	8.40*	-7.52	6.45*	0.37	12.17*	-6.29	-2.89
rrbnds	lev	- -	- -	- -	- -	- -	3.91*	5.24*
rrbndm	lev	-7.02	- -	- -	12.97*	- -	- -	35.15*
rrbndl	lev	-5.68	-7.93	-0.42	18.37*	-2.09	6.49*	35.67*
rrovnght	1d	-0.19	-0.91	1.88	-1.40	-0.68	-2.06	-2.74
rrtbill	1d	-1.79	-1.90	3.86*	-2.27	-6.31	-3.57	-10.16
rrbnds	1d	- -	- -	- -	- -	- -	-0.33	-7.39
rrbndm	1d	-0.55	- -	- -	-0.42	- -	- -	-5.81
rrbndl	1d	-0.71	-4.95	4.50*	0.32	1.49	2.89*	-5.00
rspread	lev	20.97*	10.17*	-4.62	21.57*	-3.85	-3.52	110.47*
exrate	ln1d	-1.45	- -	- -	- -	- -	- -	- -
rexrate	ln1d	-1.87	- -	- -	- -	- -	- -	- -
stockp	ln1d	- -	- -	12.02*	- -	26.47*	10.55*	- -
rstockp	ln1d	- -	- -	11.70*	- -	23.49*	16.80*	- -
rgdp	ln1d	- -	- -	- -	- -	- -	- -	- -
rgdp	gap	- -	- -	- -	- -	- -	- -	- -
ip	ln1d	- -	8.52*	-3.61	- -	12.79*	-1.33	- -
ip	gap	-2.26	10.97*	-14.04	- -	-0.89	5.92*	- -
capu	lev	- -	4.95*	31.07*	-11.16	10.74*	4.15*	- -
emp	ln1d	0.71	6.99*	13.31*	- -	-0.85	0.06	- -
emp	gap	-7.60	6.48*	5.34*	- -	-12.12	0.40	-7.11
unemp	lev	12.48*	- -	26.17*	- -	3.23*	9.97*	- -
unemp	1d	-3.31	4.17	45.44*	- -	-1.93	6.89*	- -
unemp	gap	-0.09	8.42*	9.57*	- -	-0.34	5.44*	- -
pgdp	ln1d	9.79*	-3.20	-0.13	- -	1.87	4.71*	27.55*
pgdp	ln2d	-0.88	-0.99	0.58	2.91*	-0.03	-3.04	1.98*
cpi	ln1d	25.46*	6.10*	23.38*	13.05*	12.58*	12.79*	70.85*
cpi	ln2d	-1.42	-5.62	4.37*	9.02*	-4.86	10.84*	8.71*
ppi	ln1d	3.76*	- -	18.82*	11.73*	-10.97	8.46*	34.22*
ppi	ln2d	0.11	- -	-0.17	-3.04	-8.96	0.90	-0.21
earn	ln1d	4.96*	-2.80	0.87	19.11*	-4.77	18.72*	-1.28
earn	ln2d	0.07	0.18	-1.03	13.75*	1.46	-0.14	-1.44
mon0	ln1d	2.36	- -	0.47	- -	- -	-0.15	2.76*
mon0	ln2d	2.52*	- -	0.84	- -	- -	2.11*	1.91
mon1	ln1d	17.68*	- -	23.24*	3.70*	-3.67	2.67	0.99
mon1	ln2d	0.49	- -	1.46	-0.79	-3.59	-0.82	0.43
mon2	ln1d	- -	- -	-3.21	4.78*	4.74*	-0.96	20.23*
mon2	ln2d	-0.30	- -	0.31	-0.85	1.84	0.24	0.05
mon3	ln1d	11.56*	- -	-4.10	-9.18	25.90*	3.65*	-2.29
mon3	ln2d	0.20	- -	-0.56	-0.76	3.50*	1.26	-0.20
rmon0	ln1d	24.91*	- -	0.32	- -	- -	7.77*	51.96*
rmon1	ln1d	36.10*	- -	23.16*	-0.13	-0.54	1.65	41.18*
rmon2	ln1d	-1.87	- -	9.88*	1.46	-8.41	0.19	64.26*
rmon3	ln1d	-3.56	- -	-1.40	0.93	4.16*	2.35*	38.07*

Table 5, Panel A. (Inflation) Mincer-Zarnowitz's (1969) Forecast Rationality Test
Traditional p-values versus Rossi and Sekhposyan (2011b)

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
RW	ln2d	0.98	0.85	0.85	0.80	0.72	0.89	0.98
AR	ln2d	0.93	0.81	0.48	0.85	0.87	0.70	0.88
rtbill	lev	0.48	0.84	0.44	0.17	0.48	0.15	0.61
rbnds	lev	- -	- -	- -	- -	- -	0.10	0.69
rbndm	lev	0.60	- -	- -	0.95	- -	- -	0.87
rbndl	lev	0.84	0.83	0.22	0.82	0.22	0.18	0.79
rovnght	1d	0.87	0.37	0.62*	0.08	0.83	0.64	0.74
rtbill	1d	0.93	0.73	0.62	0.85	0.80	0.64	0.79
rbnds	1d	- -	- -	- -	- -	- -	0.93	0.76
rbndm	1d	0.94	- -	- -	0.98	- -	- -	0.59
rbndl	1d	0.96	0.71	0.39	0.96	0.79	0.95	0.47
rrovnght	lev	0.14	0.25	0.60†	0.06	0.75	0.98	0.90
rrtbill	lev	0.71	0.83	0.55†	0.41	0.76†	0.14	0.95
rrbnds	lev	- -	- -	- -	- -	- -	0.39†	0.99
rrbndm	lev	0.98	- -	- -	0.99	- -	- -	0.94
rrbndl	lev	0.96	0.83	0.41	0.97	0.19	0.86	0.94
rrovnght	1d	0.81	0.41	0.64	0.23	0.92	0.61	0.71
rrtbill	1d	1.00	0.83	0.63	0.71	0.96	0.51	0.71
rrbnds	1d	- -	- -	- -	- -	- -	0.92	0.83
rrbndm	1d	0.95	- -	- -	0.93	- -	- -	0.84
rrbndl	1d	0.85	0.86	0.46	0.94	0.92	0.90	0.79
rspread	lev	0.77	0.14	0.30	0.11	0.70	0.59	0.76
exrate	ln1d	0.67	- -	- -	- -	- -	- -	- -
rexrate	ln1d	0.61	- -	- -	- -	- -	- -	- -
stockp	ln1d	0.97	0.46	0.43	0.84	0.90	0.97	0.91
rstockp	ln1d	0.97	0.46	0.35	0.85	0.84	0.99	0.95
rgdp	ln1d	0.83	0.66	0.68	0.79	0.65	0.83	0.84
rgdp	gap	0.91	0.10	0.85	0.78	0.84	0.81	0.61
ip	ln1d	0.56	0.76	0.71	0.93	0.45	0.82	0.97
ip	gap	0.79	0.75	0.33	0.89	0.80	0.81	0.71
capu	lev	0.49	0.50	0.64	0.03†	0.91	0.61	0.64
emp	ln1d	0.44	0.82	0.43	0.87	0.82	0.66	0.85
emp	gap	0.77	0.84	0.28	0.83	0.90	0.78	0.96
unemp	lev	0.75	0.69	0.31	0.75	0.74	0.84	0.85
unemp	1d	0.90	0.28	0.48	0.85	0.99	0.27	0.77
unemp	gap	0.86	0.18	0.28	0.87	0.93	0.39	0.52
pgdp	ln1d	0.64	0.10	0.66	0.87	0.86	0.50	0.52
pgdp	ln2d	0.88	0.45	0.50	0.80	0.90	0.75	0.89
cpi	ln1d	- -	- -	- -	- -	- -	- -	- -
cpi	ln2d	- -	- -	- -	- -	- -	- -	- -
ppi	ln1d	0.77	- -	0.29	0.82	0.68	0.29	0.58
ppi	ln2d	0.87	- -	0.49	0.71	0.53	0.80	0.69
earn	ln1d	0.38	0.49	0.78	0.62	0.94	0.38	0.66
earn	ln2d	0.93	0.85	0.51	0.84	0.92	0.31	0.85
mon0	ln1d	0.72	- -	0.58	- -	- -	0.80	0.78
mon0	ln2d	0.99	- -	0.31	- -	- -	0.80	0.91
mon1	ln1d	0.79	0.14	0.41	0.45	0.88	0.96	0.74
mon1	ln2d	0.91	0.66†	0.49	0.69	0.87	0.60	0.96
mon2	ln1d	0.49	0.14	0.69	0.51	0.13	0.75	0.97
mon2	ln2d	0.50	0.95†	0.44	0.69	0.83	0.81	0.80
mon3	ln1d	0.72	0.89	0.78	0.50	0.35	0.35	0.87
mon3	ln2d	0.54	0.80†	0.56	0.79	0.69	0.76	0.99
rmon0	ln1d	0.71	- -	0.56	- -	- -	0.01	0.53
rmon1	ln1d	0.85	0.75	0.27	0.53	0.60	0.98	0.64
rmon2	ln1d	0.74	0.38	0.69	0.58	0.28	0.82	0.94
rmon3	ln1d	0.80	0.50	0.81	0.65	0.01	0.73	0.95

Table 5, Panel B. (Output) Mincer-Zarnowitz's (1969) Forecast Rationality Test
Traditional p-values versus Rossi and Sekhposyan (2011b)

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
RW	ln1d	0.69†	0.64†	0.70†	0.13†	0.07†	0.10†	0.51†
AR	ln1d	0.20†	0.00†	0.00†	0.68†	0.36†	0.01†	0.05†
rtbill	lev	0.16	0.00†	0.07	0.55†	0.03†	0.04†	0.00†
rbnds	lev	- -	- -	- -	- -	- -	0.27	0.00†
rbndm	lev	0.09†	- -	- -	0.64†	- -	- -	0.00†
rbndl	lev	0.03†	0.00†	0.01†	0.63†	0.29†	0.06†	0.00†
rovnght	1d	0.11	0.00†	0.00	0.68	0.24†	0.02†	0.00†
rtbill	1d	0.17	0.00†	0.00	0.72	0.17†	0.01†	0.00†
rbnds	1d	- -	- -	- -	- -	- -	0.07†	0.01†
rbndm	1d	0.08	- -	- -	0.78†	- -	- -	0.02†
rbndl	1d	0.04†	0.00†	0.01	0.89†	0.39†	0.01†	0.02†
rrovnght	lev	0.30†	0.01	0.01	0.56†	0.31†	0.00†	0.00†
rrtbill	lev	0.09†	0.00†	0.01	0.44†	0.38†	0.00	0.00†
rrbnds	lev	- -	- -	- -	- -	- -	0.00†	0.00†
rrbndm	lev	0.02†	- -	- -	0.35†	- -	- -	0.00†
rrbndl	lev	0.03†	0.00†	0.00†	0.49†	0.56†	0.00†	0.00†
rrovnght	1d	0.09	0.00†	0.00	0.70†	0.27†	0.00†	0.00†
rrtbill	1d	0.04†	0.00†	0.01	0.76†	0.11†	0.00†	0.00†
rrbnds	1d	- -	- -	- -	- -	- -	0.01†	0.00†
rrbndm	1d	0.13†	- -	- -	0.23†	- -	- -	0.00†
rrbndl	1d	0.11†	0.00†	0.01	0.21†	0.47†	0.00†	0.00†
rspread	lev	0.32†	0.15	0.01	0.90†	0.47†	0.00†	0.22†
exrate	ln1d	0.00†	- -	- -	- -	- -	- -	- -
rexrate	ln1d	0.00†	- -	- -	- -	- -	- -	- -
stockp	ln1d	0.29	0.58	0.04†	0.13†	0.30†	0.00†	0.04†
rstockp	ln1d	0.46	0.48	0.04	0.27†	0.37†	0.01†	0.05†
rgdp	ln1d	- -	- -	- -	- -	- -	- -	- -
rgdp	gap	- -	- -	- -	- -	- -	- -	- -
ip	ln1d	0.32†	0.02	0.00†	0.86†	0.41†	0.00†	0.04†
ip	gap	0.03†	0.00	0.00†	0.49†	0.23†	0.00†	0.00†
capu	lev	0.00†	0.00	0.02†	0.47†	0.31†	0.65†	0.01†
emp	ln1d	0.12†	0.07	0.00†	0.97†	0.23†	0.00†	0.01†
emp	gap	0.01†	0.02†	0.01†	0.54†	0.01†	0.00†	0.00†
unemp	lev	0.15†	0.06†	0.02	0.16†	0.10†	0.00†	0.05†
unemp	1d	0.05†	0.00	0.00†	0.92†	0.20†	0.01†	0.02†
unemp	gap	0.04†	0.00	0.00	0.63†	0.02†	0.00†	0.00†
pgdp	ln1d	0.01†	0.00†	0.01	0.40†	0.32†	0.01†	0.00†
pgdp	ln2d	0.09†	0.00†	0.00	0.04†	0.33†	0.00†	0.00
cpi	ln1d	0.15	0.01†	0.06	0.26†	0.47†	0.01†	0.01†
cpi	ln2d	0.12†	0.00†	0.01	0.24†	0.09†	0.00†	0.00†
ppi	ln1d	0.00†	- -	0.04†	0.73†	0.06†	0.00†	0.00†
ppi	ln2d	0.12†	- -	0.00	0.71†	0.02†	0.00†	0.01†
earn	ln1d	0.11†	0.00†	0.00	0.64†	0.22†	0.50†	0.00†
earn	ln2d	0.19†	0.00†	0.00	0.44†	0.40†	0.01†	0.01†
mon0	ln1d	0.00†	- -	0.23	- -	- -	0.00†	0.00†
mon0	ln2d	0.07†	- -	0.81	- -	- -	0.01†	0.10†
mon1	ln1d	0.14	0.00†	0.01	0.83†	0.27†	0.19†	0.00†
mon1	ln2d	0.10†	0.00†	0.00	0.83†	0.17†	0.00†	0.05†
mon2	ln1d	0.00	0.00†	0.00†	0.87†	0.11†	0.00†	0.02†
mon2	ln2d	0.01	0.00†	0.00	0.89†	0.25†	0.15†	0.00†
mon3	ln1d	0.06†	0.00†	0.00	0.36†	0.45	0.19†	0.00†
mon3	ln2d	0.02†	0.00†	0.00	0.87†	0.89†	0.01†	0.00†
rmon0	ln1d	0.02†	- -	0.13	- -	- -	0.02†	0.07†
rmon1	ln1d	0.20	0.00†	0.05	0.70†	0.21†	0.07†	0.04†
rmon2	ln1d	0.00	0.00†	0.00†	0.82†	0.03†	0.06†	0.13†
rmon3	ln1d	0.00	0.00†	0.00†	0.70†	0.80†	0.04†	0.02†

Table 5, Panel C. (Inflation) Forecast Unbiasedness Test
Traditional p-values versus Rossi and Sekhposyan (2011b)

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
RW	ln2d	0.98	0.85	0.85	0.80†	0.72	0.89	0.98
AR	ln2d	0.93	0.81	0.48	0.85	0.87	0.70	0.88
rtbill	lev	0.48	0.84	0.44	0.17	0.48	0.15†	0.61
rbnds	lev	- -	- -	- -	- -	- -	0.10†	0.69
rbndm	lev	0.60	- -	- -	0.95†	- -	- -	0.87
rbndl	lev	0.84	0.83†	0.22†	0.82	0.22†	0.18†	0.79
rovnght	1d	0.87	0.37	0.62	0.08	0.83	0.64	0.74
rtbill	1d	0.93	0.73	0.62	0.85	0.80	0.64	0.79
rbnds	1d	- -	- -	- -	- -	- -	0.93	0.76
rbndm	1d	0.94	- -	- -	0.98†	- -	- -	0.59
rbndl	1d	0.96	0.71	0.39	0.96†	0.79	0.95	0.47
rrovnght	lev	0.14	0.25	0.60†	0.06	0.75	0.98	0.90
rrtbill	lev	0.71	0.83	0.55†	0.41	0.76†	0.14	0.95
rrbnds	lev	- -	- -	- -	- -	- -	0.39†	0.99
rrbndm	lev	0.98†	- -	- -	0.99	- -	- -	0.94
rrbndl	lev	0.96	0.83	0.41	0.97	0.19	0.86†	0.94
rrovnght	1d	0.81	0.41	0.64	0.23	0.92	0.61	0.71
rrtbill	1d	1.00	0.83	0.63	0.71	0.96	0.51	0.71
rrbnds	1d	- -	- -	- -	- -	- -	0.92	0.83
rrbndm	1d	0.95	- -	- -	0.93	- -	- -	0.84
rrbndl	1d	0.85	0.86	0.46	0.94	0.92	0.90	0.79
rspread	lev	0.77	0.14	0.30	0.11	0.70	0.59	0.76
exrate	ln1d	0.67	- -	- -	- -	- -	- -	- -
rexrate	ln1d	0.61	- -	- -	- -	- -	- -	- -
stockp	ln1d	0.97	0.46	0.43	0.84	0.90	0.97	0.91
rstockp	ln1d	0.97	0.46	0.35	0.85	0.84	0.99	0.95
rgdp	ln1d	0.83	0.66	0.68	0.79	0.65	0.83	0.84
rgdp	gap	0.91	0.10	0.85	0.78	0.84	0.81	0.61
ip	ln1d	0.56	0.76	0.71	0.93	0.45	0.82	0.97
ip	gap	0.79	0.75	0.33	0.89	0.80	0.81	0.71
capu	lev	0.49	0.50	0.64	0.03†	0.91	0.61	0.64
emp	ln1d	0.44	0.82	0.43	0.87	0.82	0.66	0.85
emp	gap	0.77	0.84	0.28	0.83	0.90	0.78	0.96
unemp	lev	0.75	0.69	0.31	0.75	0.74	0.84	0.85
unemp	1d	0.90	0.28	0.48	0.85	0.99	0.27	0.77
unemp	gap	0.86	0.18	0.28	0.87	0.93	0.39	0.52
pgdp	ln1d	0.64	0.10	0.66	0.87	0.86	0.50	0.52
pgdp	ln2d	0.88	0.45	0.50	0.80	0.90	0.75	0.89
cpi	ln1d	- -	- -	- -	- -	- -	- -	- -
cpi	ln2d	- -	- -	- -	- -	- -	- -	- -
ppi	ln1d	0.77	- -	0.29	0.82	0.68	0.29	0.58
ppi	ln2d	0.87	- -	0.49	0.71	0.53	0.80	0.69
earn	ln1d	0.38	0.49	0.78	0.62	0.94	0.38	0.66
earn	ln2d	0.93	0.85	0.51	0.84	0.92	0.31	0.85
mon0	ln1d	0.72	- -	0.58	- -	- -	0.80	0.78
mon0	ln2d	0.99	- -	0.31	- -	- -	0.80	0.91
mon1	ln1d	0.79	0.14†	0.41	0.45	0.88	0.96	0.74
mon1	ln2d	0.91	0.66†	0.49	0.69	0.87	0.60	0.96
mon2	ln1d	0.49	0.14†	0.69	0.51	0.13	0.75	0.97
mon2	ln2d	0.50	0.95†	0.44	0.69	0.83	0.81	0.80
mon3	ln1d	0.72†	0.89	0.78	0.50	0.35	0.35	0.87
mon3	ln2d	0.54	0.80†	0.56	0.79	0.69	0.76	0.99
rmon0	ln1d	0.71†	- -	0.56	- -	- -	0.01†	0.53
rmon1	ln1d	0.85	0.75	0.27	0.53	0.60	0.98	0.64
rmon2	ln1d	0.74†	0.38	0.69	0.58	0.28	0.82	0.94
rmon3	ln1d	0.80	0.50	0.81	0.65	0.01	0.73	0.95

Table 5, Panel D. (Output) Forecast Unbiasedness Test
Traditional p-values versus Rossi and Sekhposyan (2011b)

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
RW	ln1d	0.69†	0.64†	0.70†	0.13†	0.07†	0.10†	0.51†
AR	ln1d	0.20†	0.00†	0.00†	0.68†	0.36†	0.01†	0.05†
rtbill	lev	0.16	0.00†	0.07†	0.55†	0.03†	0.04†	0.00†
rbnds	lev	- -	- -	- -	- -	- -	0.27†	0.00†
rbndm	lev	0.09†	- -	- -	0.64†	- -	- -	0.00†
rbndl	lev	0.03†	0.00†	0.01†	0.63†	0.29†	0.06†	0.00†
rovnght	1d	0.11†	0.00†	0.00†	0.68†	0.24†	0.02†	0.00†
rtbill	1d	0.17†	0.00†	0.00†	0.72†	0.17†	0.01†	0.00†
rbnds	1d	- -	- -	- -	- -	- -	0.07†	0.01†
rbndm	1d	0.08†	- -	- -	0.78†	- -	- -	0.02†
rbndl	1d	0.04†	0.00†	0.01†	0.89†	0.39†	0.01†	0.02†
rrovnght	lev	0.30†	0.01†	0.01†	0.56†	0.31†	0.00†	0.00†
rrtbill	lev	0.09†	0.00†	0.01†	0.44†	0.38†	0.00†	0.00†
rrbnds	lev	- -	- -	- -	- -	- -	0.00†	0.00†
rrbndm	lev	0.02†	- -	- -	0.35†	- -	- -	0.00†
rrbndl	lev	0.03†	0.00†	0.00†	0.49†	0.56†	0.00†	0.00†
rrovnght	1d	0.09†	0.00†	0.00†	0.70†	0.27†	0.00†	0.00†
rrtbill	1d	0.04†	0.00†	0.01†	0.76†	0.11†	0.00†	0.00†
rrbnds	1d	- -	- -	- -	- -	- -	0.01†	0.00†
rrbndm	1d	0.13†	- -	- -	0.23†	- -	- -	0.00†
rrbndl	1d	0.11†	0.00†	0.01†	0.21†	0.47†	0.00†	0.00†
rspread	lev	0.32†	0.15†	0.01†	0.90†	0.47†	0.00†	0.22†
exrate	ln1d	0.00†	- -	- -	- -	- -	- -	- -
rexrate	ln1d	0.00†	- -	- -	- -	- -	- -	- -
stockp	ln1d	0.29†	0.58†	0.04†	0.13†	0.30†	0.00†	0.04†
rstockp	ln1d	0.46†	0.48†	0.04†	0.27†	0.37†	0.01†	0.05†
rgdp	ln1d	- -	- -	- -	- -	- -	- -	- -
rgdp	gap	- -	- -	- -	- -	- -	- -	- -
ip	ln1d	0.32†	0.02†	0.00†	0.86†	0.41†	0.00†	0.04†
ip	gap	0.03†	0.00†	0.00†	0.49†	0.23†	0.00†	0.00†
capu	lev	0.00†	0.00†	0.02†	0.47†	0.31†	0.65†	0.01†
emp	ln1d	0.12†	0.07†	0.00†	0.97†	0.23†	0.00†	0.01†
emp	gap	0.01†	0.02†	0.01†	0.54†	0.01†	0.00†	0.00†
unemp	lev	0.15†	0.06†	0.02†	0.16†	0.10†	0.00†	0.05†
unemp	1d	0.05†	0.00†	0.00†	0.92†	0.20†	0.01†	0.02†
unemp	gap	0.04†	0.00†	0.00†	0.63†	0.02†	0.00†	0.00†
pgdp	ln1d	0.01†	0.00†	0.01†	0.40†	0.32†	0.01†	0.00†
pgdp	ln2d	0.09†	0.00†	0.00†	0.04†	0.33†	0.00†	0.00†
cpi	ln1d	0.15†	0.01†	0.06	0.26†	0.47†	0.01†	0.01†
cpi	ln2d	0.12†	0.00†	0.01†	0.24†	0.09†	0.00†	0.00†
ppi	ln1d	0.00†	- -	0.04†	0.73†	0.06†	0.00†	0.00†
ppi	ln2d	0.12†	- -	0.00†	0.71†	0.02†	0.00†	0.01†
earn	ln1d	0.11†	0.00†	0.00†	0.64†	0.22†	0.50†	0.00†
earn	ln2d	0.19†	0.00†	0.00†	0.44†	0.40†	0.01†	0.01†
mon0	ln1d	0.00†	- -	0.23†	- -	- -	0.00†	0.00†
mon0	ln2d	0.07†	- -	0.81†	- -	- -	0.01†	0.10†
mon1	ln1d	0.14†	0.00†	0.01†	0.83†	0.27†	0.19†	0.00†
mon1	ln2d	0.10†	0.00†	0.00†	0.83†	0.17†	0.00†	0.05†
mon2	ln1d	0.00†	0.00†	0.00†	0.87†	0.11†	0.00†	0.02†
mon2	ln2d	0.01†	0.00†	0.00†	0.89†	0.25†	0.15†	0.00†
mon3	ln1d	0.06†	0.00†	0.00†	0.36†	0.45	0.19†	0.00†
mon3	ln2d	0.02†	0.00†	0.00†	0.87†	0.89†	0.01†	0.00†
rmon0	ln1d	0.02†	- -	0.13†	- -	- -	0.02†	0.07†
rmon1	ln1d	0.20	0.00†	0.05†	0.70†	0.21†	0.07†	0.04†
rmon2	ln1d	0.00†	0.00†	0.00†	0.82†	0.03†	0.06†	0.13†
rmon3	ln1d	0.00†	0.00†	0.00†	0.70†	0.80†	0.04†	0.02†

Table 6, Panel A. (Inflation). Pesaran and Timmermann (2007) & Inoue and Rossi (2010)

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
AR rmsfe	ln2d	1.57	1.44	1.22	2.42	2.46	2.90	1.78
rtbill	lev	0.84*	0.81*	0.88*	0.80*	0.67*	0.71*	0.66*
		(0.17)	(0.16)	(0.00)	(0.18)	(0.35)	(0.00)	(0.00)
rbnds	lev	- -	- -	- -	- -	- -	0.79*	0.73*
		(- -)	(- -)	(- -)	(- -)	(- -)	(0.01)	(0.01)
rbndm	lev	0.82*	- -	- -	0.74*	- -	- -	0.80*
		(0.07)	(- -)	(- -)	(0.09)	(- -)	(- -)	(0.05)
rbndl	lev	0.79*	0.83*	0.81*	0.81*	0.84*	0.71*	0.84*
		(0.04)	(0.17)	(0.03)	(0.21)	(0.67)	(0.04)	(0.09)
rovnght	1d	0.88*	0.83*	0.95*	1.14*	0.82*	0.79*	0.83*
		(0.05)	(0.16)	(0.09)	(0.37)	(0.05)	(0.19)	(0.01)
rtbill	1d	0.93*	0.84*	0.91*	1.04*	0.61*	0.69*	0.83*
		(0.07)	(0.12)	(0.02)	(0.56)	(0.10)	(0.01)	(0.01)
rbnds	1d	- -	- -	- -	- -	- -	0.65*	0.85*
		(- -)	(- -)	(- -)	(- -)	(- -)	(0.01)	(0.01)
rbndm	1d	0.89*	- -	- -	0.89*	- -	- -	0.89*
		(0.00)	(- -)	(- -)	(0.50)	(- -)	(- -)	(0.06)
rbndl	1d	0.87*	0.85*	0.93*	0.85*	0.83*	0.73*	0.88*
		(0.00)	(0.08)	(0.05)	(0.46)	(0.69)	(0.06)	(0.08)
rrovnght	lev	0.88*	0.91*	0.91	1.39*	0.96*	1.27*	0.83*
		(0.10)	(0.45)	(0.24)	(0.10)	(0.78)	(0.33)	(0.07)
rrtbill	lev	0.91*	1.02*	0.88*	1.07*	0.70*	1.08*	0.83*
		(0.14)	(0.85)	(0.11)	(0.63)	(0.10)	(0.66)	(0.09)
rrbnds	lev	- -	- -	- -	- -	- -	1.02*	0.86*
		(- -)	(- -)	(- -)	(- -)	(- -)	(0.91)	(0.18)
rrbndm	lev	0.89*	- -	- -	0.99*	- -	- -	0.78*
		(0.11)	(- -)	(- -)	(0.94)	(- -)	(- -)	(0.04)
rrbndl	lev	0.87*	1.03*	0.85*	0.89*	0.65*	0.82*	0.75*
		(0.10)	(0.77)	(0.13)	(0.53)	(0.33)	(0.33)	(0.03)
rrovnght	1d	0.90*	0.89*	0.97	0.96*	0.95*	0.84*	0.84*
		(0.28)	(0.23)	(0.51)	(0.55)	(0.23)	(0.20)	(0.01)
rrtbill	1d	0.91*	1.00*	0.91*	1.01*	0.74	0.85*	0.86
		(0.09)	(0.91)	(0.06)	(0.90)	(0.20)	(0.06)	(0.02)
rrbnds	1d	- -	- -	- -	- -	- -	0.86*	0.89*
		(- -)	(- -)	(- -)	(- -)	(- -)	(0.08)	(0.03)
rrbndm	1d	0.84*	- -	- -	0.86*	- -	- -	0.94*
		(0.00)	(- -)	(- -)	(0.14)	(- -)	(- -)	(0.12)
rrbndl	1d	0.85*	1.01*	0.93*	0.79*	0.63*	0.73*	0.95*
		(0.00)	(0.69)	(0.14)	(0.14)	(0.36)	(0.05)	(0.17)
rspread	lev	0.95*	0.82*	0.95*	1.01*	0.82*	0.74*	0.67*
		(0.03)	(0.19)	(0.04)	(0.97)	(0.63)	(0.03)	(0.01)
exrate	ln1d	0.94*	0.64*	0.75*	0.82*	0.70*	0.72*	0.75*
		(0.12)	(0.03)	(0.01)	(0.17)	(0.18)	(0.16)	(0.10)
rexrate	ln1d	0.95*	0.69*	0.79*	0.92*	0.69*	0.91*	0.75*
		(0.26)	(0.05)	(0.01)	(0.43)	(0.20)	(0.02)	(0.10)
stockp	ln1d	0.94*	0.96*	0.91*	0.81*	0.82*	0.82*	0.84*
		(0.34)	(0.11)	(0.07)	(0.22)	(0.17)	(0.20)	(0.02)
rstockp	ln1d	0.93*	0.96*	0.92*	0.82*	0.76*	0.87*	0.81*
		(0.36)	(0.08)	(0.03)	(0.21)	(0.17)	(0.29)	(0.02)
rgdp	ln1d	0.78*	0.93*	0.78*	0.98*	0.90*	0.91*	0.67*
		(0.05)	(0.13)	(0.06)	(0.09)	(0.45)	(0.21)	(0.01)
rgdp	gap	0.80*	0.79*	0.78*	1.00*	0.96*	0.84*	0.71*
		(0.12)	(0.02)	(0.08)	(0.98)	(0.04)	(0.10)	(0.03)

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
rgdp	gap	0.80*	0.79*	0.78*	1.00*	0.96*	0.84*	0.71*
		(0.12)	(0.02)	(0.08)	(0.98)	(0.04)	(0.10)	(0.03)
ip	ln1d	0.77*	0.93*	0.93*	0.83*	0.78*	0.95*	0.66*
		(0.08)	(0.07)	(0.07)	(0.09)	(0.17)	(0.43)	(0.00)
ip	gap	0.74*	0.89*	0.92*	0.91*	0.81*	0.88*	0.70*
		(0.07)	(0.12)	(0.22)	(0.41)	(0.27)	(0.24)	(0.01)
capu	lev	0.77*	0.69*	0.80*	0.97*	0.75*	0.66*	0.59*
		(0.23)	(0.04)	(0.04)	(0.80)	(0.28)	(0.00)	(0.00)
emp	ln1d	0.75*	0.69*	0.83*	0.96*	0.84*	0.59*	0.62*
		(0.07)	(0.01)	(0.02)	(0.11)	(0.25)	(0.04)	(0.00)
emp	gap	0.74*	0.73*	0.73*	0.94*	0.80*	0.73*	0.66*
		(0.06)	(0.07)	(0.00)	(0.27)	(0.26)	(0.21)	(0.02)
unemp	lev	0.74*	0.80*	0.77*	0.71*	0.83*	0.64*	0.66*
		(0.04)	(0.05)	(0.03)	(0.17)	(0.15)	(0.06)	(0.00)
unemp	1d	0.78*	0.88*	0.81*	0.86*	0.89*	0.63*	0.66*
		(0.08)	(0.09)	(0.05)	(0.25)	(0.34)	(0.10)	(0.01)
unemp	gap	0.78*	0.84*	0.86*	0.91*	0.91*	0.68*	0.66*
		(0.06)	(0.04)	(0.04)	(0.24)	(0.48)	(0.17)	(0.01)
pgdp	ln1d	0.93*	0.89*	0.96*	0.94*	0.98*	0.99*	0.87*
		(0.00)	(0.27)	(0.05)	(0.57)	(0.73)	(0.82)	(0.12)
pgdp	ln2d	0.96*	0.98*	0.95*	0.93*	0.99*	0.98*	0.97*
		(0.04)	(0.16)	(0.01)	(0.03)	(0.33)	(0.27)	(0.01)
ppi	ln1d	0.81*	0.56*	0.91*	0.69*	0.55*	0.68*	0.80*
		(0.02)	(0.01)	(0.06)	(0.02)	(0.10)	(0.01)	(0.02)
ppi	ln2d	0.82*	0.62*	0.95*	0.84*	0.56*	0.73*	0.87*
		(0.02)	(0.02)	(0.18)	(0.03)	(0.11)	(0.03)	(0.01)
earn	ln1d	0.93*	0.96*	0.98*	0.76*	0.90*	0.81*	0.96*
		(0.07)	(0.55)	(0.47)	(0.12)	(0.09)	(0.06)	(0.03)
earn	ln2d	0.97*	0.95*	0.98	0.87*	0.95*	0.79*	0.99
		(0.01)	(0.04)	(0.07)	(0.09)	(0.30)	(0.03)	(0.28)
mon0	ln1d	0.93*	- -	0.76*	- -	- -	0.87*	0.96*
		(0.20)	(- -)	(0.06)	(- -)	(- -)	(0.04)	(0.39)
mon0	ln2d	0.96*	- -	0.92*	- -	- -	0.93*	0.97*
		(0.05)	(- -)	(0.00)	(- -)	(- -)	(0.23)	(0.29)
mon1	ln1d	0.96	0.54*	0.90*	0.79*	0.93*	0.52*	0.92*
		(0.28)	(0.07)	(0.20)	(0.12)	(0.32)	(0.01)	(0.20)
mon1	ln2d	0.97*	0.74*	0.97*	0.97*	0.96*	0.77*	0.96*
		(0.04)	(0.26)	(0.18)	(0.54)	(0.10)	(0.15)	(0.10)
mon2	ln1d	0.82*	0.49*	0.86*	0.73*	0.89*	0.86*	0.97*
		(0.13)	(0.06)	(0.04)	(0.09)	(0.49)	(0.04)	(0.38)
mon2	ln2d	0.83*	0.71*	0.98*	0.89*	0.99*	0.86*	0.97*
		(0.04)	(0.23)	(0.07)	(0.27)	(0.66)	(0.08)	(0.05)
mon3	ln1d	0.78*	0.91*	0.93*	0.74*	1.20*	0.85*	0.91*
		(0.16)	(0.70)	(0.10)	(0.11)	(0.22)	(0.00)	(0.07)
mon3	ln2d	0.87*	1.30*	1.00*	0.89*	0.89*	0.90*	0.97*
		(0.04)	(0.23)	(0.94)	(0.25)	(0.24)	(0.14)	(0.03)
rmon0	ln1d	0.79*	- -	0.67*	- -	- -	0.58*	0.80*
		(0.03)	(- -)	(0.01)	(- -)	(- -)	(0.02)	(0.11)
rmon1	ln1d	0.91*	0.63*	0.84*	0.90*	0.72*	0.52*	0.76*
		(0.03)	(0.15)	(0.07)	(0.26)	(0.12)	(0.00)	(0.06)
rmon2	ln1d	0.61*	0.60*	0.77*	0.89*	0.61*	0.80*	0.75*
		(0.03)	(0.12)	(0.05)	(0.27)	(0.07)	(0.14)	(0.01)
rmon3	ln1d	0.61*	0.89*	0.88*	0.90*	0.90*	0.80*	0.78*
		(0.03)	(0.62)	(0.15)	(0.31)	(0.73)	(0.00)	(0.07)
AVG-AVG	na	0.72	0.73	0.78	0.68	0.60	0.57	0.71
		(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)

Table 6, Panel B. (Output). Pesaran and Timmerman (2007) & Inoue and Rossi (2010)

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
AR rmsfe	ln1d	2.17	1.50	2.72	5.27	3.34	2.13	2.22
rtbill	lev	0.71*	0.93*	0.86*	0.80*	0.71*	0.71*	0.76*
		(0.14)	(0.50)	(0.18)	(0.00)	(0.09)	(0.07)	(0.11)
rbnds	lev	- -	- -	- -	- -	- -	0.79*	0.79*
		(- -)	(- -)	(- -)	(- -)	(- -)	(0.17)	(0.16)
rbndm	lev	0.76*	- -	- -	0.91*	- -	- -	0.84*
		(0.21)	(- -)	(- -)	(0.16)	(- -)	(- -)	(0.26)
rbndl	lev	0.81*	0.94*	0.88*	0.96*	0.81*	0.82*	0.86*
		(0.30)	(0.57)	(0.26)	(0.44)	(0.13)	(0.06)	(0.32)
rovnght	1d	0.85*	0.93*	0.89*	0.94*	0.95*	0.92*	0.61*
		(0.07)	(0.38)	(0.02)	(0.09)	(0.53)	(0.01)	(0.01)
rtbill	1d	0.75*	0.99*	0.89*	0.97*	0.92*	0.82*	0.70*
		(0.06)	(0.92)	(0.07)	(0.18)	(0.45)	(0.01)	(0.02)
rbnds	1d	- -	- -	- -	- -	- -	0.90*	0.71*
		(- -)	(- -)	(- -)	(- -)	(- -)	(0.09)	(0.02)
rbndm	1d	0.81*	- -	- -	0.95*	- -	- -	0.78*
		(0.16)	(- -)	(- -)	(0.14)	(- -)	(- -)	(0.06)
rbndl	1d	0.78*	1.01*	0.82*	0.98*	0.99*	0.73*	0.77*
		(0.13)	(0.90)	(0.05)	(0.15)	(0.95)	(0.00)	(0.05)
rrovnght	lev	0.73*	0.91*	0.91*	0.78*	0.84*	0.86*	0.91*
		(0.07)	(0.18)	(0.12)	(0.07)	(0.18)	(0.02)	(0.14)
rrtbill	lev	0.89*	0.96*	0.91*	0.74*	0.82*	0.85*	0.86*
		(0.27)	(0.52)	(0.11)	(0.05)	(0.20)	(0.04)	(0.24)
rrbnds	lev	- -	- -	- -	- -	- -	0.80*	0.85*
		(- -)	(- -)	(- -)	(- -)	(- -)	(0.00)	(0.24)
rrbndm	lev	0.95*	- -	- -	0.74*	- -	- -	0.79*
		(0.26)	(- -)	(- -)	(0.05)	(- -)	(- -)	(0.20)
rrbndl	lev	0.93*	0.95*	0.88*	0.73*	0.81*	0.76*	0.77*
		(0.07)	(0.57)	(0.03)	(0.06)	(0.11)	(0.00)	(0.17)
rrovnght	1d	0.90*	0.97*	0.91*	0.96*	0.97*	0.92*	0.94*
		(0.05)	(0.02)	(0.00)	(0.03)	(0.01)	(0.06)	(0.01)
rrtbill	1d	0.92*	1.00*	0.92*	0.96*	0.94*	0.87*	0.90*
		(0.04)	(0.95)	(0.01)	(0.25)	(0.05)	(0.02)	(0.01)
rrbnds	1d	- -	- -	- -	- -	- -	0.86*	0.89*
		(- -)	(- -)	(- -)	(- -)	(- -)	(0.03)	(0.06)
rrbndm	1d	0.97*	- -	- -	0.82*	- -	- -	0.85*
		(0.14)	(- -)	(- -)	(0.16)	(- -)	(- -)	(0.08)
rrbndl	1d	0.98*	0.99*	0.95*	0.80*	0.95*	0.83*	0.84*
		(0.19)	(0.78)	(0.01)	(0.14)	(0.12)	(0.01)	(0.07)
rspread	lev	0.56*	0.87*	0.78*	0.82*	0.97*	0.86*	0.64*
		(0.03)	(0.26)	(0.01)	(0.03)	(0.46)	(0.18)	(0.02)
exrate	ln1d	0.96*	0.89*	0.90*	0.85*	0.99*	0.95*	0.92*
		(0.03)	(0.01)	(0.02)	(0.01)	(0.86)	(0.21)	(0.34)
rexrate	ln1d	0.93*	0.88*	0.87*	0.94*	1.00*	0.94*	0.92*
		(0.02)	(0.01)	(0.01)	(0.30)	(0.99)	(0.17)	(0.34)
stockp	ln1d	0.88*	0.91*	0.84*	0.90*	0.80*	0.81*	0.86*
		(0.08)	(0.11)	(0.03)	(0.20)	(0.00)	(0.00)	(0.03)
rstockp	ln1d	0.84*	0.91*	0.83*	0.92*	0.81*	0.76*	0.81*
		(0.06)	(0.14)	(0.05)	(0.20)	(0.00)	(0.00)	(0.02)

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
ip	ln1d	0.93*	0.92*	0.93*	0.94*	0.92*	0.85*	0.97*
		(0.19)	(0.12)	(0.09)	(0.40)	(0.08)	(0.03)	(0.34)
ip	gap	0.94*	0.91*	0.93*	0.92*	0.94*	0.89*	0.92*
		(0.13)	(0.15)	(0.11)	(0.15)	(0.05)	(0.13)	(0.11)
capu	lev	0.63*	0.88*	0.93*	0.77*	0.88*	0.60*	0.86*
		(0.02)	(0.14)	(0.29)	(0.12)	(0.00)	(0.00)	(0.09)
emp	ln1d	0.96*	0.86*	0.92*	0.96*	0.94*	0.90*	0.97*
		(0.11)	(0.10)	(0.06)	(0.00)	(0.01)	(0.03)	(0.11)
emp	gap	0.95*	0.80*	0.94*	0.95*	0.84*	0.82*	0.90*
		(0.13)	(0.11)	(0.20)	(0.31)	(0.07)	(0.21)	(0.15)
unemp	lev	0.86*	0.79*	0.91*	0.73*	0.89*	0.74*	0.86*
		(0.03)	(0.05)	(0.12)	(0.04)	(0.02)	(0.01)	(0.14)
unemp	1d	0.93*	0.90*	0.95*	0.97*	0.97*	0.75*	0.97*
		(0.11)	(0.11)	(0.50)	(0.20)	(0.06)	(0.02)	(0.19)
unemp	gap	0.94*	0.84*	0.93*	0.95*	0.92*	0.72*	0.87*
		(0.11)	(0.04)	(0.03)	(0.35)	(0.08)	(0.04)	(0.07)
pgdp	ln1d	0.78*	0.94*	0.98*	0.81*	0.93*	0.88*	0.82*
		(0.01)	(0.18)	(0.13)	(0.05)	(0.01)	(0.06)	(0.17)
pgdp	ln2d	0.95*	0.94*	0.99*	0.75*	0.97*	0.96*	0.85*
		(0.02)	(0.03)	(0.00)	(0.14)	(0.01)	(0.15)	(0.06)
cpi	ln1d	0.86*	0.80*	0.89*	0.72*	0.85*	0.71*	0.69*
		(0.22)	(0.09)	(0.23)	(0.04)	(0.23)	(0.00)	(0.05)
cpi	ln2d	0.98*	0.92*	0.95*	0.80*	0.93*	0.78*	0.78*
		(0.01)	(0.34)	(0.01)	(0.12)	(0.08)	(0.02)	(0.06)
ppi	ln1d	0.85*	0.30*	0.81*	0.82*	0.81*	0.77*	0.70*
		(0.03)	(0.03)	(0.07)	(0.04)	(0.23)	(0.03)	(0.04)
ppi	ln2d	0.96*	0.60*	0.93*	0.96*	0.85*	0.77*	0.96*
		(0.00)	(0.01)	(0.02)	(0.07)	(0.03)	(0.01)	(0.03)
earn	ln1d	0.91*	0.97*	0.96*	0.82*	0.89*	0.87*	0.89*
		(0.15)	(0.55)	(0.09)	(0.04)	(0.24)	(0.05)	(0.11)
earn	ln2d	0.95*	0.96*	0.98	0.81*	0.97*	0.95*	0.98
		(0.03)	(0.35)	(0.08)	(0.21)	(0.07)	(0.11)	(0.01)
mon0	ln1d	0.91*	- -	0.89*	- -	- -	0.91*	0.92*
		(0.24)	(- -)	(0.02)	(- -)	(- -)	(0.04)	(0.07)
mon0	ln2d	0.92*	- -	0.92*	- -	- -	0.96*	0.97*
		(0.08)	(- -)	(0.05)	(- -)	(- -)	(0.00)	(0.00)
mon1	ln1d	0.87*	0.74*	0.86*	0.83*	0.96*	0.46*	0.93*
		(0.13)	(0.03)	(0.06)	(0.00)	(0.19)	(0.03)	(0.10)
mon1	ln2d	0.91*	0.91*	0.95*	0.94*	0.97*	0.60*	0.91*
		(0.04)	(0.03)	(0.05)	(0.04)	(0.02)	(0.04)	(0.01)
mon2	ln1d	0.79*	0.65*	0.95*	0.82*	0.72*	0.79*	0.92*
		(0.00)	(0.00)	(0.26)	(0.00)	(0.01)	(0.01)	(0.29)
mon2	ln2d	0.96*	0.88*	0.98*	0.94*	0.88*	0.80*	0.89*
		(0.16)	(0.00)	(0.05)	(0.00)	(0.04)	(0.01)	(0.01)
mon3	ln1d	0.77*	0.80	0.95*	0.78*	0.74*	0.58*	0.95*
		(0.01)	(0.02)	(0.17)	(0.03)	(0.02)	(0.02)	(0.56)
mon3	ln2d	0.94	0.90	0.98*	0.91*	0.90*	0.81*	0.92*
		(0.13)	(0.00)	(0.01)	(0.00)	(0.03)	(0.01)	(0.04)
rmon0	ln1d	0.80*	- -	0.87*	- -	- -	0.86*	0.69*
		(0.15)	(- -)	(0.02)	(- -)	(- -)	(0.01)	(0.05)
rmon1	ln1d	0.78*	0.83	0.78*	0.83*	0.92*	0.65*	0.83*
		(0.19)	(0.02)	(0.05)	(0.01)	(0.11)	(0.07)	(0.21)
rmon2	ln1d	0.89*	0.73*	0.93*	0.92*	0.75*	0.94*	0.69*
		(0.07)	(0.00)	(0.33)	(0.01)	(0.01)	(0.05)	(0.02)
rmon3	ln1d	0.76*	0.82*	0.93*	0.86*	0.83*	0.70*	0.78*
		(0.03)	(0.03)	(0.11)	(0.21)	(0.06)	(0.03)	(0.06)
AVG-AVG	na	0.73	0.71	0.79	0.64	0.75	0.64	0.66
		(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Table 7. Relative MSFE and Equal Predictive Ability Test's P-value

Model	CN	FR	GY	IT	JP	UK	US
			Panel A. Inflation				
AR RMSFE	1.79	1.68	1.47	3.05	3.15	3.61	2.04
EWA	0.88	1.02	0.88	1.00	0.85	0.80	0.82
	(0.01)	(0.68)	(0.00)	(0.50)	(0.01)	(0.01)	(0.00)
BMA	0.96	1.22	0.93	1.20	0.87	0.99	0.94
	(0.39)	(0.98)	(0.18)	(0.80)	(0.09)	(0.48)	(0.27)
UCSV	0.93	0.93	0.80	0.82	0.83	0.94	0.84
	(0.18)	(0.15)	(0.01)	(0.03)	(0.01)	(0.25)	(0.02)
FAAR	1.06	1.32	0.91	1.66	1.95	1.17	1.30
	(0.62)	(0.97)	(0.26)	(1.00)	(0.96)	(0.81)	(0.88)
			Panel B. Output				
AR RMSFE	2.34	1.68	3.38	5.01	3.35	2.51	2.46
EWA	0.92	1.02	0.92	0.98	1.01	0.91	0.84
	(0.01)	(0.65)	(0.01)	(0.41)	(0.69)	(0.00)	(0.01)
BMA	0.96	1.00	0.87	1.13	1.27	0.95	0.98
	(0.38)	(0.49)	(0.04)	(0.77)	(0.96)	(0.27)	(0.46)
FAAR	1.20	1.22	0.97	1.14	0.98	1.36	1.06
	(0.92)	(0.82)	(0.40)	(0.75)	(0.41)	(0.97)	(0.61)

Table 8. Giacomini and Rossi's (2010a) Fluctuation Test
Critical Value = 2.624

Model	CN	FR	GY	IT	JP	UK	US
			Panel A. Inflation				
EWA	16.82	5.19	14.48	11.60	12.89	12.15	10.81
BMA	10.33	0.25	15.65	0.16	13.42	3.87	8.33
UCSV	14.15	18.09	11.31	11.28	14.16	13.72	21.15
FAAR	9.76	-0.83	7.15	-0.02	0.18	0.40	1.07
			Panel B. Output				
EWA	13.74	6.88	12.74	8.33	4.66	14.85	12.96
BMA	9.41	11.37	11.77	7.47	2.03	7.71	8.36
FAAR	3.77	8.05	9.84	7.46	9.36	1.39	6.26

Table 9, Panel A. (Inflation). Giacomini and Rossi's (2009)
Forecast Breakdown Test (P-values in Parentheses)

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
AR	ln2d	3.25	2.15	3.05	3.13	2.07	1.92	5.43
		(0.00)	(0.03)	(0.00)	(0.00)	(0.04)	(0.05)	(0.00)
rovnght	lev	3.74	3.11	4.37	-6.94	5.57	-2.92	9.44
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rtbill	lev	7.97	7.99	4.33	-0.83	5.87	4.46	9.47
		(0.00)	(0.00)	(0.00)	(0.41)	(0.00)	(0.00)	(0.00)
rbnds	lev	- -	- -	- -	- -	- -	0.11	9.23
		(- -)	(- -)	(- -)	(- -)	(- -)	(0.91)	(0.00)
rbndm	lev	9.56	- -	- -	6.52	- -	- -	10.87
		(0.00)	(- -)	(- -)	(0.00)	(- -)	(- -)	(0.00)
rbndl	lev	9.59	7.76	4.86	8.42	2.00	7.69	10.36
		(0.00)	(0.00)	(0.00)	(0.00)	(0.05)	(0.00)	(0.00)
rovnght	1d	2.99	0.66	4.17	-3.70	3.62	-5.44	5.79
		(0.00)	(0.51)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rtbill	1d	5.49	4.62	4.32	-3.73	4.00	1.54	6.02
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.12)	(0.00)
rbnds	1d	- -	- -	- -	- -	- -	-1.78	6.19
		(- -)	(- -)	(- -)	(- -)	(- -)	(0.08)	(0.00)
rbndm	1d	5.01	- -	- -	5.90	- -	- -	6.91
		(0.00)	(- -)	(- -)	(0.00)	(- -)	(- -)	(0.00)
rbndl	1d	5.26	4.86	4.10	6.83	2.55	4.56	7.30
		(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
rrovnght	lev	3.47	6.14	7.06	-3.08	4.92	-1.69	8.08
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.09)	(0.00)
rrtbill	lev	7.46	9.94	7.08	1.35	3.56	5.60	8.44
		(0.00)	(0.00)	(0.00)	(0.18)	(0.00)	(0.00)	(0.00)
rrbnds	lev	- -	- -	- -	- -	- -	3.29	7.84
		(- -)	(- -)	(- -)	(- -)	(- -)	(0.00)	(0.00)
rrbndm	lev	8.35	- -	- -	6.11	- -	- -	7.90
		(0.00)	(- -)	(- -)	(0.00)	(- -)	(- -)	(0.00)
rrbndl	lev	7.83	9.74	6.93	6.67	0.52	9.32	7.96
		(0.00)	(0.00)	(0.00)	(0.00)	(0.60)	(0.00)	(0.00)
rrovnght	1d	2.92	1.03	4.45	-10.02	1.16	-4.76	6.37
		(0.00)	(0.30)	(0.00)	(0.00)	(0.25)	(0.00)	(0.00)
rrtbill	1d	5.51	3.01	4.55	-2.34	1.96	-0.15	6.90
		(0.00)	(0.00)	(0.00)	(0.02)	(0.05)	(0.88)	(0.00)
rrbnds	1d	- -	- -	- -	- -	- -	-2.48	5.73
		(- -)	(- -)	(- -)	(- -)	(- -)	(0.01)	(0.00)
rrbndm	1d	3.00	- -	- -	6.42	- -	- -	5.57
		(0.00)	(- -)	(- -)	(0.00)	(- -)	(- -)	(0.00)
rrbndl	1d	4.37	2.65	3.52	7.03	-0.44	4.37	5.59
		(0.00)	(0.01)	(0.00)	(0.00)	(0.66)	(0.00)	(0.00)
rspread	lev	1.82	3.15	4.55	0.07	-9.80	-2.78	9.84
		(0.07)	(0.00)	(0.00)	(0.94)	(0.00)	(0.01)	(0.00)
exrate	ln1d	1.08	-4.66	2.59	-2.63	-0.56	-3.07	5.26
		(0.28)	(0.00)	(0.01)	(0.01)	(0.57)	(0.00)	(0.00)
rexrate	ln1d	1.23	-7.15	2.25	-7.92	-0.07	-3.68	5.26
		(0.22)	(0.00)	(0.02)	(0.00)	(0.94)	(0.00)	(0.00)
stockp	ln1d	5.60	4.37	5.65	6.48	5.91	4.52	7.15
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rstockp	ln1d	5.74	4.36	5.08	6.32	5.73	4.58	7.72
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rgdp	ln1d	7.50	-0.62	5.42	4.89	4.71	4.78	8.36
		(0.00)	(0.53)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
rgdp	gap	7.75	6.63	6.20	6.76	2.35	4.57	10.26
		(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)
ip	ln1d	6.08	4.17	3.59	6.23	3.74	3.47	8.10
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ip	gap	9.97	5.23	5.33	6.17	5.07	5.47	8.72
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
capu	lev	3.46	6.36	5.58	-6.62	-2.95	2.64	9.37
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
emp	ln1d	6.82	5.16	5.35	4.58	4.13	-2.16	8.71
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)	(0.00)
emp	gap	6.50	6.07	6.52	7.64	5.22	2.50	9.05
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
unemp	lev	8.05	3.91	7.17	6.28	9.01	-1.65	9.87
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.10)	(0.00)
unemp	1d	7.40	0.26	5.70	3.54	3.47	-5.08	8.39
		(0.00)	(0.80)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
unemp	gap	7.40	6.45	5.32	4.93	4.86	-0.03	10.06
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.98)	(0.00)
pgdp	ln1d	4.86	3.38	4.69	5.62	2.34	3.95	11.28
		(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)
pgdp	ln2d	4.08	-1.70	4.37	3.19	2.14	2.47	5.69
		(0.00)	(0.09)	(0.00)	(0.00)	(0.03)	(0.01)	(0.00)
cpi	ln1d	- -	- -	- -	- -	- -	- -	- -
		(- -)	(- -)	(- -)	(- -)	(- -)	(- -)	(- -)
cpi	ln2d	- -	- -	- -	- -	- -	- -	- -
		(- -)	(- -)	(- -)	(- -)	(- -)	(- -)	(- -)
ppi	ln1d	7.78	2.65	5.29	3.20	7.41	7.00	8.84
		(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ppi	ln2d	4.29	2.47	4.63	3.06	5.61	4.23	6.90
		(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
earn	ln1d	6.70	7.27	3.63	6.17	1.97	3.09	8.34
		(0.00)	(0.00)	(0.00)	(0.00)	(0.05)	(0.00)	(0.00)
earn	ln2d	4.31	3.19	3.07	3.45	1.97	3.11	6.15
		(0.00)	(0.00)	(0.00)	(0.00)	(0.05)	(0.00)	(0.00)
mon0	ln1d	6.68	- -	3.26	- -	- -	-6.93	7.63
		(0.00)	(- -)	(0.00)	(- -)	(- -)	(0.00)	(0.00)
mon0	ln2d	4.19	- -	-1.20	- -	- -	-6.30	6.05
		(0.00)	(- -)	(0.23)	(- -)	(- -)	(0.00)	(0.00)
mon1	ln1d	5.70	0.17	3.69	3.46	3.37	2.47	7.82
		(0.00)	(0.87)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
mon1	ln2d	4.13	-6.68	2.95	1.63	1.82	2.56	5.48
		(0.00)	(0.00)	(0.00)	(0.10)	(0.07)	(0.01)	(0.00)
mon2	ln1d	6.13	-1.15	5.56	4.26	5.99	1.91	7.09
		(0.00)	(0.25)	(0.00)	(0.00)	(0.00)	(0.06)	(0.00)
mon2	ln2d	3.95	-6.21	3.53	1.80	3.68	0.35	6.07
		(0.00)	(0.00)	(0.00)	(0.07)	(0.00)	(0.72)	(0.00)
mon3	ln1d	3.93	-6.95	2.64	4.56	-0.79	2.70	6.32
		(0.00)	(0.00)	(0.01)	(0.00)	(0.43)	(0.01)	(0.00)
mon3	ln2d	2.34	-6.20	2.30	1.48	-2.61	2.31	2.50
		(0.02)	(0.00)	(0.02)	(0.14)	(0.01)	(0.02)	(0.01)

Table 9, Panel B. (Output). Giacomini and Rossi's (2009)
Forecast Breakdown Test (P-values in Parentheses)

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
AR	ln1d	4.57	3.72	5.27	4.78	5.07	5.80	4.56
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rovnght	lev	8.43	7.61	4.70	9.66	8.28	4.87	8.65
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rtbill	lev	14.48	7.36	5.75	6.65	8.14	7.50	8.80
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rbnds	lev	- -	- -	- -	- -	- -	6.31	10.49
		(- -)	(- -)	(- -)	(- -)	(- -)	(0.00)	(0.00)
rbndm	lev	13.50	- -	- -	10.67	- -	- -	12.08
		(0.00)	(- -)	(- -)	(0.00)	(- -)	(- -)	(0.00)
rbndl	lev	12.20	7.43	7.92	8.71	4.64	8.63	11.90
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rovnght	1d	3.52	5.93	6.46	8.48	7.71	3.93	7.82
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rtbill	1d	6.56	6.16	6.71	4.27	4.95	6.09	6.43
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rbnds	1d	- -	- -	- -	- -	- -	4.69	7.30
		(- -)	(- -)	(- -)	(- -)	(- -)	(0.00)	(0.00)
rbndm	1d	7.82	- -	- -	7.22	- -	- -	8.78
		(0.00)	(- -)	(- -)	(0.00)	(- -)	(- -)	(0.00)
rbndl	1d	8.87	6.20	8.79	6.44	4.20	7.45	8.48
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rrovnght	lev	5.99	6.74	5.52	8.99	7.28	5.76	7.52
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rrtbill	lev	9.67	8.49	5.51	7.82	3.99	6.32	8.80
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rrbnds	lev	- -	- -	- -	- -	- -	6.10	8.45
		(- -)	(- -)	(- -)	(- -)	(- -)	(0.00)	(0.00)
rrbndm	lev	7.39	- -	- -	6.93	- -	- -	8.59
		(0.00)	(- -)	(- -)	(0.00)	(- -)	(- -)	(0.00)
rrbndl	lev	6.59	8.46	6.10	6.10	4.74	9.34	8.21
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rrovnght	1d	2.71	3.99	4.90	8.17	5.53	4.04	5.77
		(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rrtbill	1d	5.74	4.47	4.79	4.32	4.86	4.78	6.31
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rrbnds	1d	- -	- -	- -	- -	- -	4.43	7.36
		(- -)	(- -)	(- -)	(- -)	(- -)	(0.00)	(0.00)
rrbndm	1d	5.18	- -	- -	5.53	- -	- -	7.79
		(0.00)	(- -)	(- -)	(0.00)	(- -)	(- -)	(0.00)
rrbndl	1d	5.25	5.49	4.83	5.37	3.52	6.93	7.84
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rspread	lev	6.69	6.29	5.49	10.43	4.93	5.72	8.61
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
exrate	ln1d	2.14	2.35	2.90	4.20	5.61	4.15	1.21
		(0.03)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.23)
rexrate	ln1d	2.20	2.32	3.20	4.94	6.02	4.14	1.21
		(0.03)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.23)
stockp	ln1d	6.28	3.11	6.79	9.81	6.18	7.73	5.34
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rstockp	ln1d	6.09	3.34	6.83	9.16	5.96	7.87	5.88
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
rgdp	ln1d	- -	- -	- -	- -	- -	- -	- -
		(- -)	(- -)	(- -)	(- -)	(- -)	(- -)	(- -)

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
rgdp	gap	- -	- -	- -	- -	- -	- -	- -
		(- -)	(- -)	(- -)	(- -)	(- -)	(- -)	(- -)
ip	ln1d	4.83	3.95	5.75	6.43	4.65	7.26	4.96
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ip	gap	7.17	4.42	6.10	8.42	6.33	8.70	8.77
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
capu	lev	3.20	5.56	7.27	10.25	5.41	3.26	7.70
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
emp	ln1d	5.14	4.24	7.61	5.30	5.40	4.80	5.27
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
emp	gap	7.71	5.28	7.28	8.68	9.10	6.66	9.64
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
unemp	lev	7.87	6.05	7.38	10.84	7.18	6.23	7.59
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
unemp	1d	5.75	4.32	4.62	5.35	6.19	4.78	5.08
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
unemp	gap	6.68	5.17	6.56	7.54	8.11	6.11	7.42
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
pgdp	ln1d	8.93	6.22	5.53	5.10	5.46	8.44	8.61
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
pgdp	ln2d	5.53	4.23	5.33	5.93	5.27	6.43	8.18
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
cpi	ln1d	7.89	10.32	5.57	5.87	6.36	9.56	10.07
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
cpi	ln2d	5.10	6.41	4.72	5.84	7.23	7.97	7.37
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ppi	ln1d	7.49	2.68	6.97	4.58	6.96	9.03	9.56
		(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ppi	ln2d	4.88	2.49	5.89	3.71	7.64	7.05	5.16
		(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
earn	ln1d	7.14	6.46	5.45	5.94	7.02	4.96	9.83
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
earn	ln2d	4.63	4.75	5.35	5.89	4.89	3.41	5.29
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
mon0	ln1d	8.21	- -	0.33	- -	- -	4.76	9.90
		(0.00)	(- -)	(0.74)	(- -)	(- -)	(0.00)	(0.00)
mon0	ln2d	5.91	- -	-1.19	- -	- -	2.76	4.63
		(0.00)	(- -)	(0.23)	(- -)	(- -)	(0.01)	(0.00)
mon1	ln1d	7.31	3.29	3.59	3.92	6.05	2.95	6.86
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
mon1	ln2d	5.96	1.96	3.89	3.64	6.18	3.19	5.00
		(0.00)	(0.05)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
mon2	ln1d	6.82	3.55	6.73	4.11	8.65	3.71	7.07
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
mon2	ln2d	4.62	1.90	5.43	3.69	7.02	2.96	6.41
		(0.00)	(0.06)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
mon3	ln1d	4.14	2.37	5.87	5.00	4.79	3.05	6.45
		(0.00)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
mon3	ln2d	2.48	1.88	4.48	3.60	4.85	3.05	2.87
		(0.01)	(0.06)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Table 10, PanelA. (Inflation) Rossi and Sekhposyan's (2010) Test
5% Critical Values are: ±3.97, ±1.96, and ±1.96

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
rovnght	lev	5.84	2.45	8.00	2.34	4.63	4.45	5.07
		0.18	12.46	-2.14	7.15	0.87	4.28	-1.70
		5.21	-1.82	-0.79	-1.48	1.45	0.76	-0.76
rtbill	lev	4.81	3.46	14.50	5.17	6.03	1.95	5.53
		1.04	19.55	1.35	1.39	-4.91	3.37	-2.82
		2.04	1.29	-0.95	0.49	1.52	1.06	0.00
rbnds	lev	- -	- -	- -	- -	- -	5.51	4.57
		- -	- -	- -	- -	- -	2.19	-5.97
		- -	- -	- -	- -	- -	3.72	3.89
rbndm	lev	2.57	- -	- -	2.91	- -	- -	5.67
		-1.24	- -	- -	11.92	- -	- -	-4.13
		1.75	- -	- -	1.30	- -	- -	6.10
rbndl	lev	3.59	3.68	4.27	3.18	2.23	4.82	5.65
		-1.52	10.38	0.65	3.93	-1.95	-3.41	-4.02
		1.85	1.29	-0.15	1.62	1.43	3.96	5.10
rovnght	1d	4.99	2.80	5.86	2.42	7.78	7.61	3.75
		-4.63	12.03	2.72	2.82	10.08	-0.78	-1.67
		2.72	1.69	-3.00	1.21	1.29	3.78	0.73
rtbill	1d	5.44	4.87	9.91	2.26	3.00	2.78	3.38
		-0.52	8.27	2.79	-2.78	4.90	7.75	-1.53
		2.49	1.40	-3.65	0.03	1.74	0.16	1.09
rbnds	1d	- -	- -	- -	- -	- -	2.44	5.71
		- -	- -	- -	- -	- -	-1.12	-0.64
		- -	- -	- -	- -	- -	0.95	1.56
rbndm	1d	3.37	- -	- -	6.34	- -	- -	7.59
		2.42	- -	- -	-2.02	- -	- -	2.89
		0.22	- -	- -	2.01	- -	- -	2.82
rbndl	1d	4.85	4.27	5.17	6.02	2.37	4.87	3.74
		-3.04	8.65	0.22	-2.95	-6.66	-2.35	2.70
		1.86	1.94	1.09	1.98	1.54	0.05	0.22
rrovnght	lev	7.97	2.55	5.19	2.18	2.69	3.03	4.23
		1.56	5.66	2.36	12.12	-2.89	0.05	-0.12
		3.44	1.75	1.97	0.81	1.48	3.40	1.85
rrtbill	lev	3.38	3.39	5.62	2.99	2.50	7.15	3.48
		-1.91	5.38	0.76	5.26	3.93	7.96	1.17
		2.04	2.64	2.74	1.39	1.03	2.27	1.55
rrbnds	lev	- -	- -	- -	- -	- -	6.34	3.59
		- -	- -	- -	- -	- -	7.38	2.11
		- -	- -	- -	- -	- -	1.01	1.46
rrbndm	lev	3.38	- -	- -	3.18	- -	- -	3.28
		-4.38	- -	- -	0.51	- -	- -	0.52
		2.11	- -	- -	1.96	- -	- -	1.38
rrbndl	lev	3.70	3.29	4.28	3.08	2.37	7.52	3.20
		-0.99	5.06	1.76	1.96	-3.33	-3.86	0.22
		2.31	2.63	-1.90	1.74	1.28	0.94	1.25
rrovnght	1d	4.10	2.92	4.77	2.40	2.77	5.34	5.01
		-1.05	9.20	4.56	1.45	5.54	1.75	-8.95
		1.31	2.93	-1.39	0.26	-1.52	3.01	2.12
rrtbill	1d	6.26	5.23	5.80	5.18	9.38	10.66	3.99
		-2.73	-0.14	2.68	-1.98	1.79	3.08	-7.69
		0.31	0.11	-1.50	0.74	0.36	-0.43	2.06

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
rrbnds	1d	- -	- -	- -	- -	- -	6.69	4.90
		- -	- -	- -	- -	- -	-4.43	-1.42
		- -	- -	- -	- -	- -	1.22	2.49
rrbndm	1d	4.74	- -	- -	5.80	- -	- -	3.00
		-3.05	- -	- -	-3.14	- -	- -	-1.72
		-8.24	- -	- -	2.31	- -	- -	-0.18
rrbndl	1d	4.15	4.53	3.89	5.57	2.32	5.27	2.34
		-1.07	0.61	1.76	-8.46	0.35	-1.38	-1.52
		-1.28	-0.54	-1.07	2.23	1.34	-1.15	-0.18
rspread	lev	6.55	3.29	3.74	3.10	4.43	3.37	6.33
		2.96	0.43	-11.58	-2.33	4.83	3.31	0.31
		4.50	2.94	2.82	1.60	-0.73	1.19	-0.03
exrate	ln1d	2.59	4.54	7.90	2.58	6.40	3.75	2.75
		-6.35	12.87	10.24	10.50	6.84	-1.16	1.59
		3.49	1.05	-0.58	1.59	-0.65	3.20	1.56
rexrate	ln1d	4.82	8.58	6.99	3.33	7.44	2.20	2.75
		0.91	24.09	6.54	3.84	7.15	1.72	1.59
		5.20	-1.41	2.94	1.36	-0.32	0.30	1.56
stockp	ln1d	6.08	5.12	7.20	3.59	5.54	3.01	4.55
		1.84	-18.60	2.36	-1.57	2.10	-1.36	0.37
		4.13	1.40	2.95	2.09	1.82	-1.07	1.45
rstockp	ln1d	7.43	4.53	6.33	3.65	5.14	3.59	5.19
		3.12	-18.30	1.06	-0.68	0.10	1.27	-0.35
		3.97	1.39	2.68	2.00	-0.01	-0.97	2.24
rgdp	ln1d	5.77	4.47	6.97	3.30	9.03	3.96	6.68
		9.13	2.71	-8.27	3.97	3.00	2.24	-0.96
		-4.54	1.09	-2.84	2.00	-0.11	0.90	-1.90
rgdp	gap	5.03	5.05	4.42	4.27	4.10	4.62	6.68
		3.29	-3.18	-0.94	5.20	0.24	2.54	-5.61
		-3.38	2.24	-3.55	2.24	-2.10	-2.12	-0.01
ip	ln1d	10.05	3.28	2.22	7.94	3.94	4.05	6.36
		12.20	2.82	-0.99	-0.46	-4.71	0.65	1.32
		-1.58	2.42	0.59	1.98	-1.51	2.07	-4.09
ip	gap	7.42	8.83	2.70	5.80	8.37	4.04	5.09
		3.03	-8.72	-0.13	-0.39	5.33	1.98	0.54
		-0.32	3.42	1.27	2.70	0.88	1.75	-2.35
capu	lev	2.82	2.73	3.27	5.68	2.84	3.37	5.85
		2.38	6.76	-0.17	6.74	-5.59	-3.28	-2.03
		1.32	1.34	-0.72	-0.44	1.51	-0.26	-0.32
emp	ln1d	15.50	2.41	5.24	2.91	6.16	4.43	3.73
		11.47	5.76	1.16	14.31	-6.78	-0.23	-4.75
		-4.73	1.20	1.72	0.66	0.96	0.72	-0.71
emp	gap	6.22	2.72	5.39	4.51	6.06	2.25	3.25
		4.79	9.77	0.18	5.56	-9.94	-2.06	-6.72
		-2.91	1.11	-1.48	2.60	2.32	1.36	-0.03
unemp	lev	6.87	5.80	5.88	4.48	2.92	5.50	4.96
		0.12	1.30	0.34	-15.20	2.17	1.66	0.44
		1.33	2.07	-0.32	2.60	1.71	1.38	10.73
unemp	1d	7.22	6.44	3.12	4.63	6.65	4.95	4.36
		1.81	-1.22	-0.89	-15.71	1.22	0.24	1.31
		-5.13	0.72	-1.02	0.50	0.62	-1.61	-1.90

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
unemp	gap	4.99	6.24	3.72	2.50	6.35	3.45	7.35
		3.62	-0.01	-1.67	-10.28	5.00	3.31	-3.28
		-2.09	3.15	-0.87	1.41	1.37	1.23	-0.60
pgdp	ln1d	4.98	3.10	4.49	3.27	7.01	4.17	4.43
		8.89	6.51	6.74	0.76	-8.01	0.69	1.47
		2.76	1.35	2.26	3.12	-0.95	2.20	2.08
pgdp	ln2d	3.49	5.92	4.62	2.79	5.01	3.75	2.47
		-1.36	5.48	-1.56	6.79	-11.08	-6.01	-2.58
		1.19	4.16	2.64	1.58	-0.40	2.07	0.57
cpi	ln1d	- -	- -	- -	- -	- -	- -	- -
		- -	- -	- -	- -	- -	- -	- -
		- -	- -	- -	- -	- -	- -	- -
cpi	ln2d	- -	- -	- -	- -	- -	- -	- -
		- -	- -	- -	- -	- -	- -	- -
		- -	- -	- -	- -	- -	- -	- -
ppi	ln1d	2.41	6.55	2.81	4.19	4.94	6.17	2.69
		-4.88	-3.47	-0.66	-0.87	7.69	3.22	-0.50
		2.46	1.82	1.52	-1.73	1.46	1.76	2.67
ppi	ln2d	4.79	6.32	14.22	2.80	3.42	3.69	6.70
		-3.91	-1.44	6.23	-8.17	-6.37	-4.45	-2.04
		-1.21	-1.74	1.02	-0.16	-0.57	-0.96	0.09
earn	ln1d	5.60	4.63	3.52	3.06	2.85	3.90	3.29
		1.39	3.79	-1.42	2.38	-11.20	3.28	3.23
		4.17	2.05	2.47	1.92	-0.23	1.65	3.05
earn	ln2d	4.49	5.64	6.18	12.46	2.40	7.12	5.67
		0.14	17.60	-3.90	6.95	-15.18	4.49	1.85
		4.03	2.81	1.26	-4.27	-0.48	1.44	4.34
mon0	ln1d	5.70	- -	3.86	- -	- -	2.53	7.46
		-0.42	- -	-4.29	- -	- -	-0.62	2.44
		3.01	- -	2.46	- -	- -	0.45	3.54
mon0	ln2d	3.39	- -	4.56	- -	- -	4.76	5.11
		1.92	- -	3.06	- -	- -	0.42	0.04
		1.77	- -	1.22	- -	- -	2.58	1.28
mon1	ln1d	4.65	3.40	5.43	2.49	3.45	8.37	6.46
		0.13	1.97	-17.69	1.12	1.30	-4.92	9.46
		2.81	2.00	2.87	1.97	2.31	0.22	0.36
mon1	ln2d	4.87	2.99	3.71	2.58	2.43	2.34	2.68
		0.39	-0.60	-6.22	1.98	-49.96	28.66	-0.33
		2.87	1.13	1.73	1.80	0.36	0.28	-1.21
mon2	ln1d	5.94	3.45	4.01	2.91	4.87	3.02	5.45
		4.43	2.52	-4.35	-3.79	-6.84	-1.44	2.02
		3.47	2.04	1.23	1.88	2.13	2.84	2.28
mon2	ln2d	4.08	3.92	6.22	2.58	3.30	2.34	2.79
		-2.44	1.66	1.87	1.88	0.58	-5.12	-0.56
		3.11	1.20	3.31	1.97	1.95	-0.34	1.56
mon3	ln1d	3.58	3.55	3.31	2.84	2.45	4.65	5.61
		2.30	2.57	-2.62	-3.32	6.43	10.80	2.84
		1.92	-1.02	0.57	2.08	1.37	3.55	3.15
mon3	ln2d	4.31	2.84	6.83	3.14	2.27	5.23	7.45
		0.95	-0.14	22.06	0.57	-5.91	3.00	3.49
		0.07	2.86	4.31	1.21	1.65	0.00	-9.85

Table 10, Panel B. (Output) Rossi and Sekhposyan's (2010) Test
5% Critical Values are: ±3.97, ±1.96, and ±1.96

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
rovnght	lev	3.58	4.11	4.09	6.18	2.58	4.12	4.70
		0.71	-0.89	2.22	-6.16	-3.68	2.46	1.35
		3.15	1.41	-5.46	2.03	1.51	1.49	-0.99
rtbill	lev	3.93	4.04	3.88	4.40	2.78	8.92	5.94
		3.40	11.92	-4.12	6.58	3.27	-1.60	0.54
		0.26	4.89	-0.81	1.75	1.30	0.79	0.48
rbnds	lev	- -	- -	- -	- -	- -	5.21	5.66
		- -	- -	- -	- -	- -	-1.86	1.63
		- -	- -	- -	- -	- -	-0.36	0.36
rbndm	lev	4.17	- -	- -	7.08	- -	- -	6.34
		2.46	- -	- -	0.40	- -	- -	5.31
		0.87	- -	- -	4.52	- -	- -	2.16
rbndl	lev	6.92	4.40	4.30	6.78	3.00	7.78	6.41
		4.49	9.01	2.17	-0.82	3.00	-2.03	5.26
		2.14	5.00	-0.30	5.84	1.40	1.30	3.12
rovnght	1d	8.73	2.42	5.52	2.57	4.85	4.77	5.87
		0.57	3.63	2.27	7.44	-1.22	8.08	4.99
		2.78	1.29	4.74	1.18	1.79	0.06	-0.58
rtbill	1d	8.64	2.40	6.23	3.11	4.57	2.51	3.55
		10.94	3.22	4.70	1.89	-7.28	-1.62	3.14
		-6.06	1.06	4.96	0.91	2.43	0.75	0.94
rbnds	1d	- -	- -	- -	- -	- -	1.96	4.84
		- -	- -	- -	- -	- -	-3.73	1.77
		- -	- -	- -	- -	- -	0.60	1.03
rbndm	1d	5.85	- -	- -	4.04	- -	- -	5.55
		8.54	- -	- -	0.56	- -	- -	2.68
		3.20	- -	- -	2.34	- -	- -	1.06
rbndl	1d	7.93	2.27	4.49	3.89	2.88	9.59	8.05
		7.73	3.40	-1.20	4.64	4.99	-4.04	2.87
		1.69	0.96	2.39	2.43	0.58	0.42	0.64
rrovnght	lev	7.00	5.60	4.16	2.86	2.54	5.14	3.80
		18.10	-6.11	0.82	-6.44	-3.49	-0.95	14.76
		-0.12	3.62	-0.50	-0.21	1.92	3.37	2.66
rrtbill	lev	4.66	2.31	4.83	5.81	5.27	5.09	3.51
		2.42	0.88	0.73	-2.31	-11.02	-0.02	1.65
		2.67	1.93	-0.13	3.06	0.35	3.72	2.04
rrbnds	lev	- -	- -	- -	- -	- -	3.67	3.96
		- -	- -	- -	- -	- -	2.13	1.72
		- -	- -	- -	- -	- -	-1.05	2.02
rrbndm	lev	3.41	- -	- -	3.96	- -	- -	3.41
		1.19	- -	- -	14.61	- -	- -	1.39
		2.04	- -	- -	0.38	- -	- -	1.62
rrbndl	lev	3.55	2.34	3.44	4.76	2.17	6.72	3.86
		-0.31	0.72	0.19	12.26	-1.13	5.42	1.62
		1.66	1.98	2.42	0.14	0.79	-1.08	1.59
rrovnght	1d	4.18	3.73	6.16	7.32	3.20	2.30	4.75
		5.16	7.18	1.97	2.10	5.98	-2.15	3.33
		-0.21	1.67	-0.99	6.79	0.20	1.87	4.20
rrtbill	1d	7.64	3.43	2.73	2.95	2.99	2.77	3.56
		7.32	20.69	4.81	2.13	21.47	1.74	-4.22
		3.88	1.88	-3.59	1.95	1.19	2.20	2.71

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
rrbnds	1d	- -	- -	- -	- -	- -	3.74	3.60
		- -	- -	- -	- -	- -	0.93	-1.30
		- -	- -	- -	- -	- -	0.95	2.49
rrbndm	1d	3.73	- -	- -	3.38	- -	- -	3.67
		1.70	- -	- -	-4.61	- -	- -	0.22
		1.99	- -	- -	1.85	- -	- -	2.36
rrbndl	1d	3.15	2.05	4.29	3.36	5.91	3.07	3.72
		1.61	0.08	-3.63	-4.73	0.99	1.07	-0.55
		2.11	1.99	-2.08	1.92	1.25	0.63	2.37
rspread	lev	5.21	4.22	4.48	2.89	3.00	3.01	3.10
		-4.18	0.37	2.38	-3.06	-0.44	5.75	-2.52
		0.38	0.26	-0.09	0.16	1.72	2.45	-1.28
exrate	ln1d	3.25	2.01	9.75	2.58	3.49	2.62	5.42
		2.41	1.73	-5.54	5.06	-7.49	5.04	-0.54
		1.55	1.89	0.63	0.02	3.16	1.25	6.73
rexrate	ln1d	2.61	2.80	2.40	2.72	4.26	3.39	5.42
		2.82	0.23	-1.85	-3.31	-11.23	5.13	-0.54
		1.78	2.05	1.64	-0.78	3.67	1.13	6.73
stockp	ln1d	3.57	2.30	4.71	3.62	3.93	7.18	3.15
		0.93	-2.79	-1.27	2.61	-1.88	0.55	-1.46
		-0.46	-1.76	0.22	1.82	-4.55	0.94	-1.99
rstockp	ln1d	3.37	2.78	4.84	3.45	4.23	4.94	4.24
		0.69	-3.17	-2.26	14.06	-2.98	0.40	-2.02
		-0.85	-1.56	0.61	1.25	-2.22	0.02	-2.17
rgdp	ln1d	- -	- -	- -	- -	- -	- -	- -
		- -	- -	- -	- -	- -	- -	- -
		- -	- -	- -	- -	- -	- -	- -
rgdp	gap	- -	- -	- -	- -	- -	- -	- -
		- -	- -	- -	- -	- -	- -	- -
		- -	- -	- -	- -	- -	- -	- -
ip	ln1d	5.73	5.97	4.70	3.92	4.54	3.39	4.65
		1.23	2.34	2.56	4.31	-4.36	2.74	3.74
		-2.24	-3.20	1.72	2.55	-2.33	1.14	-0.39
ip	gap	5.65	6.14	4.13	3.35	4.64	5.18	3.07
		1.42	3.35	3.47	-1.67	-4.05	3.66	3.74
		4.82	-1.98	2.31	4.41	3.13	1.26	1.78
capu	lev	2.50	2.94	4.66	3.30	7.89	3.18	5.45
		7.59	6.15	5.67	6.52	2.44	0.24	0.12
		0.64	1.26	-1.32	1.75	0.00	0.55	-0.22
emp	ln1d	5.30	3.47	6.03	2.30	3.17	2.66	4.69
		-0.32	-1.19	7.22	-1.77	2.91	-0.41	5.88
		1.45	0.30	0.25	1.85	1.31	1.56	0.55
emp	gap	4.53	5.68	4.54	4.33	5.40	2.71	2.80
		3.30	-0.81	1.68	-0.15	3.48	3.50	2.79
		3.88	1.86	1.00	3.68	4.38	1.26	1.68
unemp	lev	3.65	3.79	3.21	3.79	5.53	5.86	6.60
		0.72	-3.06	-5.48	-0.14	-4.37	-3.34	-1.60
		0.35	-0.39	0.44	0.72	2.76	2.65	0.32
unemp	1d	5.36	4.77	4.56	3.29	3.69	5.27	5.48
		9.96	0.95	-6.13	1.79	4.60	-2.79	1.40
		2.16	0.63	-1.69	1.43	1.16	0.96	-2.44

Indicator	Trans.	CN	FR	GY	IT	JP	UK	US
unemp	gap	5.07	5.62	4.32	2.71	4.38	4.68	5.64
		3.94	3.56	5.10	-1.41	1.58	2.17	0.90
		4.31	0.26	-0.65	2.65	2.23	1.48	1.37
pgdp	ln1d	5.52	6.18	5.29	3.33	6.42	7.05	5.24
		0.65	-0.71	3.87	3.85	1.42	-0.82	1.80
		1.54	3.68	0.79	0.04	-0.53	2.52	1.41
pgdp	ln2d	2.35	6.58	7.05	3.49	5.49	2.47	3.20
		0.40	1.87	-1.44	0.93	-8.15	5.10	3.50
		1.63	1.79	-0.66	1.50	2.82	0.89	2.01
cpi	ln1d	6.21	5.97	5.04	3.40	5.69	5.47	4.11
		-1.16	4.29	2.07	9.14	-9.21	3.69	2.21
		0.43	2.39	0.00	0.58	1.39	0.02	-1.21
cpi	ln2d	3.61	2.26	3.58	3.44	2.64	6.74	5.90
		4.25	6.48	3.85	3.09	8.74	3.13	-5.84
		2.06	1.88	-2.91	1.37	0.80	1.05	3.26
ppi	ln1d	4.21	5.19	3.08	4.92	2.49	6.86	7.86
		-0.25	21.92	-1.97	-1.61	-9.25	5.37	6.73
		2.75	-2.29	-0.19	3.07	1.82	2.08	2.89
ppi	ln2d	6.13	10.02	8.61	3.65	2.36	2.60	2.77
		-0.94	4.59	0.83	-3.79	2.39	1.36	0.44
		3.02	0.00	5.19	4.92	0.71	0.82	1.58
earn	ln1d	5.18	3.13	7.00	3.46	2.73	6.39	3.77
		1.08	5.54	1.62	-3.06	4.42	-0.94	-6.41
		1.99	1.91	0.32	-0.53	1.39	0.08	3.95
earn	ln2d	3.92	2.52	2.69	7.47	3.79	6.19	7.34
		-1.47	2.13	1.63	2.23	-6.17	7.27	0.85
		2.35	0.98	1.84	1.10	-0.34	1.63	3.49
mon0	ln1d	6.80	- -	3.52	- -	- -	6.41	5.85
		14.28	- -	-1.73	- -	- -	-0.01	1.62
		1.99	- -	2.20	- -	- -	2.31	6.55
mon0	ln2d	8.01	- -	4.71	- -	- -	2.77	3.25
		3.30	- -	0.47	- -	- -	4.78	9.31
		1.94	- -	1.45	- -	- -	-1.04	-1.01
mon1	ln1d	5.97	3.65	6.04	7.54	2.30	9.40	3.13
		-0.75	2.95	-0.22	3.75	4.28	33.11	-4.56
		0.65	1.92	-4.50	0.73	1.89	-0.58	2.77
mon1	ln2d	4.72	2.66	7.63	2.36	4.13	4.35	6.98
		5.62	-3.28	-0.68	1.02	3.28	8.43	14.03
		2.11	1.53	-1.09	1.49	3.14	1.49	-0.67
mon2	ln1d	4.46	3.47	4.53	3.52	2.91	2.51	4.35
		3.31	0.85	3.15	-3.13	4.08	4.56	0.12
		-5.46	2.68	3.04	0.79	0.17	1.74	-0.60
mon2	ln2d	2.13	4.75	4.40	2.44	3.29	3.35	2.40
		0.61	0.26	-0.78	-3.53	3.43	0.00	1.14
		0.66	0.96	0.50	1.58	0.31	-0.11	1.68
mon3	ln1d	2.32	3.55	3.14	2.74	4.78	5.75	5.26
		-4.96	11.96	0.18	1.24	-5.58	-0.08	-0.09
		0.88	0.25	3.28	1.58	-1.16	-0.01	5.56
mon3	ln2d	6.20	2.69	5.29	3.64	4.57	3.81	2.15
		1.41	1.13	1.01	1.01	-2.55	0.00	1.20
		3.62	-1.09	2.59	1.33	-2.23	2.59	1.25

Notes to the Tables.

Table 1 reports, for each predictor and transformation (listed on the first two columns on the left) and for each country (listed in the columns), the p-value of the Granger-causality for each predictor as well as the p-values of Rossi's (2005) Granger-causality test robust to instabilities (eq. (3), reported in the first and second row for each predictor, respectively). The test statistics are reported for several countries, listed on the columns. Panel A is for predicting inflation and panel B is for predicting real GDP growth.

Table 2 reports, for each predictor and transformation (listed on the first two columns on the left) and for each country (listed in the columns), the value of the ratio of the MSFE for each predictor relative to the RMSFE of the benchmark model. The p-value of the Diebold and Mariano's (1995) test statistic, eq. (41), is reported in parenthesis. The benchmark model is the autoregressive model, whose RMSFE of the benchmark model is reported in the first row of the table. The test statistics are reported for several countries, listed on the columns. Panel A is for predicting inflation and panel B is for predicting real GDP growth.

Table 3 reports, for each predictor and transformation (listed on the first two columns on the left) and for each country (listed in the columns), the value of the Clark and McCracken's (2001) test statistic; asterisks denote significance at the 5% significance level. The benchmark model is the autoregressive model. The test statistics are reported for several countries, listed on the columns. Panel A is for predicting inflation and panel B is for predicting real GDP growth.

Table 4 reports, for each predictor and transformation (listed on the first two columns on the left) and for each country (listed in the columns), the value of the Giacomini and Rossi's (2010a) Fluctuation test statistic, eq. (13). The benchmark model is the autoregressive model. The test statistics are reported for several countries, listed on the columns. Panel A is for predicting inflation and panel B is for predicting real GDP growth. The 5% critical value is listed on top of the table.

Table 5 reports, for each predictor and transformation (listed on the first two columns on the left) and for each country (listed in the columns), the p-values of the forecast rationality test statistic (Panels A and B, for inflation and real GDP growth respectively) and those of the forecast unbiasedness test statistic (Panels C and D, for inflation and real GDP growth respectively). Daggers (†) in Panels A and B denote instead rejections at the 5% significance level using Rossi and Sekhposyan's (2011b) Fluctuation rationality test statistic, eq. (19), implemented by choosing $g_t = [1, y_{t+h,t}]$ and jointly testing both coefficients; daggers in Panels C and D denote rejections of Rossi and Sekhposyan's (2011b) Fluctuation unbiasedness test, i.e. eq. (19) implemented by choosing $g_t = 1$. The test statistics are reported for several countries, listed on the columns. The 5% critical value of the Fluctuation rationality test is 16.90, whereas that of the Fluctuation unbiasedness test is 7.1035.

Table 6 reports, for each predictor and transformation (listed on the first two columns on the left) and for each country (listed in the columns), the ratio of the MSFE of the "Average" forecast across window sizes (based on Pesaran and Timmermann's (2007) method, eq. (24)) relative to the RMSFE of the benchmark model; the p-value of the Diebold and Mariano's (1995) test statistic is reported in parenthesis. The benchmark model is the autoregressive model. Asterisks denote 5% significance of the Inoue and Rossi's (2010) sup-type test statistic across window sizes (unreported) implemented using the Clark and McCracken's (2001) method, R_T^{ε}, i.e. eq. (27). The test statistics are reported for several countries, listed on the columns. Panel A is for predicting inflation and panel B is for predicting real GDP growth.

Table 7 reports the values of the ratio of the MSFE for each model listed on the first column relative to the RMSFE of the benchmark model; the p-value of the Diebold and Mariano's (1995) test statistic is reported in parenthesis. The models are: forecast combinations with equal weights (labeled "EWA"), Bayesian model averaging (labeled "BMA"), Stock and Watson's (2007) unobserved components stochastic volatility (labeled "UCSV") and the factor-augmented autoregressive model (labeled "FAAR"). The benchmark model is the autoregressive model. The test statistics are reported for several countries, listed on the columns. Panel A is for predicting inflation and panel B is for predicting real GDP growth.

Table 8 reports the values of the Giacomini and Rossi's (2010a) Fluctuation test statistic for each model listed on the first column: forecast combinations with equal weights (labeled "EWA"), Bayesian model averaging (labeled "BMA"), Stock and Watson's (2007) unobserved components stochastic volatility (labeled "UCSV") and the factor-augmented autoregressive model (labeled "FAAR"). The benchmark is the autoregressive model. The test statistics are reported for several countries, listed on the columns. Panel A is for predicting inflation and panel B is for predicting real GDP growth. The 5% critical value is listed on top of the table.

Table 9 reports, for each predictor and transformation (listed on the first two columns on the left) and for each country (listed in the columns) the value of Giacomini and Rossi's (2009) forecast breakdown test statistic (p-values are reported in parentheses below the statistics). Panel A is for predicting inflation and panel B is for predicting real GDP growth.

Table 10 reports, for each predictor and transformation (listed on the first two columns on the left) and for each country (listed in the columns), the values of the $\Gamma_P^{(A)}$, $\Gamma_P^{(B)}$ and $\Gamma_P^{(U)}$ test statistics corresponding to the decomposition in Rossi and Sekhposyan (2010), eqs. (49). The three test statistics are listed in the first, second and third row, respectively, for each predictor. Panel A is for predicting inflation and panel B is for predicting real GDP growth.

8 Figures

Figure 1, Panel A. (Inflation) Granger-causality Tests

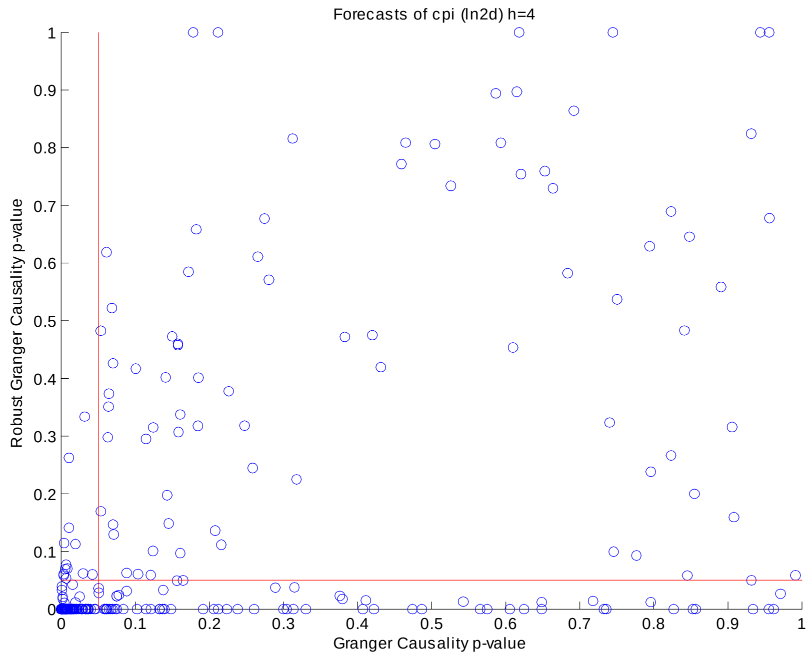


Figure 1, Panel B. (Output) Granger-causality Tests

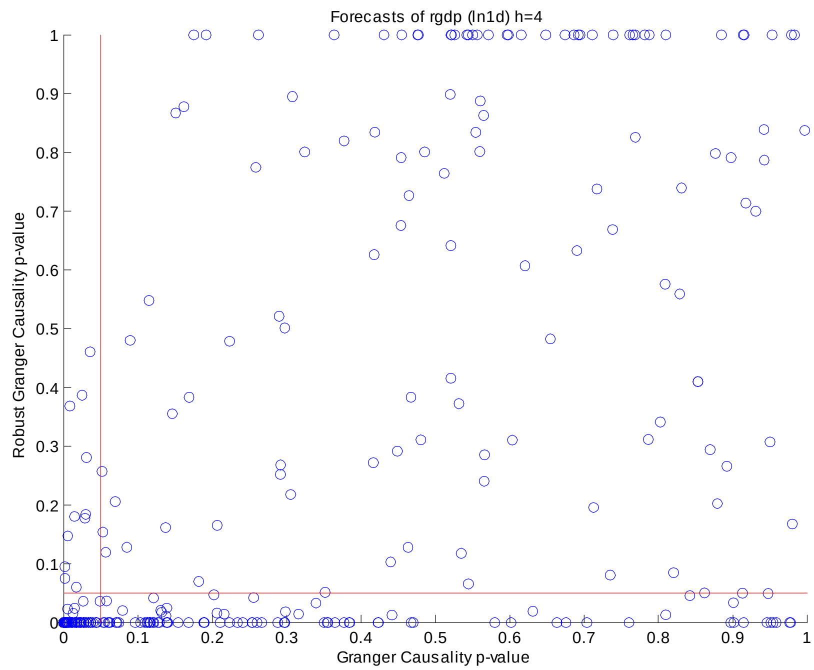


Figure 2, Panel A. (Inflation) Robust vs. Traditional Forecast Comparison Tests

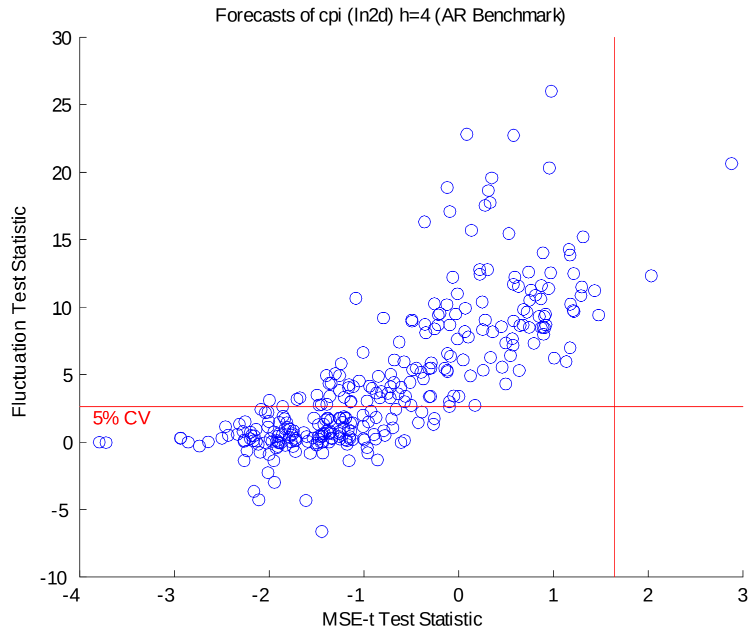


Figure 2, Panel B. (Output) Robust vs. Traditional Forecast Comparison Tests

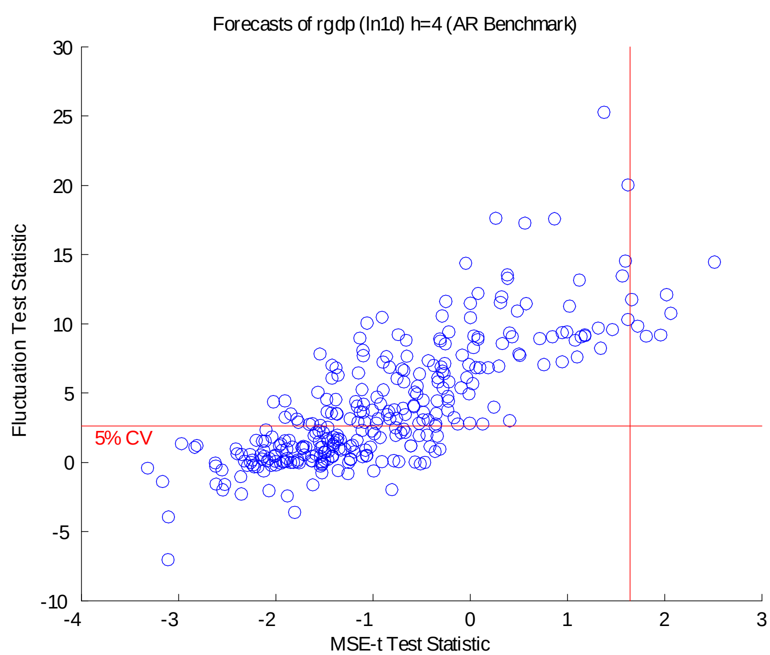


Figure 3, Panel A. (Inflation) Traditional In-sample Vs. Out-of-Sample Tests

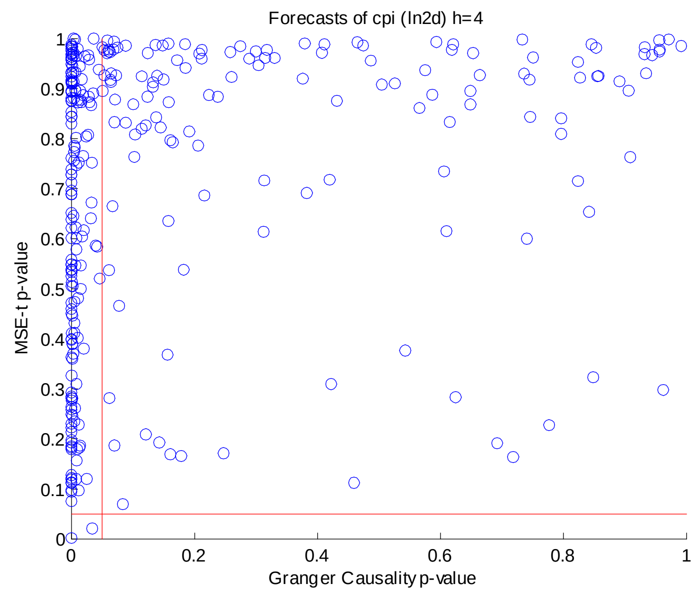


Figure 3, Panel B. (Output) Traditional In-sample Vs. Out-of-Sample Tests

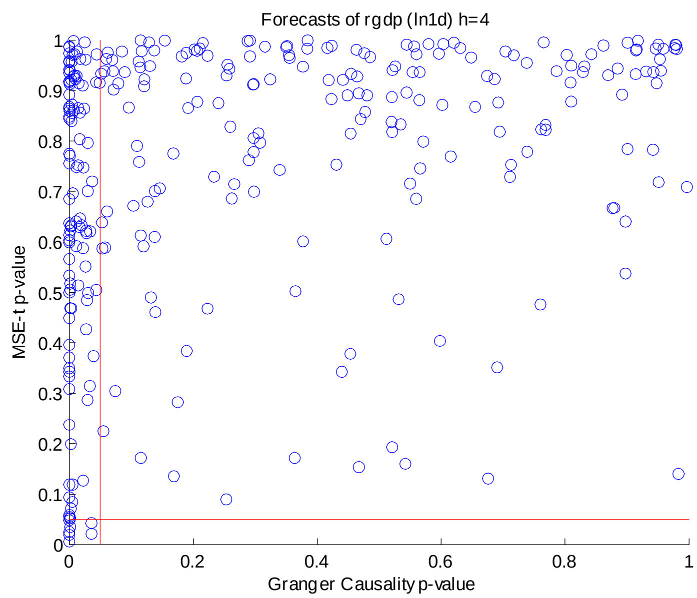


Figure 4, Panel A. (Inflation) Robust In-sample Vs. Out-of-Sample Tests

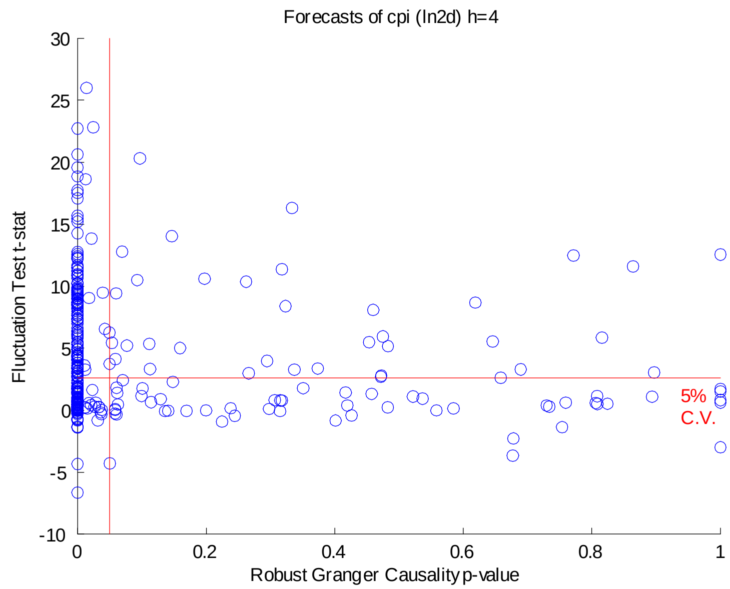


Figure 4, Panel B. (Output) Robust In-sample Vs. Out-of-Sample Tests

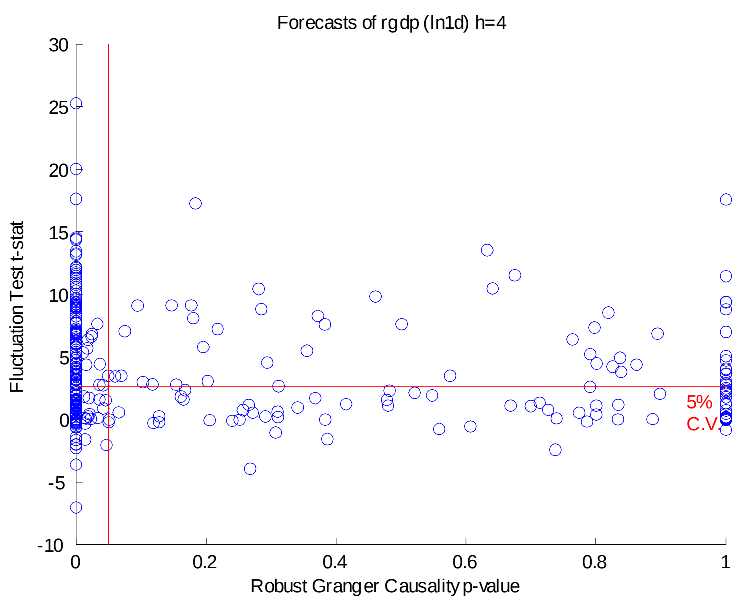


Figure 5, Panel A. (Inflation) Fluctuation Tests Across Series

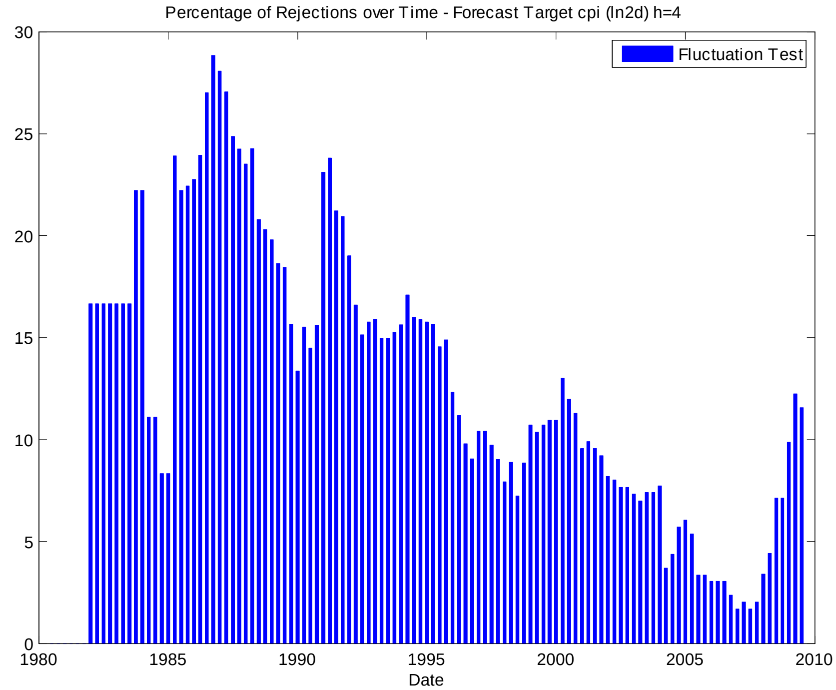


Figure 5, Panel B. (Output) Fluctuation Tests Across Series

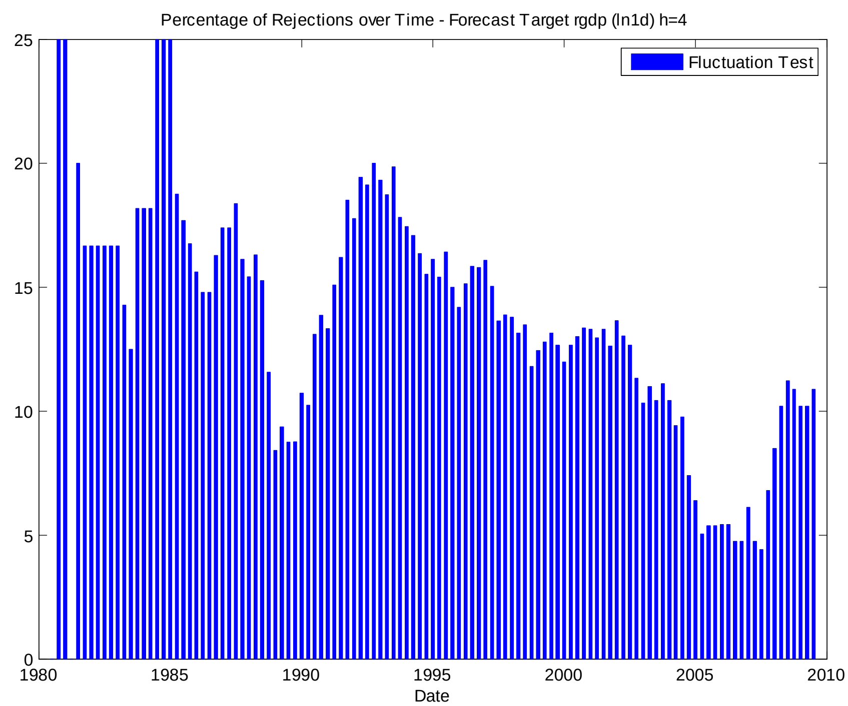


Figure 6, Panel A. (Inflation) Robust vs. Traditional Forecast Rationality Tests

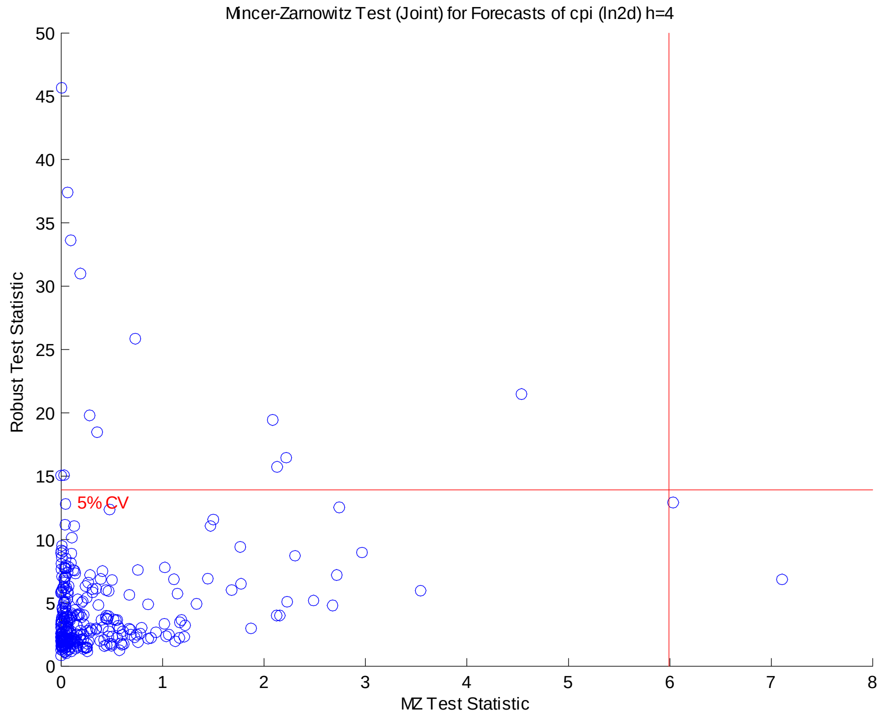


Figure 6, Panel B. (Output) Robust vs. Traditional Forecast Rationality Tests

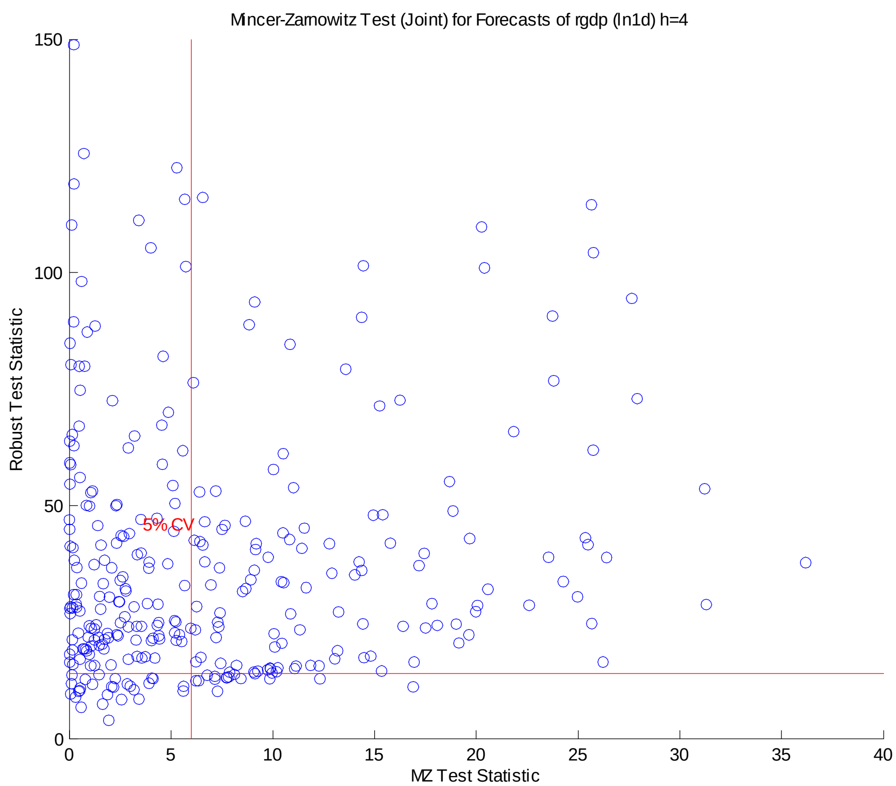


Figure 7, Panel A. (Inflation) Robust vs. Traditional Forecast Unbiasedness Tests

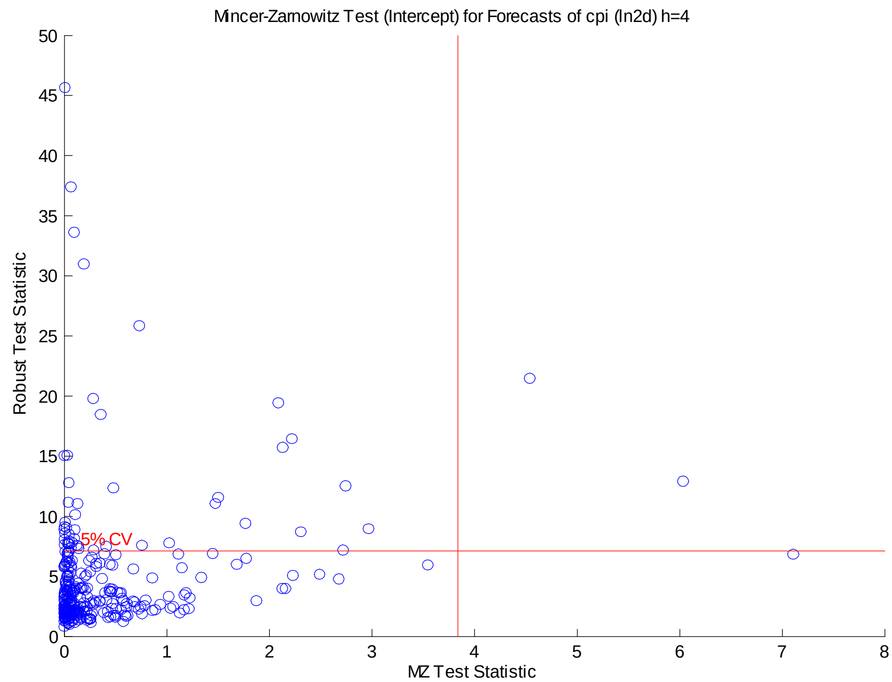


Figure 7, Panel B. (Output) Robust vs. Traditional Forecast Unbiasedness Tests

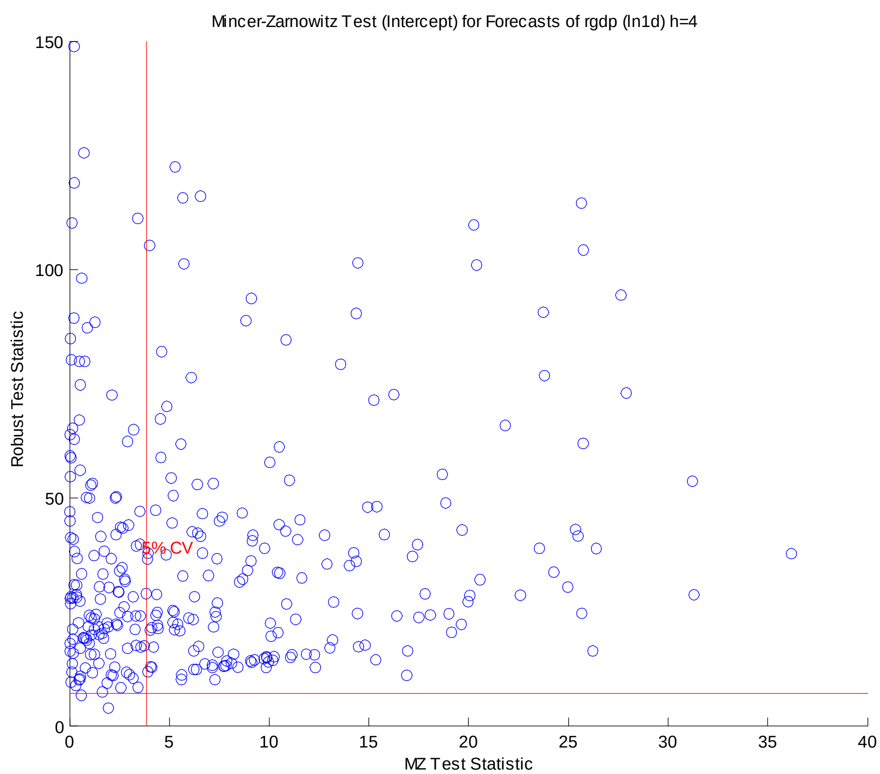


Figure 8, Panel A. (Inflation) Fluctuation Test on EWA vs. AR Model

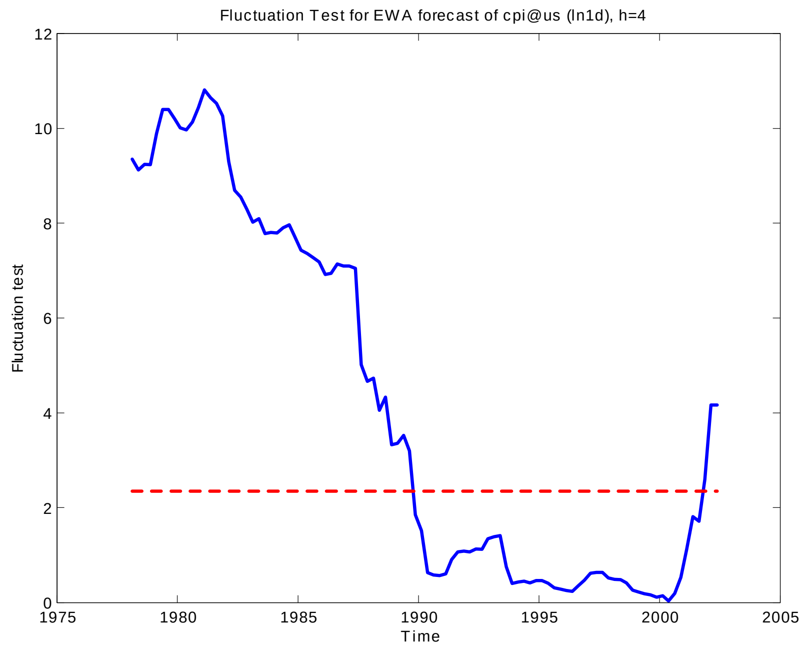


Figure 8, Panel B. (Inflation) Fluctuation Test on BMA vs. AR Model

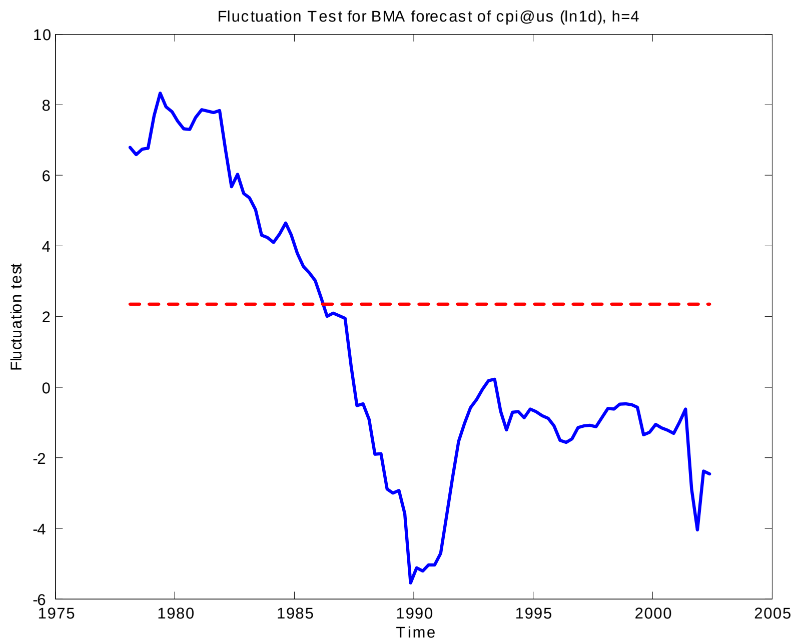


Figure 8, Panel C. (Inflation) Fluctuation Test on FAAR vs. AR Model

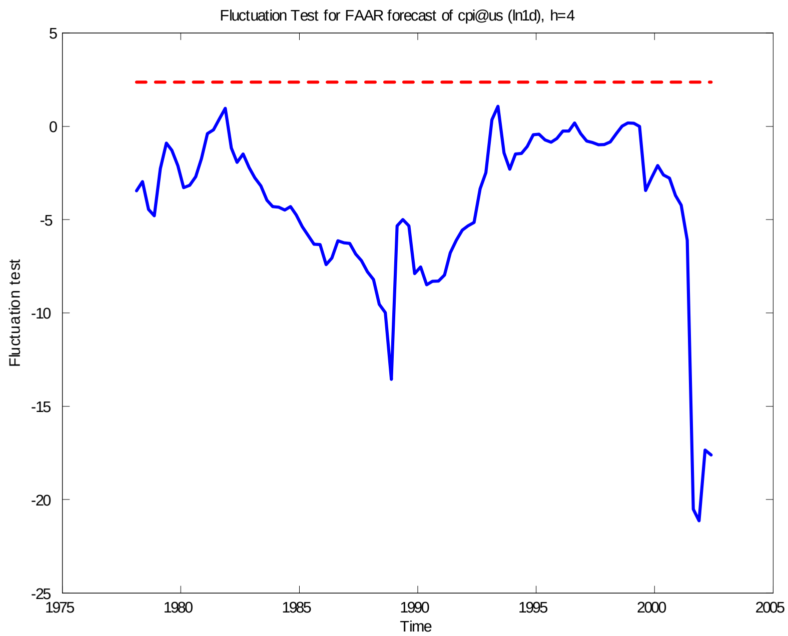


Figure 8, Panel D. (Inflation) Fluctuation Test on UCSV vs. AR Model

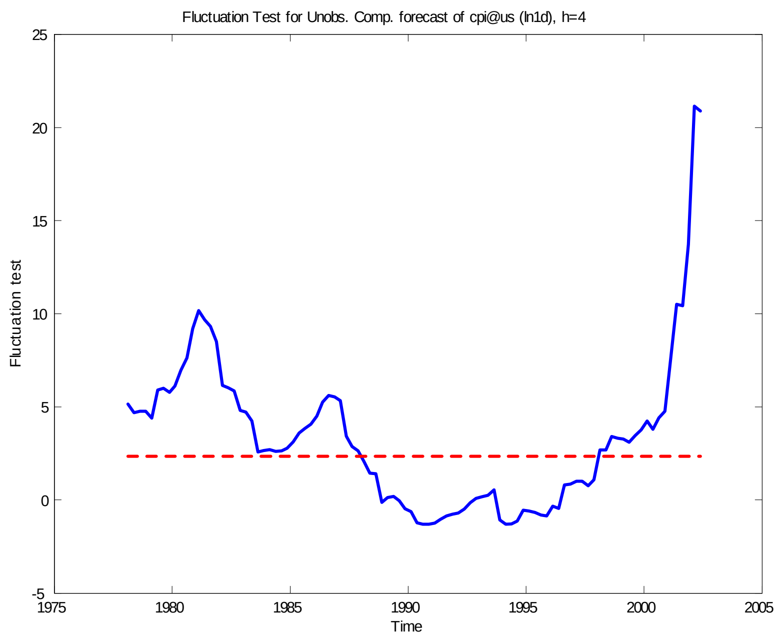


Figure 8, Panel E. (Output) Fluctuation Test on EWA vs. AR Model

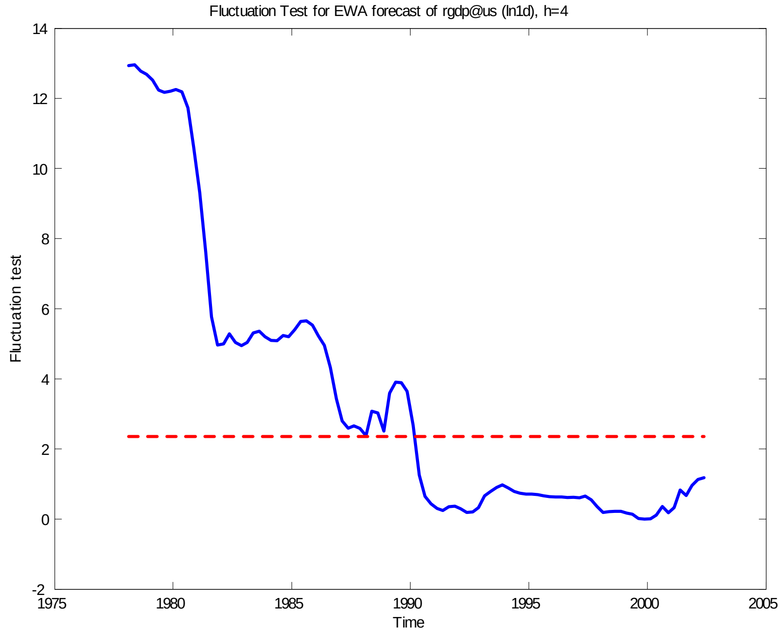


Figure 8, Panel F. (Output) Fluctuation Test on BMA vs. AR Model

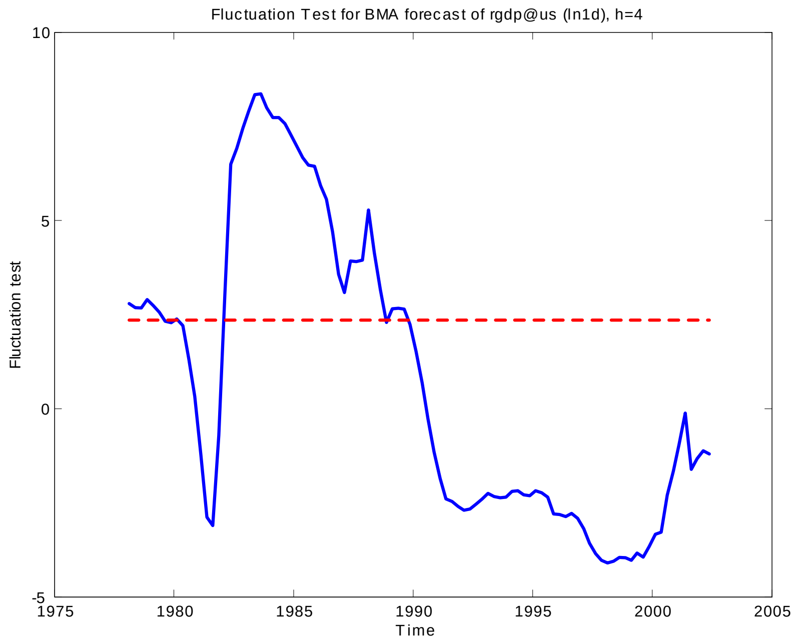
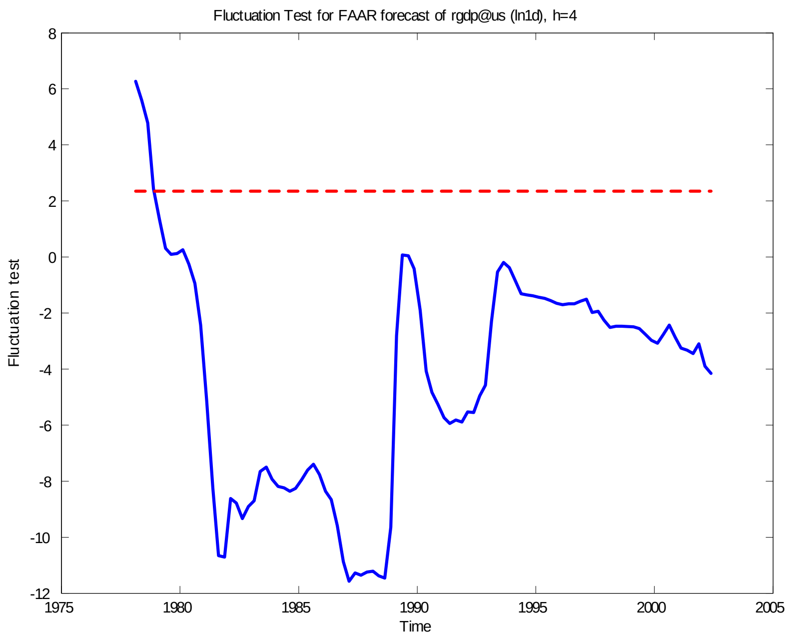


Figure 8, Panel G. (Output) Fluctuation Test on FAAR vs. AR Model



Notes to Figures.

Figure 1 reports scatterplots of the p-values of the traditional Granger-causality tests (on the horizontal axis) and of Rossi's (2005) Granger-causality test robust to instabilities (on the vertical axis). Each dot in the figure corresponds to one of the series that we consider. The dotted lines represent p-values of 5%.

Figure 2 reports a scatterplot of the p-values of the traditional MSE-t test using Giacomini and White's (2006) critical values (on the horizontal axis) and of the Giacomini and Rossi's (2010a) Fluctuation test (on the vertical axis).

Figure 3 reports scatterplots of the p-values of the traditional Granger-causality tests (on the horizontal axis) and of the traditional MSE-t test using Giacomini and White's (2006) critical values (on the vertical axis). Each dot in the figure corresponds to one of the series that we consider. The dotted lines represent p-values of 5%.

Figure 4 reports a scatterplot of the p-values of Rossi's (2005) Granger-causality test robust to instabilities (on the horizontal axis) and of Giacomini and Rossi's (2010a) Fluctuation test (on the vertical axis). Each dot in the figure corresponds to one of the series that we consider. The dotted lines represent p-values of 5%.

Figure 5 reports the percentage of predictors whose Giacomini and Rossi' (2010a) Fluctuation test is outside the critical value at each point in time.

Figure 6 reports a scatterplot of the p-values of the traditional Mincer and Zarnowitz's (1969) tests (on the horizontal axis) and of Rossi and Sekhposyan's (2011b) Fluctuation rationality test (on the vertical axis).

Figure 7 reports a scatterplot of the p-values of the traditional forecast unbiasedness tests (on the horizontal axis) and of Rossi and Sekhposyan's (2011b) Fluctuation unbiasedness test (on the vertical axis).

Figure 8 reports plots the Fluctuation test over time for each of the models that we consider. Panels A-D report results for forecasting inflation, and Panels E-G report results for forecasting output.