Robust Inference under Time-Varying Volatility: A Real-Time Evaluation of Professional Forecasters*

Matei Demetrescu[a], Christoph Hanck[b] and Robinson Kruse[c]

[a]Christian-Albrechts-University of Kiel† [b]University of Duisburg-Essen‡
[c]University of Cologne and CREATES, Aarhus University§

January 31, 2018

Abstract

Forecast evaluation is a long-standing issue in applied econometrics. Standard tests suffer however from the presence of time-varying volatility in many applications. Besides heteroskedasticity, we tackle the important issues of time-variation in relative forecast ability and estimation uncertainty. To this end, we study the fluctuation test of Giacomini and Rossi (2010) and of two new CUSUM- and Cramér-von Mises based tests. While "fixed-b" arguments (Kiefer and Vogelsang, 2005) provide refinements in the use of heteroskedasticity and autocorrelation consistent variance estimators, the resulting limiting distributions of test statistics depend on the unconditional variance changes over time for both small-b and fixed-b approaches. To restore asymptotically pivotal inference, we employ a wild bootstrap approach. After establishing necessary theoretical results, simulations quantify the size distortions from using the original fixed-b approach and show the suggested bootstrap to work reliably. The empirical part studies the (time-varying) superiority of professional forecasters relative to naive no-change predictions in real-time. We exploit the most comprehensive database of the Survey of Professional Forecasters (SPF) and analyze forecasts for several key macroeconomic variables over a sample from 1969 to 2017. Our findings suggest that not accounting for heteroskedasticity seriously affects outcomes of tests for equal predictive ability and time-variation: wild bootstrap inference yields convincing evidence for the superiority of the SPF in most cases, while tests using asymptotic critical values provide remarkably less. Moreover, we find significant evidence for time-variation of relative predictive power; the dominance of the SPF appears to weaken considerably after the "Great Moderation".

Keywords: Forecast evaluation, Hypothesis testing, HAC estimation, Structural breaks, Bootstrap

JEL classification: C12 (Hypothesis Testing), C32 (Time-Series Models), C51 (Forecasting Models)

*The authors would like to thank three anonymous referees and the editor Barbara Rossi for their very constructive comments. Seminar and conference participants in Aarhus, Bath, Cologne, Konstanz, Maastricht and Münster, in particular Ulrich Müller provided very useful comments and remarks. The first two authors gratefully acknowledge the support of the German Research Foundation (DFG) through the projects DE 1617/4-2 and HA 6766/2-1. Robinson Kruse gratefully acknowledges financial support from CREATES funded by the Danish National Research Foundation (DNRF78).

†Institute for Statistics and Econometrics, Christian-Albrechts-University of Kiel, Olshausenstr. 40-60, D-24118 Kiel, Germany, e-mail address: mdeme@stat-econ.uni-kiel.de.

‡Faculty of Economics and Business Administration, University of Duisburg-Essen, Universitätsstraße 12, D-45117 Essen, Germany, e-mail address: christoph.hanck@vwl.uni-due.de.

§**Corresponding author:** University of Cologne, Faculty of Management, Economics and Social Science, Albert-Magnus-Platz, 50923 Cologne, Germany, e-mail address: kruse-becher@wiso.uni-koeln.de and CREATES, Aarhus University, School of Economics and Management, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark, e-mail address: rkruse@creates.au.dk.

1 Introduction

Forecasting plays a crucial role in economics, finance and many other disciplines. Policy makers, firms, investors and households have various needs for macroeconomic predictions. Several macroeconomic forecasts are available to the public, e.g., from international institutions like the IMF and OECD, governmental forecasts like 'Green Book' forecasts from the Federal Reserve and those created by commercial forecasters (e.g. Blue Chip Economic Indicators, Data Resources Inc. and the Survey of Professional Forecasters). We shall focus on the Survey of Professional Forecasters (SPF), which publishes quarterly forecasts for key macroeconomic variables since 1968. It is the most comprehensive database available to assess the performance of professional forecasters and widely used in academic research. A fundamental question is then whether SPF forecasts outperform simple alternatives. An extensive study by Zarnowitz and Braun (1993) reveals e.g. that SPF forecasts perform well in comparison to standard time series models (see also Croushore, 1993; Stark, 2010). With data from 1969 to 2017, we re-evaluate SPF forecasts for US output growth, GDP deflator inflation, unemployment rate and changes in housing starts using robust inference methods.

The long evaluation period contains subsamples with structural breaks mainly in connection to the "Great Moderation", but also with respect to the "Great Financial Crisis". The "Great Moderation" is a period of considerable reduction in macroeconomic volatility, but also of a sharp decline in predictability (Campbell, 2007). The "Great Financial Crisis" did increase volatility, yet less is known about its consequences on predictability. The changing macroeconomic volatility and predictability have important implications for forecast evaluation tests. While the first feature typically leads to time-varying volatility in forecast error loss differentials, the second might imply an instability of its mean. Ignoring these features may lead to significant size distortions and power losses. There is a rich literature on forecasting in unstable environments (e.g. Giacomini and Rossi, 2010). We develop tests for the cases of (i) constant relative forecast performance and (ii) time-variation by considering CUSUM and Cramér-von Mises statistics alongside the fluctuation test by Giacomini and Rossi (2010). These tests are robust to time-varying volatility (via the wild bootstrap) and may take estimation errors into account (by appropriate modifications of the bootstrap algorithms).

All tests use the "fixed-b" paradigm as proposed by Kiefer and Vogelsang (2005). This delivers more accurately sized tests than the standard ones based on heteroskedasticity- and autocorrelation consistent [HAC] standardization. The HAC framework—see the seminal contributions of Newey and West (1987) and Andrews (1991)—permits to use critical values from standard distributions,

like the χ^2 or standard normal. These asymptotic distributions, however, turn out to be rather poor approximations to the actual finite-sample distributions. As a consequence, substantial size distortions are likely to arise in practice. In particular, test results turn out to be sensitive to the choice of bandwidth $B \to \infty$ and kernel k employed for estimating the long-run variance. The poor performance of the asymptotic approximation can be explained by the "small-b" requirement that a vanishing fraction $b := B/T \to 0$ of the number of observations T be used, while of course $b > 0$ in finite samples. To tackle this issue, Kiefer et al. (2000); Kiefer and Vogelsang (2002a,b, 2005) propose "fixed-b" asymptotics, in which it is not required that $b \to 0$. This leads to nonstandard distributions (reviewed in Section 2) for the test statistics. Conveniently and unlike in the standard small-b framework, the new distributions reflect the choice of B and k even in the limit. The above papers convincingly demonstrate that the new distributions provide substantially better approximations to actual finite-sample distributions. In fact, the usefulness of such procedures has spawned an active literature; recent contributions include Sun et al. (2008), Yang and Vogelsang (2011), Vogelsang and Wagner (2013) or Sun (2014). Choi and Kiefer (2010) suggest the use of Diebold and Mariano (1995) statistics with fixed-b critical values; see also Li and Patton (2015).

Time-varying variances, however, change fixed-b limiting distributions and thus lead to a loss of asymptotic pivotality; see Müller (2014, p. 314). This actually emphasizes the strength of the fixed-b approach, as it implies that the variability of the variances—influencing finite-sample behavior—is reflected in the limiting distribution, but comes at the cost of different critical values for the tests of interest. Such time-varying variances are pervasive in applied work in general and in our empirical application in Section 4 specifically.[1] We first show how time-varying volatilities change the fixed-b asymptotics of tests for equal predictive accuracy. The fixed-b Diebold and Mariano (1995) statistic is not pivotal in the limit, but we establish the validity of a wild bootstrap correction (Section 2.1). Then, we address the issue of changes in the relative predictive power of competing forecasts and discuss the small-b and fixed-b asymptotics of the fluctuation test of Giacomini and Rossi (2010) as well as of two new CUSUM- and Cramér-von Mises based test statistics (Section 2.2). We also show how to implement the wild bootstrap to obtain asymptotically correct critical values under time-varying variances, as neither the small-b nor the fixed-b approach deliver pivotality here. Section 2.3 discusses the issue of estimation error. In particular, we characterize the additional terms affecting

[1] Indeed, Groen et al. (2013) find that structural breaks in the variance play an important role for real-time inflation forecasting. More generally, time-varying volatility is present in many macroeconomic (see e.g. Stock and Watson, 2002; Sensier and van Dijk, 2004; Justiniano and Primiceri, 2008; Clark and Ravazzolo, 2015) and financial (see among others Guidolin and Timmermann, 2006; Rapach and Strauss, 2008; Amado and Teräsvirta, 2013) time series such as economic growth, inflation rates and excess returns.

the fixed-b distribution of the tests for a fairly general class of GMM estimators and develop suitable wild bootstrap algorithms replicating these features of the asymptotic distribution.

Section 3 quantifies finite-sample distortions due to time-varying volatility. They are considerable, even as the sample size increases. At the same time, the bootstrap is shown to work well. Moreover, we find the CUSUM- and Cramér-von Mises tests to be more powerful than the fluctuation test.

Section 4 returns to the empirical research question. We compare the predictive ability of survey forecasts to a naive no-change approach. We focus on nowcasts, one-quarter and one-year ahead forecasts and evaluate these by considering the first and the final release of data. Overall, we find forecast error loss differentials to exhibit substantial heteroskedasticity. This is expected to have a direct impact on test decisions when comparing outcomes of traditional and our new robust tests. While the bootstrap provides strong evidence for the superiority of SPF forecasts (especially for nowcasts), there are notably fewer rejections when using asymptotic critical values. Our findings strongly suggest that SPF forecasts perform better early in the sample, but also that this advantage shrank considerably in the 1980s, leading to equal predictive ability starting in the mid-1980s. There are signs of recoveries of forecast superiority around 2000 for unemployment and GDP deflator inflation. We discuss our findings in relation to the literature on SPF accuracy, in general as well as with emphasis on the loss in relative predictability related to the "Great Moderation".

Section 5 concludes. The online appendices A-D collect proofs (unless indicated otherwise in the main text), other derivations, additional simulation results and further empirical results.

2 Fixed-b inference under time-varying volatility

2.1 The baseline case

We test the null of equal predictive ability of two competing forecasts, either generated by models or derived from surveys. We follow Giacomini and White (2006) and examine the loss differentials

$$\Delta\mathcal{L}_{t,h} = \mathcal{L}_{t,h}\left(z_{t+h}, f_{1,t,h}\right) - \mathcal{L}_{t,h}\left(z_{t+h}, f_{2,t,h}\right)$$

where $f_{i,t,h}$, $i = 1, 2$, denote the competing h-step ahead forecasts at time t for the target series z and $\mathcal{L}_{t,h}$ the loss function relevant at time t for horizon h, $t = 1, \ldots, P$. We shall not assume a specific loss function but work with generic differentials directly. Here, $\Delta\mathcal{L}_{t,h}$ may be nonstationary

due to nonstationarity in the series to be forecasted, the forecast series itself, or even changes in the loss function (such as different weights attached to the losses at different t).

The forecasts $f_{i,t,h}$ depend on the observed sample of z and on parameters $\boldsymbol{\theta}_i$. Should $\boldsymbol{\theta}_i$ be unknown, we have $\hat{f}_{i,t,h} = f_{i,h}\big(\boldsymbol{x}_{i,t}, \hat{\boldsymbol{\theta}}_{i,h,(t)}\big)$ where $\boldsymbol{x}_{i,t}$ are a vector of predictors and the notation $\hat{\boldsymbol{\theta}}_{i,h,(t)}$ emphasizes that one may update the estimators as t increases. Often, though, one focusses on one horizon h at a time, and we shall suppress the dependence of $f_{i,t,h}$, $\mathcal{L}_{t,h}$ and $\hat{\boldsymbol{\theta}}_{i,h,(t)}$ on h.

Tests of equal conditional predictive ability are derived by leveraging the observed loss differentials with a vector $\boldsymbol{h}_t$ of K suitable test functions (which are measurable w.r.t. the relevant information set; see Giacomini and White, 2006); in this framework, $\boldsymbol{h}_t = 1$ leads to testing unconditional predictive ability. Let $\boldsymbol{y}_t = \boldsymbol{h}_t \Delta\mathcal{L}_t$ for all relevant t, such that the null becomes

$$H_0 : \mathrm{E}\left(\boldsymbol{y}_t\right) = \mathrm{E}\left(\boldsymbol{h}_t \Delta\mathcal{L}_t\right) \equiv \boldsymbol{\mu} = \mathbf{0}.$$

The baseline alternative is $\boldsymbol{\mu} \neq \mathbf{0}$; in the univariate case we shall also consider one-sided alternatives. In Section 2.2, we follow Giacomini and Rossi (2010) and allow for time-variation in relative forecasting ability under the alternative, i.e. for changes in $\boldsymbol{\mu}$ over time. In the simplest case, $\boldsymbol{\theta}_i$ is known, and we write $f_{i,t} = f_i\left(\boldsymbol{x}_{i,t}, \boldsymbol{\theta}_i\right)$ as an "ideal forecast" vs. $\hat{f}_{i,t} = f_i\big(\boldsymbol{x}_{i,t}, \hat{\boldsymbol{\theta}}_{i,(t)}\big)$ based on estimated parameters. We return to the issue of estimation error in Section 2.3.

Testing the null $\boldsymbol{\mu} = \mathbf{0}$ is done in the baseline case via a Wald-type statistic (Diebold and Mariano, 1995). With P denoting the number of (pseudo) out-of-sample predictions available, define

$$\mathcal{T}_K = \frac{1}{P}\left(\sum_{t=1}^{P} \boldsymbol{y}_t\right)' \hat{\boldsymbol{\Omega}}^{-1} \left(\sum_{t=1}^{P} \boldsymbol{y}_t\right). \tag{1}$$

Here, given suitable choices for the kernel k and the bandwidth B (see Newey and West, 1987; Andrews, 1991), $\hat{\boldsymbol{\Omega}} = \sum_{j=-P+1}^{P-1} k\left(j/B\right)\hat{\boldsymbol{\Gamma}}_j$ is a HAC covariance matrix estimator with $\hat{\boldsymbol{\Gamma}}_{|j|} = P^{-1}\sum_{t=|j|+1}^{P}\left(\boldsymbol{y}_t - \bar{\boldsymbol{y}}\right)\left(\boldsymbol{y}_{t-|j|} - \bar{\boldsymbol{y}}\right)'$ and $\hat{\boldsymbol{\Gamma}}_{-|j|} = \hat{\boldsymbol{\Gamma}}'_{|j|}$. Standard regularity conditions assumed, the small-b limiting distribution is χ^2_K; see Diebold and Mariano (1995). Although this asymptotic result does not depend on the particular choice of k and b, Kiefer and Vogelsang (2005) show that the finite-sample dependence on k and b translates into poor finite-sample behavior of $\mathcal{T}_1$. This leads Choi and Kiefer (2010) to develop fixed-b asymptotics for Diebold and Mariano type tests.

To make the dependence of the distribution of $\mathcal{T}_K$ on k and B explicit, Kiefer and Vogelsang (2005) let $b \in (0, 1]$ in the limit. The resulting limiting distribution is free of nuisance parameters (any

scale matrix cancelling out), but is nonstandard, depending on k and B (via b). E.g. for $K = 1$,

$$\mathcal{T}_1 \xrightarrow{d} \mathcal{B}_{k,b} \qquad \text{with} \qquad \mathcal{B}_{k,b} = W^2(1)/\Theta_{k,b}(W) \quad \text{and}$$

$$\Theta_{k,b}(W) \equiv \begin{cases} -\int_0^1 \int_0^1 \frac{1}{b^2} k''\left(\frac{r-s}{b}\right) \bar{W}(r)\bar{W}(s)\,\mathrm{d}r\mathrm{d}s & \text{for kernels with smooth derivatives} \\ \frac{2}{b}\int_0^1 \bar{W}(r)^2\mathrm{d}r - \frac{2}{b}\int_0^{1-b} \bar{W}(r+b)\bar{W}(r)\mathrm{d}r & \text{for the Bartlett kernel,} \end{cases}$$

where $\bar{W}(s) \equiv W(s) - sW(1)$ for a standard Wiener process $W(s)$. The corresponding critical values for $\mathcal{T}_1$ are tabulated as a function of k and b in Kiefer and Vogelsang (2005). For $b \to 0$, we have $\Theta_{k,b}(W) \xrightarrow{d} 1$ and $\mathcal{B}_{k,b} \xrightarrow{d} \chi^2_1$. In this sense, small-$b$ asymptotics are a particular case.

The functional $\mathcal{B}_{k,b}$ depends on the entire path of the Wiener process $W(s)$ – and not only on $W(1)$, like for small-b. As we show below, this has important consequences for fixed-b when the volatility of $\boldsymbol{y}_t$ varies in time. The following assumption states the variance non-stationarities we allow for.

Assumption 1. *Let $\boldsymbol{y}_t = \boldsymbol{\mu} + \mathbf{G}(t/T)\,\boldsymbol{v}_t$ where $\mathbf{G}(s)$ is a matrix of deterministic, piecewise Lipschitz functions, full-rank at all $s \in [0,1]$. Furthermore, $\boldsymbol{v}_t$ has zero mean and unit long-run covariance matrix, and is $L_{2+\delta}$-bounded for some $\delta > 0$, strictly stationary and strong mixing with mixing coefficients $\alpha(j)$ satisfying $\sum_{j\geq 0} \alpha(j)^{1/p-1/(2+\delta)} < \infty$ for some $2 < p < 2+\delta$.*

The following lemma, whose proof follows from Smeekes and Urbain (2014, Lemma 1) and is omitted, describes the resulting limiting behavior of $P^{-1/2}\sum_{t=1}^{[sP]} \boldsymbol{y}_t$.

Lemma 1. *Let $\boldsymbol{W}$ a vector of independent standard Wiener processes. Under Assumption 1 and the null $\boldsymbol{\mu} = \mathbf{0}$, $P^{-1/2}\sum_{t=1}^{[sP]} \boldsymbol{y}_t \Rightarrow \int_0^s \mathbf{G}(r)\mathrm{d}\boldsymbol{W}(r) \equiv \boldsymbol{B}_{\mathbf{G}}(s)$ as $P \to \infty$.*

The process $\boldsymbol{B}_{\mathbf{G}}(s)$ is Gaussian with independent, zero-mean increments, but not a Brownian motion as its quadratic variation process $[\boldsymbol{B}_{\mathbf{G}}](s) = \int_0^s \mathbf{G}(r)\mathbf{G}'(r)\mathrm{d}r$ is nonlinear due to time-variation of $\mathbf{G}(\cdot)$. Still, $\bar{\boldsymbol{\Omega}} = \int_0^1 \mathbf{G}(s)\mathbf{G}'(s)\mathrm{d}s$ may be interpreted as "average" long-run covariance matrix of $\boldsymbol{y}_t$. The process $\boldsymbol{B}_{\mathbf{G}}$ being Gaussian, it then holds true that $\boldsymbol{B}_{\mathbf{G}}(1) \sim \mathcal{N}\left(\mathbf{0}, \bar{\boldsymbol{\Omega}}\right)$, ensuring asymptotic pivotality of the small-b approach (since it can be shown that $\hat{\boldsymbol{\Omega}} \xrightarrow{p} \bar{\boldsymbol{\Omega}}$ for small-b; see Cavaliere, 2004, Assumption K and Lemma 4 for the univariate case). The finite-sample influence of $\mathbf{G}(s)$ is not removed, though. This is more accurately reflected by the fixed-b approach, as shown in

Proposition 1. *Under the assumptions of Lemma 1, it holds for $P \to \infty$, $B/P \to b \in (0,1]$, that*

$$\mathcal{T}_K \xrightarrow{d} \boldsymbol{B}'_{\mathbf{G}}(1)\,\boldsymbol{\Theta}^{-1}_{k,b}(\boldsymbol{B}_{\mathbf{G}})\,\boldsymbol{B}_{\mathbf{G}}(1),$$

where, for any process $\boldsymbol{X}(s)$ *with a.s. continuous paths and* $\bar{\boldsymbol{X}}(s) = \boldsymbol{X}(s) - s\boldsymbol{X}(1)$,

$$\boldsymbol{\Theta}_{k,b}(\boldsymbol{X}) \equiv \begin{cases} -\frac{1}{b^2}\int_0^1\int_0^1 k''\left(\frac{r-s}{b}\right)\bar{\boldsymbol{X}}(r)\bar{\boldsymbol{X}}(s)'\,\mathrm{d}r\mathrm{d}s \\ \frac{1}{b}\left(2\int_0^1 \bar{\boldsymbol{X}}(r)\bar{\boldsymbol{X}}(r)'\mathrm{d}r - \int_0^{1-b}\bar{\boldsymbol{X}}(r+b)\bar{\boldsymbol{X}}(r)'\,\mathrm{d}r - \int_0^{1-b}\bar{\boldsymbol{X}}(r)\bar{\boldsymbol{X}}(r+b)'\,\mathrm{d}r\right) \end{cases}$$

for kernels with smooth derivatives and the Bartlett kernel, respectively.

We employ a wild bootstrap procedure to account for heteroskedasticity. While there are alternative ways to deal with time-varying (co)variances, we find in related work (Demetrescu et al., 2017) that the wild bootstrap's performance is superior. The algorithm is as follows:

Algorithm 1

1. Generate T iid standardized random variables r_t^*, where $\mathrm{E}\left(|r_t|^k\right) < \infty\ \forall k \in \mathbb{N}$. Typical choices are the Gaussian, Rademacher, or Mammen (1993) distributions.
2. Generate the wild bootstrap sample $\boldsymbol{y}_t^* = (\boldsymbol{y}_t - \bar{\boldsymbol{y}})\, r_t^*$ (a *scalar* r_t^* preserves the covariance matrix of $\boldsymbol{y}_t$ in the bootstrap world). Compute the bootstrap statistic $\mathcal{T}_K^*$ based on $\boldsymbol{y}_t^*$, i.e.

$$\mathcal{T}_K^* = \frac{1}{P}\left(\sum_{t=1}^P \boldsymbol{y}_t^*\right)' \hat{\boldsymbol{\Omega}}^{*-1}\left(\sum_{t=1}^P \boldsymbol{y}_t^*\right),$$

   $\hat{\boldsymbol{\Omega}}^* = \sum_{j=-P+1}^{P-1} k\left(j/B\right)\hat{\boldsymbol{\Gamma}}_j^*$, $\hat{\boldsymbol{\Gamma}}_{|j|}^* = \frac{1}{P}\sum_{t=|j|+1}^P (\boldsymbol{y}_t^* - \bar{\boldsymbol{y}}^*)(\boldsymbol{y}_{t-|j|}^* - \bar{\boldsymbol{y}}^*)'$ and $\hat{\boldsymbol{\Gamma}}_{-|j|}^* = \hat{\boldsymbol{\Gamma}}_{|j|}^{*\prime}$.
3. Repeat Steps 1-2 to obtain a set of M resampled statistics $\left\{\mathcal{T}_{K,m}^*\right\}_{m=1,\ldots,M}$ and use their $(1-\alpha)$-quantile, say $q_{1-\alpha}^*$, as critical value.

Proposition 2. *Under Assumption 1, the null* $\boldsymbol{\mu} = \boldsymbol{0}$ *and* $\mathrm{E}\left(\boldsymbol{v}_t\boldsymbol{v}_t'\right) = c \cdot I_K$ *with* $c > 0$, *it holds as* $M, P \to \infty$ *that* $P\left(\mathcal{T}_K \geq q_{1-\alpha}^*\right) \xrightarrow{p} \alpha$.

Remark 1. The wild bootstrap is asymptotically valid under the additional condition that $\mathrm{E}\left(\boldsymbol{v}_t\boldsymbol{v}_t'\right) = c \cdot I_K$, namely that the covariance and long-run covariance matrices of $\boldsymbol{v}_t$ are proportional. This is trivially fulfilled in the case $K = 1$ of comparing unconditional predictive accuracy, and may often be justified in the multivariate case as well. Should it be violated, one should resort to a sieve wild bootstrap (see e.g. Cavaliere et al., 2010, for an implementation in co-integrated models with time-varying volatility) or, in a less parametric vein, to a block wild bootstrap (Smeekes and Urbain, 2014). Our focus being on univariate tests in Sections 3 and 4, we omit the details.

Remark 2. An examination of the proof in Appendix B reveals that, under $\boldsymbol{\mu} \neq \boldsymbol{0}$, $q_{1-\alpha}^*$ is bounded in probability while $\mathcal{T}_K \xrightarrow{p} \infty$, so that consistency of the bootstrap test is given.

Remark 3. We focus on deterministic variance changes here; nonstationary stochastic volatility leads to a representation similar to that in Lemma 1, $\int_0^s \mathbf{G}(r)\mathrm{d}\boldsymbol{W}(r)$, but with random $\mathbf{G}$. Cavaliere and Taylor (2009) show the wild bootstrap to be valid in several such situations as well.

2.2 Time variation in relative forecast ability

We now turn our attention to the situation in which $\mathrm{E}(\boldsymbol{y}_t) = \boldsymbol{\mu}_t$ is time-varying under the alternative. As pointed out by Giacomini and Rossi (2010) in a univariate unconditional setup, one may expect some loss of power and reduced interpretability of the outcomes based on $\mathcal{T}_K$.

The first method used here for such situations is the fluctuations test of Giacomini and Rossi (2010). In a K-variate setup, the procedure is as follows. Compute for each $t = S/2+1, \ldots, P-S/2+1$ a moving-window based version of (1),

$$F_{t,S} = \frac{1}{S}\left(\sum_{j=t-S/2}^{t+S/2-1} \boldsymbol{y}_j\right)' \hat{\boldsymbol{\Omega}}^{-1} \left(\sum_{j=t-S/2}^{t+S/2-1} \boldsymbol{y}_j\right),$$

with $\hat{\boldsymbol{\Omega}}$ based on all P observations available (see also Giacomini and Rossi, 2010). Consider

$$\mathcal{T}_K^F = \max_{t\in\{S/2+1,\ldots,P-S/2+1\}} F_{t,S}, \qquad S/P = \nu \in (0,1), \tag{2}$$

rejecting for large values of the test statistic. The limiting distribution is given under the assumption that S is a fixed fraction ν of P. For $K > 1$, the test in (2) is two-sided by construction, a consequence of the employed quadratic form; one may then use the procedure suggested by Giacomini and White (2006) to decide which forecast is better in which period. Yet, more importantly, the limit distribution is affected by time-varying volatility, as shown in Proposition 3 below.

We consider two additional solutions for dealing with situations where the relative predictive power varies in time, a CUSUM-type and a Cramér-von Mises functional.[2] The statistics are computed as follows. First, the CUSUM-type statistic is directly based on the partial sums of $\boldsymbol{y}_t$,[3]

$$\mathcal{T}_K^Q = \max_{1\le t\le P} \frac{1}{\sqrt{P}} \sqrt{\boldsymbol{S}_t' \hat{\boldsymbol{\Omega}}^{-1} \boldsymbol{S}_t} \qquad \text{with} \qquad \boldsymbol{S}_t = \sum_{j=1}^{t} \boldsymbol{y}_j. \tag{3}$$

[2] These appear to be more popular in the statistical literature, with prominent econometric exceptions such as the KPSS test for stationarity, and have the advantage of not requiring specification of the tuning parameter ν.

[3] The (perhaps more familiar) CUSUM statistic for a break in mean involves $\boldsymbol{S}_t/t - \boldsymbol{S}_P/P$. This effectively demeans the series, and such a test is rather for a break in relative predictive power. We however test for departures from the null $\mathbf{0}$ rather than from a constant unknown mean, so centering $\boldsymbol{S}_t$ at $\mathbf{0}$ is more natural here.

Also, for $K = 1$, one may work directly with the maximum and minimum of the normalized partial sums to obtain one-sided tests. Second, the Cramér-von Mises statistic: with the same $\boldsymbol{S}_t = \sum_{j=1}^{t} \boldsymbol{y}_j$ (and the same remark on demeaning), compute

$$\mathcal{T}_K^C = \frac{1}{P^2} \sum_{t=1}^{P} \boldsymbol{S}_t' \hat{\boldsymbol{\Omega}}^{-1} \boldsymbol{S}_t. \tag{4}$$

For some vector process $\boldsymbol{X}$ and (stochastic) matrix $\mathbf{T}$, (a.s.) invertible, define the functionals

$$\mathcal{F}(\boldsymbol{X}, \mathbf{T}) = \sup_{s \in [\nu/2; 1-\nu/2]} \frac{1}{\nu} \left(\boldsymbol{X}\left(s + \frac{\nu}{2}\right) - \boldsymbol{X}\left(s - \frac{\nu}{2}\right) \right)' \mathbf{T}^{-1} \left(\boldsymbol{X}\left(s + \frac{\nu}{2}\right) - \boldsymbol{X}\left(s - \frac{\nu}{2}\right) \right),$$

$$\mathcal{Q}(\boldsymbol{X}, \mathbf{T}) = \sup_{s \in [0,1]} \sqrt{\boldsymbol{X}'(s)\, \mathbf{T}^{-1} \boldsymbol{X}(s)} \qquad \text{and} \qquad \mathcal{C}(\boldsymbol{X}, \mathbf{T}) = \int_0^1 \boldsymbol{X}'(s)\, \mathbf{T}^{-1} \boldsymbol{X}(s)\, \mathrm{d}s.$$

Under small-b asymptotics, Lemma 1 and the continuous mapping theorem yield, analogously to Giacomini and Rossi (2010), $\mathcal{T}_K^F \Rightarrow \mathcal{F}(\boldsymbol{B}_{\mathbf{G}}, \bar{\boldsymbol{\Omega}})$. Moreover, $\mathcal{T}_K^Q \Rightarrow \mathcal{Q}(\boldsymbol{B}_{\mathbf{G}}, \bar{\boldsymbol{\Omega}})$ and $\mathcal{T}_K^C \Rightarrow \mathcal{C}(\boldsymbol{B}_{\mathbf{G}}, \bar{\boldsymbol{\Omega}})$. Hence, under time-varying $\mathbf{G}(\cdot)$, all three limiting distributions differ from the homoskedastic case even under small-b asymptotics. Proposition 3, whose proof is analogous to that of Proposition 1 and omitted, shows that the distortions do not disappear for fixed-b versions of the tests:

Proposition 3. *Under the assumptions of Proposition 1,*

$$\mathcal{T}_K^F \Rightarrow \mathcal{F}\left(\boldsymbol{B}_{\mathbf{G}}, \boldsymbol{\Theta}_{k,b}(\boldsymbol{B}_{\mathbf{G}})\right), \quad \mathcal{T}_K^Q \Rightarrow \mathcal{Q}\left(\boldsymbol{B}_{\mathbf{G}}, \boldsymbol{\Theta}_{k,b}(\boldsymbol{B}_{\mathbf{G}})\right) \quad \textit{and} \quad \mathcal{T}_K^C \Rightarrow \mathcal{C}\left(\boldsymbol{B}_{\mathbf{G}}, \boldsymbol{\Theta}_{k,b}(\boldsymbol{B}_{\mathbf{G}})\right).$$

Given the dependence on time-varying variances, the application of a wild bootstrap suggests itself. (See also Zhou, 2013 for a related test of constant means.) For $\mathcal{T}_K^F$, $\mathcal{T}_K^Q$ and $\mathcal{T}_K^C$, the implementation of the wild bootstrap is analogous to that of the $\mathcal{T}_K$ test. Concretely, obtain M resamples $\boldsymbol{y}_{t,m}^*$ as in Algorithm 1, compute M bootstrap test statistics $\{\mathcal{T}_{K,m}^{x*}\}_{m=1,\ldots,M}$ for $x = \{F, Q, C\}$ based on $\boldsymbol{y}_{t,m}^*$ and use their $(1-\alpha)$-quantile, say $q_{1-\alpha}^{x*}$, as critical value for the test based on $\mathcal{T}_K^x$. Then, we obtain Proposition 4, whose proof is analogous to that of Proposition 2 and omitted:

Proposition 4. *Under the assumptions of Proposition 3, the null and* $\mathrm{E}\left(\boldsymbol{v}_t \boldsymbol{v}_t'\right) = c \cdot I_K$ *with* $c > 0$*, it holds as* $M, P \to \infty$ *for* $x = \{F, Q, C\}$ *that* $P\left(\mathcal{T}_K^x \geq q_{1-\alpha}^{x*}\right) \xrightarrow{p} \alpha$.

Remark 4. As in Remark 1, one should use a sieve or a block wild bootstrap if the condition $\mathrm{E}\left(\boldsymbol{v}_t \boldsymbol{v}_t'\right) = c \cdot I_K$ is not met.

2.3 Estimation error

As a leading case in forecasting practice, we now consider the case where the unknown $\boldsymbol{\theta}_i$ are recursively estimated.[4] We follow closely the setup pioneered by West (1996).[5] There are R preliminary observations available, which are used for obtaining estimates $\hat{\boldsymbol{\theta}}_{1,(R)}$ and $\hat{\boldsymbol{\theta}}_{2,(R)}$. These are used to set up the forecasts $\hat{f}_{1,R+1}$ and $\hat{f}_{2,R+1}$ which are compared with z_{t+h} for $t = R+1$. The estimation sample is expanded by one observation and repeat the procedure for $t = R+2$. In total, P observations are available for forecast comparison, $z_{R+1+h}, \ldots, z_{R+P+h}$ together with $\hat{f}_{i,R+1}, \ldots, \hat{f}_{i,R+P}$. In this setup, R and P go to infinity jointly, with $P/R \to \pi > 0$ to ensure, as is well known, that the estimation effect is reflected in the asymptotics.

The forecast losses are given by $\mathcal{L}_t\big(z_{t+h}, \hat{f}_{i,t}\big) = \mathcal{L}_t\big(z_{t+h}, f_i\big(\boldsymbol{x}_{i,t}, \hat{\boldsymbol{\theta}}_{i,(t)}\big)\big)$, so one uses

$$\hat{\boldsymbol{y}}_t = \boldsymbol{h}_t\big(\mathcal{L}_t\big(z_{t+h}, \hat{f}_{1,t}\big) - \mathcal{L}_t\big(z_{t+h}, \hat{f}_{2,t}\big)\big), \qquad t = R+1, \ldots, R+P, \tag{5}$$

for testing. We set $\hat{\boldsymbol{y}}_t = \mathbf{0}$ for $1 \leq t \leq R$ since they do not enter the test statistics. Following Giacomini and White (2006) we assume that pseudo-true value $\boldsymbol{\theta}_i$ exist, such that, as $R, P \to \infty$ with $P/R \to \pi$, $\hat{\boldsymbol{\theta}}_{i,(t)} \xrightarrow{p} \boldsymbol{\theta}_i$ for all $t > R$ (see also Assumption 4 below). The "ideal" loss at time t is

$$\mathcal{L}_t\left(z_{t+h}, f_i\left(\boldsymbol{x}_{i,t}, \boldsymbol{\theta}_i\right)\right) = \mathcal{L}_t\left(z_{t+h}, f_{i,t}\right) \ , \ i = 1, 2.$$

In line with the literature (again, see West, 1996), we assume $\mathcal{L}_t$ and f_i to be smooth enough to allow for an evaluation of the estimation noise. The following assumption differs slightly e.g. from the analogous one of West, but is convenient for dealing with the bootstrap later on. Let

$$\mathbf{D}_i(a, \boldsymbol{b}) = \boldsymbol{h}_t \cdot \left.\frac{\partial \mathcal{L}_t}{\partial u_2}\right|_{\substack{u_1 = z_{t+h} \\ u_2 = a}} \left.\frac{\partial f_i}{\partial \boldsymbol{\theta}'}\right|_{\substack{\boldsymbol{x}_{i,t} \\ \boldsymbol{\theta} = \boldsymbol{b}}}.$$

Assumption 2. *There exists $0 < \epsilon < 1/2$ such that, for the neighbourhood $\Phi_P = \times_{i=1,2}\big\{\tilde{\boldsymbol{\theta}}_i : \big\|\tilde{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\big\| < CP^{-1/2+\epsilon}, C > 0\big\}$ of $\left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)'$, it holds as $R, P \to \infty$ with $P/R \to \pi$ that*

$$\sup_{\tilde{\boldsymbol{\theta}}_{1,2} \in \Phi_P; t = R+1, \ldots, P} \left\|\mathbf{D}_i(\tilde{f}_{i,t}, \tilde{\boldsymbol{\theta}}_i) - \mathbf{D}_i(f_{i,t}, \boldsymbol{\theta}_i)\right\| \xrightarrow{p} 0$$

where $\tilde{f}_{i,t} = f_i\big(\boldsymbol{x}_{i,t}, \tilde{\boldsymbol{\theta}}_i\big)$, $i = 1, 2$.

[4] Rolling windows estimation is dealt with analogously; we mention the differences whenever needed.

[5] For (approximately) finite-memory estimators, one may also use the approach of Giacomini and White (2006).

As a consequence, we may write for $i = 1, 2$

$$\mathcal{L}_t\big(z_{t+h}, \hat{f}_{i,t}\big) = \mathcal{L}_t\left(z_{t+h}, f_{i,t}\right) - \mathbf{D}_i(f_{i,t}, \boldsymbol{\theta}_i) \cdot \big(\hat{\boldsymbol{\theta}}_{i,(t)} - \boldsymbol{\theta}_i\big) + o_p\left(1\right), \qquad t = R+1, \ldots, R+P, \quad (6)$$

where the $o_p\left(1\right)$ term is negligible uniformly in t (see the proof of Lemma 3). To describe the effect of the additional terms $\mathbf{D}_i(f_{i,t}, \boldsymbol{\theta}_i) \cdot \big(\hat{\boldsymbol{\theta}}_{i,(t)} - \boldsymbol{\theta}_i\big)$, we make the following assumption.

Assumption 3. *We have as $R, P \to \infty$ with $P/R \to \pi$ that $R^{-1} \sum_{t=1}^{[uR]} \mathbf{D}_i(f_{i,t}, \boldsymbol{\theta}_i) \Rightarrow \mathbf{T}_i(u)$ on $[0, 1+\pi]$, where $\mathbf{T}_i$ is deterministic and Lipschitz.*

The quantities $\mathbf{T}_i$ are defined for convenience for the full sample $t = 1, \ldots, R+P$, but, as can be seen in (6), one only needs observations at times $R+1, \ldots, R+P$. So let

$$\mathbf{H}_i(s) = \left(\mathbf{T}_i(1+s\pi) - \mathbf{T}_i(1)\right)/\pi, \quad s \in (0,1), \qquad \text{such that} \qquad \frac{1}{P}\sum_{t=R+1}^{R+[sP]} \mathbf{D}_i(f_{i,t}, \boldsymbol{\theta}_i) \Rightarrow \mathbf{H}_i(s).$$

A sufficient condition for the negligibility of the estimation effect is that $\mathbf{H}_i(s) = \mathbf{0}$ for all s (see e.g. West, 1996). Verifying whether this holds in a particular application requires additional information beyond the observed forecast errors: the estimation effect depends the examined forecasting methods via $\frac{\partial f_i}{\partial \boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{i,(t)}$. In order to compare two forecasts, one therefore requires information regarding their construction, i.e., information in addition to the point forecasts and the actual realizations.

We next describe suitable corrections when such information is available. We let $\hat{\boldsymbol{\theta}}_{i,(t)}$ be (overidentified) GMM estimators, with at least as many moment conditions N_i as parameters M_i.

Assumption 4. *For $t = R+1, \ldots, R+P$, let the following decomposition hold:*

$$\hat{\boldsymbol{\theta}}_{i,(t)} = \boldsymbol{\theta}_i + \left(\sum_{j=1}^{t} \mathbf{C}'_{i,j,(\boldsymbol{\theta}_i)} \mathbf{W}_{i,(\boldsymbol{\theta}_i)} \sum_{j=1}^{t} \mathbf{C}_{i,j,(\boldsymbol{\theta}_i)}\right)^{-1} \sum_{j=1}^{t} \mathbf{C}'_{i,j,(\boldsymbol{\theta}_i)} \mathbf{W}_{i,(\boldsymbol{\theta}_i)} \sum_{j=1}^{t} \boldsymbol{a}_{i,j,(\boldsymbol{\theta}_i)} + \boldsymbol{r}_{i,t}$$

with symmetric $\mathbf{W}_{i,(\boldsymbol{\theta}_i)} > 0$, where $\sup_{R<t\leq R+P} \|\boldsymbol{r}_{i,t}\| = o_p\big(R^{-1/2}\big)$ as $R, P \to \infty$ with $P/R \to \pi$.

The moment conditions, *jointly* with $\boldsymbol{y}_t$, are specified in a time-varying framework as follows.

Assumption 5. *For $\boldsymbol{\xi}_t = (\boldsymbol{a}'_{1,t,(\boldsymbol{\theta}_1)}, \boldsymbol{a}'_{2,t,(\boldsymbol{\theta}_2)}, (\boldsymbol{y}_t - \boldsymbol{\mu})')' \in \mathbb{R}^{N_1+N_2+K}$, let $\boldsymbol{\xi}_t = \tilde{\mathbf{G}}(t/T)\tilde{\boldsymbol{v}}_t$, where*

$$\tilde{\mathbf{G}}(t/T) = \begin{pmatrix} \mathbf{G}_{11}\left(t/T\right) & \mathbf{G}_{12}\left(t/T\right) & \mathbf{G}_{1y}\left(t/T\right) \\ \mathbf{0} & \mathbf{G}_{22}\left(t/T\right) & \mathbf{G}_{2y}\left(t/T\right) \\ \mathbf{0} & \mathbf{0} & \mathbf{G}\left(t/T\right) \end{pmatrix}$$

and $\tilde{\boldsymbol{v}}_t = \left(\boldsymbol{u}_{1,t}', \boldsymbol{u}_{2,t}', \boldsymbol{v}_t'\right)'$*, with* $\boldsymbol{v}_t$ *and* $\mathbf{G}$ *from Assumption 1, and* $\mathbf{G}_{11}(s)$ *(*$N_1 \times N_1$*) and* $\mathbf{G}_{22}(s)$ *(*$N_2 \times N_2$*) matrices of piecewise Lipschitz functions, full-rank at all* $s \in [0,1]$*. Furthermore,* $\tilde{\boldsymbol{v}}_t$ *satisfies the same conditions as* $\boldsymbol{v}_t$ *in Assumption 1. Finally, there exist matrices* $\mathbf{C}_i(u)$ *of deterministic Lipschitz functions, full-rank for all* $u > 0$*, such that* $R^{-1}\sum_{t=1}^{[uR]} \mathbf{C}_{i,t,(\boldsymbol{\theta}_i)} \Rightarrow \mathbf{C}_i(u)$ *on* $[0, 1+\pi]$*.*

The assumption simply extends Assumption 1 to cover all random components. The upper triangular structure of $\tilde{\mathbf{G}}$ is not restrictive, since its role is to generate arbitrary symmetric, positive definite localized (long-run) covariance matrices for $\boldsymbol{\xi}_t$; this is easily checked to be the case. That we require $\mathrm{E}\left(\boldsymbol{a}_{i,t,(\boldsymbol{\theta}_i)}\right) = \mathbf{0}$ is nothing else than specifying moment conditions for the estimation of $\boldsymbol{\theta}_i$. The dependence on $\boldsymbol{\theta}_i$ comes from possibly having nonlinear moment conditions which are linearized for obtaining the decomposition. Just like in Lemma 1, we obtain (after setting w.l.o.g. $\boldsymbol{y}_t = \boldsymbol{\mu}$ for $1 \leq t \leq R$, since they do not enter the statistics of interest) the following partial sum behavior:

Lemma 2. *Under Assumptions 4 and 5 with* $\boldsymbol{B}_{\mathbf{G}}$ *from Lemma 1 and* $\boldsymbol{y}_t = \boldsymbol{\mu}$ *for* $1 \leq t \leq R$*,* $R^{-1/2}\sum_{t=1}^{[uR]} \boldsymbol{\xi}_t \Rightarrow \int_0^u \tilde{\mathbf{G}}(s)\mathrm{d}\boldsymbol{W}(s) \equiv \left(\boldsymbol{A}_1'(u), \boldsymbol{A}_2'(u), \sqrt{\pi}\boldsymbol{B}_{\mathbf{G}}'\left((\max(r,1)-1)/\pi\right)\right)'$ *on* $[0, 1+\pi]$*.*

This implies a different behavior of the relevant partial sums.

Lemma 3. *Under Assumptions 1-5, under the null and as* $R, P \to \infty$ *with* $P/R \to \pi$*,*

$$\frac{1}{\sqrt{P}}\sum_{t=R+1}^{R+[sP]} \hat{\boldsymbol{y}}_t \Rightarrow \boldsymbol{B}_{\mathbf{G}}(s) + \sqrt{\pi}\sum_{i=1}^{2}(-1)^i\left(\int_0^s \boldsymbol{N}_i'(r)\mathbf{M}_i(r)^{-1}\mathrm{d}\mathbf{H}_i'(r)\right)' \equiv \boldsymbol{B}_{\mathbf{G},\pi}(s)$$

on $[0,1]$*, where* $\mathbf{M}_i(s) \equiv \mathbf{C}_i'(1+\pi s)\,\mathbf{W}_{i,(\boldsymbol{\theta}_i)}\,\mathbf{C}_i(1+\pi s)$ *and* $\boldsymbol{N}_i(s) \equiv \mathbf{C}_i'(1+\pi s)\,\mathbf{W}_{i,(\boldsymbol{\theta}_i)}\,\boldsymbol{A}_i(1+\pi s)$*.*

Remark 5. Note that $\boldsymbol{B}_{\mathbf{G},0}(s) \equiv \boldsymbol{B}_{\mathbf{G}}(s)$, and one recovers the case without estimation for $\pi \to 0$. At the same time, for $\pi \to \infty$, the estimation effect dominates.

Remark 6. Analogous arguments lead for rolling window estimation of $\boldsymbol{\theta}_i$,

$$\hat{\boldsymbol{\theta}}_{i,(t)}^{rol} = \boldsymbol{\theta}_i + \left(\sum_{j=t-R+1}^{t} \mathbf{C}_{i,j,(\boldsymbol{\theta}_i)}'\,\mathbf{W}_{i,(\boldsymbol{\theta}_i)}\sum_{j=t-R+1}^{t}\mathbf{C}_{i,j,(\boldsymbol{\theta}_i)}\right)^{-1}\sum_{j=t-R+1}^{t}\mathbf{C}_{i,j,(\boldsymbol{\theta}_i)}'\,\mathbf{W}_{i,(\boldsymbol{\theta}_i)}\sum_{j=t-R+1}^{t}\boldsymbol{a}_{i,j,(\boldsymbol{\theta}_i)} + \boldsymbol{r}_{i,t}^{rol} \tag{7}$$

with $\boldsymbol{r}_{i,t}^{rol}$ negligible in the sense of Assumption 4, to the following result under the null:

$$\frac{1}{\sqrt{P}}\sum_{t=R+1}^{R+[sP]} \hat{\boldsymbol{y}}_t \Rightarrow \boldsymbol{B}_{\mathbf{G}}(s) + \sqrt{\pi}\sum_{i=1}^{2}(-1)^i\left(\int_0^s \tilde{\boldsymbol{N}}_i'(r)\,\tilde{\mathbf{M}}_i^{-1}(r)\,\mathrm{d}\mathbf{H}_i'(r)\right)' \equiv \boldsymbol{B}_{\mathbf{G},\pi}^{rol}(s)$$

on $[0,1]$, where we define $\tilde{\mathbf{M}}_i(s) \equiv (\mathbf{C}_i(1+\pi s) - \mathbf{C}_i(\pi s))' \, \mathbf{W}_{i,(\boldsymbol{\theta}_i)} \, (\mathbf{C}_i(1+\pi s) - \mathbf{C}_i(\pi s))$ and $\tilde{\boldsymbol{N}}_i(s) \equiv (\mathbf{C}_i(1+\pi s) - \mathbf{C}_i(\pi s))' \, \mathbf{W}_{i,(\boldsymbol{\theta}_i)} \, (\boldsymbol{A}_i(1+\pi s) - \boldsymbol{A}_i(\pi s))$.

Lemma 3 implies, along the lines of the proof of Proposition 3, non-pivotal null distributions:

Proposition 5. *Under the assumptions of Lemma 3 and the null* $\boldsymbol{\mu} = \mathbf{0}$*, we have that*

$$\mathcal{T}_K \xrightarrow{d} \boldsymbol{B}'_{\mathbf{G},\pi}(1) \, \boldsymbol{\Theta}_{k,b}^{-1}(\boldsymbol{B}_{\mathbf{G},\pi}) \, \boldsymbol{B}_{\mathbf{G},\pi}(1) \quad \textit{and}$$

$$\mathcal{T}_K^F \Rightarrow \mathcal{F}\left(\boldsymbol{B}_{\mathbf{G},\pi}, \boldsymbol{\Theta}_{k,b}(\boldsymbol{B}_{\mathbf{G},\pi})\right), \quad \mathcal{T}_K^Q \Rightarrow \mathcal{Q}\left(\boldsymbol{B}_{\mathbf{G},\pi}, \boldsymbol{\Theta}_{k,b}(\boldsymbol{B}_{\mathbf{G},\pi})\right), \quad \mathcal{T}_K^C \Rightarrow \mathcal{C}\left(\boldsymbol{B}_{\mathbf{G},\pi}, \boldsymbol{\Theta}_{k,b}(\boldsymbol{B}_{\mathbf{G},\pi})\right).$$

To correct for inherent non-pivotality via the bootstrap, one must replicate the properties of $\boldsymbol{B}_{\mathbf{G},\pi}(s)$, which depends, among others, on $\mathbf{H}(\cdot)$ and the joint behavior of $\boldsymbol{B}_{\mathbf{G}}$ and $\boldsymbol{A}_i$. Given such information, a wild bootstrap can be used to replicate the above limiting null distributions. One must resort to estimated quantities since $\boldsymbol{\theta}_i$ are unknown, though. While $\hat{\boldsymbol{y}}_t$ is a natural estimator for $\boldsymbol{y}_t$, estimates of $\boldsymbol{a}_{i,t,(\boldsymbol{\theta}_i)}$, $\mathbf{W}_{i,(\boldsymbol{\theta}_i)}$ and $\mathbf{C}_{i,t,\left(\hat{\boldsymbol{\theta}}_{i,(t)}\right)}$, say $\hat{\boldsymbol{a}}_{i,t}$, $\hat{\mathbf{W}}_i$ and $\hat{\mathbf{C}}_{i,t}$, require plugging-in:

Algorithm 2

1. Compute $\hat{\boldsymbol{y}}_t$ from (5) (and, for $t = 1, \ldots, R$, $\hat{\boldsymbol{y}}_t = \mathbf{0}$); for $t = 1, \ldots, R+P$, compute $\hat{\mathbf{C}}_{i,t}$, $\hat{\mathbf{W}}_i$ and $\hat{\boldsymbol{a}}_{i,t}$ as $\mathbf{C}_{i,t,\left(\hat{\boldsymbol{\theta}}_{i,(t)}\right)}$, $\mathbf{W}_{i,\left(\hat{\boldsymbol{\theta}}_{i,(t)}\right)}$ and $\boldsymbol{a}_{i,t,\left(\hat{\boldsymbol{\theta}}_{i,(t)}\right)}$ (alternatively, one may evaluate at $\hat{\boldsymbol{\theta}}_{i,(R+P)}$).
2. For $t = 1, \ldots, R+P$, draw multipliers r_t^* and construct $\left(\boldsymbol{a}_{1,t}^{*\prime}, \boldsymbol{a}_{2,t}^{*\prime}, \boldsymbol{y}_t^{*\prime}\right)'$ as $\left(\hat{\boldsymbol{a}}_{1,t}', \hat{\boldsymbol{a}}_{2,t}', \hat{\boldsymbol{y}}_t'\right)' r_t^*$.
3. Compute for $t = R+1, \ldots, R+P$

$$\hat{\boldsymbol{\theta}}_{i,(t)}^* = \left(\sum_{j=1}^{t} \hat{\mathbf{C}}_{i,j}' \, \hat{\mathbf{W}}_i \sum_{j=1}^{t} \hat{\mathbf{C}}_{i,j} \right)^{-1} \sum_{j=1}^{t} \hat{\mathbf{C}}_{i,j}' \, \hat{\mathbf{W}}_i \sum_{j=1}^{t} \boldsymbol{a}_{i,j}^* + \hat{\boldsymbol{\theta}}_{i,(R+P)}.$$

4. Letting $\hat{f}_{i,t}^* = f_i\left(\boldsymbol{x}_{i,t}, \hat{\boldsymbol{\theta}}_{i,(t)}^*\right)$, compute for $t = R+1, \ldots, R+P$

$$\hat{\boldsymbol{y}}_t^* = \boldsymbol{y}_t^* - \mathbf{D}_1(\hat{f}_{1,t}^*, \hat{\boldsymbol{\theta}}_{1,(t)}^*) \cdot \left(\hat{\boldsymbol{\theta}}_{1,(t)}^* - \hat{\boldsymbol{\theta}}_{1,(R+P)}\right) + \mathbf{D}_2(\hat{f}_{2,t}^*, \hat{\boldsymbol{\theta}}_{2,(t)}^*) \cdot \left(\hat{\boldsymbol{\theta}}_{2,(t)}^* - \hat{\boldsymbol{\theta}}_{2,(R+P)}\right).$$

5. Compute the test statistics of interest using the bootstrap sample $\hat{\boldsymbol{y}}_t^*$, $t = R+1, \ldots, R+P$.
6. Repeat steps 2–5 M times and obtain the desired quantile(s).

Some mild additional conditions are required for establishing the validity of the bootstrap.

Proposition 6. *Let* $\mathbf{W}_{i,(\boldsymbol{\theta}_i)}$ *be continuous in* $\boldsymbol{\theta}_i$*, and, for* $1 \leq t \leq R+P$, $\sup_t \left\| \hat{\mathbf{C}}_{i,t} - \mathbf{C}_{i,t,(\boldsymbol{\theta}_i)} \right\| \xrightarrow{p} 0$. *Moreover,* $\exists \gamma > 0$ *such that* $\sup_t \left\| \mathbf{D}_i(f_{i,t}, \boldsymbol{\theta}_i) \right\| = O_p\left(P^{1/2-\gamma}\right)$ *and* $\sup_t \left\| \hat{\boldsymbol{a}}_{i,t} - \boldsymbol{a}_{i,t,(\boldsymbol{\theta}_i)} \right\| = O_p\left(P^{-\gamma}\right)$.

Under the assumptions of Proposition 5, under the null and if $\mathrm{E}\left(\tilde{\boldsymbol{v}}_t\tilde{\boldsymbol{v}}_t'\right) = c \cdot I_K$ *with* $c > 0$,

$$P\left(\mathcal{T}_K^x \geq q_{1-\alpha}^{x*}\right) \xrightarrow{p} \alpha, \quad x = \{F, Q, C\}, \qquad as \quad M, P \to \infty \text{ with } P/R \to \pi.$$

Remark 7. While the above corrections are feasible when a researcher possesses all the necessary information regarding the construction of the forecast, some external sources (cf. Section 4) only publish point forecasts and actual realizations. Such information is not sufficient to assess the relative strengths of privately constructed forecast *models*. Essentially, the covariance of $\boldsymbol{A}_i$ and $\boldsymbol{B}_{\mathbf{G}}$ is often not known to "outsiders", making it impossible to apply a suitable bootstrap.

Remark 8. In the case of rolling windows estimation errors $\hat{\boldsymbol{\theta}}_{i,(t)}^{rol} - \boldsymbol{\theta}_i$ from (7) with analogous conditions on the components, we compute the estimated components as $\mathbf{C}_{i,t,\left(\hat{\boldsymbol{\theta}}_{i,(t)}^{rol}\right)}$, $\mathbf{W}_{i,\left(\hat{\boldsymbol{\theta}}_{i,(t)}^{rol}\right)}$ and $\boldsymbol{a}_{i,t,\left(\hat{\boldsymbol{\theta}}_{i,(t)}^{rol}\right)}$; for $t \leq R$ we set $\hat{\boldsymbol{\theta}}_{i,(t)}^{rol} = \hat{\boldsymbol{\theta}}_{i,(R)}^{rol}$. Steps 4 and 5 must be modified accordingly.

3 Numerical Evidence

3.1 Setup

This section investigates the finite-sample properties of the different statistics, in view of the asymptotic arguments from Sections 2.1-2.3. Here, we focus on the most general case discussed: we consider both potential time-varying forecasting ability and estimation uncertainty. For concreteness, we shall investigate the simple and widely relevant case of regression-based prediction through competing univariate predictors. Algorithm 3 in Appendix A summarizes the corresponding bootstrap procedure for replicating the non-pivotal distributions from Propositions 3 and 5.

Our main DGP is as follows. We aim to predict an ARMA(1,1)-process $z_t = 0.4z_{t-1} + \epsilon_t + 0.3\epsilon_{t-1}$ through two competing AR(1)-processes $x_{i,t} = 0.5x_{i,t-1} + u_{i,t}$, $i = 1, 2$. The predictions of z_t via the $x_{i,t}$ are obtained by simple (recursive) OLS as in (9), taking $h = 0$ for simplicity. Here, $t = 1, \ldots, R+P$, where $R \in \{100, 200, 300\}$ (estimation sample) and $P \in \{50, 100, 200\}$ (prediction sample) for the size experiments. For power, $P \in \{50, 100, 200, 500\}$, including 500 to shed more light on test consistency. Let $\boldsymbol{u}_t = (\epsilon_t, u_{1,t}, u_{2,t})'$ denote the vector of innovations of z_t and $x_{j,t}$, generated from a multivariate normal distribution. Its correlation matrix is labeled as $\boldsymbol{\Upsilon}_t$. The size experiments take $\boldsymbol{\Upsilon}_t = \boldsymbol{\Upsilon}$ to be an equicorrelation matrix with identical off-diagonal elements $\sigma_u = 0.5$. This yields a scenario in which $x_{1,t}$ and $x_{2,t}$ have equal predictive ability for z_t so that

the null hypothesis of the tests is true. For conciseness, we take $\boldsymbol{h}_t = 1$ throughout and thereby investigate the case of unconditional predictive ability.

In our power experiments we specify two distinct scenarios. First, we consider a time-invariant so-called "Toeplitz" structure for $\boldsymbol{\Upsilon}_t$. More specifically, both ϵ_t and $u_{1,t}$ as well as $u_{1,t}$ and $u_{2,t}$ are correlated (with a correlation coefficient of $\sigma_u \in \{-0.4, -0.2, \ldots, 0.6\}$) while ϵ_t and $u_{2,t}$ are uncorrelated. Thus, $x_{2,t}$ is independent of z_t and therefore has no predictive power, in contrast to $x_{1,t}$. In order to generate time-varying forecasting ability, we specify a simple switch from an equicorrelated matrix to a "Toeplitz" matrix at time $\tau := [R + P/4]$. Thus, the structural break in predictive power emerges from a time-varying correlation matrix $\boldsymbol{\Upsilon}_t$. The break date is located in the first quarter of the prediction sample and renders the DGP practically relevant.[6]

Time-varying variance is introduced by generating a structural break in the covariance matrix by scaling $\boldsymbol{\Upsilon}_t$ by $\delta_1 \in \{1/3, 1, 3\}$ at time $[\zeta \cdot (R+P)]$, where $\zeta \in \{0.3, 0.6, 0.9\}$ (for the size experiments) and at time point $\tau = [R + P/4]$ for the time-varying forecasting ability power experiment. For instance, $\delta_1 = 1/3$ and $\zeta = 0.9$ yield a late downward break in variance.

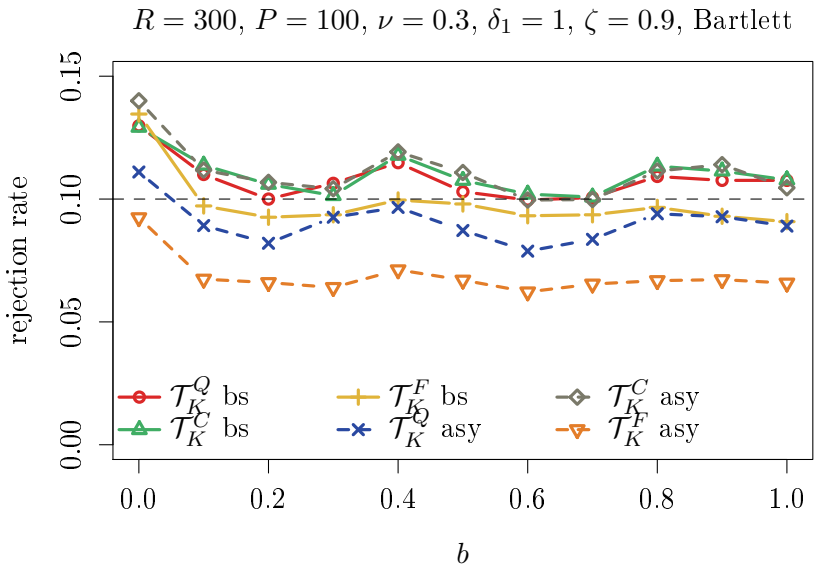
We consider the Bartlett and Quadratic Spectral (QS) kernel, the fixed-bandwidth parameter $b \in \{0, 0.1, 0.2, \ldots, 1\}$ for size experiments and, to reduce the computational burden, $b \in \{0, 0.4, 0.8\}$ for power.[7] Large values of the test statistics provide evidence against equal predictive ability, so that we test against right-tailed alternatives at a nominal significance level of $\alpha = 0.1$. The number of replications equals $5,000$ (size) or $2,500$ (power). We use $M = 500$ bootstrap replications for the wild bootstrap tests. In step 4 of Algorithm 3, we draw r_t^* from the Mammen (1993) two-point distribution. These choices are common in the literature. In view of the findings of Giacomini and Rossi (2010, Table II), we choose a relative window size of $\nu = 0.3$ for $\mathcal{T}_K^F$.

We also investigate "asymptotic" fixed-b tests for completeness. These are fixed-b versions of $\mathcal{T}_K^Q$, $\mathcal{T}_K^C$ and $\mathcal{T}_K^F$ ignoring time-varying variances which thus suffer from non-pivotality. Directly extending the approach of Kiefer and Vogelsang (2005), we obtain fixed-b critical values for these tests from simulating the distributions from Proposition 3 under the traditional assumption that $\mathbf{G}(r) = \boldsymbol{I}$.[8]

[6] We also experimented with later values of τ. Of course, a smaller sample in which the predictors' forecasting ability differs translates into lower power, but the general qualitative conclusions of our study remain the same.

[7] For $b = 0$, we use the automatic estimator for B, $\hat{B} = d(4\hat{\rho}^2(1-\hat{\rho})^{-4}T)^{1/g}$ with $\hat{\rho}$ from an approximating $AR(1)$ model for the series (see Andrews, 1991, eqs. (6.2) and (6.4)). Here, $d = 1.1447$, $g = 3$ for the Bartlett and $d = 1.3221$, $g = 5$ for the QS kernel.

[8] Table 19 in the Appendix reports these critical values.



See (3), (4) and (2) for the CUSUM ($\mathcal{T}_K^Q$), Cramér-von Mises ($\mathcal{T}_K^C$) and fluctuation ($\mathcal{T}_K^F$) statistics. "bs" abbreviates the bootstrap versions, cf. Algorithm 3. "asy" uses standard non-robust critical values, cf. the last paragraph of Section 3.1. R denotes the estimation sample, P the prediction sample, ν the relative window width of $\mathcal{T}_K^F$ (see (2)), δ_1 the post-break variance and ζ the breakfraction. See Section 3.1 for further details. Bartlett denotes the Bartlett kernel, QS is short for quadratic spectral (in later figures).

Figure 1: Size under homoskedasticity, asymptotic and bootstrap tests

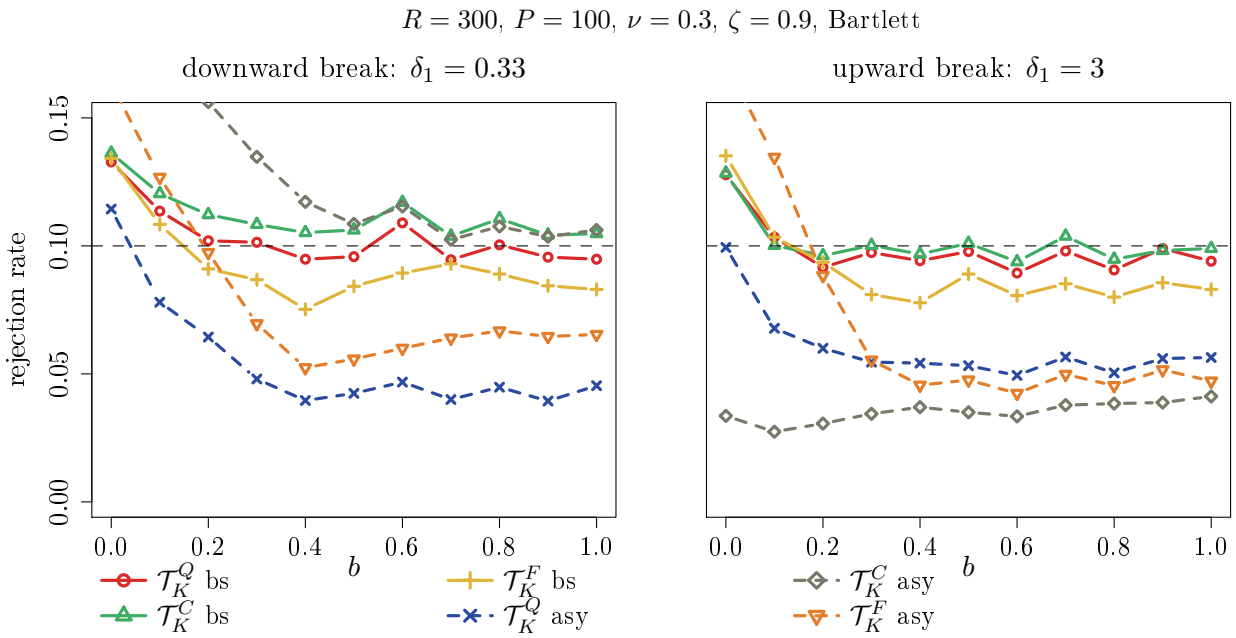
3.2 Size results

When considering the full amount of experiments, we obtain (with $|A|$ the number of elements of a set A) $|R| \cdot |P| \cdot |\delta_1| \cdot |\zeta| \cdot |b| \cdot |\text{kernels}| = 1782$ experiments in total. In the following, we report some selected and representative results from the above grid of parameter and sample sizes.[9]

First, Figure 1 shows that both asymptotic and bootstrap fixed-b tests perform well across b under the benchmark case of homoskedasticity ($\delta_1 = 1$). The asymptotic fluctuation test $\mathcal{T}_K^F$ is a possible exception (but unreported simulations for larger P reveal this to be, as expected, a small-sample phenomenon). Also, the findings are reminiscent of Kiefer and Vogelsang (2005)—while fixed-b tests yield good finite-sample size, there are finite-sample size distortions for the small-b versions (Newey and West, 1987; Andrews, 1991), i.e., for $b = 0$.

The non-pivotality of the asymptotic tests under heteroskedasticity becomes apparent in Figure 2. Here, the break occurs at observation $[0.9 \cdot (300 + 100)] = 360$. The first $R = 300$ preliminary observations have been used for parameter estimation. In particular, the asymptotic tests are undersized as soon as b takes moderate or large values. That is, fixed-b versions of the tests, as

[9] TO THE REFEREES: The full results are of course available upon request. In case of publication, we shall make these available in MonteCarlo objects (R Core Team, 2017; Leschinski, 2017) on the JAE data archive.



See notes to Figure 1.

Figure 2: Size under heteroskedasticity (late break), asymptotic and bootstrap tests

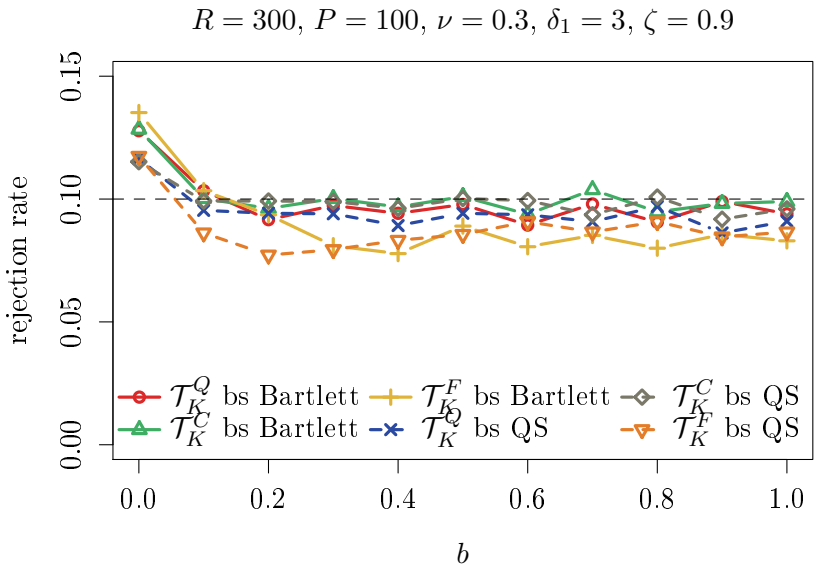
predicted by Proposition 5, no longer provide accurate finite-sample size in the presence of time-varying variance. In turn and as a result of Proposition 6, the bootstrap fixed-b versions maintain good size. Again, they are slightly less successful at correcting the well-known small-sample small-b size distortions. Also, the fixed-b approximations work slightly less well for smaller b, still being close to the standard small-b case ($b = 0$) which is in line with Kiefer and Vogelsang (2005). For about $b > 0.2$, the bootstrap tests generally perform very well.[10]

Focussing on the robust bootstrap tests, Figure 3 reveals that there is little to choose between the Bartlett and QS kernel in terms of size. Both perform similarly well. Figure 4 demonstrates, for $\mathcal{T}_K^Q$, that P has a minor effect on the bootstrap tests. Size appears to improve for the asymptotic tests, but this finding is not robust with respect to other ζ and δ_1.

3.3 Power results

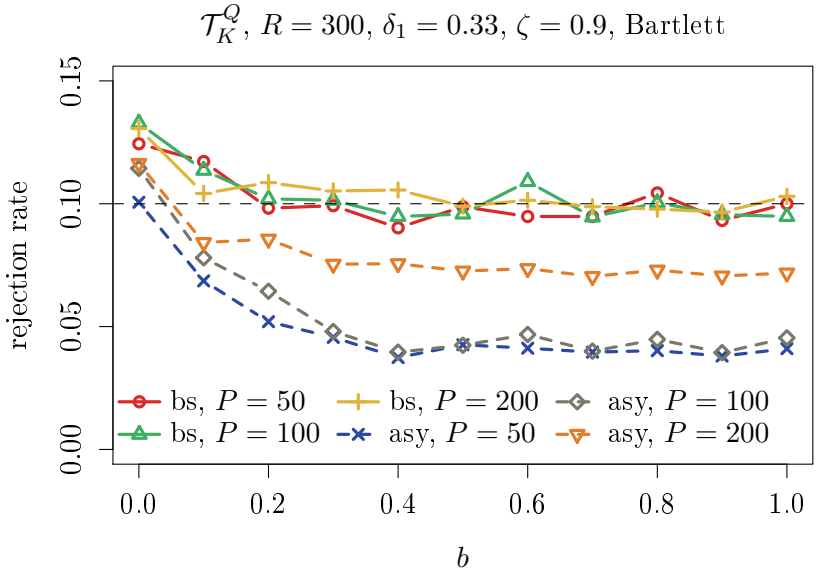
We now discuss the power of the procedures described above. We first consider a few of the $|R| \cdot |P| \cdot |\delta_1| \cdot |\zeta| \cdot |b| \cdot |\text{kernels}| \cdot |\sigma_u| = 3240$ "Toeplitz" experiments. Here, $x_{2,t}$ has no predictive power for z_t over the full sample, unlike $x_{1,t}$.

[10] There is some small and unsystematic variation in the empirical sizes when varying b. We consider this to be due to simulation variability given the relatively small number of Monte Carlo replications for each case.



See notes to Figure 1.

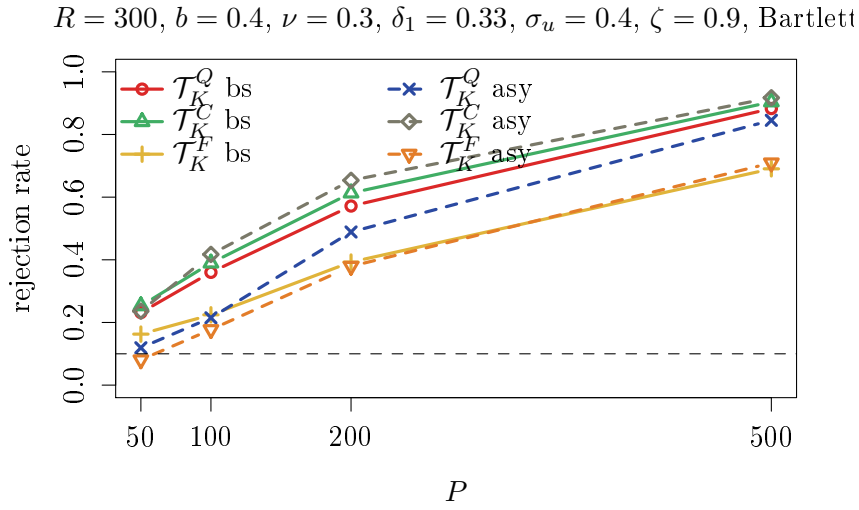
Figure 3: Bootstrap size under heteroskedasticity, different kernels



See notes to Figure 1.

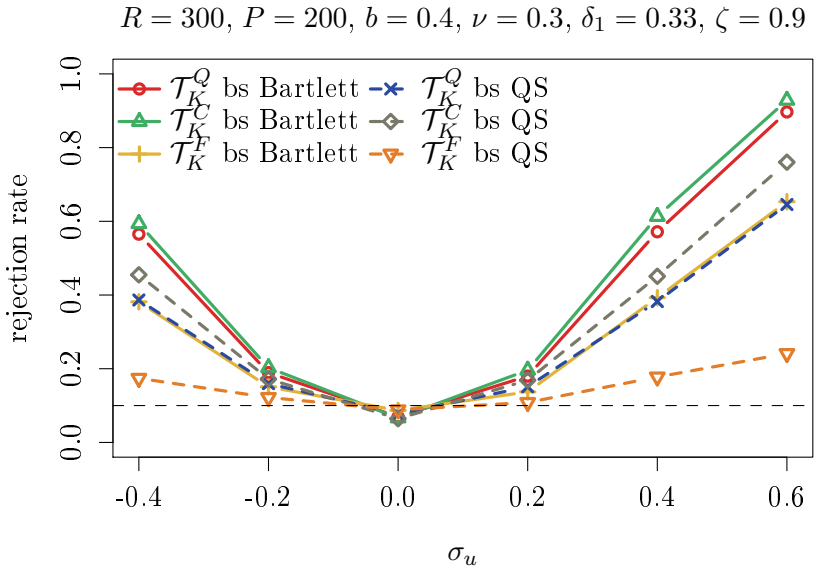
Figure 4: Size of $\mathcal{T}_K^Q$ under heteroskedasticity for different P, asymptotic and bootstrap tests

First, Figure 5 shows that the power of both bootstrap and asymptotic tests increases in P. Second, observing that this power experiment corresponds to the size study reported in Figure 2 (right panel), it comes as no surprise that the power ranking is strongly affected by whether a tests accurately exhausts or even exceeds nominal size. For example, the asymptotic CUSUM statistic $\mathcal{T}_K^Q$ is fairly undersized for $b = 0.4$, negatively affecting its power. The asymptotic Cramér-von Mises statistic $\mathcal{T}_K^C$ is slightly oversized, with corresponding positive impact on power. Recall, however, that Section 3.2 revealed that the bootstrap tests generally effectively exhaust nominal size, implying that their



See notes to Figure 1.

Figure 5: Power vs. P, constant relative forecasting ability, asymptotic and bootstrap tests

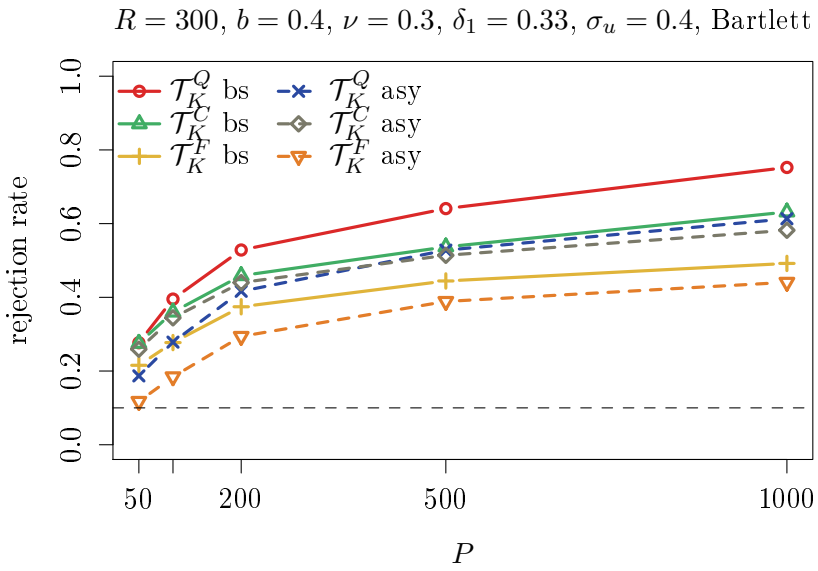


See notes to Figure 1.

Figure 6: Power bootstrap tests vs. σ_u, both kernels

power is either better than that of the asymptotic tests when the latter are undersized, or more credible when the latter are oversized. Third, $\mathcal{T}_K^Q$ and $\mathcal{T}_K^C$ perform quite similar in terms of power, while the power of $\mathcal{T}_K^F$ is less convincing.

Figure 6 compares the power for the two kernels in the bootstrap case. Here, we plot power against σ_u. First and as expected, the power of all tests increases in $|\sigma_u|$. This is because the predictive power of $x_{1,t}$ for z_t then increases, while $x_{2,t}$ has none throughout. The Bartlett kernel leads to



See notes to Figure 1.

Figure 7: Power vs. P, time-varying relative forecasting ability, asymptotic and bootstrap tests

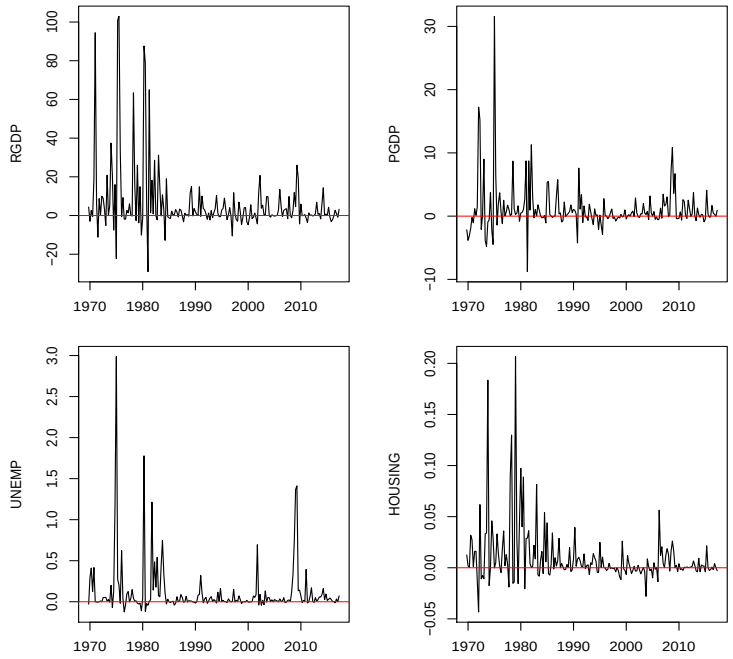


Figure 8: Loss differential series (no-change versus SPF) for output growth (RGDP), GDP deflator inflation (PGDP), unemployment rate (UNEMP) and the growth rate of housing starts (HOUSING). Nowcasts are evaluated against the first release for mean squared error loss.

more powerful tests. This is noteworthy, as both variants fairly effectively exhaust nominal size (cf. the entries at $\sigma_u = 0$). Figure 6 also reveal that $\mathcal{T}_K^Q$ and $\mathcal{T}_K^C$ outperform $\mathcal{T}_K^F$ for any σ_u.

We now present some of the results for the time-varying forecasting ability case in which the predictive power of $x_{2,t}$ for z_t is identical to that of $x_{1,t}$ until τ. After τ, the predictive power of $x_{2,t}$ for z_t

vanishes. First, Figure 7 demonstrates (for the Bartlett kernel) that the power increases in P (here also including $P = 1000$), the reason being that the time span during which a change in forecasting ability can be detected also increases. A comparison of Figures 5 and 7 reveals that power of the tests is lower in the time-varying forecasting ability case, as there is a smaller period $(R + P) - \tau$ during which they may detect differences in forecasting power.

4 Empirical results

4.1 The SPF data

The survey started in 1968 (conducted by the American Statistical Association and the National Bureau for Economic Research) and is administered by the Federal Reserve Bank of Philadelphia since 1990. Participants are asked to predict key US macroeconomic variables in the middle of each quarter for the current and the following four quarters. We consider four variables from the SPF forecast error statistics database:[11] real GDP growth (RGDP), GDP price deflator inflation (PGDP), the unemployment rate (UNEMP) and the growth rate of housing starts (HOUSING). These four series reflect different key aspects of the US economy and are available from the start of the SPF up to 2017Q2, yielding the longest possible series of forecasts to evaluate from this database. Given our focus on time-variation in relative forecast performance, a long sample is particularly interesting as different episodes in relative forecast performance might be identified.

Our sample includes several important economic phases. Among these are the 1970s with severe oil price shocks leading to increases in macroeconomic volatility and conversely, the "Great Moderation" lasting until the mid-1980s which exhibited a sharp decline in volatility and predictability (see Campbell, 2007). It is well-documented that the "Great Moderation" led to enhanced macroeconomic stability which eased forecasting in general, but also made it more difficult to beat simple time series models (see, e.g., Stock and Watson, 2007). Similarly, Groen et al. (2013) find that structural breaks in the variance play an important role for real-time inflation forecasting. In addition, the "Great Financial Crisis" in 2007/2008 with further volatility changes (and possibly changes in predictability) is also included.

We consider three horizons (nowcasting ($h = 0$), one-quarter ahead ($h = 1$) and one-year ahead ($h = 4$) forecasts) and two vintages. Macroeconomic data is often revised significantly, see Croushore and

[11] The exact data files are located at https://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/data-files/error-statistics.

Table 1: Summary statistics for output growth (RGDP), GDP deflator inflation (PGDP), unemployment rate (UNEMP) and the growth rate of housing starts (HOUSING) using the first data release and the MSE loss function. RelLoss(NC/SPF) denotes the relative RMSE loss of the no-change and the SPF forecasts. SD(·) labels the standard deviation of the loss differentials in the subsample I (1969-1984), II (1985-2006) or III (2007-2017). AC(1) denotes the empirical first-order autocorrelation coefficient of the loss differential series.

Statistic		RelLoss(NC/SPF)	SD(I)	SD(II)	SD(III)	AC(1)
Sample		1969-2017	1969-1984	1985-2006	2007-2017	1969-2017
RGDP	$h=0$	1.69	28.67	4.95	6.02	0.24
	$h=1$	1.51	60.61	5.49	14.02	0.14
	$h=4$	1.40	55.26	8.17	15.76	0.44
PGDP	$h=0$	1.38	5.88	1.68	2.41	0.08
	$h=1$	1.23	9.82	2.01	1.91	0.26
	$h=4$	1.12	6.62	2.33	2.57	0.29
UNEMP	$h=0$	2.38	0.50	0.09	0.31	0.33
	$h=1$	1.76	1.15	0.17	0.95	0.58
	$h=4$	1.43	2.21	0.48	2.16	0.67
HOUSING	$h=0$	1.40	0.05	0.01	0.01	0.07
	$h=1$	1.32	0.08	0.02	0.02	0.26
	$h=4$	1.20	0.25	0.05	0.07	0.61

Stark (2001). Faust et al. (2013) and Stark (2010) discuss and demonstrate the importance of the vintage structure when evaluating SPF (inflation) nowcasts and forecasts. We consider the first and the final data release. We compare SPF forecasts to no-change forecasts using the first data release to enable a fair comparison with regard to the available information in real-time; see also Stark (2010), D'Agostino et al. (2006) and Coroneo and Iacone (2016). The involvement of professional judgment might be expected to lead to advantages over uninformed no-change predictions.

Figure 8 displays some representative loss differentials for nowcasting evaluated against the first data release. It reveals that (i) loss differentials are mostly, but not always, positive indicating superiority of SPF forecasts, (ii) there is potentially some time-variation in the mean, (iii) there are some striking volatility changes and (iv) there is some mild to intermediate autocorrelation.

Table 1 provides some summary statistics for the full sample which covers $P = 191$ quarterly observations from 1969Q4 to 2017Q2.[12] We report RMSE ratios of no-change versus SPF, such that values above unity indicate superiority of SPF. The ratio always exceeds one, suggesting a better performance of the SPF over the full sample. However, the ratios take values in a range of 1.12 to 2.38, indicating notable heterogeneity. We also observe a clear monotonicity with regard to the horizon: SPF is particularly successful at nowcasting (most strongly for unemployment with a ratio of 2.38 and least, but still considerable, for GDP deflator inflation with 1.38). The advantages

[12]Some series contain a few missing values. Details on imputation are provided in Appendix C. As there are relatively many missing values in the first year of the survey, we decided to start in 1969Q4.

shrink with an increasing forecast horizon. This result, using the latest observations available, strengthens earlier findings (see the discussion in Section 4.4). The monotonicity is robust with respect to the vintage structure.[13] For output growth and GDP deflator inflation rates, professional forecasts seem to be more successful at predicting the first than the final release.

Next, we report estimated unconditional standard deviation for three subsamples merely for the purpose of illustration: subsample I (1969Q1-1984Q4, $T = 61$), II (1985Q1-2006Q4, $T = 88$) and III (2007Q1-2017Q2, $T = 42$). Volatility breaks associated with the "Great Moderation" are strongest for real GDP growth (with break factors even smaller than 1/5), followed by unemployment, housing starts and GDP deflator changes. In all cases, volatility changes are considerable. Comparing the relatively low volatility regime II to the one including the recent financial crisis (subsample III) reveals that volatility is either almost stable (for housing starts and PGDP) or increasing (slightly so for output and noticeably for unemployment). Differences between releases are negligible. Such substantial changes in unconditional volatility underline the need for suitable inferential procedures.

Finally, the first-order autocorrelation coefficient in the loss differentials unsurprisingly increases with h. Crucially for autocorrelation-robust fixed-b inference, there is rather strong dependence for unemployment and housing starts (up to 0.67 for one-year ahead forecasts). For nowcasts, autocorrelation coefficients range between 0.07 and 0.42.

4.2 Tests for equal predictive ability and time-variation

For all statistics $\mathcal{T}, \mathcal{T}_K^F$, $\mathcal{T}_K^Q$ and $\mathcal{T}_K^C$ (here with $K = 1$) we consider $b \in \{0, 0.1, \ldots, 1\}$ for the fixed-b bandwidth parameter. We thus include a classic Newey-West type statistic ($b = 0$, see also fn. 7) and also the fixed-b versions proposed by Choi and Kiefer (2010). We focus on the Bartlett kernel due to its higher power relative to the Quadratic Spectral kernel, where both have similar size (cf. Section 3). We consider tests based on asymptotic and wild bootstrap critical values. The latter are robust to time-varying volatility as observed for the loss differentials (cf. Table 1), while the former are not. Hence, allowing for changes in volatility may have important implications for the test results. Section 3 demonstrated that the tests work well for $P = 200$ observations.

As we compare the SPF to a simple no-change competitor, we apply statistics without a correction for estimation error. For no-change forecasts we do not estimate any parameters as the predictions are past values of first data releases of the respective variable. For the SPF, the estimation error

[13] The only slight exception is output growth, *final release*, one-quarter (1.39) and one-year forecasts (1.41).

Table 2: Test decisions for the full-sample $\mathcal{T}$-statistic either based on wild bootstrap ('bs') or asymptotic critical values ('asy'). Nowcasts ($h = 0$), one-quarter ($h = 1$) and one-year ahead forecasts ($h = 4$) are evaluated against the first data release under MSE loss. Evaluation sample runs from 1969Q4 to 2017Q2.

RGDP	$h=0$		$h=1$		$h=4$		PGDP	$h=0$		$h=1$		$h=4$	
b	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$		$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$
0	***	***	***	***	***	***		***	***	*	**		
0.1	***	***	***	**	***	**		***	***	***	***		
0.2	***	**	***	*	**	*		***	***	***	***		
0.3	**	*	***		**			***	**	***	**	*	
0.4	**		**		*			***	**	***	**	*	
0.5	**		**		*			***	**	***	**	*	
0.6	**		**		*			***	**	***	**	*	
0.7	**		**		*			***	**	***	**	**	
0.8	**		**		*			***	**	***	**	**	
0.9	**		**		*			***	**	***	**	*	
1	**		**		*			***	**	***	**	*	

UNEMP	$h=0$		$h=1$		$h=4$		HOUSING	$h=0$		$h=1$		$h=4$	
b	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$		$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$
0	***	***	***	***	***	***		***	***	***	***	**	**
0.1	***	***	***	***	***	***		***	**	***	**	**	*
0.2	***	***	***	**	***	***		**	*	**	*	*	
0.3	***	**	**	**	***	**		**		**			
0.4	***	**	**	**	**	**		*		*			
0.5	***	**	***	**	**	**		*					
0.6	**	**	**	**	**	**		*					
0.7	**	**	***	**	**	**		*					
0.8	***	**	***	**	***	**		*					
0.9	**	**	**	**	**	**		*					
1	**	**	**	**	***	**		*					

is not available and therefore, no correction of estimation error is applied, see the discussion in Giacomini and Rossi (2010) and Rossi and Sekhposyan (2016). Therefore, we employ the bootstrap algorithm 1 modified for the $\mathcal{T}^x$ statistics using $M = 5,000$ replications.

First, we test for equal predictive ability using the full-sample statistic $\mathcal{T}$, which is not explicitly designed to capture potential time-variation in mean loss differentials. We compare $\mathcal{T}$ to asymptotic ("asy") and wild bootstrap ("bs") critical values. Table 2 reports rejections at significance levels of one, five and ten percent. These are labeled as '***', '**' and '*' to ease the presentation of the many results and to conserve space by not reporting six different critical values for each statistic.

In general, the tests are two-sided due to their quadratic form. As all RMSE ratios exceed one (cf. Table 1), the corresponding sample means of loss differentials are positive. Hence, a rejection of a two-sided test implies that SPF significantly outperforms the competing no-change approach.

Starting with output growth, the bootstrap version (subscript 'bs') rejects equal predictive ability across the full sample in all cases — at least at the nominal ten percent level, but mostly at the five percent level or lower. This finding holds for all horizons h (albeit not as strong for $h = 4$) and

all values of the bandwidth-parameter b. It thereby clearly suggests the superiority of the SPF. On the contrary, using asymptotic critical values leads to far less rejections.

For GDP deflator inflation, bootstrap inference leads to rejections at the one percent level in nearly all cases for the shortest horizons $h = 0$ and $h = 1$. The traditional approach relying on asymptotic critical values mainly rejects at the five percent level. We find a clear difference in test decisions for one-year ahead forecasts ($h = 4$): while the bootstrap finds significant differences, asymptotic inference does not indicate any significant deviation from equal predictive ability.

We see a similar pattern for unemployment, with the exception of additional rejections for $h = 4$. Finally, we find evidence in favor of the SPF for nowcasts of changes in housing starts. For longer horizons, the evidence is weaker. In any case, the usage of bootstrap critical values robust to time-varying volatility leads to rejections at lower significance levels.[14]

We next consider tests suitable for time-variation in the relative forecast performance. To this end, we employ the $\mathcal{T}^F$ (with $\nu = 0.3$ as suggested in Giacomini and Rossi (2010)), $\mathcal{T}^Q$ and $\mathcal{T}^C$ statistics presented in Section 2. We see similar test decisions for output growth as for the full-sample statistics. Bootstrapped versions of the test statistics provide stronger rejections than their asymptotic counterparts. This is especially true for the $\mathcal{T}_{bs}^C$ statistic for all h. For GDP deflator inflation, we also find an almost identical pattern of rejections across the different statistics in comparison to the preceding full-sample analysis. For $h = 4$, only the $\mathcal{T}_{bs}^C$ statistic provides evidence against the null. For unemployment and housing starts, the general conclusion that bootstrap inference provides more and stronger rejections holds true as well.[15]

The time-varying components of the fluctuation and the CUSUM statistic may explain the previous findings. In particular, we look at the squared (i) rolling standardized MSE difference and (ii) scaled partial sum of the loss differential.[16] One explanation for the strong agreement between the full-sample and the time-variation tests might be that the SPF outperforms the benchmark at all time points. Even though the full sample tests might indicate this, it is not clear if the advantages of the SPF really existed over the entire period from 1969 to 2017. Looking at the statistics (i) and (ii) over time helps identifying different episodes of relative predictability, if present.

Figure 9 (nowcasting, first release, $b = 0.2$, $\nu = 0.3$) visualizes the time-varying components of

[14] These conclusions generally hold as well for the evaluation against the *final* release of data, as reported in Table 8. The differences in the results are relatively minor. The robustness with respect to the vintage structure is in line with Stark (2010), who finds data revisions to be of less importance in the relative performance of the SPF.

[15] The test outcomes are quite similar for the final data release, see Tables 9 and 10.

[16] The fluctuation and CUSUM statistics search for the maximum of these statistics over the evaluation period.

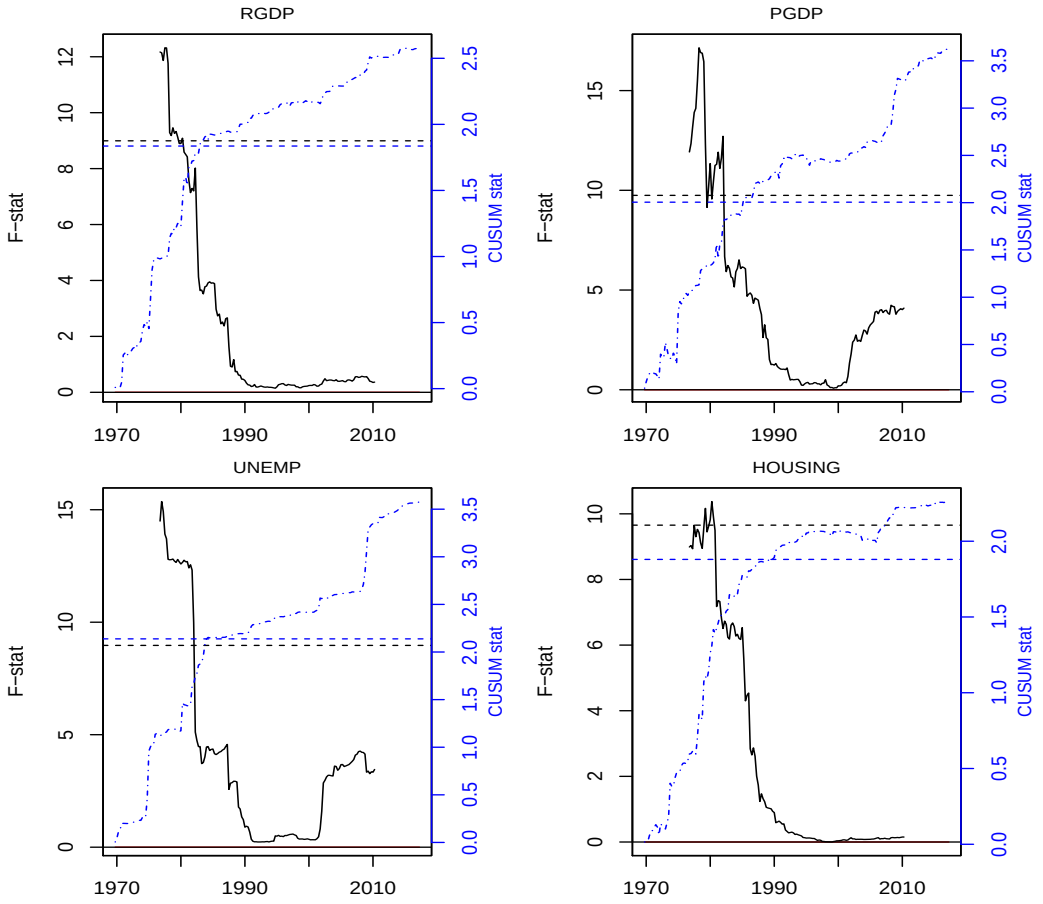


Figure 9: The plots show the time-varying components of the fluctuation statistic (left axis, solid black line) and the CUSUM statistic (right axis, dashed-dotted blue line), see equations (2) and (3). Horizontal dashed lines are the corresponding five percent critical values for the *maximum* of the displayed statistics. Nowcasts are evaluated against the first release for MSE loss; $b = 0.2$, $\nu = 0.3$.

the tests. The fluctuation statistic reveals two striking patterns: Judging from the statistic and the associated five percent bootstrap critical value, there is a sizable deterioration in predictability in the early 1980s associated with the "Great Moderation". This breakdown is significant, while the recoveries observed for GDP deflator inflation and unemployment in the early 2000s are too weak for a rejection. For output growth and housing starts, the results suggest that there is no comeback in relative predictive ability of the SPF. Interestingly, relative forecast performance did not change a lot during the "Great Financial Crisis" even though volatility increased somewhat, but to a much lesser extent when compared to the "Great Moderation". These results holds more generally for other horizons and the final data release, see Figures 14-18 in the Appendix. Figures 19-24 show the unscaled rolling window MSE difference between the SPF and no-change forecasts. They support the previous interpretation and reveal that, at least, no-change forecasts never significantly

Table 3: Test decisions for the time-variation $\mathcal{T}^{\{Q,C,F\}}$-statistics either based on wild bootstrap ('bs') or asymptotic critical values ('asy'). Nowcasts ($h = 0$), one-quarter ($h = 1$) and one-year ahead forecasts ($h = 4$) are evaluated against the first data release under MSE loss. Evaluation sample runs from 1969Q4 to 2017Q2.

RGDP	$h=0$						$h=1$						$h=4$					
b	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$
0	***	***	***	***	***	***	***	***	***	***	**	***	***	**	***	***	**	***
0.1	***	**	***	***	***	***	***	**	***	***	***	***	**	*	***	**	**	**
0.2	***	*	***	**	**	**	***	*	***	**	*	*	**		**	**		*
0.3	**		**	**	*		**		**	*			*		**	*		
0.4	**		**	*			**		**	*					**	*		
0.5	*		**	*			*		**	*					*			
0.6	*		**				*		**						*			
0.7	*		**				*		**						*			
0.8			**				*		**						*			
0.9			**				*		**						*			
1			**				*		*									

PGDP	$h=0$						$h=1$						$h=4$					
b	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$
0	***	***	***	***	***	***	**	*	*	**		*						
0.1	***	***	***	***	***	***	***	***	***	***	***	***						
0.2	***	**	***	***	***	***	***	**	***	***	***	***						
0.3	***	**	***	**	**	**	***	**	***	***	***	***						
0.4	***	**	***	**	**	*	***	**	***	**	**	**			*			
0.5	***	**	***	**	**	*	***	**	***	**	**	**			*			
0.6	***	**	***	**	**	*	***	**	***	**	**	**			*			
0.7	***	**	***	**	**	*	***	*	***	**	**	**			**			
0.8	**	**	***	**	*	*	***	*	***	**	**	**			*			
0.9	***	**	***	**	**	*	***	*	***	**	**	**			*			
1	***	**	***	**	**	*	***	*	***	**	**	**			*			

Table 4: continued from Table 3.

UNEMP	$h=0$						$h=1$						$h=4$					
b	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$
0	***	***	***	***	***	***	***	***	***	***	***	**	***	***	***	***	***	***
0.1	***	***	***	***	***	***	***	***	***	***	**	***	***	***	***	***	***	***
0.2	***	**	***	***	***	**	***	**	***	***	**	**	***	***	***	***	**	**
0.3	***	**	***	**	**	*	**	**	***	**	**	*	**	**	***	**	*	
0.4	**	**	**	**	**	*	**	**	**	**	**		**	**	**	**	*	
0.5	**	**	**	**	*		**	**	**	**	*		**	**	**	**		
0.6	**	**	**	**	*		**	**	**	**	*		**	**	**	**		
0.7	**	**	**	**	*		**	**	**	**	*	*	**	**	***	**	*	
0.8	**	**	***	**	*		**	**	***	**	**		**	**	***	**	*	*
0.9	**	*	**	**	*		**	**	**	**	**		**	**	***	**	*	*
1	**	*	**	**	*		**	**	***	**	**		**	**	**	**	*	*

HOUSING	$h=0$						$h=1$						$h=4$					
b	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$
0	***	***	***	***	***	***	***	***	***	***	***	***	**	**	***	***	**	***
0.1	***	**	***	***	***	***	***	**	***	***	***	***	**		**	**	**	**
0.2	**		**	**	*	*	**		**	**	*	*			*	*		
0.3	*		**	*					**	*								
0.4			*	*					*									
0.5			*															
0.6			*															
0.7			*															
0.8			*															
0.9																		
1			*															

outperform the SPF. The CUSUM statistic indicates a breakdown in relative forecast performance as it also turns significant in the 1980s, implying that the accumulated changes are large enough for a rejection.[17]

Our interpretation is that the full sample results are mainly driven by first of part of the sample (until the mid-1980s) in which the SPF clearly performed better. As the statistics for time-variation further indicate clearly and robustly, the advantages in relative predictability largely disappear in the mid-1980s. Most of the evidence for time-variation, however, would not have been detected by a traditional analysis using asymptotic critical values.

4.3 Asymmetric loss

We additionally consider an asymmetric loss function (e.g., Elliott et al., 2005):

$$\mathcal{L}_{t,h} = [\alpha + (1 - 2\alpha) \cdot 1(z_{t+h} - f_{t+h} < 0)](z_{t+h} - f_{t+h})^2 \, .$$

For $\alpha = 0.5$, a symmetric loss function arises as a special case. In line with Rudebusch and Williams (2009) and Wang and Lee (2014), we consider a value of $\alpha = 0.7$. Such a value reflects an intermediate asymmetry where over-predictions receive higher weight than under-predictions. The appendix provides detailed results. Here, we give a brief comparison to the symmetric case. The summary statistics in Tables 11 and 12 show the same patterns as reported for MSE loss. The full-sample test decisions are broadly similar. The only noticeable difference is the lack of evidence against the null for one-year GDP deflator inflation. We find very similar results (with the same exception) regarding relative time-varying forecasting performance. Overall, the vintage still plays a minor role, except for GDP price deflator inflation for which we find a lack of any evidence against the null not only for $h = 4$, but also for $h = 1$. Notably, the strong evidence persists for $h = 0$. Moreover, patterns of time-variation (Figures 25—30 in the Appendix) based on asymmetric loss are fully consistent with the ones for MSE loss for output growth, housing starts and unemployment. For GDP deflator inflation, however, the general movement is the same, but the stronger weighting of forecast errors in the 1970s (due to large oil price shocks) introduces more erratic behavior.

[17] Its behavior at the beginning and end of the sample provides additional information which $\mathcal{T}_K^F$ cannot provide due to trimming. Before 1976, there are signs for time-variation in all series. Unemployment and GDP deflator inflation apparently exhibit some further time-variation after 2010 (less for output growth), unlike housing starts.

4.4 Discussion of our results in light of the related literature

We now provide a comparison of our findings and those of previous studies on the performance of the SPF. Most of these use the Diebold and Mariano (1995) test for differences in MSE. One strand of the literature deals with the accuracy of the SPF in general, while a second smaller one focusses on the decline of predictability in connection to the "Great Moderation". A comparison is generally complicated by the fact that studies obviously use different variables (and partly also definitions thereof), benchmarks, vintages, horizons, samples etc. However, two articles, viz. D'Agostino et al. (2006) and Coroneo and Iacone (2016), are particularly close to the scope of our work. Therefore, we provide a more extensive treatment of these studies towards the end of this discussion.

The general notion in the literature is that the SPF provides accurate forecasts, especially nowcasts, for real output growth, inflation and unemployment, but less so for housing starts. Zarnowitz and Braun (1993) and Croushore (1993) (see also references therein) provide early evidence on the good performance of SPF forecasts for real GDP and inflation. Ang et al. (2007) find that surveys (including the SPF) forecast inflation better than macro variables, time series models and asset markets. They also find that when allowing for time-variation, the SPF dominates throughout the whole sample. Croushore (2010) uses real-time data instead and finds confirmatory evidence.

The superiority of SPF nowcasts has been documented in several influential studies, e.g. Giannone et al. (2008). Liebermann (2014) considers real-time nowcasting for output growth and compares the performance of professional forecasters and a dynamic factor model to simple autoregressive and no-change forecasts. She finds that gains in forecasting accuracy are pronounced for $h = 0$ and decrease in h. For a sample from 1985Q1 to 2007Q4, Stark (2010) similarly finds that the accuracy of the SPF declines significantly for $h > 1$. Moreover, he finds the SPF to outperform no-change forecasts (except for housing starts). For unemployment, Montgomery et al. (1998) provide early evidence for the superiority of SPF forecasts. Interestingly, the authors find that the relative performance (evaluated against time series models) is particularly good at short horizons up to $h = 3$, supporting the general finding that the SPF has relative advantages for short horizons.

We now turn to the discussion of D'Agostino et al. (2006) and Coroneo and Iacone (2016). Both use a naive benchmark under MSE loss and deal with time-variation by running tests on subsamples. In contrast to our approach, the applied tests are not robust to time-varying volatility and do not exploit the full sample to formally test for time-variation in an endogenous way.

Coroneo and Iacone (2016) apply a Diebold and Mariano (1995) statistic with fixed-smoothing

asymptotics. Their full-sample test has good size even in samples of only 40 observations, while tests using standard asymptotics are oversized. In addition, they consider a stationary block-bootstrap version of the test and find it to yield better size than standard asymptotics, and to be equally powerful as the fixed-smoothing approach. In a sample ranging from 1985Q1 to 2014Q4 for real GNP/GDP growth, GNP/GDP inflation, unemployment rate and the three-month Treasury bill rate, the SPF sometimes significantly outperforms a naive random walk. For output growth, there is virtually no evidence against the null. The SPF seems to provide more accurate inflation forecasts at horizons $h = \{0, 1, 2\}$, but not beyond. The evidence for unemployment suggests the superiority of the SPF in particular for $h = 0$. For other horizons, the evidence is relatively weak. For the short-term interest rate, SPF is clearly found to perform better for all horizons except the longest one ($h = 4$). In a subsample analysis with blocks of ten years of data, the authors investigate time-variation and find: (i) for output growth, there is no single decade in which the SPF performs better; (ii) relative advantages of the SPF observed for the period from 1985 to 1994 vanished at all horizons for inflation, unemployment and the interest rate between 1995 and 2004; (iii) in the most recent subsample, the superiority of the SPF is re-established mostly for inflation, and partly for unemployment and interest rates; in particular when considering the bootstrap version of the test. The fixed-smoothing test, however, provides much less evidence against the null.

Thus, our findings partly corroborate those of Coroneo and Iacone (2016), with some notable differences. Unlike Stark (2010), we and Coroneo and Iacone (2016) do not find that the SPF easily outperforms naive forecasts after 1985. This might be due to the different tests applied in the analysis. An important difference to the results of Coroneo and Iacone (2016) regards the period from 1985 to 1995, for which we do not find evidence for superiority of the SPF at any h.

D'Agostino et al. (2006) find a significant decline in relative predictive accuracy of the SPF for inflation and output growth for $h = 1$ to $h = 4$. Their full sample (1975Q1 to 1999Q4) results indicating the superiority of the SPF seem to be driven by the period prior to 1985 in which the SPF outperformed the naive benchmark. After 1985, the results indicate that the SPF no longer has a significant advantage. This strongly suggests instabilities in relative forecast performance. Our findings corroborate their results and sharpen them in showing that this phenomenon carries over to unemployment and housing starts as well and that is also holds for the case of nowcasting. In addition, Campbell (2007), D'Agostino and Whelan (2008) and Gamber and Smith (2009) find by analyses of various subsamples, consistent with our results, declining predictability of the SPF after the "Great Moderation" for output growth and inflation. Explanations about the causes of the

forecast breakdown differ across these studies and remain an open issue.

By applying tests (which are agnostic regarding structural breaks and robust to time-varying volatility and autocorrelation) to a comparably long sample of more than 40 years of data, we obtain results which support several previous findings. Among these are (i) the superiority of the SPF for shortest horizons, but less advantages for one-year ahead forecasts; (ii) a significant decline in relative predictability during the 1980s; (iii) the robustness of the relative performance of the SPF to data revisions. Our results yield the following new insights: (i) results for the full-sample analysis and time-variation continue to hold under an asymmetric loss function; (ii) advantages of the SPF forecasts are minimal in the 1990s, with weak signs of recoveries for GDP deflator inflation and unemployment later on and (iii) relative forecast performance did not change during the "Great Financial Crisis", even though volatility increased.

These recoveries possibly turn into a significant comeback of SPF forecasts in the future. In this case, the exact timing would be certainly unknown, rendering a subsample analysis inappropriate. In general, the ad hoc choice of break points may easily lead to biases. Moreover, it is not always possible (especially in view of recent developments) to invoke economic reasons like the well-studied "Great Moderation". In contrast, the methods proposed here are suitable for data containing possibly multiple unknown breakpoints in forecast performance alongside changes in volatility.

5 Concluding remarks

This paper proposes wild bootstrap tests for equal predictive ability that can be applied when volatility and relative forecast performance are time-varying, and proves their validity. Both data features are present in many macroeconomic and financial forecast comparisons. We also provide suitable rolling and recursive estimation adjustments of the procedures when estimation error is relevant. The considered tests are either full sample tests (similar to existing ones like Diebold and Mariano (1995)) or CUSUM, Cramér-von Mises and fluctuation statistics when testing for time-variation. All employ fixed-b asymptotics which deliver more accurately sized tests in finite-samples. Our simulation study demonstrates that the tests work well in empirically relevant situations.

Our empirical application investigates the (time-varying) forecast performance of professional forecasters obtained from the SPF relative to simple no-change forecasts in real-time. The analysis suggests that ignoring time-varying variance seriously affects conclusions regarding the null of equal

predictive ability. Traditional tests provide considerably weaker evidence against the null than the wild bootstrap versions. Tests allowing for time-variation indicate that the SPF had significant advantages until the mid-1980s, but not thereafter. Further research might address to what extent the time-varying relative forecast performance can be explained (e.g. Campbell, 2007). Another interesting avenue is to investigate the Fed's 'Green Book' forecasts which receive a lot of attention (e.g. Romer and Romer, 2000; D'Agostino and Whelan, 2008; Rossi and Sekhposyan, 2016).

References

Amado, C. and T. Teräsvirta (2013). Modelling volatility by variance decomposition. *Journal of Econometrics 175*(2), 142–153.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*(3), 817–858.

Ang, A., G. Bekaert, and M. Wei (2007). Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics 54*(4), 1163–1212.

Campbell, S. D. (2007). Macroeconomic volatility, predictability, and uncertainty in the great moderation: evidence from the survey of professional forecasters. *Journal of Business & Economic Statistics 25*(2), 191–200.

Cavaliere, G. (2004). Unit root tests under time-varying variances. *Econometric Reviews 23*(3), 259–292.

Cavaliere, G., A. Rahbek, and A. M. R. Taylor (2010). Testing for co-integration in vector autoregressions with non-stationary volatility. *Journal of Econometrics 158*(1), 7–24.

Cavaliere, G. and A. M. R. Taylor (2009). Heteroskedastic time series with a unit root. *Econometric Theory 25*(5), 1228–1276.

Choi, H. S. and N. M. Kiefer (2010). Improving robust model selection tests for dynamic models. *The Econometrics Journal 13*(2), 177–204.

Clark, T. E. and F. Ravazzolo (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics 30*(4), 551–575.

Coroneo, L. and F. Iacone (2016). Comparing predictive accuracy in small samples using fixed-smoothing asymptotics. *Mimeo*.

Croushore, D. (1993). Introducing: the survey of professional forecasters. *Business Review-Federal Reserve Bank of Philadelphia 6*.

Croushore, D. (2010). An evaluation of inflation forecasts from surveys using real-time data. *The BE Journal of Macroeconomics 10*(1).

Croushore, D. and T. Stark (2001). A real-time data set for macroeconomists. *Journal of Econometrics 105*(1), 111–130.

D'Agostino, A., D. Giannone, and P. Surico (2006). (Un)Predictability and macroeconomic stability. *Working Paper Series 605, European Central Bank*.

D'Agostino, A. and K. Whelan (2008). Federal reserve information during the great moderation. *Journal of the European Economic Association 6*(2-3), 609–620.

Davidson, J. (1994). *Stochastic Limit Theory*. Oxford University Press.

Demetrescu, M., C. Hanck, and R. Kruse (2017). Robust fixed-b testing under time-varying volatility. *Mimeo*.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–263.

Elliott, G., A. Timmermann, and I. Komunjer (2005). Estimation and testing of forecast rationality under flexible loss. *The Review of Economic Studies 72*(4), 1107–1125.

Faust, J., J. H. Wright, et al. (2013). Forecasting inflation. *Handbook of Economic Forecasting 2*(Part A), 3–56.

Gamber, E. N. and J. K. Smith (2009). Are the fed's inflation forecasts still superior to the private sector's? *Journal of Macroeconomics 31*(2), 240–251.

Giacomini, R. and B. Rossi (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics 25*(4), 595–620.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*(6), 1545–1578.

Giannone, D., L. Reichlin, and D. Small (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics 55*(4), 665–676.

Groen, J. J. J., R. Paap, and F. Ravazzolo (2013). Real-time inflation forecasting in a changing world. *Journal of Business & Economic Statistics 31*(1), 29–44.

Guidolin, M. and A. Timmermann (2006). An econometric model of nonlinear dynamics in the joint distribution of stock and bond returns. *Journal of Applied Econometrics 21*(1), 1–22.

Honaker, J., G. King, and M. Blackwell (2011). Amelia II: A program for missing data. *Journal of Statistical Software 45*(7), 1–47.

Justiniano, A. and G. Primiceri (2008). The time-varying volatility of macroeconomic fluctuations. *American Economic Review 98*(3), 604–641.

Kiefer, N. M. and T. J. Vogelsang (2002a). Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica 70*(5), 2093–2095.

Kiefer, N. M. and T. J. Vogelsang (2002b). Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory 18*(6), 1350–1366.

Kiefer, N. M. and T. J. Vogelsang (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory 21*(6), 1130–1164.

Kiefer, N. M., T. J. Vogelsang, and H. Bunzel (2000). Simple robust testing of regression hypotheses. *Econometrica 68*(3), 695–714.

Leschinski, C. H. (2017). *MonteCarlo: Automatic Parallelized Monte Carlo Simulations*. R package version 1.0.2.

Li, J. and A. J. Patton (2015). Asymptotic inference about predictive accuracy using high frequency data. *Mimeo*.

Liebermann, J. (2014). Real-time nowcasting of GDP: A factor model vs. professional forecasters. *Oxford Bulletin of Economics and Statistics 76*(6), 783–811.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics 21*, 255–285.

Montgomery, A. L., V. Zarnowitz, R. S. Tsay, and G. C. Tiao (1998). Forecasting the us unemployment rate. *Journal of the American Statistical Association 93*(442), 478–493.

Müller, U. K. (2014). HAC corrections for strongly autocorrelated time series. *Journal of Business & Economic Statistics 32*(3), 311–322.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rapach, D. E. and J. K. Strauss (2008). Structural breaks and GARCH models of exchange rate volatility. *Journal of Applied Econometrics 23*(1), 65–90.

Romer, C. D. and D. H. Romer (2000). Federal reserve information and the behavior of interest rates. *American Economic Review*, 429–457.

Rossi, B. and T. Sekhposyan (2016). Forecast rationality tests in the presence of instabilities, with applications to federal reserve and survey forecasts. *Journal of Applied Econometrics 31*(3), 507–532.

Rudebusch, G. D. and J. C. Williams (2009). Forecasting recessions: the puzzle of the enduring power of the yield curve. *Journal of Business & Economic Statistics 27*(4), 492–503.

Sensier, M. and D. van Dijk (2004). Testing for volatility changes in U.S. macroeconomic time series. *The Review of Economics and Statistics 86*(3), 833–839.

Smeekes, S. and J.-P. Urbain (2014). A multivariate invariance principle for modified wild bootstrap methods with an application to unit root testing. *Maastricht University GSBE Research Memoranda RM/14/008*.

Stark, T. (2010). Realistic evaluation of real-time forecasts in the survey of professional forecasters. *Federal Reserve Bank of Philadelphia, Research Department* (Special Report), 726–740.

Stock, J. H. and M. W. Watson (2002). Has the business cycle changed and why? *NBER Macroeconomics Annual 17*(1), 159–218.

Stock, J. H. and M. W. Watson (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking 39*(S1), 3–33.

Sun, Y. (2014). Fixed-smoothing asymptotics in a two-step generalized method of moments framework. *Econometrica 82*(6), 2327–2370.

Sun, Y., P. C. B. Phillips, and S. Jin (2008). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica 76*(1), 175–194.

Vogelsang, T. J. and M. Wagner (2013). A fixed-b perspective on the Phillips-Perron unit root tests. *Econometric Theory 29*(3), 609–628.

Wang, Y. and T.-H. Lee (2014). Asymmetric loss in the greenbook and the survey of professional forecasters. *International Journal of Forecasting 30*(2), 235–245.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica 64*(5), 1067–1084.

Yang, J. and T. J. Vogelsang (2011). Fixed-b analysis of LM-type tests for a shift in mean. *The Econometrics Journal 14*(3), 438–456.

Zarnowitz, V. and P. Braun (1993). Twenty-two years of the NBER-ASA quarterly economic outlook surveys: aspects and comparisons of forecasting performance. In *Business cycles, indicators and forecasting*, pp. 11–94. University of Chicago Press.

Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association 108*(502), 726–740.

Appendix

A Bootstrap implementation for linear regression forecasts

Here, we work out the corresponding wild bootstrap algorithm for the simple, but important case of a regression-based prediction using two different sets of predictors, $\boldsymbol{x}_{1,t}$ and $\boldsymbol{x}_{2,t}$. Let us consider the following linear predictive models

$$z_{t+h} = \boldsymbol{\theta}_i' \boldsymbol{x}_{i,t} + \epsilon_{i,t}\,, \quad t = 1, \ldots, R+P, \qquad i = 1, 2, \tag{8}$$

which we estimate by OLS in an either recursive or rolling manner. The theoretical forecasts for z_{t+h} are given by

$$f_{i,t} = \boldsymbol{\theta}_i' \boldsymbol{x}_{i,t}, \qquad i = 1, 2,$$

at each step t. The line version gradient of $f_{i,t}$ is given by $\boldsymbol{x}_{i,t}'$. Note that $\boldsymbol{x}_{1,t}$ and $\boldsymbol{x}_{2,t}$ (and thus $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$) need not have the same dimensionality.

Then, at each t, the forecasts are generated as $\hat{f}_{i,t} = \hat{\boldsymbol{\theta}}_{i,(t)}' \boldsymbol{x}_{i,t}$ with $\hat{\boldsymbol{\theta}}_{i,(t)}$ computed recursively (or in a rolling manner, $\hat{\boldsymbol{\theta}}_{i,(t)}^{rol}$). The weighted loss differentials are computed using weights $\boldsymbol{h}_t = 1$ in this example, so $\hat{\boldsymbol{y}}_t$ is a scalar, $\hat{y}_t$. We use a quadratic $\mathcal{L}(u) = u^2$ with derivative $2u$. In the linear regression case, the estimation effect depends on

$$\mathbf{C}_{i,t} = \boldsymbol{x}_{i,t} \boldsymbol{x}_{i,t}' \quad \text{and} \quad \boldsymbol{a}_{i,t,(\boldsymbol{\theta}_i)} = \boldsymbol{x}_{i,t} \left(z_{t+h} - \boldsymbol{\theta}_i' \boldsymbol{x}_{i,t} \right) = \boldsymbol{x}_{i,t} \epsilon_{i,t}.$$

To account for the estimation effect, one needs to replicate the behavior of partial sums of $(\boldsymbol{a}_{i,t}, y_t)'$; to be more precise, we need estimates of these quantities since they are not observed directly. While $\hat{y}_t$ is the natural estimator for y_t, computing estimates $\hat{\boldsymbol{a}}_{i,t}$ requires a set of residuals, say $\hat{\epsilon}_{i,t}$.

For each $t = R+1, \ldots, R+P$, estimate the LS regression

$$z_{j+h} = \hat{\boldsymbol{\theta}}_{i,(t)}' \boldsymbol{x}_{i,j} + \hat{e}_{i,j,(t)}\,, \quad j = 1, \ldots, t, \tag{9}$$

where the extra index (t) in Equation (9) indicates the dependence of the estimates on the time at which estimation is conducted. Moreover, residuals are denoted by $\hat{e}_{i,j,(t)}$ to emphasize that a full set of residuals is computed at each time t in a recursive manner; therefore, we do not have one single set of residuals which we could call $\hat{\epsilon}_{i,t}$.

For the recursive setup, we use

$$\hat{\epsilon}_{i,t} = \hat{e}_{i,t,(t)} \tag{10}$$

for all $t = 1, \ldots, R+P$, i.e. we use the most precise residuals available as bootstrap population.

We employ the following bootstrap algorithm:

Algorithm 3

1. Compute $\hat{y}_t$ from (5) (and set $\hat{y}_t = 0$ for $1 \leq t \leq R$)
2. For all $t = 1, \ldots, R+P$, compute $\mathbf{C}_{i,t} = \boldsymbol{x}_{i,t} \boldsymbol{x}_{i,t}'$ and $\hat{\boldsymbol{a}}_{i,t} = \boldsymbol{x}_{i,t} \hat{\epsilon}_{i,t}$ with $\hat{\epsilon}_{i,t}$ from (10).
3. Generate r_t^* wild bootstrap draws, $t = 1, \ldots, R+P$.
4. Construct $\left(\boldsymbol{a}_{1,t}^{*\prime}, \boldsymbol{a}_{2,t}^{*\prime}, y_t^* \right)'$ as $\left(\hat{\boldsymbol{a}}_{1,t}', \hat{\boldsymbol{a}}_{2,t}', \hat{y}_t \right)' r_t^*$ for $t = 1, \ldots, R+P$.
5. Compute for $t = R+1, \ldots, R+P$

$$\hat{\boldsymbol{\theta}}_{i,(t)}^* = \left(\sum_{j=1}^{t} \mathbf{C}_{i,j} \right)^{-1} \sum_{j=1}^{t} \boldsymbol{a}_{i,j}^* + \hat{\boldsymbol{\theta}}_{i,(R+P)}.$$

6. Compute for $t = R+1, \ldots, R+P$

$$\begin{aligned}\hat{y}_t^* &= y_t^* - 2\left(z_{t+h} - \hat{\boldsymbol{\theta}}_{1,(t)}^{*\prime}\boldsymbol{x}_{1,t}\right)\boldsymbol{x}_{1,t}'\left(\hat{\boldsymbol{\theta}}_{1,(t)}^* - \hat{\boldsymbol{\theta}}_{1,(R+P)}\right) \\ &\quad + 2\left(z_{t+h} - \hat{\boldsymbol{\theta}}_{2,(t)}^{*\prime}\boldsymbol{x}_{2,t}\right)\boldsymbol{x}_{2,t}' \cdot \left(\hat{\boldsymbol{\theta}}_{2,(t)}^* - \hat{\boldsymbol{\theta}}_{2,(R+P)}\right).\end{aligned}$$

7. Compute the test statistics using the bootstrap sample $\hat{y}_t^*$, $t = R+1, \ldots, R+P$.

8. Repeat the steps M times and obtain the desired quantile(s).

The wild bootstrap provides asymptotically pivotal inference if $\sup_{t=1,\ldots,R+P} \mathrm{E}\left(\|\boldsymbol{x}_{i,t}\|^4\right) < \infty$, which suffices to verify the additional conditions required by Proposition 6.

The procedure is similar for rolling window estimation. For each $t = R+1, \ldots, R+P$,

$$z_{j+h} = \left(\hat{\boldsymbol{\theta}}_{i,(t)}^{rol}\right)' \boldsymbol{x}_{i,j} + \hat{e}_{i,j,(t)}^{rol}, \quad j = t-R+1, \ldots, t, \tag{11}$$

At each time t, the forecasts are generated as $\hat{f}_{i,t} = \left(\hat{\boldsymbol{\theta}}_{i,(t)}^{rol}\right)' \boldsymbol{x}_{i,t}$.

The bootstrap algorithm for the rolling windows case is very similar, but takes into account that we only resort to estimates from the current window at each t. The biggest change is how we get the residuals $\hat{\epsilon}_{i,t}^{rol}$ entering $\hat{\boldsymbol{a}}_{i,t}^{rol}$ (the rolling version estimate of $\boldsymbol{a}_{i,t}$). Again, we have multiple variants to choose $\hat{\epsilon}_{i,t}^{rol}$, given the multitude of computed residuals $\hat{e}_{i,j,(t)}^{rol}$. The natural choice is for the rolling window scheme to take

$$\hat{\epsilon}_{i,t}^{rol} = \begin{cases} \hat{e}_{i,t,(R)}^{rol} & t = 1, \ldots, R \\ \hat{e}_{i,t,(t)}^{rol} & t = R+1, \ldots, R+P. \end{cases} \tag{12}$$

That is, the last residual from each window is added to the series of residuals as the window rolls on and the first R are the residuals from the first window. The changes in the algorithm are as follows. First, compute $\hat{\boldsymbol{a}}_{i,t}^{rol} = \boldsymbol{x}_{i,t}\hat{\epsilon}_{i,t}^{rol}$ for all $t = 1, \ldots, R+P$, with $\hat{\epsilon}_{i,t}^{rol}$ from (12). Second, compute for $t = R+1, \ldots, R+P$

$$\hat{\boldsymbol{\theta}}_{i,(t)}^{*\,rol} = \left(\sum_{j=t-R+1}^{t} \mathbf{C}_{i,j}\right)^{-1} \sum_{j=t-R+1}^{t} \boldsymbol{a}_{i,j}^* + \hat{\boldsymbol{\theta}}_{i,(R+P)}^{rol}.$$

Finally, compute

$$\begin{aligned}\hat{y}_t^* &= y_t^* - 2\Big(z_{t+h} - \Big(\hat{\boldsymbol{\theta}}_{1,(t)}^{*\,rol}\Big)'\boldsymbol{x}_{1,t}\Big)\boldsymbol{x}_{1,t}'\big(\hat{\boldsymbol{\theta}}_{1,(t)}^{*\,rol} - \hat{\boldsymbol{\theta}}_{1,(t)}^{rol}\big) \\ &\quad + 2\Big(z_{t+h} - \Big(\hat{\boldsymbol{\theta}}_{2,(t)}^{*\,rol}\Big)'\boldsymbol{x}_{2,t}\Big)\boldsymbol{x}_{2,t}' \cdot \big(\hat{\boldsymbol{\theta}}_{2,(t)}^{*\,rol} - \hat{\boldsymbol{\theta}}_{2,(t)}^{rol}\big) \qquad \text{for } t = R+1, \ldots, R+P.\end{aligned}$$

B Proofs

Proof of Proposition 1

The arguments in the proof of Theorem 2 in Kiefer and Vogelsang (2005) indicate that

$$\hat{\boldsymbol{\Omega}} = -\frac{1}{P^2}\sum_{i=1}^{P-1}\sum_{j=1}^{P-1}\frac{P^2}{B^2}k''\left(\frac{i-j}{B}\right)\frac{1}{\sqrt{P}}\sum_{t=1}^{i}(\boldsymbol{y}_t - \overline{\boldsymbol{y}})\frac{1}{\sqrt{P}}\sum_{t=1}^{j}(\boldsymbol{y}_t - \overline{\boldsymbol{y}})' + o_p(1)$$

for kernels with smooth derivatives, or

$$\begin{aligned}\hat{\boldsymbol{\Omega}} &= \frac{2}{bP}\sum_{i=1}^{P}\left(\frac{1}{\sqrt{P}}\sum_{t=1}^{i}(\boldsymbol{y}_t-\overline{\boldsymbol{y}})\right)\left(\frac{1}{\sqrt{P}}\sum_{t=1}^{i}(\boldsymbol{y}_t-\overline{\boldsymbol{y}})\right)' \\ &\quad -\frac{1}{bP}\sum_{i=1}^{[(1-b)P]}\left(\frac{1}{\sqrt{P}}\sum_{t=1}^{i+[bP]}(\boldsymbol{y}_t-\overline{\boldsymbol{y}})\right)\left(\frac{1}{\sqrt{P}}\sum_{t=1}^{i}(\boldsymbol{y}_t-\overline{\boldsymbol{y}})\right)' \\ &\quad -\frac{1}{bP}\sum_{i=1}^{[(1-b)P]}\left(\frac{1}{\sqrt{P}}\sum_{t=1}^{i}(\boldsymbol{y}_t-\overline{\boldsymbol{y}})\right)\left(\frac{1}{\sqrt{P}}\sum_{t=1}^{i+[bP]}(\boldsymbol{y}_t-\overline{\boldsymbol{y}})\right)' + o_p(1)\end{aligned}$$

for the Bartlett kernel. The continuous mapping theorem [CMT] implies together with Lemma 1 that

$$\frac{1}{\sqrt{P}}\sum_{t=1}^{[rP]}(\boldsymbol{y}_t-\overline{\boldsymbol{y}}) = \frac{1}{\sqrt{P}}\sum_{t=1}^{[rP]}\boldsymbol{y}_t - \frac{[rP]}{P}\frac{1}{\sqrt{P}}\sum_{t=1}^{[rP]}\boldsymbol{y}_t \Rightarrow \boldsymbol{B}_{\mathbf{G}}(r) - r\boldsymbol{B}_{\mathbf{G}}(1);$$

a second applicaiton of the CMT leads to the desired limiting null distribution.

Proof of Proposition 2

Conditionally on the data, the bootstrap variables $\boldsymbol{y}_t^*$ are independent so they obey the same moment and serial dependence restrictions as $\boldsymbol{y}_t$. Therefore, we may obtain a representation of the bootstrap long-run covariance estimator parallelling to the one in the proof of Proposition 1,

$$\hat{\boldsymbol{\Omega}}^* = -\frac{1}{P^2}\sum_{i=1}^{P-1}\sum_{j=1}^{P-1}\frac{P^2}{B^2}k''\left(\frac{i-j}{B}\right)\frac{1}{\sqrt{P}}\sum_{t=1}^{i}\left(\boldsymbol{y}_t^*-\overline{\boldsymbol{y}^*}\right)\frac{1}{\sqrt{P}}\sum_{t=1}^{j}\left(\boldsymbol{y}_t^*-\overline{\boldsymbol{y}^*}\right)' + o_p(1)$$

for kernels with smooth derivatives, or

$$\begin{aligned}\hat{\boldsymbol{\Omega}}^* &= \frac{2}{bP}\sum_{i=1}^{P}\left(\frac{1}{\sqrt{P}}\sum_{t=1}^{i}\left(\boldsymbol{y}_t^*-\overline{\boldsymbol{y}^*}\right)\right)\left(\frac{1}{\sqrt{P}}\sum_{t=1}^{i}\left(\boldsymbol{y}_t^*-\overline{\boldsymbol{y}^*}\right)\right)' \\ &\quad -\frac{1}{bP}\sum_{i=1}^{[(1-b)P]}\left(\frac{1}{\sqrt{P}}\sum_{t=1}^{i+[bP]}\left(\boldsymbol{y}_t^*-\overline{\boldsymbol{y}^*}\right)\right)\left(\frac{1}{\sqrt{P}}\sum_{t=1}^{i}\left(\boldsymbol{y}_t^*-\overline{\boldsymbol{y}^*}\right)\right)' \\ &\quad -\frac{1}{bP}\sum_{i=1}^{[(1-b)P]}\left(\frac{1}{\sqrt{P}}\sum_{t=1}^{i}\left(\boldsymbol{y}_t^*-\overline{\boldsymbol{y}^*}\right)\right)\left(\frac{1}{\sqrt{P}}\sum_{t=1}^{i+[bP]}\left(\boldsymbol{y}_t^*-\overline{\boldsymbol{y}^*}\right)\right)' + o_p(1)\end{aligned}$$

for the Bartlett kernel. Note that $M \to \infty$ for any T, such that the bootstrapped quantiles converge in probability to the quantiles of the bootstrap distribution, and it then suffices to establish weak convergence in probability of the partial sums of $\boldsymbol{y}_t^*$ to $\boldsymbol{B}_{\mathbf{G}}(s)$

$$\frac{1}{\sqrt{P}}\sum_{t=R+1}^{R+[sP]}\boldsymbol{y}_t^* \overset{p}{\Rightarrow} \boldsymbol{B}_{\mathbf{G}}(s).$$

We now examine the case where the bootstrap variables r_t^* are standard normal. Let $\boldsymbol{S}_P^*(s)$ denote

the normalized partial sums of the bootstrapped sample,

$$\boldsymbol{S}_P^*(s) = \frac{1}{\sqrt{P}} \sum_{t=1}^{[sP]} \boldsymbol{y}_t^* = \frac{1}{\sqrt{P}} \sum_{t=1}^{[sP]} (\boldsymbol{y}_t - \bar{\boldsymbol{y}})\, r_t^*,$$

which, conditional on the sample $\boldsymbol{y}_t$, $t = 1, \ldots, T$, is a Gaussian process with independent increments. Its covariance kernel is given by

$$\mathrm{Cov}\left(\boldsymbol{S}_P^*(s), \boldsymbol{S}_P^*(r)\right) = \frac{1}{P} \sum_{t=1}^{[\min\{s,r\}P]} (\boldsymbol{y}_t - \bar{\boldsymbol{y}})\,(\boldsymbol{y}_t - \bar{\boldsymbol{y}})'\, \mathrm{E}\left((r_t^*)^2\right) = \frac{1}{P} \sum_{t=1}^{[\min\{s,r\}P]} (\boldsymbol{y}_t - \bar{\boldsymbol{y}})\,(\boldsymbol{y}_t - \bar{\boldsymbol{y}})'.$$

Note that, under Assumption 1, we obtain pointwise in s

$$\frac{1}{T} \sum_{j=1}^{[sT]} (\boldsymbol{y}_t - \bar{\boldsymbol{y}})\,(\boldsymbol{y}_t - \bar{\boldsymbol{y}})' \xrightarrow{p} \int_0^s \mathbf{G}(r)\, \mathrm{E}\left(\boldsymbol{v}_t \boldsymbol{v}_t'\right) \mathbf{G}'(r)\mathrm{d}r = c \int_0^s \mathbf{G}(r)\mathbf{G}'(r)\mathrm{d}r \tag{13}$$

via a Law of Large Numbers for strong mixing processes (see Davidson, 1994, Section 20.6).

Recall that the quadratic covariation process of the relevant Gaussian process $\boldsymbol{B}_{\mathbf{G}}$ is $c \int_0^s \mathbf{G}(r)\mathbf{G}'(r)\mathrm{d}r$. Then, like in the proof of Lemma A.5 in Cavaliere et al. (2010), weak convergence in probability of the bootstrap partial sums to a Gaussian process with independent increments and quadratic covariation process $c \int_0^s \mathbf{G}(r)\mathbf{G}'(r)\mathrm{d}r$ follows from uniformity of the convergence in (13).

Uniformity is indeed given, since the increments of the limit $c \int_0^s \mathbf{G}(r)\mathbf{G}'(r)\mathrm{d}r$ are positive semidefinite by construction, so any quadratic form thereof would be a continuous, nondecreasing function, hence leading to uniform convergence of the corresponding quadratic forms of the l.h.s. of (13). Given such univariate uniform convergence of any quadratic form, it follows that convergence in probability in (13) must be uniform.

To complete the argument for Gaussian bootstrap multiplicators, note that c cancels out in the expressions of the considered statistics as required for the result.

In the case where the bootstrap multiplicators r_t^* are not standard normal but follow the Mammen distribution, say, $\boldsymbol{S}_P^*(s)$ is not Gaussian, but weak convergence to a Gaussian process holds conditional on the sample (see e.g. Davidson, 1994, Corollary 29.14, with r_t^* being iid and having finite moments of any order). The result follows along the lines of the Gaussian argument above.

Proof of Lemma 3

Note that

$$\begin{aligned}\frac{1}{\sqrt{P}} \sum_{t=R+1}^{R+[sP]} \hat{\boldsymbol{y}}_t \;=\; & \frac{1}{\sqrt{P}} \sum_{t=R+1}^{R+[sP]} \boldsymbol{y}_t - \frac{1}{P} \sum_{t=R+1}^{R+[sP]} \mathbf{D}_1(f_{1,t}, \boldsymbol{\theta}_1) \cdot \sqrt{P}\left(\hat{\boldsymbol{\theta}}_{1,(t)} - \boldsymbol{\theta}_1\right) \\ & + \frac{1}{P} \sum_{t=R+1}^{R+[sP]} \mathbf{D}_2(f_{2,t}, \boldsymbol{\theta}_2) \cdot \sqrt{P}\left(\hat{\boldsymbol{\theta}}_{2,(t)} - \boldsymbol{\theta}_2\right) + Q_{s,P}\end{aligned}$$

where

$$Q_{s,P} = \sum_{i=1}^{2} (-1)^{i+1}\, \frac{1}{P} \sum_{t=R+1}^{R+[sP]} \left(\mathbf{D}_i(\hat{f}_{i,t}, \hat{\boldsymbol{\theta}}_{i,(t)}) - \mathbf{D}_i(f_{i,t}, \boldsymbol{\theta}_i)\right) \sqrt{P}\left(\hat{\boldsymbol{\theta}}_{i,(t)} - \boldsymbol{\theta}_i\right),$$

such that, for $1 \leq t \leq P+R$,

$$|Q_{s,P}| \leq 2 \sup_{i,t,\tilde{\boldsymbol{\theta}}_i} \left\| \mathbf{D}_i(\tilde{f}_{i,t}, \tilde{\boldsymbol{\theta}}_i) - \mathbf{D}_i(f_{i,t}, \boldsymbol{\theta}_i) \right\| \sup_{i,t} \sqrt{P} \left\| \hat{\boldsymbol{\theta}}_{i,(t)} - \boldsymbol{\theta}_i \right\|.$$

Furthermore, it follows from Assumption 4 that, for some $0 < \epsilon < 1$, the following weak convergence

$$\sqrt{R}\left(\hat{\boldsymbol{\theta}}_{i,[u(R+P)]} - \boldsymbol{\theta}_i\right) \Rightarrow \left(\mathbf{C}_i'(u)\,\mathbf{W}_{i,(\boldsymbol{\theta}_i)}\mathbf{C}_i(u)\right)^{-1}\mathbf{C}_i'(u)\,\mathbf{W}_{i,(\boldsymbol{\theta}_i)}\boldsymbol{A}_i(u)$$

holds on $[\epsilon, 1+\pi]$. Hence, with $P/R \to \pi > 0$ and $t > R$, $\sqrt{P}\left\|\hat{\boldsymbol{\theta}}_{i,(t)} - \boldsymbol{\theta}_i\right\|$ is uniformly (in t) bounded in probability, such that $\hat{\boldsymbol{\theta}}_{i,(t)} \in \Phi_P$ for all t w.p.1 and Assumption 2 ensures $\sup_{s\in[0,1]} |Q_{s,P}| \xrightarrow{p} 0$. Since the limit processes $\mathbf{H}_i$ are Lipschitz and deterministic, the result follows with the CMT.

Proof of Proposition 6

To establish the result, it suffices to show that weak convergence in probability to $\boldsymbol{B}_{\mathbf{G},\pi}(s)$ holds,

$$\frac{1}{\sqrt{P}} \sum_{t=R+1}^{R+[sP]} \hat{\boldsymbol{y}}_t^* \overset{p}{\Rightarrow} \boldsymbol{B}_{\mathbf{G},\pi}(s).$$

Taking the supremums over $1 \leq t \leq R+P$, note first that $\sup_t \|\mathbf{D}_i(f_{i,t}, \boldsymbol{\theta}_i)\| = O_p\left(P^{1/2-\gamma}\right)$ implies for $\gamma > 0$ that $\sup_t \|\hat{\boldsymbol{y}}_t - \boldsymbol{y}_t\| \xrightarrow{p} 0$ – given the behavior of $\hat{\boldsymbol{\theta}}_{i,(R+P)}$; see the proof of Lemma 3.

We then study the behavior (under the null $\boldsymbol{\mu} = \mathbf{0}$) of

$$\frac{1}{\sqrt{P}} \sum_{t=1}^{[uR]} \begin{pmatrix} \boldsymbol{a}_{1,t}^* \\ \boldsymbol{a}_{2,t}^* \\ \boldsymbol{y}_t^* \end{pmatrix} = \frac{1}{\sqrt{P}} \sum_{t=1}^{[uR]} \begin{pmatrix} \hat{\boldsymbol{a}}_{1,t} - \boldsymbol{a}_{1,t,(\boldsymbol{\theta}_1)} \\ \hat{\boldsymbol{a}}_{2,t} - \boldsymbol{a}_{2,t,(\boldsymbol{\theta}_2)} \\ \hat{\boldsymbol{y}}_t - \boldsymbol{y}_t \end{pmatrix} r_t^* + \frac{1}{\sqrt{P}} \sum_{t=1}^{[uR]} \begin{pmatrix} \boldsymbol{a}_{1,t,(\boldsymbol{\theta}_1)} \\ \boldsymbol{a}_{2,t,(\boldsymbol{\theta}_2)} \\ \boldsymbol{y}_t \end{pmatrix} r_t^*$$

for $u \leq 1+\pi$, where $\sup_t \left\| \begin{pmatrix} \hat{\boldsymbol{a}}_{i,t} - \boldsymbol{a}_{i,t,(\boldsymbol{\theta}_i)} \\ \hat{\boldsymbol{y}}_t - \boldsymbol{y}_t \end{pmatrix} \right\| \xrightarrow{p} 0$ implies that $\frac{1}{P}\sum_{t=R+1}^{R+[sP]} \left\| \begin{pmatrix} \hat{\boldsymbol{a}}_{i,t} - \boldsymbol{a}_{i,t,(\boldsymbol{\theta}_i)} \\ \hat{\boldsymbol{y}}_t - \boldsymbol{y}_t \end{pmatrix} r_t^* \right\| \overset{p}{\Rightarrow} 0$ since $\sup_t |r_t^*| = o\left(P^{-\gamma}\right)$ for any $\gamma > 0$ whenever r_t^* has finite moments of any order. Furthermore, like in the proof of Proposition 2, the change of variable $s = (u-1)/\pi$ for $1 \leq u \leq \pi$ gives

$$\frac{1}{\sqrt{P}} \sum_{t=R+1}^{R+[sP]} \begin{pmatrix} \boldsymbol{a}_{1,t,(\boldsymbol{\theta}_1)} \\ \boldsymbol{a}_{2,t,(\boldsymbol{\theta}_2)} \\ \boldsymbol{y}_t \end{pmatrix} r_t^* \overset{p}{\Rightarrow} \begin{pmatrix} \boldsymbol{A}_1(1+s\pi) \\ \boldsymbol{A}_2(1+s\pi) \\ \boldsymbol{B}_{\mathbf{G}}(s) \end{pmatrix}$$

on $[0,1]$. Moreover, $\sup_t \left\| \hat{\mathbf{C}}_{i,t} - \mathbf{C}_{i,t,(\boldsymbol{\theta}_i)} \right\| \xrightarrow{p} 0$ implies that

$$\frac{1}{R} \sum_{t=1}^{[uR]} \hat{\mathbf{C}}_{i,j} \Rightarrow \mathbf{C}_i(u)$$

such that, with $\mathbf{W}_i$ continuous,

$$\sqrt{R}\left(\hat{\boldsymbol{\theta}}_{i,(t)}^* - \hat{\boldsymbol{\theta}}_{i,(R+P)}\right) \overset{p}{\Rightarrow} \left(\mathbf{C}_i'(u)\,\mathbf{W}_{i,(\boldsymbol{\theta}_i)}\mathbf{C}_i(u)\right)^{-1}\mathbf{C}_i'(u)\,\mathbf{W}_{i,(\boldsymbol{\theta}_i)}\boldsymbol{A}_i(u)$$

on $[\epsilon, 1+\pi]$ for any $0 < \epsilon < 1$. Given the behavior of $\hat{\boldsymbol{\theta}}_{i,(R+P)}$ from the proof of Lemma 3, this implies that $\sup_{R\leq t\leq R+P} \left\|\hat{\boldsymbol{\theta}}_{i,(t)}^*\right\| = O_p\left(R^{-1/2}\right)$, and both $\hat{\boldsymbol{\theta}}_{i,(R+P)}$ and $\hat{\boldsymbol{\theta}}_{i,(t)}^*$ belong w.p.1 to the set Φ_P. The result follows along the lines of the proof of Lemma 3.

C Imputation

This appendix contains details on the imputed values for the missing observations in the SPF data set from the "Forecast Error Statistics for the Survey of Professional Forecasters" obtained from the Federal Reserve Bank of Philadelphia.

A few missing values in the SPF series and the are imputed via a bootstrap based expectation maximization [EM] algorithm, see Honaker et al. (2011). The algorithm makes use of the standard EM algorithm on multiple bootstrapped samples of the original data set (containing missing values) to obtain imputed values. We use 10,000 bootstrap replications for the EM algorithm. The code is written in `R` (by using the `Amelia` package) and available upon request from the authors. Tables 5–6 contain the imputed values (underlined) in connection to neighboring values. The obtained bootstrap averages serve as imputed values which are plausible.

Table 5: Data entries for the first release of RGDP and PGDP series. #MV gives the number of missing values in total. For underlined dates imputed values are obtained from the bootstrap-based EM algorithm. Neighboring values are reported for comparison.

Date	RGDP	PGDP
1995Q3	4.20481	0.58927
1995Q4	2.41452	2.26685
1996Q1	2.80932	2.60573
#MV	1	1

Table 6: Data entries for four-quarters ahead SPF forecasts. #MV gives the number of missing values in total. For underlined dates imputed values are obtained from the bootstrap-based EM algorithm. Neighboring values are reported for comparison.

Date	RGDP	PGDP	UNEMP	HOUSING
1969Q4	4.03701	3.21260	3.9	1.7
1970Q1	3.55115	3.56122	4.51615	1.38393
1970Q2	3.90855	3.56355	4.32870	1.35874
1970Q3	4.05961	3.90631	4.68596	1.42407
1970Q4	3.10037	3.01866	4.3	1.6
1971Q1	4.54798	4.15267	5.40173	1.50692
1971Q2	4.26233	2.95183	4.4	1.52
1975Q2	5.40498	3.50332	5.55	1.895
1975Q3	5.33554	6.52413	8.88112	1.37777
1975Q4	5.02638	6.57499	7.0	1.6
#MV	5	5	5	5

D Additional empirical results

This appendix contains additional empirical results. First, it reports evaluations against the final release, starting with summary statistics. Next, full-sample and time-variation test results are given. They are followed by similar tables computed for the case of asymmetric loss and being evaluated against the first and final release. The appendix ends with plots of forecast error loss differentials and graphs for the analysis of time-variation in the relative forecast performance.

Table 7: Summary statistics for output growth (RGDP), GDP deflator inflation (PGDP), unemployment rate (UNEMP) and the growth rate of housing starts (HOUSING) using the final data release and the MSE loss function. RelLoss(NC/SPF) denotes the relative RMSE loss of the no-change and the SPF forecasts. SD(·) labels the standard deviation of the loss differentials in the subsample I (1969-1984), II (1985-2006) or III (2007-2017). AC(1) denotes the empirical first-order autocorrelation coefficient of the loss differential series.

Statistic		RelLoss(NC/SPF)	SD(I)	SD(II)	SD(III)	AC(1)
Sample		1969-2017	1969-1984	1985-2006	2007-2017	1969-2017
RGDP						
	$h=0$	1.54	44.66	5.83	8.05	0.11
	$h=1$	1.39	51.28	7.42	10.57	0.21
	$h=4$	1.41	62.13	10.63	18.70	0.28
PGDP						
	$h=0$	1.35	4.79	1.44	2.30	0.22
	$h=1$	1.18	9.04	1.61	2.29	0.14
	$h=4$	1.08	15.47	2.55	2.12	0.33
UNEMP						
	$h=0$	2.45	0.44	0.08	0.36	0.42
	$h=1$	1.81	1.11	0.19	0.99	0.60
	$h=4$	1.43	2.22	0.50	2.16	0.67
HOUSING						
	$h=0$	1.48	0.04	0.01	0.01	0.13
	$h=1$	1.35	0.08	0.02	0.02	0.34
	$h=4$	1.20	0.24	0.05	0.07	0.62

Table 8: Test decisions for the full-sample $\mathcal{T}$-statistic either based on wild bootstrap ('bs') or asymptotic critical values ('asy'). Nowcasts ($h = 0$), one-quarter ($h = 1$) and one-year ahead forecasts ($h = 4$) are evaluated against the final data release under MSE loss. Evaluation sample runs from 1969Q4 to 2017Q2.

RGDP	$h=0$		$h=1$		$h=4$	
b	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$
0	***	***	***	***	***	***
0.1	***	***	***	**	***	**
0.2	***	*	***	*	***	*
0.3	**		**		**	
0.4	**		**		**	
0.5	**		**		*	
0.6	**		**		*	
0.7	**		*		*	
0.8	**		*		*	
0.9	**		*			
1	**		*		*	

PGDP	$h=0$		$h=1$		$h=4$	
b	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$
0	**	***				
0.1	***	***	*	**		
0.2	***	***	**	**		
0.3	***	**	**	**		
0.4	***	**	***	**		
0.5	***	**	***	**		
0.6	***	**	***	***	*	
0.7	***	**	***	**	**	
0.8	***	**	***	**	**	
0.9	***	**	***	**	**	
1	***	**	***	**	**	

UNEMP	$h=0$		$h=1$		$h=4$	
b	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$
0	***	***	***	***	***	***
0.1	***	***	***	***	***	***
0.2	***	***	***	***	***	***
0.3	***	**	***	**	**	**
0.4	***	**	***	**	**	**
0.5	***	**	***	**	**	**
0.6	***	**	***	**	**	**
0.7	***	**	***	**	***	**
0.8	***	**	***	**	***	**
0.9	***	**	***	**	***	**
1	***	**	***	**	***	**

HOUSING	$h=0$		$h=1$		$h=4$	
b	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$	$\mathcal{T}_{\text{bs}}$	$\mathcal{T}_{\text{asy}}$
0	***	***	***	***	***	**
0.1	***	**	***	**	**	*
0.2	**	*	**	*	*	
0.3	**		*			
0.4	*		*			
0.5	*					
0.6						
0.7						
0.8						
0.9						
1						

Table 9: Test decisions for the time-variation $\mathcal{T}^{\{Q,C,F\}}$-statistics either based on wild bootstrap ('bs') or asymptotic critical values ('asy'). Nowcasts ($h = 0$), one-quarter ($h = 1$) and one-year ahead forecasts ($h = 4$) are evaluated against the final data release under MSE loss. Evaluation sample runs from 1969Q4 to 2017Q2.

RGDP	$h=0$						$h=1$						$h=4$					
b	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$
0	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	
0.1	***	**	***	***	***	***	***	**	***	***	***	***	***	**	***	***	***	
0.2	**	*	***	**	**	*	**		***	**	*	*	**	*	**	**	**	
0.3	**		**	*	*		**		**	*			*		**	*		
0.4	**		**	*			*		**	*			*		**	*		
0.5	**		**				*		**						*			
0.6	*		**						**						*			
0.7	*		**						*						*			
0.8	*		**						**									
0.9	*		**						*						*			
1	*		**						*						*			

PGDP	$h=0$						$h=1$						$h=4$					
b	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$
0	***	***	**	***	**	***												
0.1	***	***	***	***	***	***	***	**	*	*	***	***						
0.2	***	***	***	***	***	***	***	**	**	**	***	***						
0.3	***	**	***	***	***	**	***	**	**	**	***	***					*	
0.4	***	**	***	**	**	**	***	**	***	**	***	***					*	
0.5	***	**	***	**	**	**	***	**	***	**	***	***					**	
0.6	***	**	***	**	**	**	***	**	***	**	***	***	*		*		**	
0.7	***	**	***	***	**	**	***	**	***	**	***	***	*		**		**	
0.8	***	**	***	***	**	**	***	**	***	**	***	***			**		**	
0.9	***	**	***	**	**	**	***	**	***	**	***	***	*		**		**	
1	***	**	***	***	***	**	***	**	***	**	***	***	*		**		**	

Table 10: continued from Table 9.

UNEMP	$h=0$						$h=1$						$h=4$					
b	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$
0	***	***	***	***	***	***	***	***	***	***	***	**	***	***	***	***	***	
0.1	***	***	***	***	***	***	***	***	***	***	**	***	***	***	***	***	***	
0.2	***	**	***	***	***	**	***	**	***	***	**	**	***	***	***	***	**	
0.3	***	**	***	**	**	**	**	**	***	**	**	*	**	**	***	**	*	
0.4	**	**	***	**	**	*	**	**	***	**	**	*	**	**	**	**	*	
0.5	**	**	***	**	*	*	**	**	***	**	*		**	**	**	**		
0.6	**	**	**	**	*	*	**	**	***	**	*	*	**	**	**	**	*	
0.7	**	**	***	**	*	*	**	**	**	**	**	*	**	**	***	**	*	
0.8	**	**	***	**	**	*	**	**	***	**	**	*	**	**	***	**	*	
0.9	**	**	***	**	**	*	**	**	***	**	**	*	***	**	***	**	*	
1	**	**	***	**	**	*	**	**	**	**	**	*	**	**	***	**	*	

HOUSING	$h=0$						$h=1$						$h=4$					
b	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$
0	***	***	***	***	***	***	***	***	***	***	***	***	**	**	***	***	**	
0.1	***	**	***	***	***	***	***	**	***	***	***	***	**		***	**	**	
0.2	**		**	**	*	*	**		**	**	*	*			*	*		
0.3	*		**	*			*		**	*								
0.4			*	*					*	*								
0.5																		
0.6			*															
0.7																		
0.8																		
0.9																		
1																		

Table 11: Summary statistics for output growth (RGDP), GDP deflator inflation (PGDP), unemployment rate (UNEMP) and the growth rate of housing starts (HOUSING) using the first data release and the asymmetric loss function. RelLoss(NC/SPF) denotes the relative asymmetric RMSE loss of the no-change and the SPF forecasts. SD(·) labels the standard deviation of the loss differentials in the subsample I (1969-1984), II (1985-2006) or III (2007-2017). AC(1) denotes the empirical first-order autocorrelation coefficient of the loss differential series.

Statistic		RelLoss(NC/SPF)	SD(I)	SD(II)	SD(III)	AC(1)
Sample		1969-2017	1969-1984	1985-2006	2007-2017	1969-2017
RGDP						
	$h=0$	1.69	18.02	2.57	3.99	0.26
	$h=1$	1.62	41.12	3.13	9.88	0.17
	$h=4$	1.65	36.07	3.80	10.11	0.52
PGDP						
	$h=0$	1.37	2.59	0.94	1.15	0.06
	$h=1$	1.17	3.92	1.04	0.83	0.33
	$h=4$	1.06	7.46	1.08	1.05	0.36
UNEMP						
	$h=0$	2.71	0.34	0.06	0.22	0.36
	$h=1$	1.86	0.78	0.12	0.68	0.61
	$h=4$	1.40	1.45	0.29	1.58	0.67
HOUSING						
	$h=0$	1.30	0.02	0.01	0.01	-0.01
	$h=1$	1.25	0.04	0.01	0.01	0.11
	$h=4$	1.27	0.12	0.03	0.02	0.72

Table 12: Summary statistics for output growth (RGDP), GDP deflator inflation (PGDP), unemployment rate (UNEMP) and the growth rate of housing starts (HOUSING) using the final data release and the asymmetric loss function. RelLoss(NC/SPF) denotes the relative asymmetric RMSE loss of the no-change and the SPF forecasts. SD(·) labels the standard deviation of the loss differentials in the subsample I (1969-1984), II (1985-2006) or III (2007-2017). AC(1) denotes the empirical first-order autocorrelation coefficient of the loss differential series.

Statistic		RelLoss(NC/SPF)	SD(I)	SD(II)	SD(III)	AC(1)
Sample		1969-2017	1969-1984	1985-2006	2007-2017	1969-2017
RGDP						
	$h=0$	1.53	30.32	3.32	4.36	0.09
	$h=1$	1.44	34.78	4.52	6.88	0.26
	$h=4$	1.57	41.89	6.27	11.9	0.37
PGDP						
	$h=0$	1.30	2.58	0.70	0.91	0.20
	$h=1$	1.10	4.26	0.82	0.93	0.25
	$h=4$	1.01	7.47	1.11	0.98	0.42
UNEMP						
	$h=0$	2.73	0.30	0.05	0.26	0.44
	$h=1$	1.91	0.75	0.13	0.71	0.62
	$h=4$	1.40	1.46	0.31	1.58	0.67
HOUSING						
	$h=0$	1.37	0.02	0.01	0.01	0.05
	$h=1$	1.28	0.04	0.01	0.01	0.20
	$h=4$	1.28	0.12	0.03	0.02	0.74

Table 13: Test decisions for the full-sample $\mathcal{T}$-statistic either based on wild bootstrap ('bs') or asymptotic critical values ('asy'). Nowcasts ($h = 0$), one-quarter ($h = 1$) and one-year ahead forecasts ($h = 4$) are evaluated against the first data release under asymmetric loss. Evaluation sample runs from 1969Q4 to 2017Q2.

RGDP	$h=0$		$h=1$		$h=4$	
b	$\mathcal{T}$bs	$\mathcal{T}$asy	$\mathcal{T}$bs	$\mathcal{T}$asy	$\mathcal{T}$bs	$\mathcal{T}$asy
0	***	***	***	**	***	**
0.1	***	**	***	**	***	**
0.2	***	*	***	*	**	*
0.3	***		**		**	
0.4	**		**		**	
0.5	**		**		*	
0.6	**		**		*	
0.7	**		*		*	
0.8	**		*			
0.9	**		*			
1	**		*		*	

PGDP	$h=0$		$h=1$		$h=4$	
b	$\mathcal{T}$bs	$\mathcal{T}$asy	$\mathcal{T}$bs	$\mathcal{T}$asy	$\mathcal{T}$bs	$\mathcal{T}$asy
0	***	***		*		
0.1	***	***	**	***		
0.2	***	***	***	**		
0.3	***	***	**	**		
0.4	***	***	***	**		
0.5	***	***	***	**		
0.6	***	***	***	**		
0.7	***	***	***	**		
0.8	***	***	***	**		
0.9	***	***	***	**		
1	***	***	***	**		

UNEMP	$h=0$		$h=1$		$h=4$	
b	$\mathcal{T}$bs	$\mathcal{T}$asy	$\mathcal{T}$bs	$\mathcal{T}$asy	$\mathcal{T}$bs	$\mathcal{T}$asy
0	***	***	***	***	***	***
0.1	***	***	***	***	***	***
0.2	***	**	***	**	***	**
0.3	***	**	***	**	**	**
0.4	**	**	***	**	**	**
0.5	**	**	***	**	**	**
0.6	**	**	**	**	**	**
0.7	**	**	**	**	**	**
0.8	**	**	**	**	**	**
0.9	**	**	**	**	**	**
1	**	**	***	**	***	**

HOUSING	$h=0$		$h=1$		$h=4$	
b	$\mathcal{T}$bs	$\mathcal{T}$asy	$\mathcal{T}$bs	$\mathcal{T}$asy	$\mathcal{T}$bs	$\mathcal{T}$asy
0	***	***	***	***	**	**
0.1	***	**	***	**	**	*
0.2	**		**		*	
0.3	*		*			
0.4						
0.5						
0.6						
0.7						
0.8						
0.9						
1						

Table 14: Test decisions for the full-sample $\mathcal{T}$-statistic either based on wild bootstrap ('bs') or asymptotic critical values ('asy'). Nowcasts ($h = 0$), one-quarter ($h = 1$) and one-year ahead forecasts ($h = 4$) are evaluated against the final data release under asymmetric loss. Evaluation sample runs from 1969Q4 to 2017Q2.

RGDP	$h=0$		$h=1$		$h=4$	
b	$\mathcal{T}$bs	$\mathcal{T}$asy	$\mathcal{T}$bs	$\mathcal{T}$asy	$\mathcal{T}$bs	$\mathcal{T}$asy
0	***	***	***	***	***	***
0.1	***	**	***	**	***	**
0.2	***	*	**	*	**	*
0.3	**		**		**	
0.4	**		**		*	
0.5	**		*		*	
0.6	**		*		*	
0.7	**		*		*	
0.8	**		*		*	
0.9	**		*		*	
1	**		*		*	

PGDP	$h=0$		$h=1$		$h=4$	
b	$\mathcal{T}$bs	$\mathcal{T}$asy	$\mathcal{T}$bs	$\mathcal{T}$asy	$\mathcal{T}$bs	$\mathcal{T}$asy
0	**	***				
0.1	***	***				
0.2	***	***				
0.3	***	***				
0.4	***	***				
0.5	***	***				
0.6	***	***				
0.7	***	***				
0.8	***	***				
0.9	***	***				
1	***	***				

UNEMP	$h=0$		$h=1$		$h=4$	
b	$\mathcal{T}$bs	$\mathcal{T}$asy	$\mathcal{T}$bs	$\mathcal{T}$asy	$\mathcal{T}$bs	$\mathcal{T}$asy
0	***	***	***	***	***	***
0.1	***	***	***	***	***	***
0.2	***	**	***	**	***	**
0.3	***	**	**	**	***	**
0.4	***	**	***	**	**	**
0.5	**	**	**	**	**	**
0.6	**	**	**	**	**	**
0.7	**	**	**	**	**	**
0.8	***	**	***	**	***	**
0.9	**	**	**	**	**	**
1	**	**	**	**	***	**

HOUSING	$h=0$		$h=1$		$h=4$	
b	$\mathcal{T}$bs	$\mathcal{T}$asy	$\mathcal{T}$bs	$\mathcal{T}$asy	$\mathcal{T}$bs	$\mathcal{T}$asy
0	***	***	***	***	**	**
0.1	***	**	***	**	**	*
0.2	**	*	**		*	
0.3	*		*			
0.4						
0.5						
0.6						
0.7						
0.8						
0.9						
1						

Table 15: Test decisions for the time-variation $\mathcal{T}^{\{Q,C,F\}}$-statistics either based on wild bootstrap ('bs') or asymptotic critical values ('asy'). Nowcasts ($h = 0$), one-quarter ($h = 1$) and one-year ahead forecasts ($h = 4$) are evaluated against the first data release under asymmetric loss. Evaluation sample runs from 1969Q4 to 2017Q2.

RGDP	$h=0$						$h=1$						$h=4$					
b	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$
0	***	***	***	***	***	***	***	**	***	***	**	**	***	**	***	***	***	***
0.1	***	**	***	***	***	***	**	**	***	***	**	**	**	*	***	**	**	**
0.2	***	*	***	**	**	*	**		**	**	*	*	**		**	**		*
0.3	**		**	*			*		**	*					**	*		
0.4	*		**	*			*		**	*					*			
0.5	*		**				*		*						*			
0.6	*		**				*		*						*			
0.7			**				*		*						*			
0.8			*						*						*			
0.9			**				*		*						*			
1			**						*									

PGDP	$h=0$						$h=1$						$h=4$					
b	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$	$\mathcal{T}^Q_{\text{bs}}$	$\mathcal{T}^Q_{\text{asy}}$	$\mathcal{T}^C_{\text{bs}}$	$\mathcal{T}^C_{\text{asy}}$	$\mathcal{T}^F_{\text{bs}}$	$\mathcal{T}^F_{\text{asy}}$
0	***	***	***	***	***	***	*											
0.1	***	***	***	***	***	***	***	**	**	**	***	***						*
0.2	***	***	***	***	***	***	***	**	**	**	***	***					**	**
0.3	***	***	***	***	***	***	***	**	**	**	**	**	*				**	***
0.4	***	***	***	***	***	**	***	**	**	**	**	**	*				**	**
0.5	***	***	***	***	***	**	***	**	***	**	**	**	*				**	**
0.6	***	***	***	***	***	**	***	**	***	**	**	**	*				**	**
0.7	***	***	***	***	***	**	***	**	***	**	**	**	*				**	**
0.8	***	***	***	***	***	**	***	**	***	**	**	**					**	**
0.9	***	***	***	***	***	**	***	**	***	**	**	**					**	**
1	***	***	***	***	***	**	***	**	***	**	**	**					*	**

Table 16: continued from Table 15.

UNEMP	$h=0$						$h=1$						$h=4$					
b	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$
0	***	***	***	***	***	**	***	***	***	***	**	*	***	***	***	***	**	**
0.1	***	***	***	***	***	***	***	**	***	***	**	**	***	***	***	***	**	**
0.2	***	**	***	**	***	**	***	**	***	**	**	*	***	**	***	**	*	*
0.3	**	**	***	**	**	*	**	**	***	**	*		**	**	**	**		
0.4	**	*	***	**	*		**	*	***	**	*		**	**	**	**		
0.5	**	*	**	**	*		**	*	**	**	*		**	*	**	**		
0.6	**	*	**	**	*		**	*	**	**	*		**	**	**	**		
0.7	**	*	**	**	*		**	**	**	**	*		**	**	**	**		
0.8	**	*	**	**	*		**	**	**	**	**		**	**	***	**	*	
0.9	**	*	**	**	*		**	*	***	**	*		**	**	**	**	*	
1	**	*	**	**	*		**	*	**	**	*		**	**	**	**	*	

HOUSING	$h=0$						$h=1$						$h=4$					
b	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$
0	***	***	***	***	***	***	***	***	***	***	***	***	**	**	**	***	**	***
0.1	**	**	***	***	**	***	***	**	***	***	***	***	**	*	***	***	**	***
0.2	*		**	**			*		**	**					*	*		
0.3			*	*						*								
0.4																		
0.5																		
0.6																		
0.7																		
0.8																		
0.9																		
1																		

Table 17: Test decisions for the time-variation $\mathcal{T}^{\{Q,C,F\}}$-statistics either based on wild bootstrap ('bs') or asymptotic critical values ('asy'). Nowcasts ($h = 0$), one-quarter ($h = 1$) and one-year ahead forecasts ($h = 4$) are evaluated against the final data release under asymmetric loss. Evaluation sample runs from 1969Q4 to 2017Q2.

RGDP

b	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$
0	***	***	***	***	***	***	***	**	***	***	**	***	***	**	***	***	***	***
0.1	***	**	***	***	***	***	***	**	***	***	**	***	**	**	***	***	**	***
0.2	**		**	**	**	*	**		**	**	*	*	**		**	**	*	**
0.3	**		**	*	*		*		**	*			*		**	*		
0.4	**		**				*		**						*	*		
0.5	*		**						**						*			
0.6			**						*						*			
0.7			**						*						*			
0.8			**						*						*			
0.9			**						*									
1			**						*									

PGDP

b	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$
0	***	***	**	***	*	**												
0.1	***	***	***	***	***	***												
0.2	***	***	***	***	***	***												
0.3	***	***	***	***	***	***												
0.4	***	***	***	***	***	***												
0.5	***	***	***	***	***	***												
0.6	***	***	***	***	***	***												
0.7	***	***	***	***	***	***												
0.8	***	***	***	***	***	***												
0.9	***	***	***	***	***	***												
1	***	***	***	***	***	***												

Table 18: continued from Table 17.

b	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$
0	***	***	***	***	***	**	***	***	***	***	**	*	***	***	***	***	**	**
0.1	***	***	***	***	***	**	***	**	***	***	**	**	***	***	***	***	**	**
0.2	***	**	***	***	***	**	***	**	***	**	**	*	***	**	***	**	**	*
0.3	**	**	***	**	**	*	**	**	***	**	**		**	**	**	**	*	
0.4	**	**	***	**	**		**	**	***	**	**		**	**	**	**		
0.5	**	**	**	**	*		**	**	***	**	*		**	**	**	**		
0.6	**	**	**	**	*		**	**	**	**	*		**	**	**	**		
0.7	**	**	**	**	*		**	**	**	**	*		**	**	***	**	*	
0.8	**	**	**	**	*		**	**	***	**	*		**	**	***	**	*	
0.9	**	**	**	**	*		**	**	***	**	*		**	**	***	**		
1	**	**	**	**	*		**	**	***	**	*		**	**	**	**		

HOUSING

b	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$	$\mathcal{T}^{Q}_{\text{bs}}$	$\mathcal{T}^{Q}_{\text{asy}}$	$\mathcal{T}^{C}_{\text{bs}}$	$\mathcal{T}^{C}_{\text{asy}}$	$\mathcal{T}^{F}_{\text{bs}}$	$\mathcal{T}^{F}_{\text{asy}}$
0	***	***	***	***	***	***	***	***	***	***	***	***	**	**	**	***	**	***
0.1	**	**	***	***	**	***	**	**	***	***	***	***	**	*	**	***	**	***
0.2	*		**	**			*		**	**					*	*		
0.3			*	*					*	*								
0.4																		
0.5																		
0.6																		
0.7																		
0.8																		
0.9																		
1																		

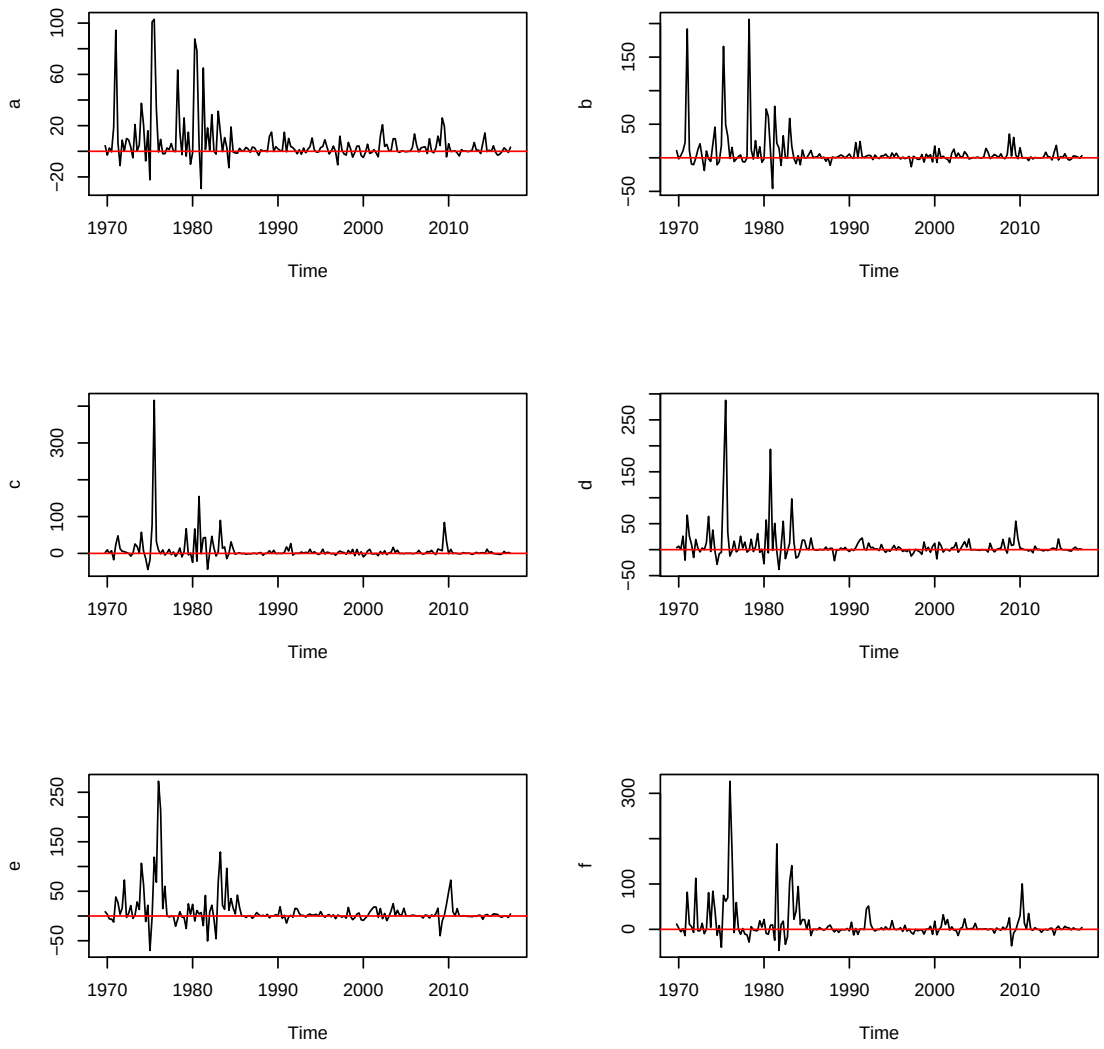


Figure 10: RGDP loss differentials; (a/b) nowcast, evaluated against first/final release; (c/d) one-quarter ahead forecast, evaluated against first/final release; (e/f) one-year ahead forecast, evaluated against first/final release.

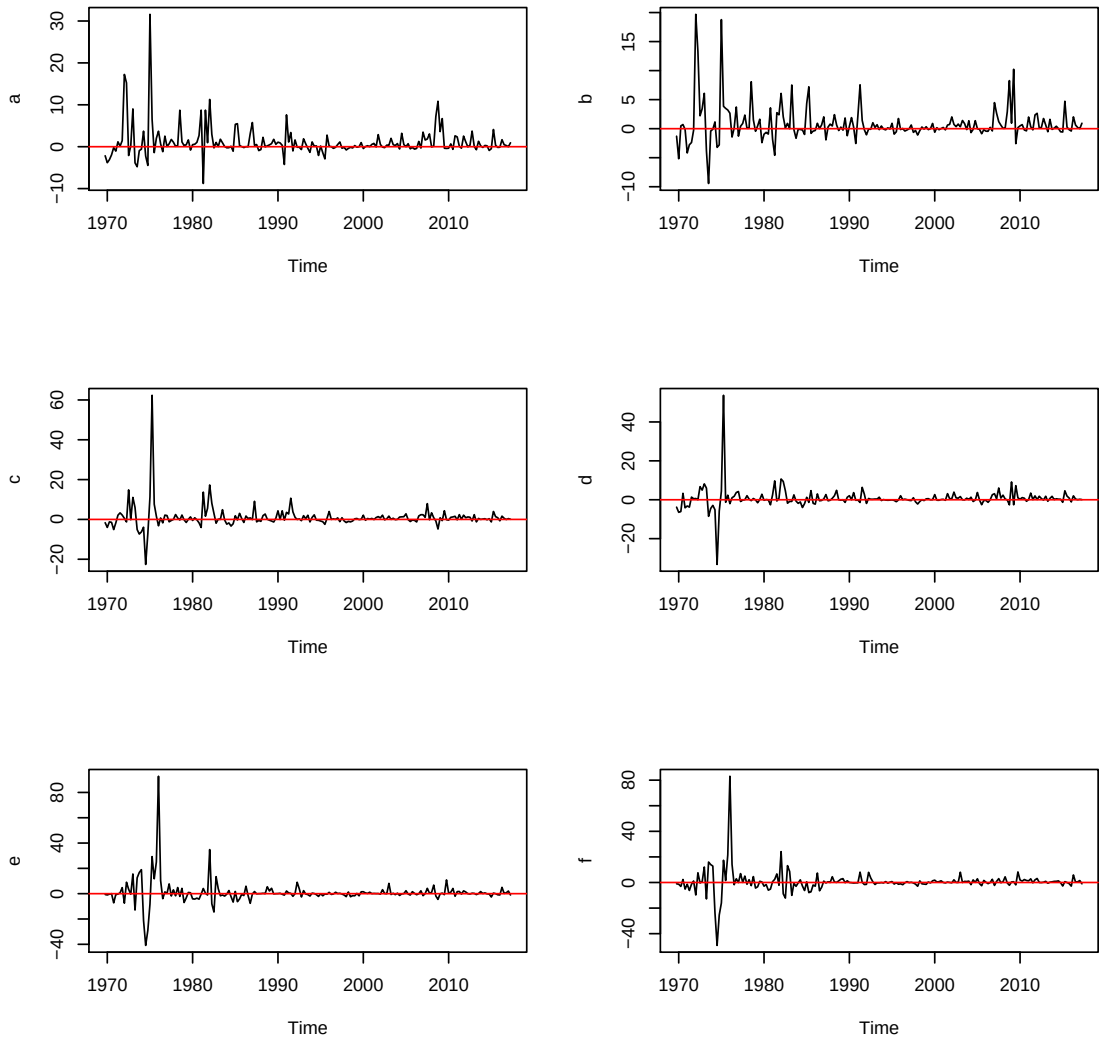


Figure 11: PGDP loss differentials; (a/b) nowcast, evaluated against first/final release; (c/d) one-quarter ahead forecast, evaluated against first/final release; (e/f) one-year ahead forecast, evaluated against first/final release.

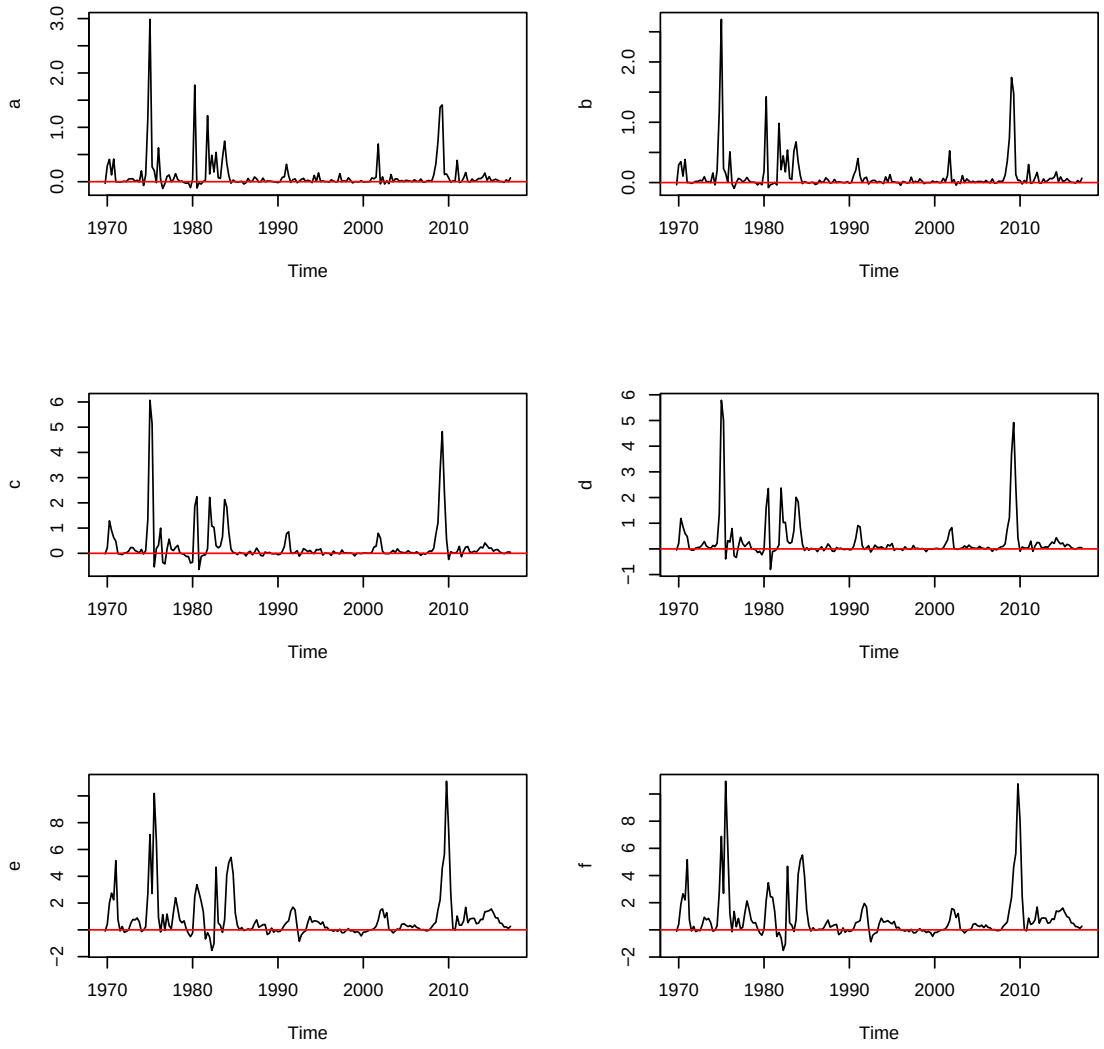


Figure 12: UNEMP loss differentials; (a/b) nowcast, evaluated against first/final release; (c/d) one-quarter ahead forecast, evaluated against first/final release; (e/f) one-year ahead forecast, evaluated against first/final release.

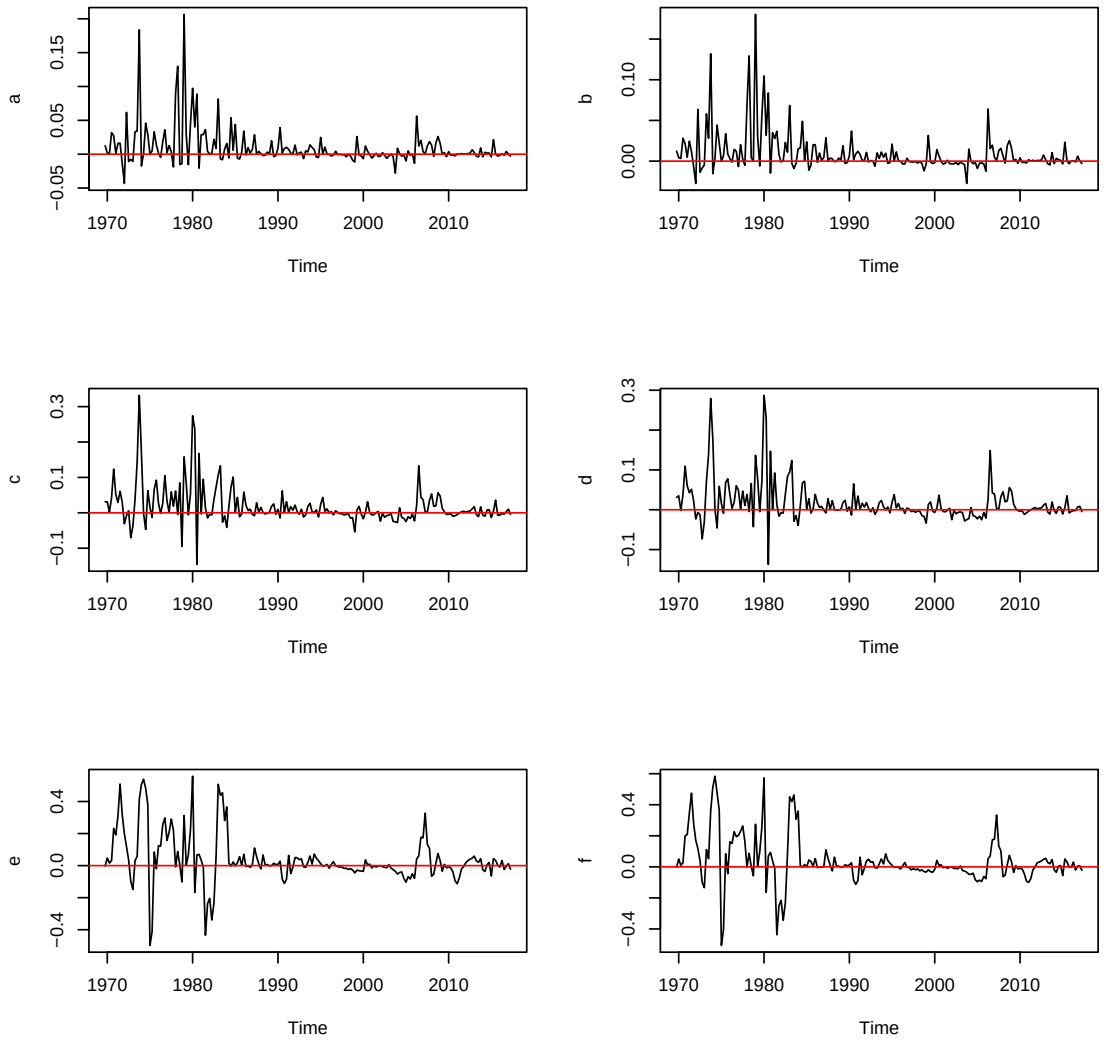


Figure 13: HOUSING loss differentials; (a/b) nowcast, evaluated against first/final release; (c/d) one-quarter ahead forecast, evaluated against first/final release; (e/f) one-year ahead forecast, evaluated against first/final release.

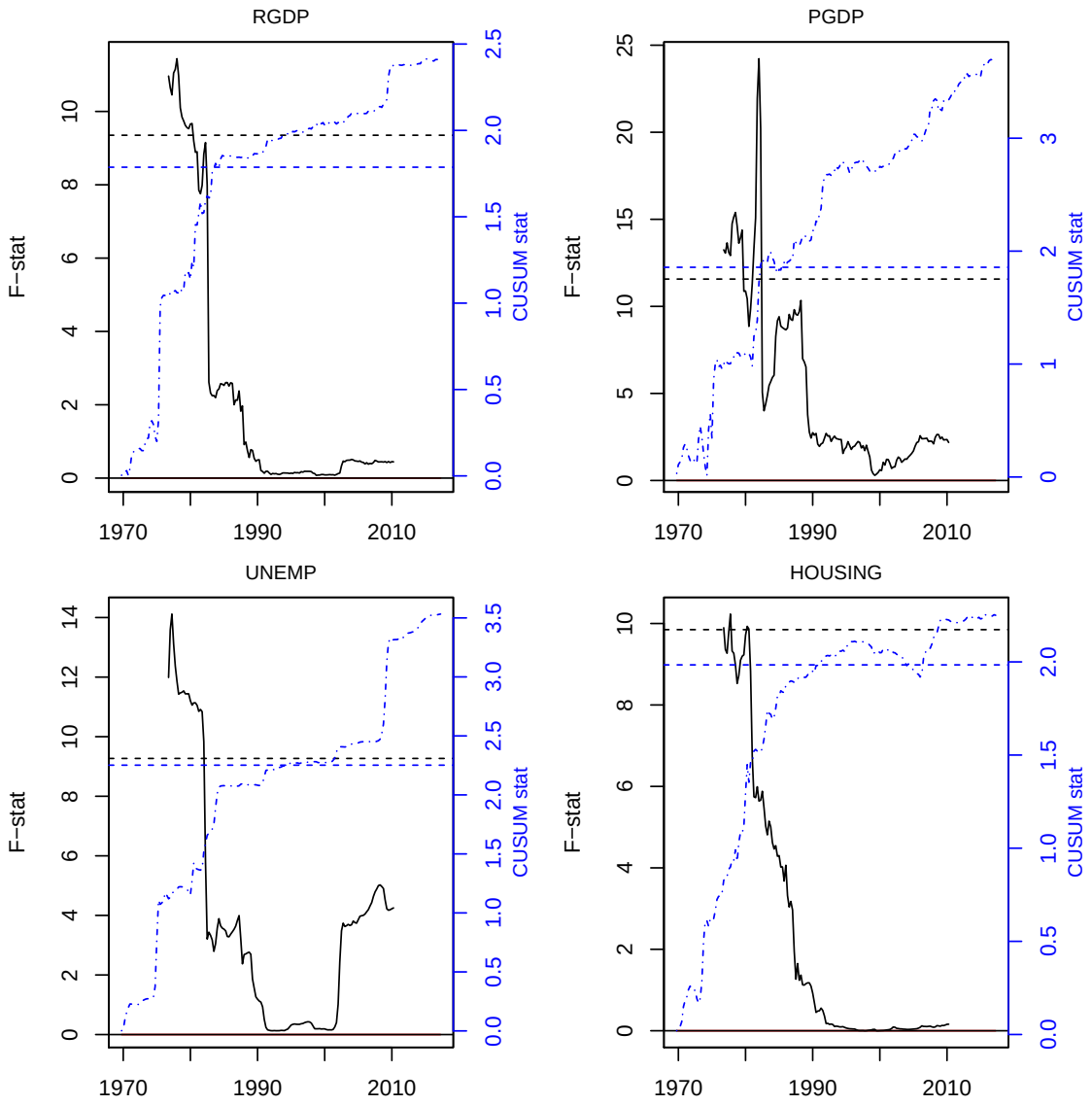


Figure 14: The plots show the time-varying components of the fluctuation statistic (left axis, solid black line) and the CUSUM statistic (right axis, dashed-dotted blue line), see equations (2) and (3). Horizontal dashed lines are the corresponding five percent critical values for the *maximum* of the displayed statistics. One-quarter ahead forecasts are evaluated against the first release for mean squared error loss; $b = 0.2$, $\nu = 0.3$.

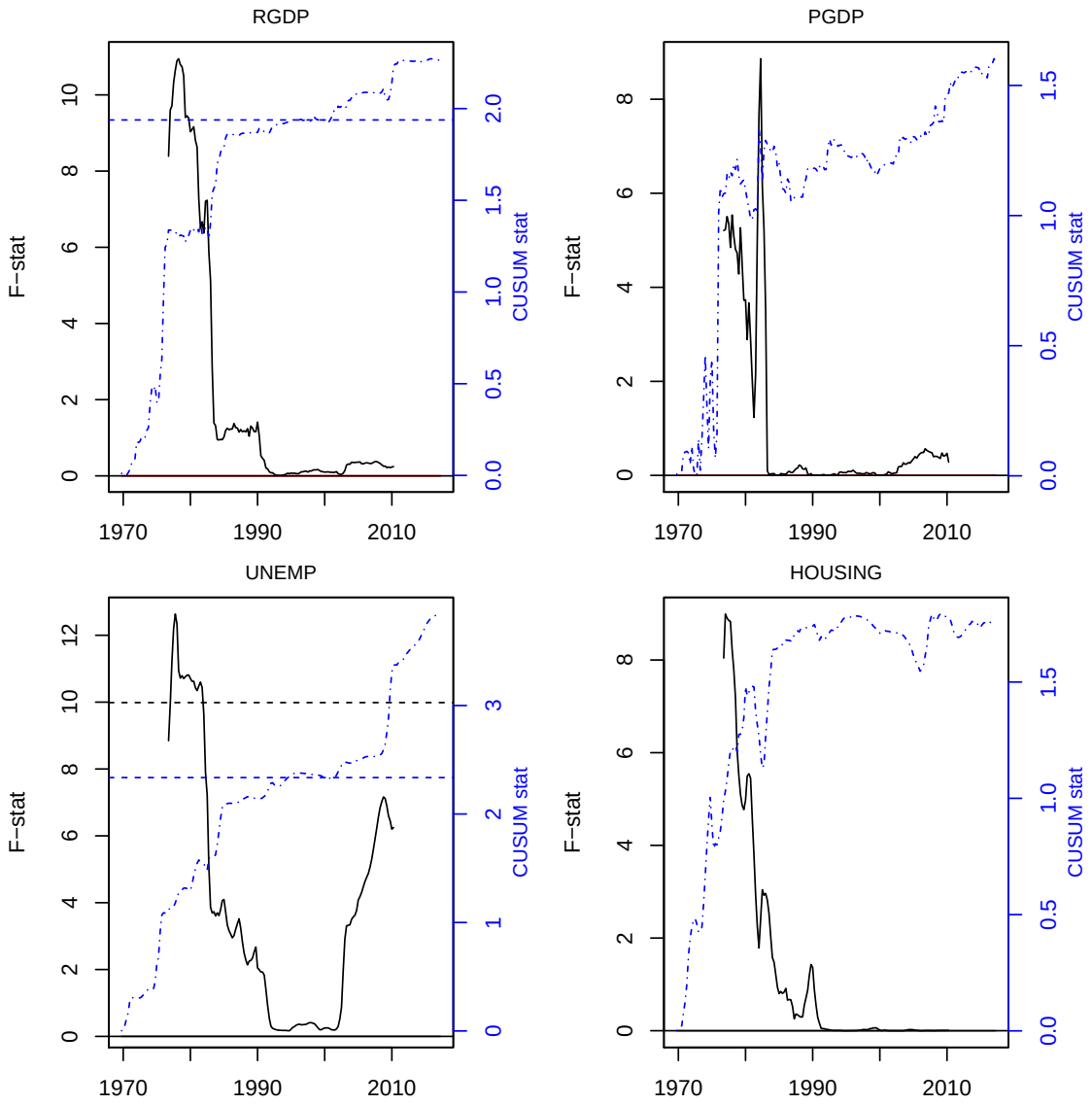


Figure 15: The plots show the time-varying components of the fluctuation statistic (left axis, solid black line) and the CUSUM statistic (right axis, dashed-dotted blue line), see equations (2) and (3). Horizontal dashed lines are the corresponding five percent critical values for the *maximum* of the displayed statistics. One-year ahead forecasts are evaluated against the first release for mean squared error loss; $b = 0.2$, $\nu = 0.3$.

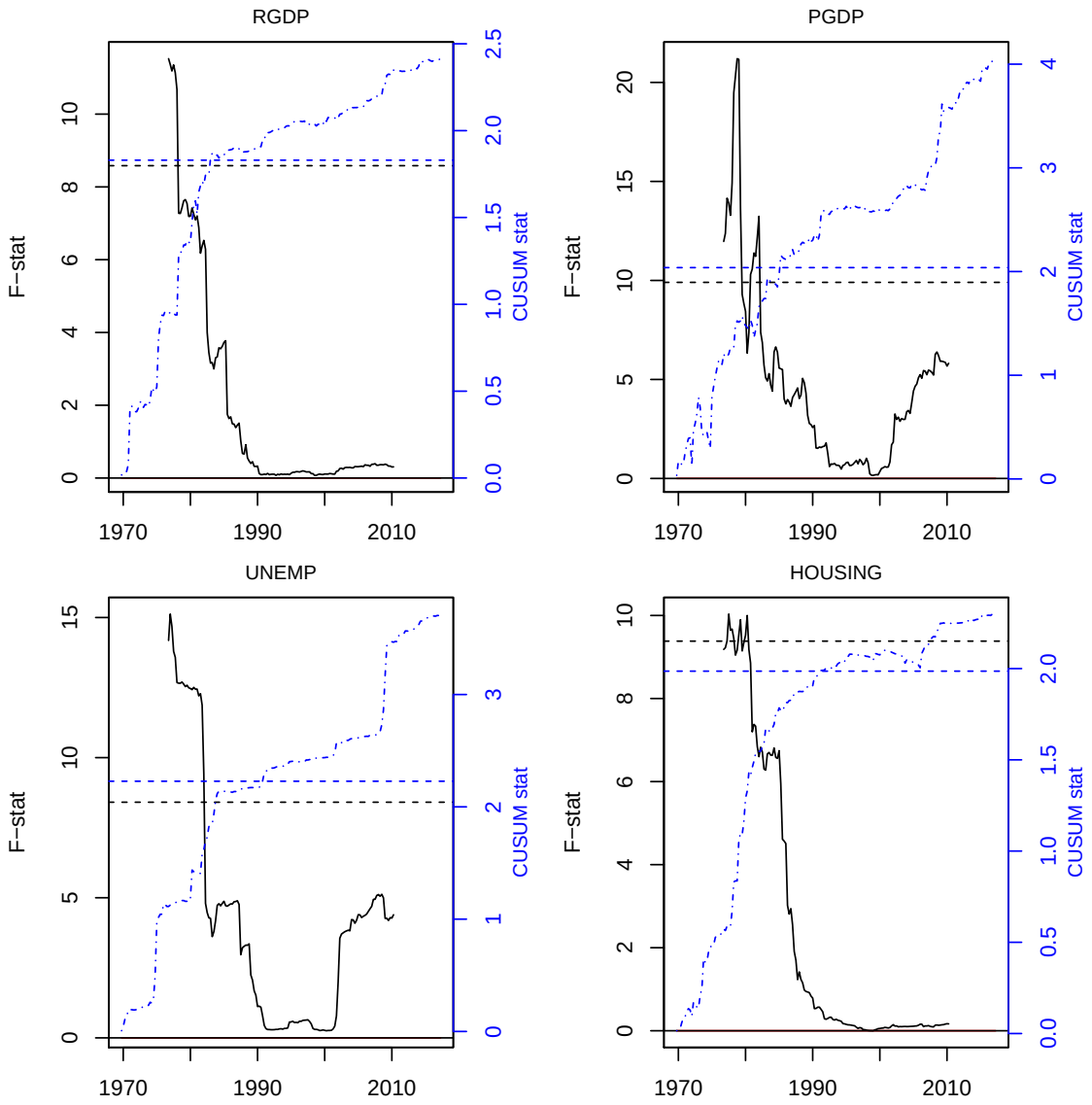


Figure 16: The plots show the time-varying components of the fluctuation statistic (left axis, solid black line) and the CUSUM statistic (right axis, dashed-dotted blue line), see equations (2) and (3). Horizontal dashed lines are the corresponding five percent critical values for the *maximum* of the displayed statistics. Nowcasts are evaluated against the final release for mean squared error loss; $b = 0.2$, $\nu = 0.3$.

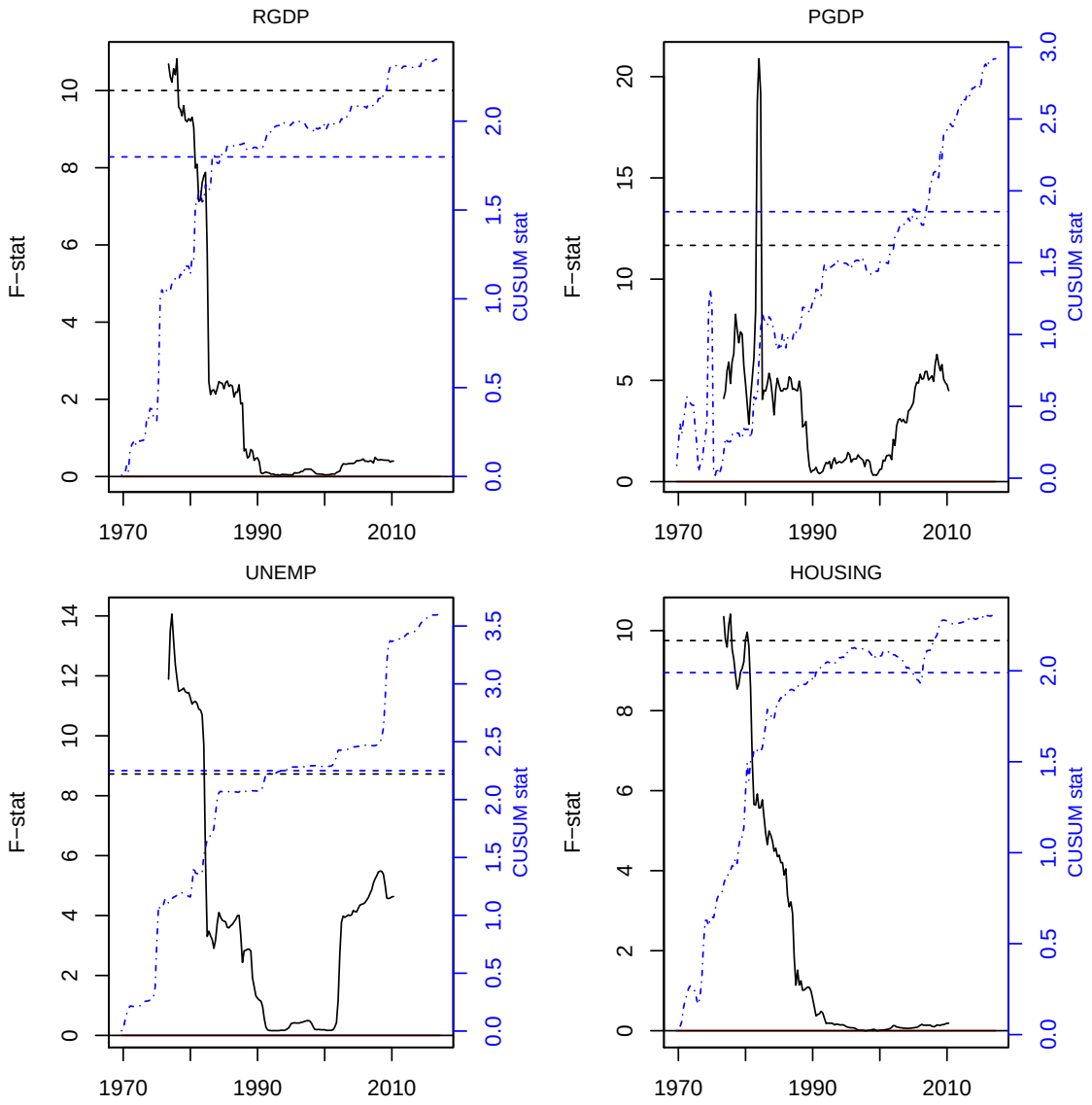


Figure 17: The plots show the time-varying components of the fluctuation statistic (left axis, solid black line) and the CUSUM statistic (right axis, dashed-dotted blue line), see equations (2) and (3). Horizontal dashed lines are the corresponding five percent critical values for the *maximum* of the displayed statistics. One-quarter ahead forecasts are evaluated against the final release for mean squared error loss; $b = 0.2$, $\nu = 0.3$.

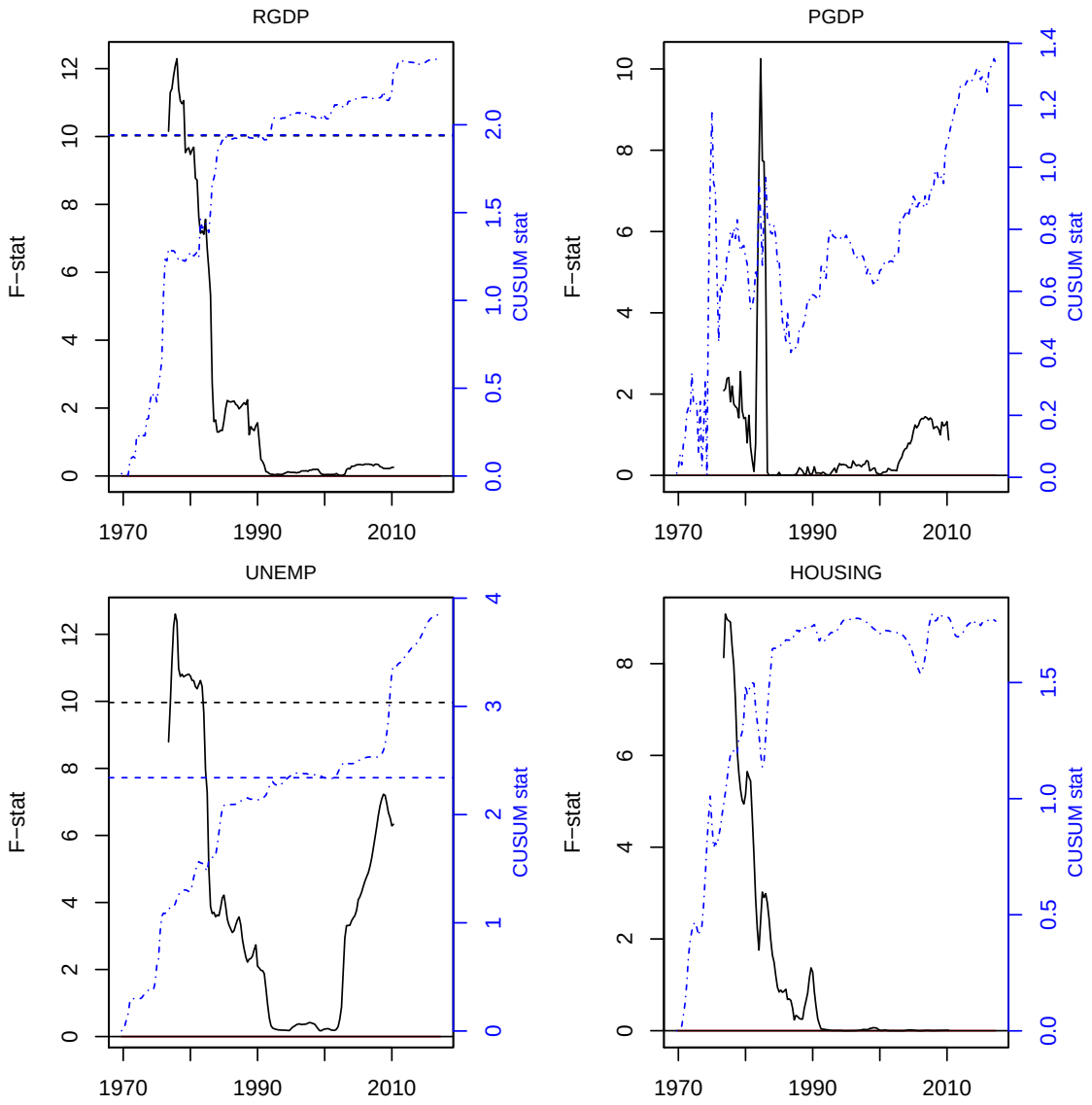


Figure 18: The plots show the time-varying components of the fluctuation statistic (left axis, solid black line) and the CUSUM statistic (right axis, dashed-dotted blue line), see equations (2) and (3). Horizontal dashed lines are the corresponding five percent critical values for the *maximum* of the displayed statistics. One-year ahead forecasts are evaluated against the final release for mean squared error loss; $b = 0.2$, $\nu = 0.3$.

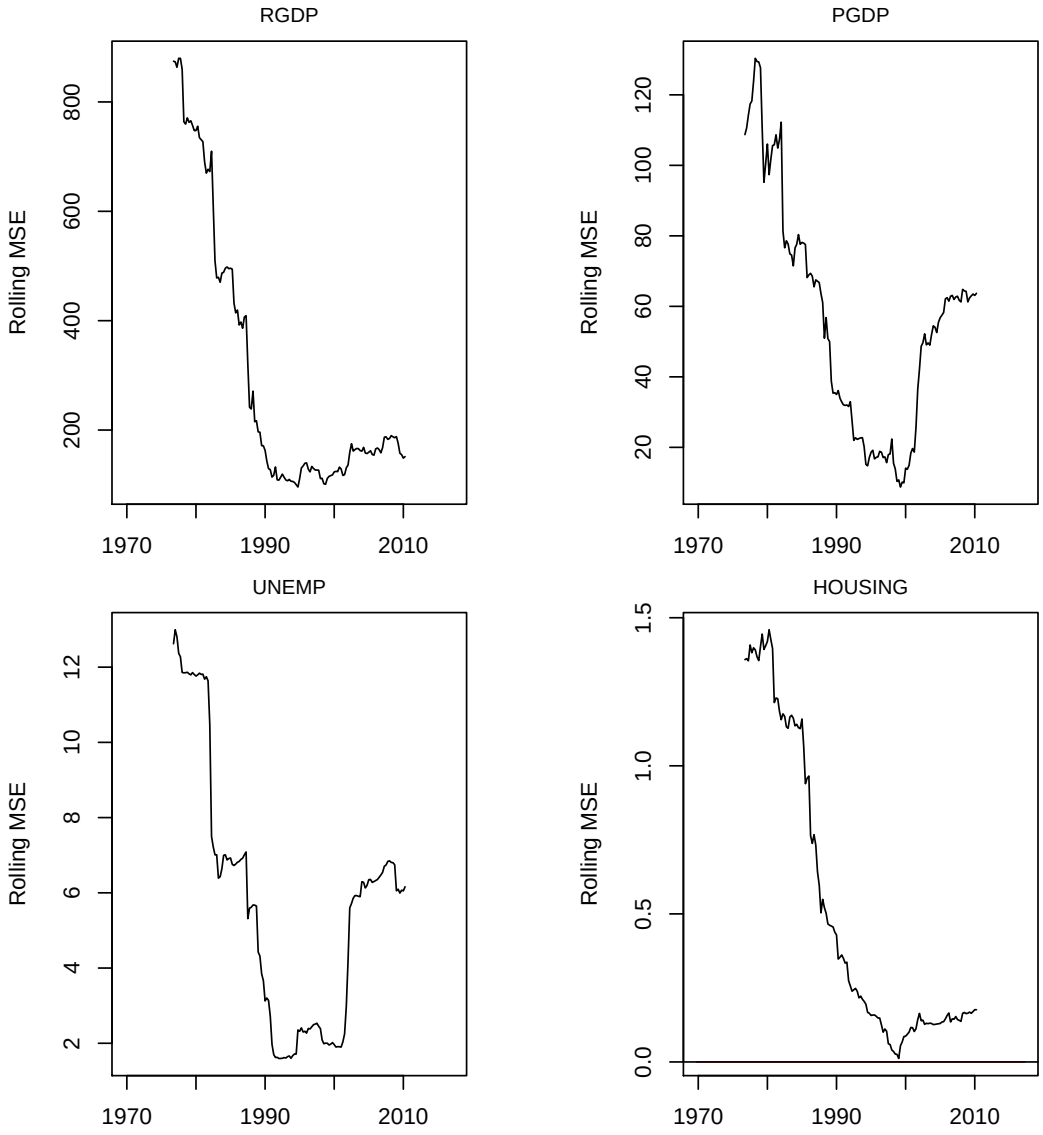


Figure 19: The plots show the rolling MSE difference (unscaled, $\nu = 0.3$). One-quarter ahead forecasts are evaluated against the first release.

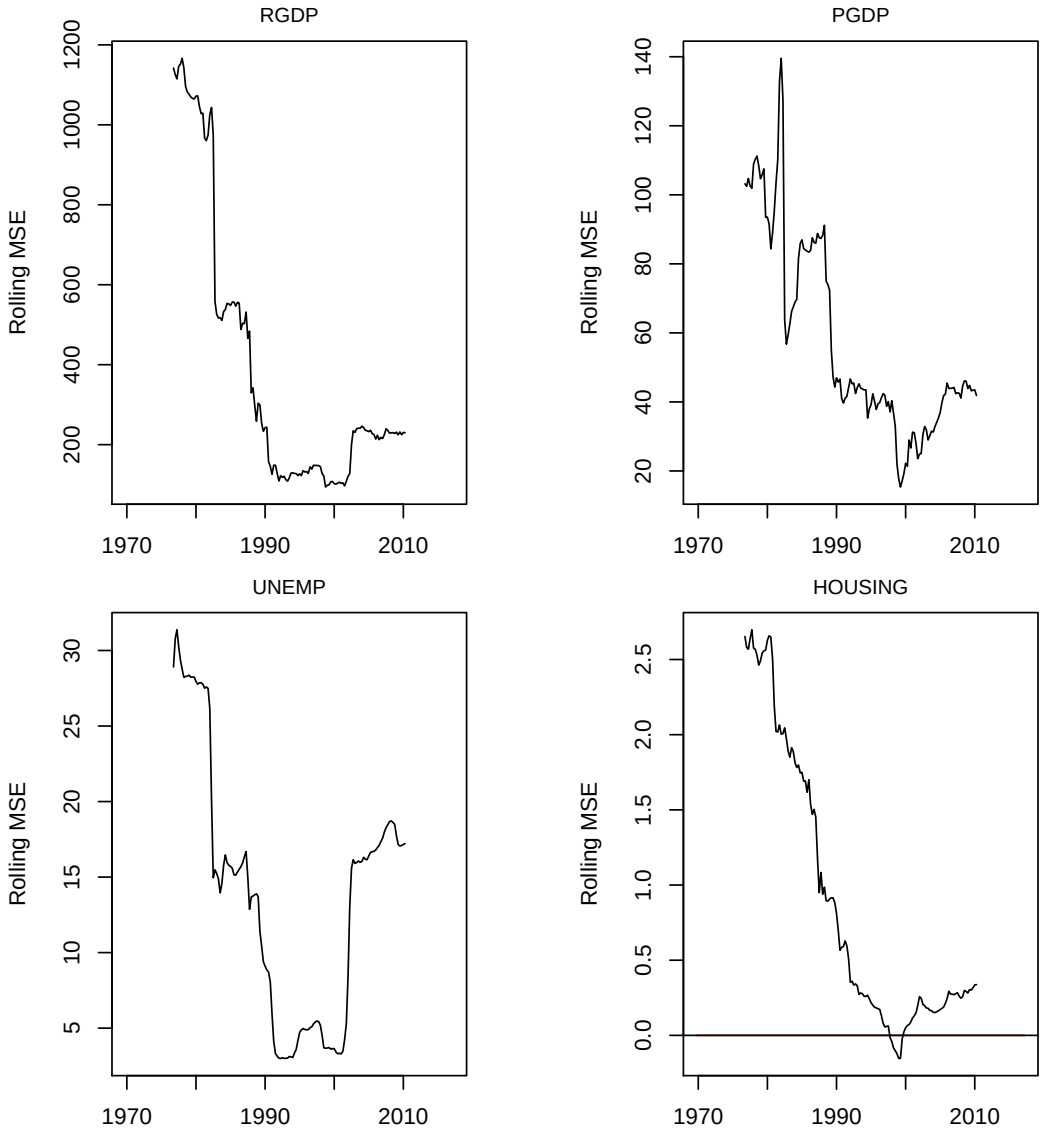


Figure 20: The plots show the rolling MSE difference (unscaled, $\nu = 0.3$). One-quarter ahead forecasts are evaluated against the first release.

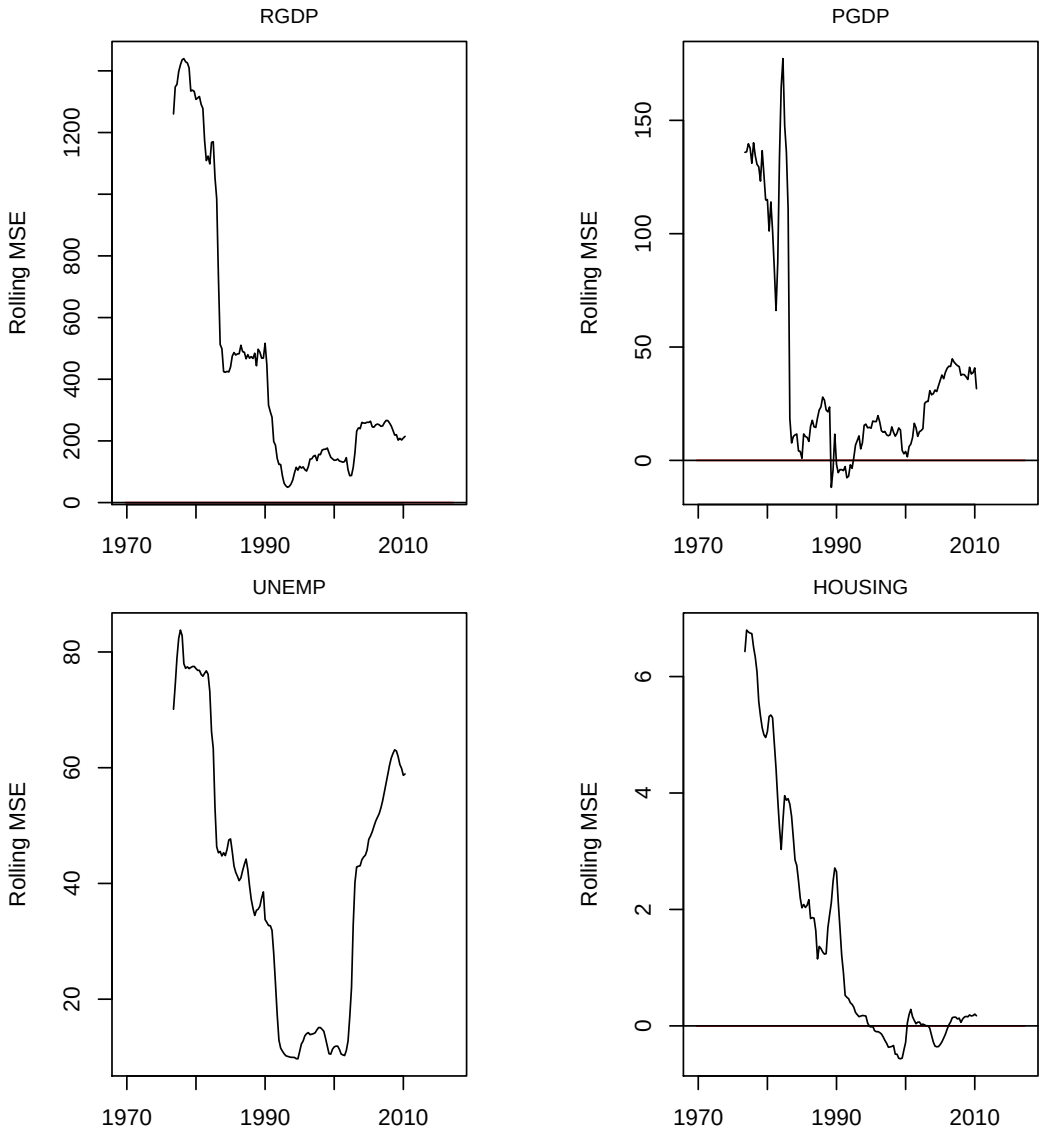


Figure 21: The plots show the rolling MSE difference (unscaled, $\nu = 0.3$). One-year ahead forecasts are evaluated against the first release.

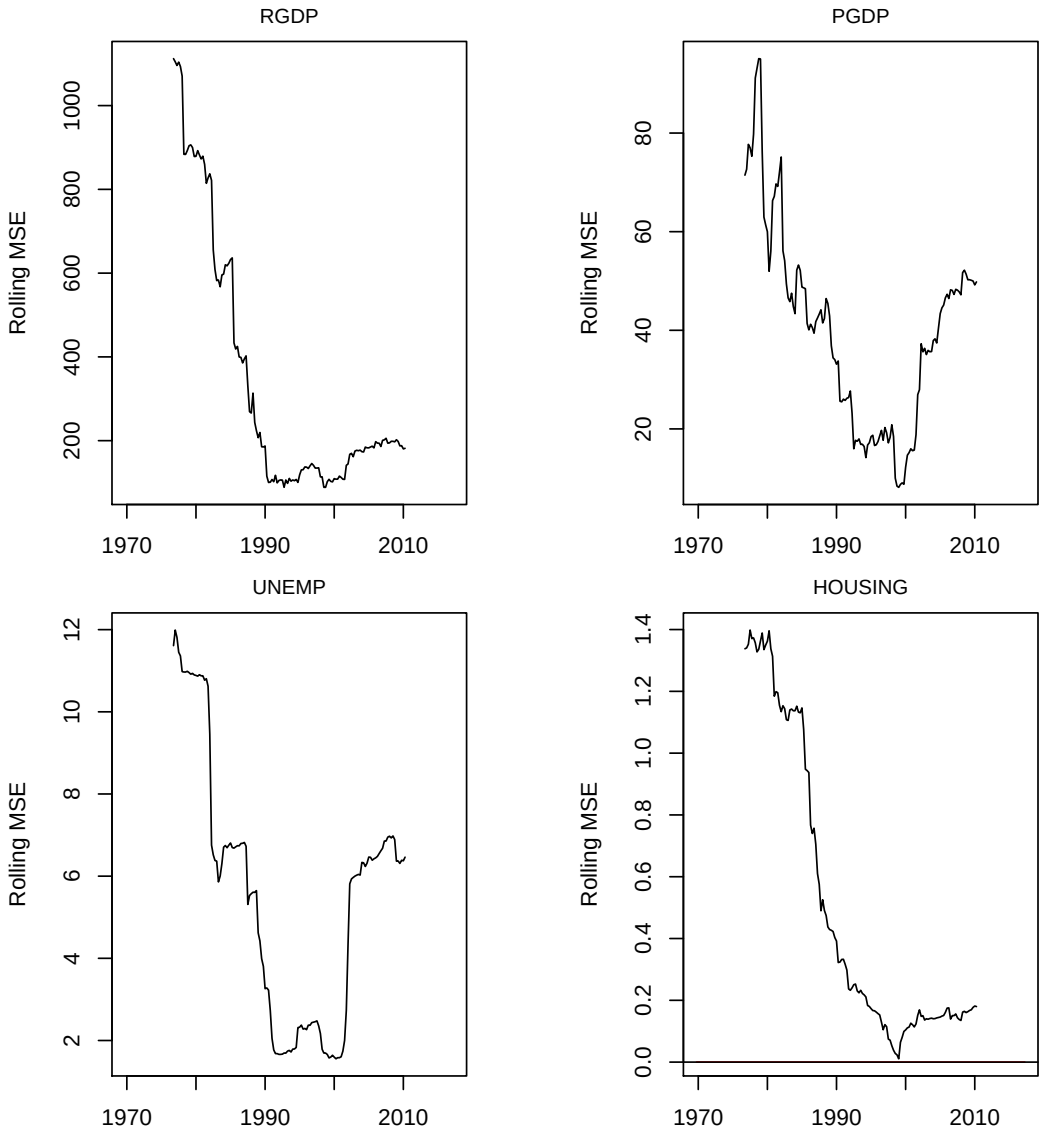


Figure 22: The plots show the rolling MSE difference (unscaled, $\nu = 0.3$). Nowcasts are evaluated against the final release.

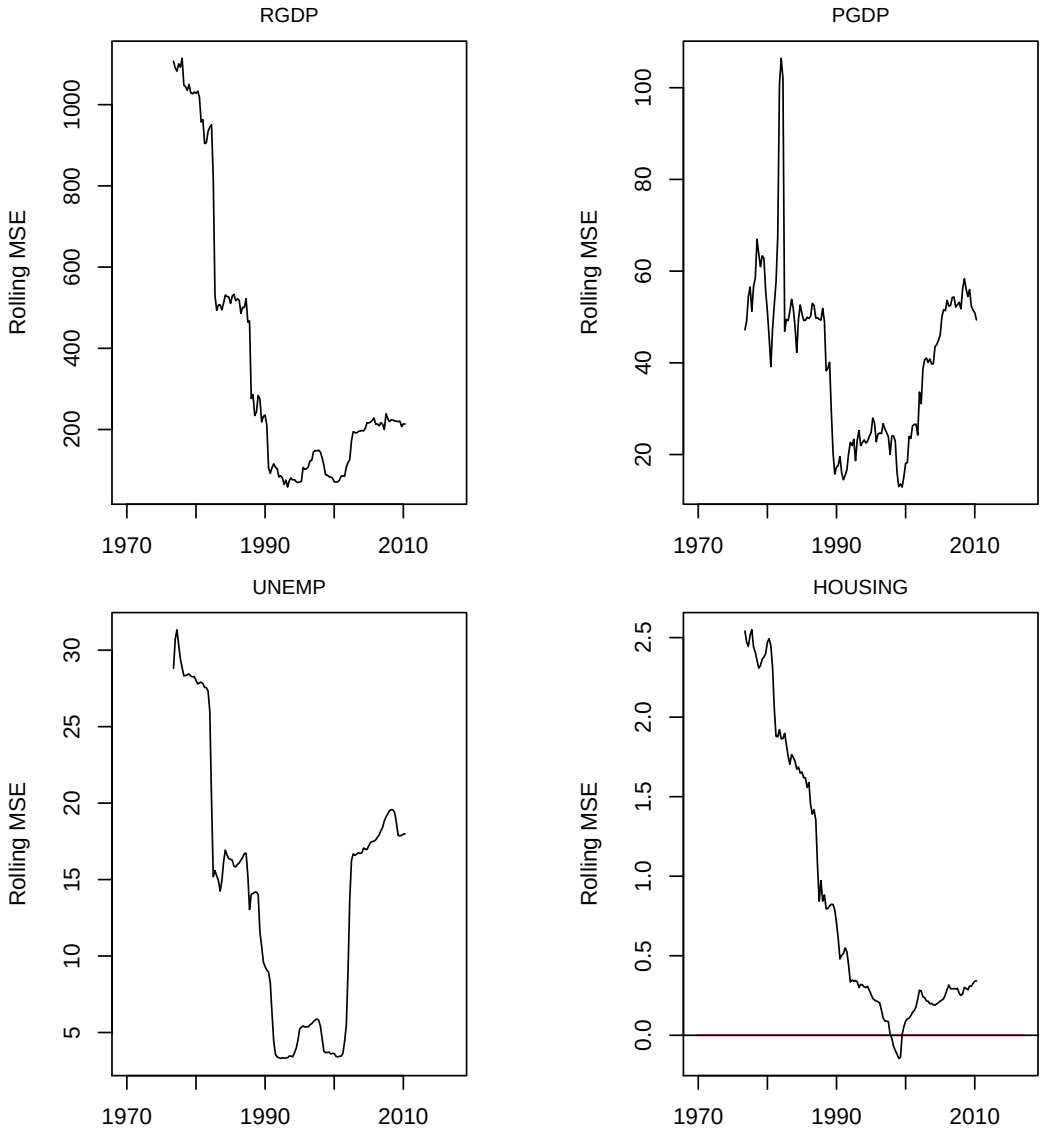


Figure 23: The plots show the rolling MSE difference (unscaled, $\nu = 0.3$). One-quarter ahead forecasts are evaluated against the final release.

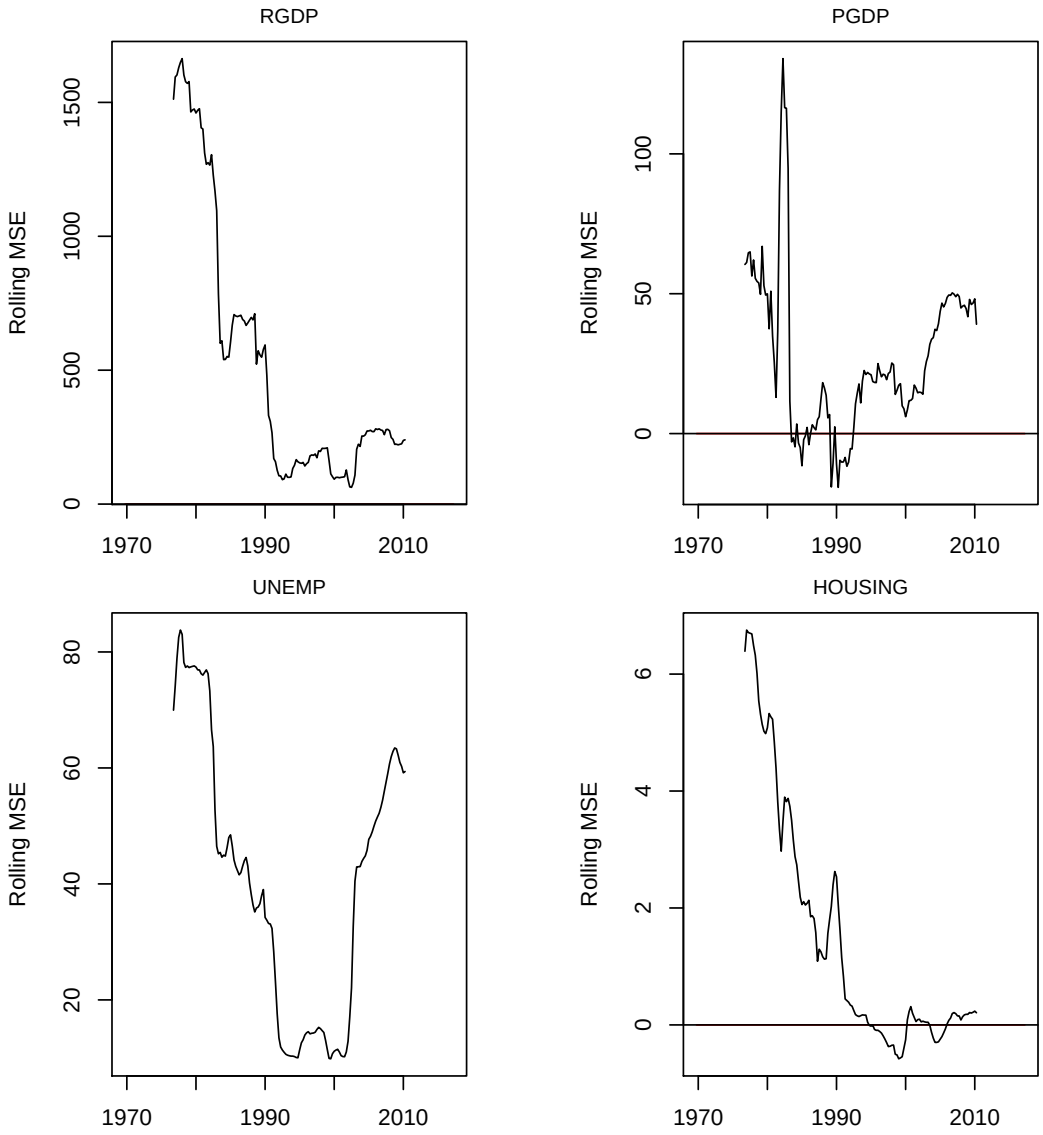


Figure 24: The plots show the rolling MSE difference (unscaled, $\nu = 0.3$). One-year ahead forecasts are evaluated against the final release.

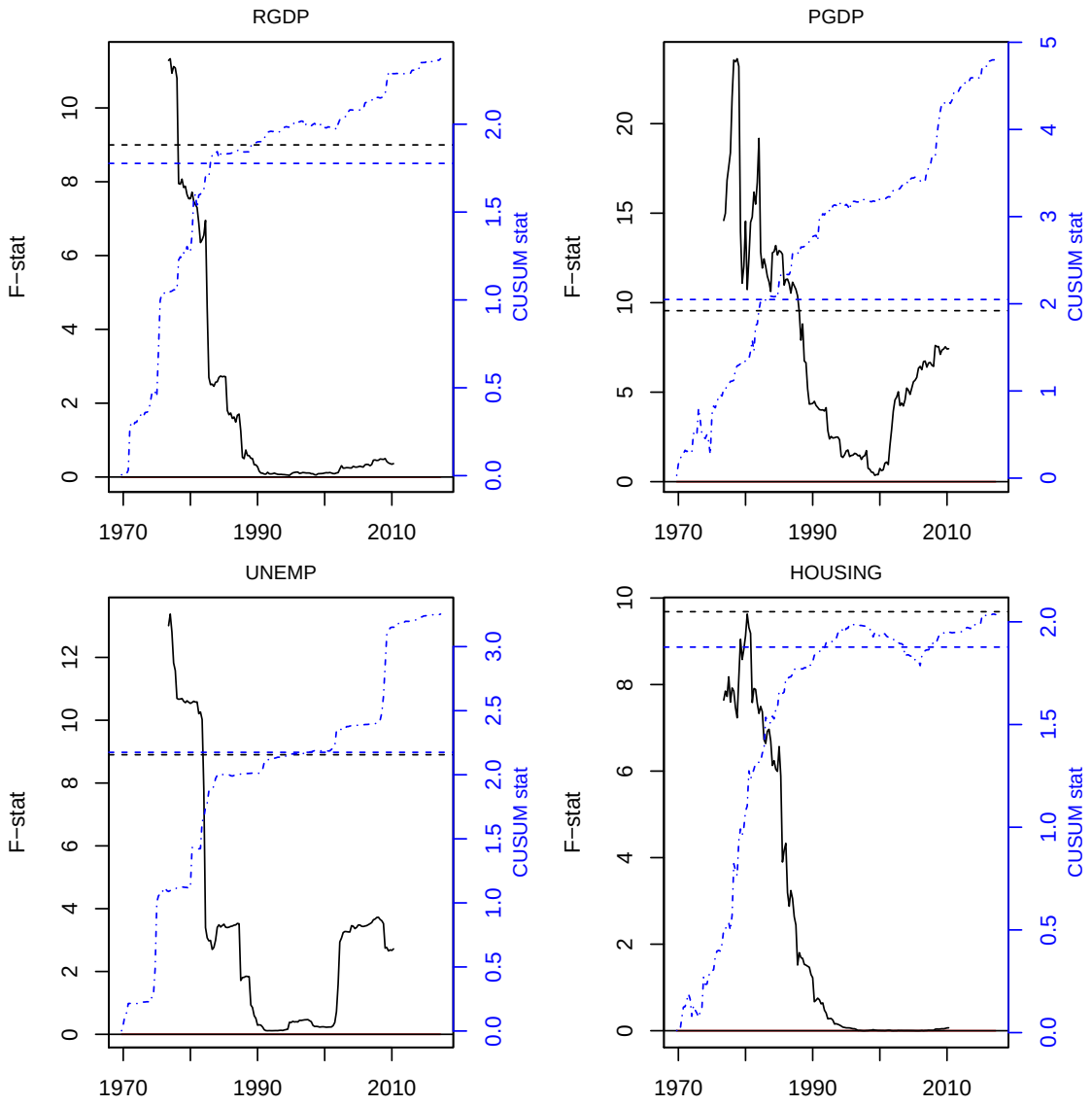


Figure 25: The plots show the time-varying components of the fluctuation statistic (left axis, solid black line) and the CUSUM statistic (right axis, dashed-dotted blue line), see equations (2) and (3). Horizontal dashed lines are the corresponding five percent critical values for the *maximum* of the displayed statistics. Nowcasts are evaluated against the first release for asymmetric loss; $b = 0.2$, $\nu = 0.3$.

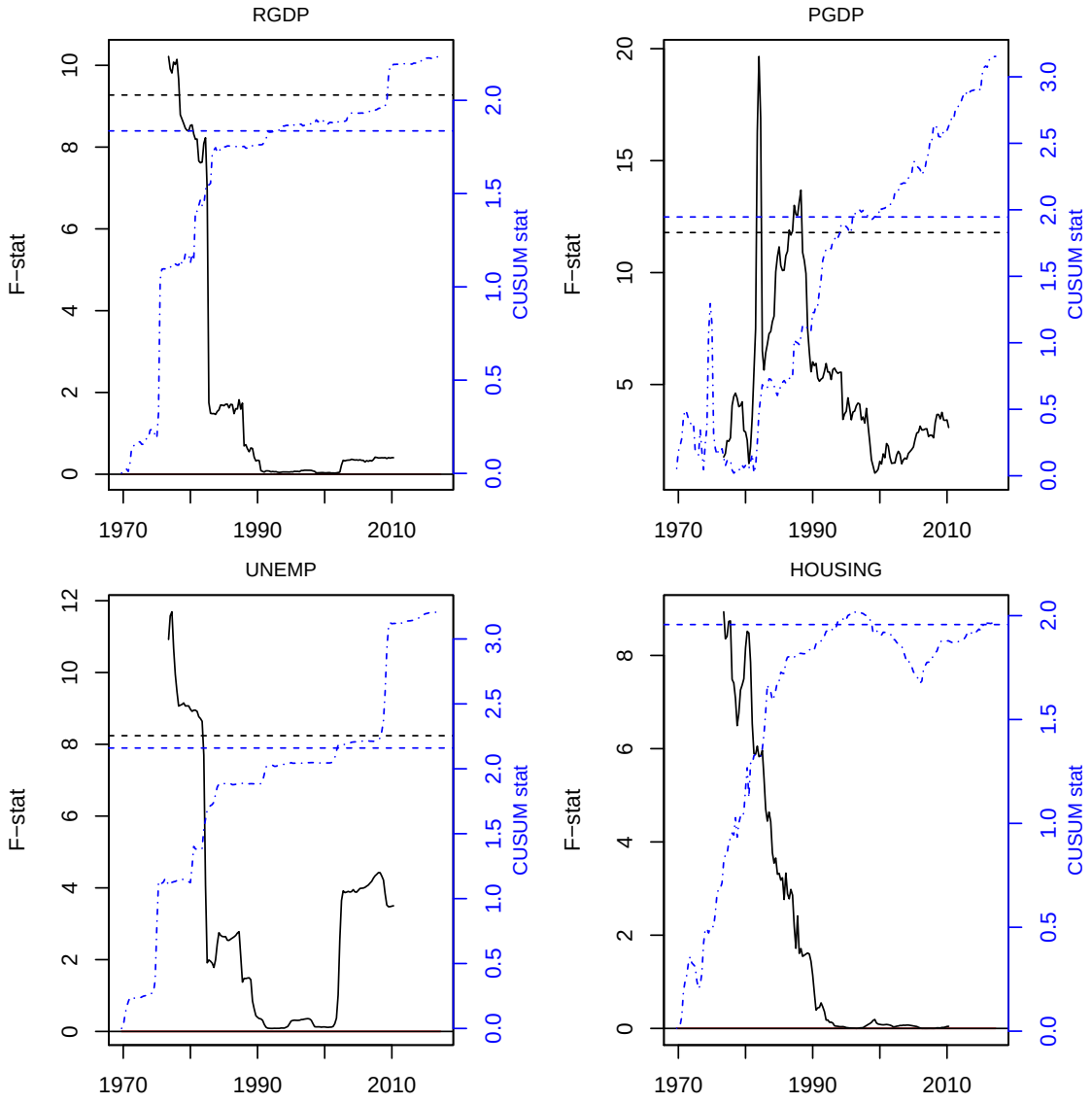


Figure 26: The plots show the time-varying components of the fluctuation statistic (left axis, solid black line) and the CUSUM statistic (right axis, dashed-dotted blue line), see equations (2) and (3). Horizontal dashed lines are the corresponding five percent critical values for the *maximum* of the displayed statistics. One-quarter ahead forecasts are evaluated against the first release for asymmetric loss; $b = 0.2$, $\nu = 0.3$.

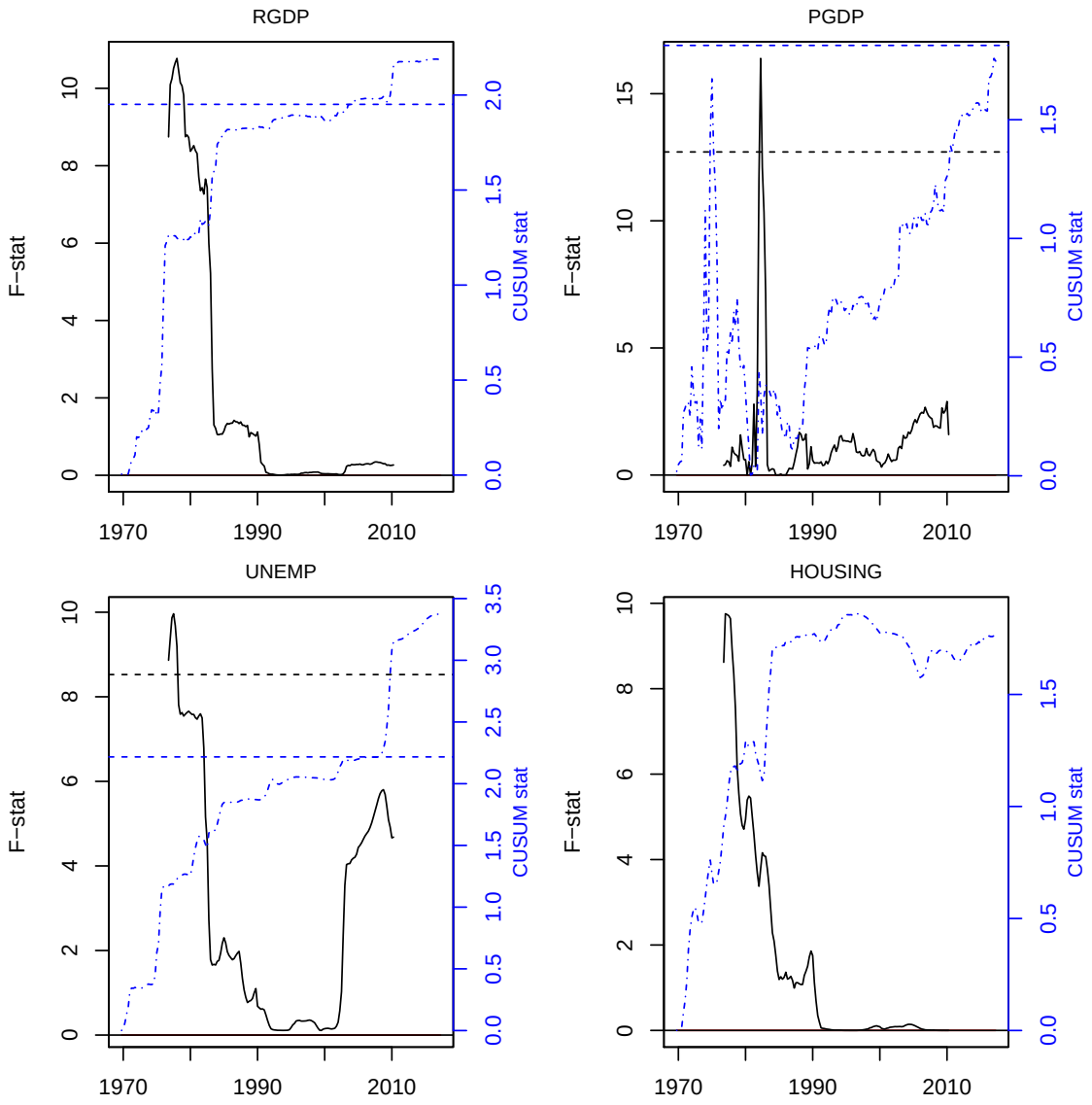


Figure 27: The plots show the time-varying components of the fluctuation statistic (left axis, solid black line) and the CUSUM statistic (right axis, dashed-dotted blue line), see equations (2) and (3). Horizontal dashed lines are the corresponding five percent critical values for the *maximum* of the displayed statistics. One-year ahead forecasts are evaluated against the first release for asymmetric loss; $b = 0.2$, $\nu = 0.3$.

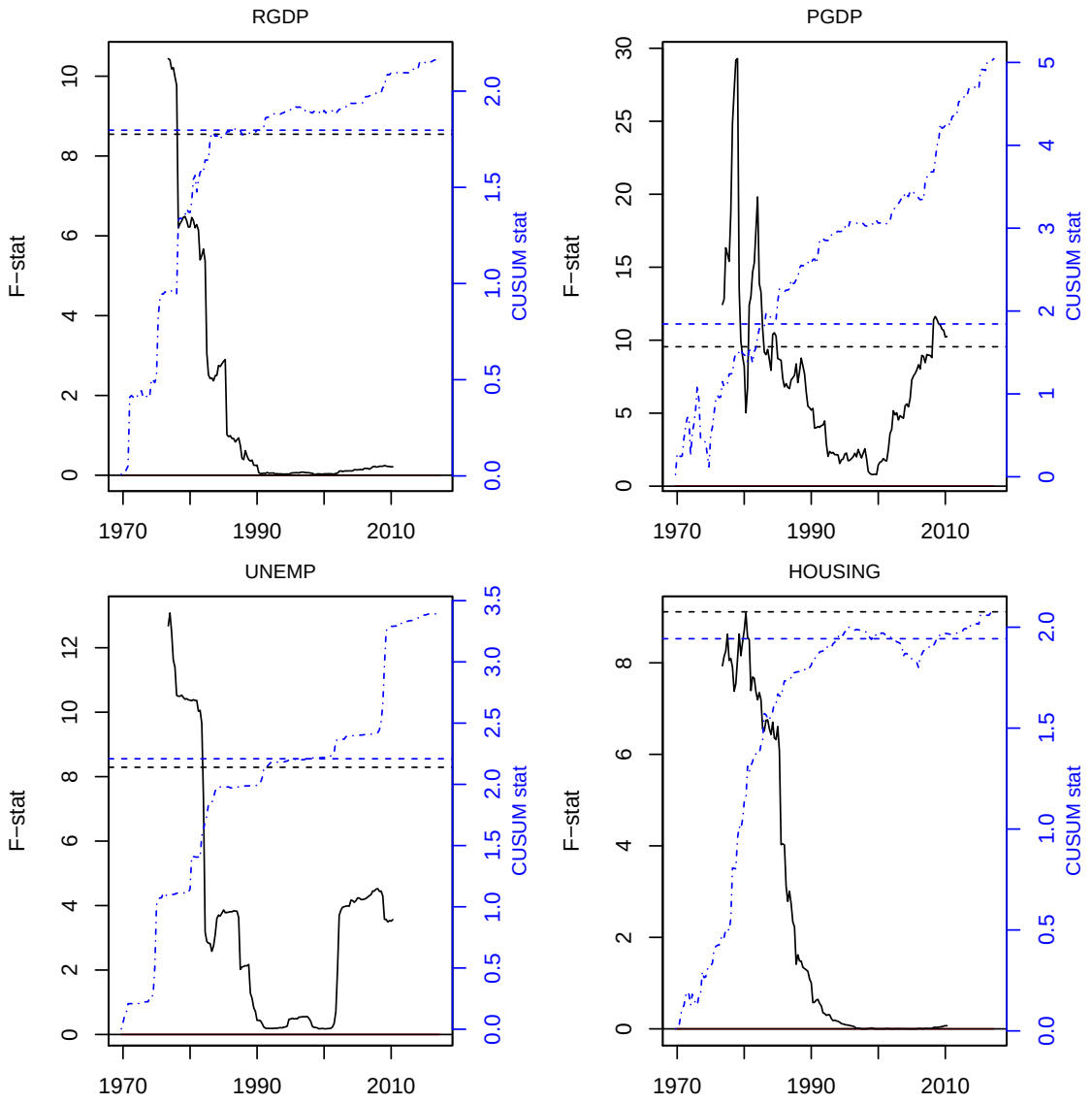


Figure 28: The plots show the time-varying components of the fluctuation statistic (left axis, solid black line) and the CUSUM statistic (right axis, dashed-dotted blue line), see equations (2) and (3). Horizontal dashed lines are the corresponding five percent critical values for the *maximum* of the displayed statistics. Nowcasts are evaluated against the final release for asymmetric loss; $b = 0.2$, $\nu = 0.3$.

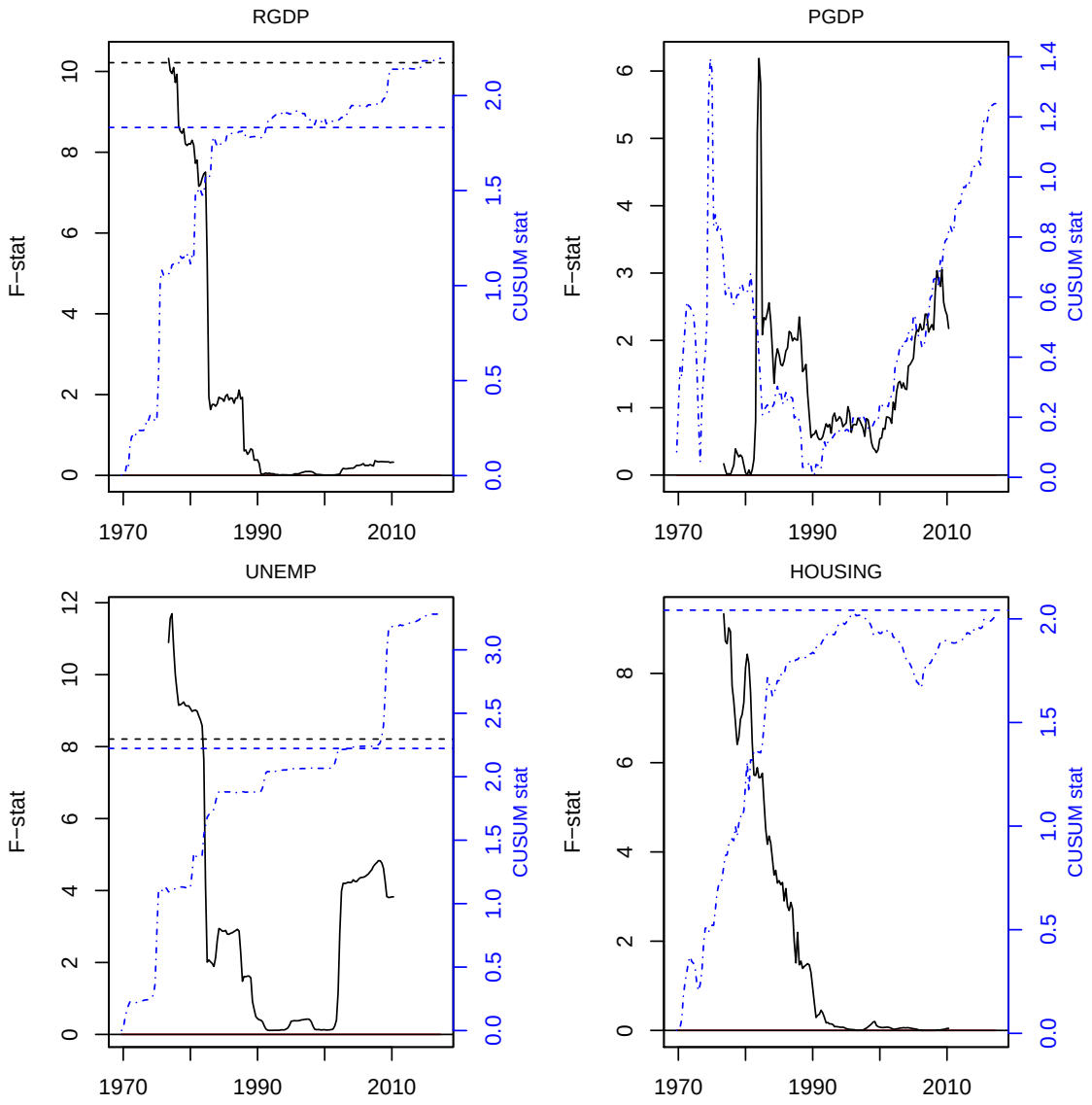


Figure 29: The plots show the time-varying components of the fluctuation statistic (left axis, solid black line) and the CUSUM statistic (right axis, dashed-dotted blue line), see equations (2) and (3). Horizontal dashed lines are the corresponding five percent critical values for the *maximum* of the displayed statistics. One-quarter ahead forecasts are evaluated against the final release for asymmetric loss; $b = 0.2$, $\nu = 0.3$.

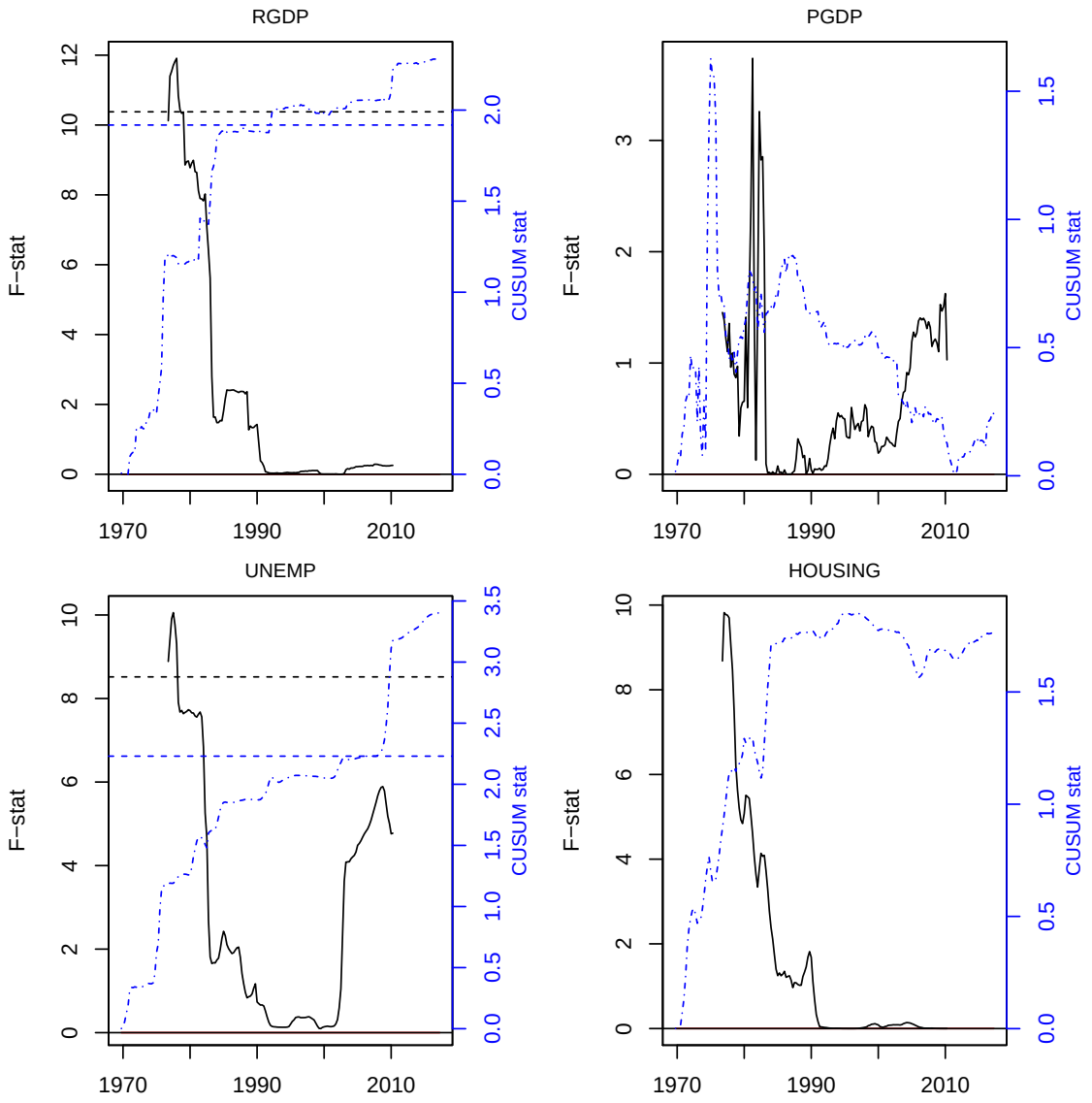


Figure 30: The plots show the time-varying components of the fluctuation statistic (left axis, solid black line) and the CUSUM statistic (right axis, dashed-dotted blue line), see equations (2) and (3). Horizontal dashed lines are the corresponding five percent critical values for the *maximum* of the displayed statistics. One-year ahead forecasts are evaluated against the final release for asymmetric loss; $b = 0.2$, $\nu = 0.3$.

E Asymptotic critical values

b	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
					10% critical values						
$\mathcal{T}_1^Q$											
Bartlett	1.97	2.14	2.37	2.63	2.92	3.19	3.46	3.70	3.92	4.14	4.36
QS	1.97	2.25	2.71	3.30	4.08	4.99	5.97	7.02	8.17	9.38	10.68
$\mathcal{T}_1^C$											
Bartlett	1.21	1.43	1.71	2.06	2.47	2.91	3.42	3.91	4.35	4.89	5.42
QS	1.22	1.57	2.14	3.08	4.47	6.38	9.09	12.29	16.47	21.75	28.04
$\mathcal{T}_1^F, \nu = 0.3$											
Bartlett	8.05	8.46	9.87	12.13	15.50	19.38	23.27	27.00	30.46	34.01	37.76
QS	8.05	9.30	13.00	20.79	35.50	56.94	85.95	121.04	164.70	218.95	282.40
$\mathcal{T}_1^F, \nu = 0.5$											
Bartlett	6.52	7.19	8.55	10.44	12.52	14.86	17.86	20.94	23.83	26.69	29.56
QS	6.53	7.95	11.34	16.57	25.37	39.12	58.27	83.59	115.38	153.14	197.37
$\mathcal{T}_1$											
Bartlett	2.71	3.39	4.20	5.19	6.33	7.59	8.91	10.11	11.40	12.75	14.16
QS	2.71	3.76	5.31	7.83	11.52	16.47	22.92	30.83	41.03	53.50	68.53
					5% critical values						
$\mathcal{T}_1^Q$											
Bartlett	2.25	2.49	2.81	3.17	3.50	3.87	4.19	4.49	4.76	5.03	5.30
QS	2.25	2.65	3.32	4.21	5.36	6.76	8.29	9.94	11.65	13.44	15.35
$\mathcal{T}_1^C$											
Bartlett	1.69	2.03	2.46	3.07	3.69	4.44	5.16	5.94	6.67	7.44	8.24
QS	1.69	2.26	3.31	5.00	7.86	11.95	17.58	25.19	34.80	46.25	59.28
$\mathcal{T}_1^F, \nu = 0.3$											
Bartlett	9.58	9.85	11.79	14.80	19.30	24.53	29.33	33.99	37.96	42.41	47.06
QS	9.59	11.17	16.96	29.94	54.87	94.61	150.47	222.97	317.35	428.42	559.59
$\mathcal{T}_1^F, \nu = 0.5$											
Bartlett	8.14	8.92	10.87	13.57	16.49	19.52	23.79	28.14	31.83	35.77	39.47
QS	8.14	10.08	15.48	24.40	40.58	68.10	105.25	151.32	212.99	288.50	380.02
$\mathcal{T}_1$											
Bartlett	3.83	4.97	6.45	8.04	9.79	11.90	13.92	15.91	17.96	20.12	22.26
QS	3.83	5.68	8.64	13.38	21.02	31.57	46.04	65.35	89.22	119.31	151.89

Table 19: Asymptotic critical values

Table 19 reports asymptotic critical values ignoring possible time-varying variance. "Small-b" χ^2_1 quantiles are recovered as special cases for the squared Diebold and Mariano (1995) statistic $\mathcal{T}_1$ for $b = 0$. Also note that, under small-b asymptotics, the critical values are independent of the kernel (up to simulation variability).