Conditional Predictive Density Evaluation in the Presence of Instabilities

Barbara Rossi
Tatevik Sekhposyan

February 2013

Conditional Predictive Density Evaluation in the Presence of Instabilities

Barbara Rossi[*] and Tatevik Sekhposyan[†]

February 11, 2013

Abstract

We propose new methods for evaluating predictive densities. The methods include Kolmogorov-Smirnov and Cramér-von Mises-type tests for the correct specification of predictive densities robust to dynamic mis-specification. The novelty is that the tests can detect mis-specification in the predictive densities even if it appears only over a fraction of the sample, due to the presence of instabilities. Our results indicate that our tests are well sized and have good power in detecting mis-specification in predictive densities, even when it is time-varying. An application to density forecasts of the Survey of Professional Forecasters demonstrates the usefulness of the proposed methodologies.

Keywords: Predictive Density, Dynamic Mis-specification, Instability, Structural Change, Forecast Evaluation.

J.E.L. Codes: C22, C52, C53

[*]ICREA-UPF, Barcelona GSE, and CREI. Carrer Ramon Trias Fargas, 25-27, Mercè Rodoreda bldg., 08005 Barcelona, Spain. Phone: (+34) 93 542 1655. Fax: (+34) 93 542 2826. E-mail: barbara.rossi@upf.edu

[†]Bank of Canada. 234 Wellington Street, Ottawa, ON K1A 0G9, Canada. Phone: (+1) 613 782 7239. Fax: (+1) 613 782 7658. E-mail: tsekhposyan@bankofcanada.ca

1   Introduction

Predictive densities provide a measure of uncertainty around mean forecasts, thus enabling researchers to quantify the risk in forecast-based decisions. For example, predictive densities are useful tools for central banks and policymakers, as they allow to take into account forecast uncertainty in economic decisions. It is therefore important to have a framework for evaluating whether predictive densities are correctly specified. Diebold et al. (1998, 1999) introduced the probability integral transform (PIT, Rosenblatt, 1952) to economics and finance as a tool to test whether a predictive distribution matches that of the true (and unobserved) distribution that generates the data. If the forecasting model is dynamically correctly specified, Diebold et al. (1998, 1999) show that the PITs based on the forecasts are Uniform, independent and identically distributed. They propose to test two of these implications, namely the uniformity and the serial correlation properties of the PITs, in order to detect mis-specification in density forecasts. Subsequent contributions extended the PIT framework to account for parameter estimation error and dynamic mis-specification. The former requires an adjustment to the PIT to account for the uncertainty associated with parameter estimation. Dynamic mis-specification implies that the information available to a researcher spans only a subset of the information the true model is in fact conditioned upon. Among recent contributions, Bai (2003) proposes tests for correct specification aimed at correcting for parameter estimation error based on martingalization techniques. Hong and Li (2005) suggest a non-parametric test robust to dynamic mis-specification and parameter estimation error using the generalized cross-spectrum. Corradi and Swanson (2006a), on the other hand, propose tests robust to both parameter estimation error as well as dynamic mis-specification that have the advantage of a parametric rate of convergence. See Corradi and Swanson (2006b) for an extensive overview of estimation and inference for predictive densities, and Corradi and Swanson (2006c, 2007) for empirical applications. However, none of these approaches has considered testing the "identical distribution" of the PITs; the latter becomes especially important when the mis-specification of the forecast density appears only in a sub-sample, say, due to instabilities.

The main objective of this paper is to provide new methodologies for testing the correct specification of density forecasts that are robust to both dynamic mis-specification as well as instabilities. Regarding the robustness to instabilities, one of the important assumptions for the validity of the tests proposed by Diebold et al. (1998, 1999), Bai (2003) and Corradi and Swanson (2006a) is stationarity (i.e. absence of structural breaks), which we relax in this

paper. In particular, we propose a specification test robust to instabilities by extending the PIT approach to test whether the predictive density is correctly specified *at each point in time*. Our proposed test detects distributional change in the predictive densities even if the densities are mis-specified by building on the framework proposed in Corradi and Swanson (2006a,b), although we derive our tests not only within the Kolmogorov-Smirnov class of tests that they consider, but also the Cramér-von Mises class. A special case of our test is a test for the constancy of predictive densities over time, which we also analyze. We investigate the small sample properties of our proposed tests in Monte Carlo simulation exercises. In addition, we also show that the proposed tests have good power to detect mis-specification in the predictive distribution even when the mis-specification affects only a sub-sample.

Our approach is primarily related to Diebold et al. (1998, 1999) and especially to Corradi and Swanson (2006a): we test the null hypothesis of correct specification of density forecasts, although in a way robust not only to the presence of dynamic mis-specification and parameter estimation error, but also time-varying mis-specification in the conditional density over time. Our approach is also related to Inoue (2001). Inoue (2001) develops techniques to test whether the in-sample empirical distribution of a model is constant over time. There are two important differences between Inoue's (2001) approach and ours: we focus on the out-of-sample evaluation of densities (as opposed to in-sample tests) and our null hypothesis is different: it involves testing whether the true predictive distribution matches that implied by a model at each point in time (rather than whether the predictive distribution has changed over time, as in Inoue, 2001). However, we also discuss a modified statistic for testing the constancy of the predictive density over time. Our approach is more distantly related to Rossi (2005): she jointly tests the hypothesis of stability of the parameters as well as that the parameters satisfy a certain restriction in-sample. The approach taken in this paper is similar in that we focus on testing a joint null hypothesis of stability in the predictive distribution as well as correct specification of the predictive distribution. However, it is very different for two reasons: first, because it focuses on prediction, which requires a different approach than in-sample tests; second, because it focuses on predictive density tests, which are very different from tests on parameters. Our approach is more distantly related to Amisano and Giacomini (2007), who instead focus on density forecast tests of relative predictive ability in a framework where parameter estimation error is maintained under the null; we instead focus on tests of absolute predictive ability and we derive density forecast tests where the asymptotic distribution is corrected for parameter estimation error.

We provide an empirical application of our proposed tests to the density forecasts pro-

vided in the Survey of Professional Forecasters (SPF). Our test uncovers that the predictive densities of both output growth and inflation are mis-specified and that there is evidence of time variation in the mis-specification. We detect instabilities in the correct specification of current year forecast (nowcasts) of inflation and output growth in 1979:II and 1985:IV, right at the onset of the Great Recession. The instability in the specification of the current year inflation (for which we have the longest span of data available, among our data) coincides with the beginning of Volker's chairmanship of the Federal Reserve, and suggests a significant change in the way forecasters formed inflation expectations at the time of a major change in monetary policy. For the one-year-ahead inflation and output growth, the break is in mid- to late-1990s, though statistically insignificant for the case of output growth. In general, our results suggest that the densities have been mis-specified both before and after the estimated break dates, although the nature of the mis-specification varies over time, an empirical fact that we investigate in detail.

The paper is organized as follows. Section 2 introduces the notation and definitions. Section 3 presents our test of correct specification of the density forecasts robust to dynamic mis-specification in the presence of instabilities. Section 4 provides Monte Carlo evidence on the performance of our test in small samples, and Section 5 presents the empirical results. Section 6 concludes.

2 Notation and Definitions

We first introduce the notation and discuss the assumptions about the data, the models and the estimation procedure.

We are interested in the true but unknown $h-$step-ahead conditional predictive densities for the scalar variable y_t, denoted by $\phi_0(.)$, where h is fixed and finite.[1] We assume that the researcher has divided the available sample of size $T + h$ into an in-sample portion of size R and an out-of-sample portion of size P, and obtained a sequence of $h-$step-ahead out-of-sample predictive densities, such that $R + P - 1 + h = T + h$, to evaluate at the ex-post realizations. Let $\mathcal{F}_t$ be the true information set available at time t, however the forecaster might observe only a subset of the information set, $\Im_t \subseteq \mathcal{F}_t$. Further, let $Z^t \in \Im_t$ denote the predictors used in the conditional forecast exercise.

Let the sequence of P out-of-sample estimated conditional predictive densities evaluated

---

[1]The true conditional forecast density may depend on the forecast horizon. To simplify notation, we omit this dependence without loss of generality given that the forecast horizon is fixed.

at the ex-post realizations y_{t+h} be denoted by $\left\{ \phi \left(y_{t+h}|\Im_t, \widehat{\theta}_{t,R} \right) \right\}_{t=R}^{T}$, which depend on the in-sample parameter estimates, $\widehat{\theta}_{t,R}$, assumed to be a $p \times 1$ vector. These parameters are re-estimated over time using a sample including data indexed $1, ..., t$ (recursive scheme), where $t = R, ..., T$. The latter mimics a forecasting environment where a researcher starts estimating the model using a window of size R, and then progressively adds an additional observation to the estimation sample in each subsequent time period as new data become available.[2]

We consider the probability integral transform (PIT), i.e. the cumulative density function corresponding to the density $\phi(.)$, evaluated at the realized value y_{t+h}:

$$z_{t+h} = \int_{-\infty}^{y_{t+h}} \phi \left(u|\Im_t, \widehat{\theta}_{t,R} \right) du \equiv \Phi \left(y_{t+h}|\Im_t, \widehat{\theta}_{t,R} \right).$$

Let

$$\widehat{\xi}_{t+h}(r) \equiv \left(1 \left\{ \Phi \left(y_{t+h}|\Im_t, \widehat{\theta}_{t,R} \right) \leq r \right\} - r \right), \tag{1}$$

where $r \in [0,1]$ denotes quantiles of the cumulative density function, $1\{.\}$ denotes the indicator function, and $\widehat{\xi}_{t+h}(r)$ measures the distance between the empirical cumulative distribution function and that of the uniform distribution (which is the 45-degree line, and hence, r itself). We consider the out-of-sample partial sum of $\hat{\xi}_{t+h}(r)$, defined as:

$$\Psi_P(\tau, r) \equiv P^{-1/2} \sum_{t=R}^{R+[\tau P]} \widehat{\xi}_{t+h}(r), \tag{2}$$

and the full out-of-sample average:

$$\Psi_P(1, r) \equiv P^{-1/2} \sum_{t=R}^{T} \widehat{\xi}_{t+h}(r), \tag{3}$$

where $\tau \in \Upsilon \subset (0,1)$.

In order to derive the asymptotic distribution of the proposed tests we first describe the asymptotic behavior of the empirical processes $\Psi_P(1, r)$ and $\Psi_P(\tau, r)$ above; then, we describe our proposed tests and derive their asymptotic distributions. Finally, we describe how to implement our proposed tests in practice using a detailed step-by-step procedure. In what follows, $||.||$ denotes the Euclidean norm, $[\tau P]$ denotes the integer of τP, $\underset{p}{\rightarrow}$ denotes convergence in probability, and $\Rightarrow$ denotes weak convergence in the space of $D\left([0,1] \times \Re\right)$ under the Skorohod topology.

---

[2]Results can be generalized to the rolling window estimation scheme where the size of the window is a fixed fraction of the total sample size.

3   Predictive Density Specification Tests in the Presence of Instabilities

This section discusses our proposed tests for the correct specification of predictive densities that allow for dynamic mis-specification under the null hypothesis and that can detect mis-specification in the predictive density even if it arises only in a sub-sample.

3.1   Assumptions and Asymptotic Results

Let the parameter vector θ be estimated recursively, such that:

$$\widehat{\theta}_{t,R} = \arg\max_{\theta} \frac{1}{t} \sum_{j=1}^{t} \ln \phi\left(y_j | \Im_{j-h}, \theta\right), t = R, R+1, ..., T.$$

Note that the parameter is estimated directly on h lags of the predictors (in order to obtain a direct h-step-ahead forecast). Let $q_j(\theta) \equiv \nabla_\theta \ln \phi\left(y_j | \Im_{j-h}, \theta\right)$ be a $p \times 1$ vector of scores. Our interest lies in testing whether $\phi_t\left(y_{t+h} | \Im_t\right) = \phi_0\left(y_{t+h} | \Im_t, \theta^\dagger\right)$ at any point in time t over the out-of-sample portion of the data, that is:

$$H_0 : \Phi_t\left(y_{t+h} | \Im_t\right) = \Phi_0\left(y_{t+h} | \Im_t, \theta^\dagger\right) \text{ for all } t = R, ..., T, \tag{4}$$

where $\Phi_0\left(y_{t+h} | \Im_t, \theta^\dagger\right) \equiv \Pr\left(y_{t+h} \leq y | \Im_t, \theta^\dagger\right)$ and $\theta^\dagger$ is the probability limit of $\widehat{\theta}_{t,R}$. Note that, under H_0, both the density and the parameters are constant.

We first consider the convergence properties of the empirical process $\Psi_P(\tau, r)$ defined in (2). We derive our results under the following assumption:

Assumption 1.

Assumption 1a: *(i) for $\theta \in \Theta$, $\Phi\left(y_{t+h} | \Im_t, \theta\right)$ is twice continuously differentiable on the interior of $\Theta \subset \Re^p$, Θ compact; (ii) each of the elements of $E[\sup_{\theta \in \Theta} |\nabla_\theta \Phi\left(y_{t+h} | \Im_t\right)|^{5+\psi}]$, a $(p \times 1)$ vector, are bounded by a finite constant, $0 < \psi < 1/2$;*

Assumption 1b: *(i) $\theta^\dagger = \arg\max_{\theta \in \Theta} E\left[\ln \phi\left(y_t | \Im_{t-h}, \theta\right)\right]$ is uniquely identified; (ii) $\ln \phi\left(y_t | \Im_{t-h}, \theta\right)$ is twice continuously differentiable in θ in the interior of Θ and, for some $\varepsilon > 0$, $\sup_{t \leq T, T \geq 1} E \sup_\theta ||\nabla_\theta q_{t,T}||^{1+\varepsilon} < \infty$; (iii) $Var\left[\frac{1}{\sqrt{T}} \sum_{t=1}^{[\widetilde{\tau}T]} q_t\left(\theta^\dagger\right)\right] \to \widetilde{\tau} S_{qq} \; \forall \widetilde{\tau} \in \Upsilon \subset (0,1)$ for some positive definite matrix S_{qq}; (iv) $\lim_{T \to \infty} T^{-1} \sum_{t=1}^{[\widetilde{\tau}T]} E\left(-\nabla_\theta q_t\left(\theta^\dagger\right)\right)$ exists uniformly over $\widetilde{\tau} \in \Upsilon \subset (0,1)$ and equals $\widetilde{\tau} A\left(\theta^\dagger\right)^{-1}$, which has full rank; (v) $\sup_\tau ||\widehat{\theta}_{R+[\tau P],R} - \theta^\dagger|| \underset{p}{\to} 0$ for some $\theta^\dagger$ in the interior of Θ.*

Assumption 1c: $P, R \rightarrow \infty$ as $T \rightarrow \infty$, and $\lim_{T\to\infty} \frac{P}{R} = \pi$, $0 < \pi < \infty$.[3]

Assumption 1d: $\left(y_t, Z^{t-h}\right)$ is a strong mixing (α-mixing) triangular array with mixing coefficients $\sum_{j=1}^{\infty} j^2 \alpha\left(j\right)^{\frac{\gamma}{\gamma+4}} < \infty$ for some $\gamma \in (0,2)$.

Assumption 1e: Let $q_t \equiv q_t\left(\theta^{\dagger}\right)$, $\nabla_\theta q_s \equiv \nabla_\theta q_s\left(\theta^{\dagger}\right)$, (a $p \times p$ matrix), and

$$\xi_{t+h}\left(r\right) \equiv 1\left\{\Phi\left(y_{t+h}|\Im_t, \theta^{\dagger}\right) \leq r\right\} - r. \tag{5}$$

(i) For some $\varsigma > 1$, $\sup_t E \left\| \left[vec\left(\nabla_\theta \Phi\left(y_{t+h}|\Im_t, \theta^{\dagger}\right)\right)', \, \xi_{t+h}\left(r\right)', \, q_t' \right] \right\|^{4\varsigma} < \infty$;

(ii) $\left[vec\left(\nabla_\theta \Phi\left(y_{t+h}|\Im_t, \theta^{\dagger}\right) - E\left(\nabla_\theta \Phi\left(y_{t+h}|\Im_t, \theta^{\dagger}\right)\right)\right)', \, \left(\xi_{t+h}\left(r\right) - E\left(\xi_{t+h}\left(r\right)\right)\right)', \, q_t' \right]$ is strong mixing with mixing coefficients of size $-3d/\left(d-1\right)$;

(iii) under H_0, $\left[vec\left(\nabla_\theta \Phi\left(y_{t+h}|\Im_t, \theta^{\dagger}\right)\right)', \, \xi'_{t+h}\left(r\right), q_t' \right]$ is covariance stationary;

(iv) $E\left[\sum_{s=-\infty}^{\infty} \left(1\left\{\Phi\left(y_h|\Im_0, \theta^{\dagger}\right) \leq r_1\right\} - r_1\right) \times \left(1\left\{\Phi\left(y_s|\Im_{s-h}, \theta^{\dagger}\right) \leq r_2\right\} - r_2\right)\right]$ is positive definite.

Assumption 1f: *For every $M > 0$, in some open neighborhood around $\theta^{\dagger}$, $N\left(\theta^{\dagger}, M\right)$,*

$\sup_{\tau \in \Upsilon \subset (0,1)} \sup_{u,v \in N\left(\theta^{\dagger}, M\right)} \left\| \frac{1}{P} \sum_{t=R}^{R+[\tau P]} \nabla_\theta \Phi\left(\Phi^{-1}(r|\Im_t, u)|\Im_t, v\right) - E\left[\nabla_\theta \Phi\left(\Phi^{-1}(r|\Im_t, \theta^{\dagger})|\Im_t, \theta^{\dagger}\right)\right] \right\|$
$= o_p\left(1\right)$, *where the $o_p\left(1\right)$ term is uniform in r.*

Notes to the Assumptions. Assumption 1a is similar to Corradi and Swanson's (2006a) Assumption A2, and imposes mild smoothness as well as moment restrictions on the cumulative distribution function under the null hypothesis.

Assumption 1b is sufficient to obtain the weak convergence of the parameter estimate based on the partial sum of moment conditions, and follows Andrews (1993). Assumption 1b(i) is similar to Corradi and Swanson's (2006b, p.272) Assumption CS3(i) and guarantees identification of the parameter estimate; Assumption 1b(ii) is similar to Corradi and Swanson's (2006b, p. 272) Assumption CS3(ii) and imposes that the objective function is sufficiently continuous; it also imposes that the gradient is sufficiently smooth, as in Andrews (1993, Assumption 1f); 1b(iii,iv) are asymptotic covariance stationarity conditions, as in Andrews (1993, Assumptions 1c and 1g); 1b(v) is necessary to prove weak convergence of the estimator, and is adapted from Andrews (1993, Assumption 1d).

Assumptions 1c and 1e are adapted from West (1996, Assumptions 4,3,1, respectively) and are necessary to describe the asymptotic behavior of parameter estimation error.

---

[3]Note that the contribution of parameter estimation error is negligible asymptotically when $\pi = 0$. See West (1996).

Assumption 1d is similar to Inoue's (2001) Assumption A, and restricts the dependence of the data. Note that Assumption 1d implies that $\left(y_t, Z^{t-h}\right)$ is near epoch dependent on an α-mixing array of size $\varsigma/\left(\varsigma - 2\right)$, $\varsigma > 2$ – see Inoue (2001, p. 177). The latter is the same as Andrews's (1993) Assumption 1(a).

Assumption 1f is adapted from Bai (2003) and ensures that certain terms in the mean value expansion are negligible.

The following theorem derives the asymptotic distribution of $\Psi_P\left(\tau, r\right)$ under the null hypothesis defined in (4).

Theorem 1 (Asymptotic Distribution of $\Psi_P\left(\tau, r\right)$) *Under Assumption 1 and H_0 in (4): (i) $\left\{z_{t+h}\right\}_{t=R}^{T}$ is $U\left(0, 1\right)$; (ii) $\Psi_P\left(\tau, r\right)$ weakly converges (considered as variables in the space $\left(\left[0, 1\right]^2 \times \mathbb{R}\right)$ to the Gaussian process $\Psi\left(., .\right)$, with mean zero and auto-covariance function:*[4]

$$E\left[\Psi\left(\tau_1, r_1\right)\Psi\left(\tau_2, r_2\right)\right] = \inf\left(\tau_1, \tau_2\right)\Omega\left(r_1, r_2\right), \;\; where \tag{6}$$

$$
\begin{aligned}
\Omega\left(r_1, r_2\right) \equiv\ & S_{\Phi\Phi}\left(r_1, r_2\right) \\
& + 2\Pi E\left[\nabla_\theta \Phi\left(\kappa(r_1)|\Im_t, \theta^\dagger\right)\right]' A\left(\theta^\dagger\right) S_{qq} A\left(\theta^\dagger\right) E\left[\nabla_\theta \Phi\left(\kappa(r_2)|\Im_t, \theta^\dagger\right)\right] \\
& - \Pi E\left[\nabla_\theta \Phi\left(\kappa(r_1)|\Im_t, \theta^\dagger\right)\right]' A\left(\theta^\dagger\right)\left(S_{q\Phi}(r_1)\right) - \Pi S'_{q\Phi}(r_2)A\left(\theta^\dagger\right) E\left[\nabla_\theta \Phi\left(\kappa(r_2)|\Im_t, \theta^\dagger\right)\right],
\end{aligned}
\tag{7}
$$

$\Pi = 1 - \pi^{-1}\ln\left(1 + \pi\right)$, $A\left(\theta^\dagger\right) \equiv E\left(-\nabla_\theta q_s\right)^{-1}$, $\kappa(r) = \Phi^{-1}(r|\Im_t, \theta^\dagger)$, $S_{qq} \equiv \sum_{d=-\infty}^{\infty} E\left(q_1 q'_d\right)$,

$S_{\Phi\Phi}\left(r_1, r_2\right) \equiv E\left[\sum_{d=-\infty}^{\infty}\left(1\left\{\Phi\left(y_h|\Im_0, \theta^\dagger\right) \le r_1\right\} - r_1\right)\left(1\left\{\Phi\left(y_d|\Im_{d-h}, \theta^\dagger\right) \le r_2\right\} - r_2\right)\right]$, and

$S_{q\Phi}(r) = \sum_{d=-\infty}^{\infty} E\left[q_1\left(1\left\{\Phi\left(y_{d+h}|\Im_d, \theta^\dagger\right) \le r\right\} - r\right)\right]$.

Note that the asymptotic distribution in the Theorem 1 is corrected for the presence of parameter estimation uncertainty. The correction follows from the fact that the forecast densities are evaluated at the probability limit of the parameter values under the null hypothesis (4), as in Corradi and Swanson (2006b). Note that when parameter estimation error is negligible, i.e. when $\Pi = 0$, the variance-covariance matrix of the limiting distribution in (7) simplifies to the first term on the right hand side, as in West (1996). The simplification may occur when the researcher has a much larger number of observations to estimate the model's parameters (R) than the number of observations used to evaluate the forecast densities out-of-sample (P). In the latter case, the parameters can be estimated with sufficient precision

---

[4] $\Psi\left(., .\right)$ is a Kiefer process.

and the contribution of parameter estimation error becomes asymptotically negligible. Also note that, by construction, the results are robust to serial correlation and, hence, would also apply for the case of multi-step-ahead forecasts, where serial correlation of order $h - 1$ is built in the forecasts.

3.2   Test Statistics And Their Distributions

We next discuss the tests that we propose. Let

$$Q_P\left(\tau, r\right) \equiv \left[R_P(\tau) F_P\left(\tau, r\right)\right]' \left[R_P(\tau) F_P\left(\tau, r\right)\right], \tag{8}$$

where

$$F_P\left(\tau, r\right) \equiv \begin{bmatrix} P^{-1/2} \sum_{t=R}^{R+[\tau P]} \left(1\left\{\Phi\left(y_{t+h}|\Im_t, \widehat{\theta}_{t,R}\right) \leq r\right\} - r\right) \\ P^{-1/2} \sum_{t=R}^{T} \left(1\left\{\Phi\left(y_{t+h}|\Im_t, \widehat{\theta}_{t,R}\right) \leq r\right\} - r\right) \end{bmatrix} \tag{9}$$

$$R_P(\tau) = \begin{bmatrix} 1 & -\tau \\ 0 & 1 \end{bmatrix} \tag{10}$$

We consider two types of test statistics: the first is a weighted Kolmogorov-Smirnov-type statistic and the second is a weighted Cramér-von Mises-type statistic:

$$\kappa_P \equiv \sup_{\tau \in \Upsilon} \sup_{r \in [0,1]} Q_P\left(\tau, r\right) \tag{11}$$

$$C_P \equiv \int_\tau \int_r Q_P\left(\tau, r\right) d\tau dr. \tag{12}$$

The following Theorem describes the asymptotic distribution of the tests we propose.

Theorem 2 (Predictive Density Tests Robust to Instabilities) *Under Assumption 1 and H_0 in (4):*

$$\kappa_P = \sup_{\tau \in \Upsilon} \sup_{r \in [0,1]} Q_P\left(\tau, r\right) \tag{13}$$

$$\Rightarrow \sup_{\tau \in \Upsilon} \sup_{r \in [0,1]} \left\{\Psi^\circ\left(\tau, r\right)' \Psi^\circ\left(\tau, r\right) + \Psi\left(1, r\right)' \Psi\left(1, r\right)\right\},$$

and

$$C_P \equiv \int_\Upsilon \int_0^1 Q_P\left(\tau, r\right) d\tau dr \Rightarrow \int_\Upsilon \int_0^1 \left\{\Psi^\circ\left(\tau, r\right)' \Psi^\circ\left(\tau, r\right) + \Psi\left(1, r\right)' \Psi\left(1, r\right)\right\} d\tau dr, \tag{14}$$

where $\Psi^\circ\left(\tau, r\right) \equiv \Psi\left(\tau, r\right) - \tau \Psi\left(1, r\right)$ is a Gaussian process with zero mean and covariance function $E\left\{\Psi^\circ\left(\tau_1, r_1\right) \Psi^\circ\left(\tau_2, r_2\right)\right\} = \left[\inf\left(\tau_1, \tau_2\right) - \tau_1 \tau_2\right] \Omega\left(r_1, r_2\right)$. Reject H_0 at the $\alpha \cdot 100\%$ significance level if $\kappa_P > \kappa_{\alpha;P}$ and $C_P > C_{\alpha;P}$.

For a given estimate of $\Omega(r_1, r_2)$, the critical values of κ_P and C_P can be obtained via Monte Carlo simulation. We suggest to estimate $\Omega(r_1, r_2)$ using a HAC covariance matrix, such as the Newey and West (1987) HAC estimator with a Bartlett kernel.[5] We discuss this further in the next sub-section, where we describe a step-by-step procedure that illustrates how to implement our tests in practice. We should emphasize that our tests are, by construction, robust to dynamic mis-specification. We demonstrate this property in our Monte Carlo simulations as well. As previously discussed, our tests are also robust to the presence of serial correlation in the PITs and can be applied for the case of multi-step-ahead forecasts, $h > 1$, as well.

In addition, it might also be of interest to test for correct specification in specific parts of the distribution.[6] For example, one might be interested in the tails of the distribution, which correspond to outliers; for example the left tail, where $r \in [0, 0.25]$, or the right tail, where $r \in [0.75, 1]$, or both: $r \in \{[0, 0.25] \cup [0.75, 1]\}$. Or, one might be interested in the central part of the distribution, for example $r \in [0.25, 0.75]$. Of course, this is possible in our framework: it merely requires defining the appropriate grid for r when implementing (11) and (12).

Our proposed tests differ from those existing in the literature in several ways. In particular, the test proposed by Corradi and Swanson (2006a) would be a special case of our approach for $R_P(\tau) = R_P^{CS}(\tau)$, where $R_P^{CS}(\tau) \equiv \begin{bmatrix} 0 & 1 \end{bmatrix}$ instead of (10) and for the case of a Kolmogorov-Smirnov-type test implemented on the absolute value (rather than the square) of $R_P(\tau)F_P(\tau, r)$.[7] Let κ_P^{CS} and C_P^{CS} be defined as in (11) and (12), where $Q_P(\tau, r)$ is constructed using $R_P^{CS}(\tau)$. Thus, the test proposed by Corradi and Swanson (2006a,b) is equivalent to:

$$\kappa_P^{CS} = \sup_{r \in [0,1]} \left| \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left(1\left\{ \Phi\left(y_{t+h} | \Im_t, \widehat{\theta}_{t,R}\right) \leq r \right\} - r \right) \right|. \tag{15}$$

The the main difference between our tests and theirs is that we are concerned with testing

---

[5] HAC estimation is also suggested in Inoue (2001) as an alternative to the bootstrap he proposes. Inoue (2001) discusses the computational challenges involved in using this type of procedure. With the increase in computing power in the last decade, we find this constraint less binding. While it might be possible to design a bootstrap methodology to use in this context along the lines of Inoue (2001) or Corradi and Swanson (2006a), we leave this for future research. Preliminary Monte Carlo simulation results suggest that a moving block bootstrap does not perform well in terms of size in our framework.

[6] See van Dijk and Franses (2003), Amisano and Giacomini (2007) and Diks et al. (2011) for a similar idea in the context of point forecast and density forecast comparisons.

[7] Note that, by construction, the test built on $|R_P(\tau)F_P(\tau, r)|$ and that built on $[R_P(\tau)F_P(\tau, r)]'$ $[R_P(\tau)F_P(\tau, r)]$ are equivalent, and differ only in their critical values.

for the correct specification of the predictive density in the presence of instabilities whereas they assume stationarity. In other words, we test the null hypothesis that the PIT is uniform at each point in time; they test the null hypothesis that the PIT is uniform on average over the out-of-sample portion of the data.

Note that our tests are also different from those proposed by Inoue (2001), who tests the null hypothesis of constancy of in-sample densities over time, and does not address the issue of correct specification. While Inoue's (2001) approach is designed for in-sample densities, our approach is for out-of-sample forecast densities. At the same time, Inoue's (2001) null hypothesis in our out-of-sample context becomes a null hypothesis on the constancy of the PITs over time, and one could consider tests for instabilities in the density forecasts inspired by Inoue (2001). Thus, the out-of-sample version of Inoue's (2001) test is another special case of our approach with $R_P(\tau) = R_P^I(\tau)$, where $R_P^I(\tau) \equiv \begin{bmatrix} 1 & -\tau \end{bmatrix}$ instead of (10). Let κ_P^I and C_P^I be tests for the stability of the PITs defined as in (11) and (12), where $Q_P(\tau, r)$ is constructed using $R_P^I(\tau)$. The Kolmogorov-Smirnov test in Inoue (2001) can be rewritten as:[8]

$$\kappa_P^I = \sup_{\tau \in \Upsilon} \sup_{r \in [0,1]} \left| \frac{1}{\sqrt{P}} \sum_{t=R}^{R+[\tau P]} 1 \left\{ \Phi\left(y_{t+h} | \Im_t, \widehat{\theta}_{t,R} \right) \leq r \right\} - \frac{\tau}{\sqrt{P}} \sum_{t=R}^{T} 1 \left\{ \Phi\left(y_{t+h} | \Im_t, \widehat{\theta}_{t,R} \right) \leq r \right\} \right|^2,$$

and the Cramér-von Mises test statistic is instead:

$$C_P^I = \int_\Upsilon \int_0^1 \left\{ \frac{1}{\sqrt{P}} \sum_{t=R}^{R+[\tau P]} 1 \left\{ \Phi\left(y_{t+h} | \Im_t, \widehat{\theta}_{t,R} \right) \leq r \right\} - \frac{\tau}{\sqrt{P}} \sum_{t=R}^{T} 1 \left\{ \Phi\left(y_{t+h} | \Im_t, \widehat{\theta}_{t,R} \right) \leq r \right\} \right\}^2 d\tau dr.$$

Note that our proposed κ_P and C_P statistics focus on testing the *joint* null hypothesis of constancy of forecast densities over time as well as their correct specification, whereas κ_P^I and C_P^I would only test constancy.

3.3 How to Implement the Proposed Tests

In this sub-section, we provide a step-by-step description of how to implement our proposed tests in practice in the leading case of a linear forecast model,

$$y_t = x'_{t-h}\beta + \varepsilon_{t,h}, \tag{16}$$

---

[8]Note that we take the square function of Inoue's (2001) statistic $d_n(m/n, t)$ in the Kolmogorov-Smirnov version of the test, whereas he considers its absolute value. As suggested previously, this will not affect the rejection frequency of the test statistic.

where $Var\left(\varepsilon_{t,h}|\mathfrak{F}_{t-h}\right) = \sigma^2$ and the density used for evaluation is a Normal. Let β be a $(p-1) \times 1$ vector of parameters, and $\theta \equiv [\beta', \sigma^2]'$, a vector of dimension $p \times 1$.

1. Obtain the test statistics κ_p or C_P as in equations (11) and (12) as follows:

    1a. Recursively estimate the parameters $\{\widehat{\theta}_{t,R}\}$ of the conditional predictive distribution of the variable y_{t+h}, for $t = R, ..., T$.

    1b. Calculate the cumulative distribution function (cdf) $\Phi_t\left(y_{t+h}|\mathfrak{F}_t, \widehat{\theta}_{t,R}\right)$ specified under H_0 – in this example, a Normal distribution.

    1c. Consider a discretized grid for r and τ. For example, in our Monte Carlo we consider $\underline{r} = [0.1, 0.2, ..., 0.9]'$, of dimension $(n_r \times 1)$, and $\underline{\tau} = [0.15, 0.16, ..., 0.85]' \equiv \Upsilon$, of dimension $(n_\tau \times 1)$. For every $r \in \underline{r}$ and $\tau \in \underline{\tau}$, calculate $\Psi_P(\tau, r)$ as in equation (2). The latter gives the first component in $F_P(\tau, r)$ in equation (9). The second component in $F_P(\tau, r)$ is obtained similarly by setting $\tau = 1$.

2. Construct the critical values by simulating the limiting distribution that includes the correction for parameter estimation error, as follows:

    2a. First, estimate the various components of the variance $\Omega\left(r_1, r_2\right)$. Let $\Omega\left(\underline{r}\right)$ be the HAC estimate of the covariance of the $((n_r + p) \times 1)$ vector $\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\begin{pmatrix} \widehat{\xi}_{t+h}\left(\underline{r}\right) \\ q_t(\widehat{\theta}_{t,R}) \end{pmatrix}$, where $q_t(\widehat{\theta}_{t,R})$ is the $(p \times 1)$ vector of scores and $\widehat{\xi}_{t+h}\left(\underline{r}\right)$ is the $(n_r \times 1)$ vector of the distance between the empirical cumulative distribution of the PITs and the uniform distribution evaluated at the vector $\underline{r}$, also of dimension $(n_r \times 1)$.[9] In the case of a Normal density and the linear regression model in equation (16), the score is a $p \times 1$ vector with the following elements

    $$q_t\left(\widehat{\theta}\right) = \left[\frac{(y_t - x'_{t-h}\widehat{\beta}_{t,R})x'_{t-h}}{\widehat{\sigma}^2} \qquad \frac{(y_t - x'_{t-h}\widehat{\beta}_{t,R})^2}{2\widehat{\sigma}^4} - \frac{1}{2\widehat{\sigma}^2}\right]'. \qquad (17)$$

    Let $S\left(\underline{r}\right) \equiv \begin{pmatrix} S_{\Phi\Phi}\left(\underline{r}\right) & S_{q\Phi}\left(\underline{r}\right)' \\ S_{q\Phi}\left(\underline{r}\right) & S_{qq} \end{pmatrix}$ be the HAC estimate of the covariance of $\begin{pmatrix} \widehat{\xi}_{t+h}\left(\underline{r}\right) \\ q_t(\widehat{\theta}_{t,R}) \end{pmatrix}$, for $S_{\Phi\Phi}\left(\underline{r}\right)$ of dimension $(n_r \times n_r)$ and S_{qq} of dimension $(p \times p)$. Note that $S_{\Phi\Phi}\left(r_1, r_2\right)$ is the $(r_1, r_2) - th$ component of $S_{\Phi\Phi}\left(\underline{r}\right)$.

---

[9]Note that $q_t(\widehat{\theta}_{t,R})$ does not depend on r.

2b. Then, estimate the $(n_r \times n_r)$ matrix $\Omega(\underline{r})$ as $\Xi(\underline{r}) \begin{pmatrix} S_{\Phi\Phi}(\underline{r}) & \Pi S'_{q\Phi}(\underline{r}) \\ \Pi S_{q\Phi}(\underline{r}) & 2\Pi S_{qq} \end{pmatrix} \Xi(\underline{r})'$,

where $\Pi = 1 - \left(\frac{P}{R}\right)^{-1} \ln\left(1 + \frac{P}{R}\right)$ and:

$$\Xi(\underline{r}) = \left[I_{n_r} \quad\quad - \left[\frac{1}{P}\sum_{t=R}^{T} \nabla_\theta \Phi\left(\kappa(\underline{r})|\Im_t, \hat{\theta}_{t,R}\right)\right]' \left[\frac{1}{P}\sum_{t=R}^{T}\left(-\nabla_\theta q_t\left(\hat{\theta}_{t,R}\right)\right)\right]^{-1} \right],$$

where: the first element of the $(n_r \times (n_r + p))$ vector $\Xi(\underline{r})$ is the $(n_r \times n_r)$ identity matrix, I_{n_r}; the second element is the product of two components: the first is the average gradient of the PIT, i.e. the average of $\nabla_\theta \Phi\left(\kappa(r)|\Im_t, \hat{\theta}_{t,R}\right)'$, of dimension $(1 \times p)$, stacked across the values for $\underline{r}$ into the $(n_r \times p)$ matrix $\nabla_\theta \Phi\left(\kappa(\underline{r})|\Im_t, \hat{\theta}_{t,R}\right)'$; and the second is the inverse of the average gradient of the score, $\nabla_\theta q_t\left(\hat{\theta}_{t,R}\right)$, of dimension $(p \times p)$. It is important to note that the gradient of the PIT is evaluated not at the realized values of the data, but rather at the values that would have yielded cumulative probabilities equal to the values of the discretized grid of r. That is, the gradient is evaluated at $\kappa(\underline{r}) = \Phi^{-1}(\underline{r}|\Im_t, \hat{\theta}_{t,R})$. For the Normal distribution, we have

$$\nabla_\theta q_t\left(\hat{\theta}_{t,R}\right) = \begin{bmatrix} -\dfrac{x_{t-h}x'_{t-h}}{\hat{\sigma}^2_{t,R}} & -\dfrac{(y_t - x'_{t-h}\hat{\beta}_{t,R})x_{t-h}}{\hat{\sigma}^4_{t,R}} \\[2mm] -\dfrac{(y_t - x'_{t-h}\hat{\beta}_{t,R})x_t}{\hat{\sigma}^4_{t,R}} & \dfrac{1}{2\hat{\sigma}^4_{t,R}} - \dfrac{(y_t - x'_{t-h}\hat{\beta}_{t,R})^2}{\hat{\sigma}^6_{t,R}} \end{bmatrix},$$

and thus the estimate of $A\left(\theta^\dagger\right)$ is $\left[-\frac{1}{P}\sum_{t=R}^{T}\nabla_\theta q_t\left(\hat{\theta}_{t,R}\right)\right]^{-1}$.

Since in our examples we have used a Normal distribution for both the in-sample estimation as well as for the out-of-sample evaluation, we can derive the following formulas for the gradient of the PIT:

$$\nabla_\beta \Phi\left(\kappa(r)|\Im_t, \hat{\theta}_{t,R}\right) = \frac{x_{t-h}}{\hat{\sigma}^2_{t,R}}E(u|u \leq \kappa(r))\Phi(\kappa(r)|\Im_t, \hat{\theta}_{t,R}) - \frac{x'_{t-h}\hat{\beta}_{t,R}x_t}{\hat{\sigma}^2_{t,R}}\Phi(\kappa(r)|\Im_t, \hat{\theta}_{t,R})$$

$$\nabla_{\sigma^2}\Phi\left(\kappa(r)|\Im_t, \hat{\theta}_{t,R}\right) = -\frac{1}{2\hat{\sigma}^2_{t,R}}\Phi(\kappa(r)|\Im_t, \hat{\theta}_{t,R}) + \frac{1}{2\hat{\sigma}^4_{t,R}}\{E(u^2|u \leq \kappa(r))\Phi(\kappa(r)|\Im_t, \hat{\theta}_{t,R}) - ...$$

$$2x'_{t-h}\hat{\beta}_{t,R}E(u|u \leq \kappa(r))\Phi(\kappa(r)|\Im_t, \hat{\theta}_{t,R}) + \left(x'_{t-h}\hat{\beta}_{t,R}\right)^2 \Phi(\kappa(r)|\Im_t, \hat{\theta}_{t,R})\},$$

and $\nabla_\theta \Phi\left(\kappa(r)|\Im_t, \hat{\theta}_{t,R}\right) = \left[\nabla_\beta \Phi\left(\kappa(r)|\Im_t, \hat{\theta}_{t,R}\right)', \nabla_{\sigma^2}\Phi\left(\kappa(r)|\Im_t, \hat{\theta}_{t,R}\right)\right]'$, which is of dimension $(p \times 1)$ for any $r \in \underline{r}$. In the above, $E(u|u \leq y_{t+h})$ and $E(u^2|u \leq$

y_{t+h}) are the conditional non-central moments for a truncated Normal distribution. Let $\tilde{y}_{t+h} = (y_{t+h} - x_t'\hat{\beta}_{t,R})/\hat{\sigma}_{t,R}$, then for the truncated Normal distribution we can write

$$E(u|u \leq y_{t+h}) = x_t'\hat{\beta}_{t,R} - \hat{\sigma}_{t,R}\frac{\phi(\tilde{y}_{t+h})}{\Phi(\tilde{y}_{t+h})}$$

$$Var(u|u \leq y_{t+h}) = \hat{\sigma}_{t,R}^2\left[1 - \tilde{y}_{t+h}\frac{\phi(\tilde{y}_{t+h})}{\Phi(\tilde{y}_{t+h})} - \left(\frac{\phi(\tilde{y}_{t+h})}{\Phi(\tilde{y}_{t+h})}\right)^2\right]$$

$\phi(\tilde{y}_{t+h})$ being the standard normal distribution evaluated at $\tilde{y}_{t+h}$ and $\Phi(\tilde{y}_{t+h})$ its cumulative. Furthermore, $E(u^2|u \leq y_{t+h}) = Var(u|u \leq y_{t+h}) + E(u|u \leq y_{t+h})^2$.

2c. Finally, for each Monte Carlo replication, we simulate the test statistics. Let Γ be the matrix of n_τ row-replications of $\underline{\tau}$, of dimension of $n_\tau \times n_\tau$.[10] Also, define $\Omega_\tau(\underline{\tau}, \underline{r}) \equiv (\min(\Gamma, \Gamma') - \Gamma \odot \Gamma') \otimes \Omega(\underline{r})$, where $\odot$ denotes the element by element product of two matrices, which has dimension $(n_\tau \times n_r) \times (n_\tau \times n_r)$.[11] Also let $\Omega_\tau(\underline{\tau}, \underline{r})^{1/2}$ denote its Cholesky factor, such that $\Omega_\tau(\underline{\tau}, \underline{r}) = \Omega_\tau(\underline{\tau}, \underline{r})^{1/2}\Omega_\tau(\underline{\tau}, \underline{r})^{1/2'}$. Let ν be a vector of $(n_\tau \times n_r) \times 1$ draws from a standard normal distribution; we generate $\widetilde{\Psi}^\circ(\underline{\tau}, \underline{r}) \equiv \Omega_\tau^{1/2}(\underline{\tau}, \underline{r})\nu$, which corresponds to a simulation of the vector $\Psi^\circ(\underline{\tau}, \underline{r}) = \Xi(\underline{r})\frac{1}{\sqrt{P}}[\sum_{t=R}^{R+[\tau P]}\begin{pmatrix}\xi_{t+h}(\underline{r}) \\ q_t\end{pmatrix} - \tau\sum_{t=R}^{T}\begin{pmatrix}\xi_{t+h}(\underline{r}) \\ q_t\end{pmatrix}]$ directly over time.[12] We then reshape the latter into an $(n_\tau \times n_r)$ matrix, for convenience, whose $(\tau, r) - th$ component is the simulated value of $\Psi^\circ(\tau, r)$. We also generate the $(n_r \times 1)$ vector $\widetilde{\Psi}(1, \underline{r}) \equiv \Omega(\underline{r})^{1/2}\eta$, where η is a $(n_r \times 1)$ standard normal, independent of ν, whose r-th component is denoted by $\widetilde{\Psi}(1, r)$. The latter is the simulated $\Psi(1, \underline{r})$. Note that $\widetilde{\Psi}^\circ(\underline{\tau}, \underline{r})$ and $\widetilde{\Psi}(1, \underline{r})$ correspond to the components of the limiting distribution of the proposed tests, (11) and (12). We then construct the test statistics, $\kappa_P \equiv \sup_{\tau\in\underline{\tau}}\sup_{r\in\underline{r}}\left[\widetilde{\Psi}^\circ(\tau, r)^2 + \widetilde{\Psi}(1, r)^2\right]$ and $C_P \equiv \sum_{\tau\in\underline{\tau}}\sum_{r\in\underline{r}}\left[\widetilde{\Psi}^\circ(\tau, r)^2 + \widetilde{\Psi}(1, r)^2\right]$. Repeat the simulations for a large number of replications. The critical values at the $10, 5$, and 1 percent significance levels correspond to the $90, 95$, and 99 percentile values of the test statistics across Monte Carlo simulations, respectively.

---

[10] For example, such set can be generated by the operation "$[\Gamma, \Gamma'] = meshgrid(\underline{\tau}, \underline{\tau})$" in Matlab.

[11] Each element of Ω_τ corresponds to $E[\Psi^\circ(\tau_1, r_1)\Psi^\circ(\tau_2, r_2)]$ defined in equation (6), over $\underline{\tau}$ and $\underline{r}$.

[12] In fact, rather than generating the $((n_r + p) \times 1)$ vector $\begin{pmatrix}\xi_{t+h}(\underline{r}) \\ q_t\end{pmatrix}$, we focus on $\Xi(\underline{r})\begin{pmatrix}\xi_{t+h}(\underline{r}) \\ q_t\end{pmatrix}$, for $\Xi(\underline{r})$ defined in step 2b, and we generate it as $\widetilde{\Psi}_0(\underline{\tau}, \underline{r}) \equiv \Omega_\tau^{1/2}\nu$, where ν is an $(n_\tau \times n_r) \times 1$ vector.

4   Monte Carlo Evidence

In this section we analyze the size and power properties of our proposed tests in small samples for both correctly specified and mis-specified forecasting models. In our Monte Carlo analysis, we compare the performance of the following tests: the test that we propose, κ_P and C_P defined in Theorem (2), which focus on testing whether a Normal predictive density is correctly specified at each point in time; the κ_P^{CS} and C_P^{CS} tests, which focus on whether the Normal predictive density is correctly specified on average over the out-of-sample period (κ_P^{CS} is Corradi and Swanson's (2006a,b) test); the κ_P^I and C_P^I tests, which focus on testing whether the Normal predictive density is constant over time.

4.1   Size Analysis

To investigate the size properties of our tests we consider two Data Generating Processes (DGPs). The forecasts are based on model parameters estimated recursively for $t = R, ..., T + h$. We consider several values of $R = [1000, 500, 200, 100]$ and $P = [1000, 500, 200, 100]$ to evaluate the performance of the proposed procedures in finite samples for $h = 1$. The DGPs, inspired by Corradi and Swanson (2006a), are the following:

DGP S1 (Autoregressive Model): Let $y_t = \rho y_{t-1} + \varepsilon_t, \varepsilon_t \sim iidN(0,1)$, $\rho = 0.2$. The estimated model is an autoregressive model with one lag: $E_t y_{t+1} = \rho y_t$.

DGP S2 (Lag Mis-specification): The data generating process is $y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \varepsilon_t$, where $\varepsilon_t \sim iidN(0,1)$ and $\rho_1 = \rho_2 = 0.2$. The estimated model is: $E_t y_{t+1} = \rho y_t$.

The estimated model in DGP S1 is correctly specified whereas that in DGP S2 is mis-specified. In all cases, the parameters are estimated by OLS using a recursive estimation window scheme. The HAC bandwidth used to estimate the variances is one.

Table 1 shows the results. Panel A in Table 1 reports results for correct specification tests of the density forecasts robust to instabilities (κ_P, C_P); panel B reports results for tests for the traditional tests of correct specification of the density forecast (κ_P^{CS}, C_P^{CS}); and panel C reports tests for instabilities in the density forecasts over time (κ_P^I, C_P^I). Table 1 shows that our tests perform very well in finite samples, with mild over-rejections in the lag mis-specification case.

4.2   Power Analysis

To investigate the power properties of our tests, we consider three DGPs. The DGPs are:

DGP P1 (Constant Mis-specification). The data are generated as: $y_t = \mu + \eta_{1,t} + c\left(\eta_{2,t}^2 - 1\right)\sqrt{2}$, where $\eta_{1,t}$ and $\eta_{2,t}$ are $iidN(0,1)$ and independent of each other, and $\mu = 1$.[13] We report results for various values of c. The case $c = 0$ corresponds to the Normal density case; when c is positive, the density becomes a convolution of a Normal and a χ_1^2 distribution, where the weight on the latter becomes bigger as c increases.

DGP P2 (Parameter Instability). The data are generated as: $y_t = \mu_t + \sigma_t \varepsilon_t$, where $\varepsilon_t \sim iidN(0,1)$ and $\mu_t = \mu_1 = 1$, $\sigma_t = \sigma_1 = 1$ for $t = 1, 2, ..., R$. Further, $\mu_t = (1 + c)\mu_1$, $\sigma_t = (1 + c)\sigma_1$ for $t = R + 1, ...T_1$, and $\mu_t = (1 - c)\mu_1$, $\sigma_t = (1 - c)\sigma_1$ for $t = T_1 + 1, ..., T$; $T_1 = [0.8P]$. We report results for various values of c. The case $c = 0$ corresponds to the constant parameter case; the larger c is, the more instability there is in the parameters.

DGP P3 (Time-varying Mis-specification). The data are generated as: $y_t = \mu + \eta_{1,t} + \sigma\left[\eta_{2,t}^2/\sqrt{2}\right] \cdot 1\,(t > T_1)$, where $\eta_{1,t}$ and $\eta_{2,t}$ are $iidN(0,1)$, independent of each other, and $\mu = 1$, $\sigma = 2$. When $T_1 = T$, the distribution of y_t is Normal; as T_1 decreases, a χ_1^2 distribution is added to the error term starting at time T_1; as a consequence, the shape of the distribution in the data changes over time.

The results are shown in Table 2. Recall that the mis-specification tests evaluate whether the predictive density is Normal. The table shows that, across all designs, our proposed density tests robust to instabilities (κ_P, C_P) have good power properties in detecting mis-specification, even when it only appears in parts of the sample. On the other hand, the other tests may lack power: Panel A shows that the instability tests (κ_P^I, C_P^I) have no power to detect mis-specification in the predictive density when the mis-specification is constant over time; the correct specification tests $(\kappa_P^{CS}, C_P^{CS})$ do instead detect mis-specification, and, by construction, have higher power than the density forecast tests robust to instabilities. At the same time, Panel B shows that the correct specification tests have less power than our proposed tests in detecting mis-specification in the moments of the predictive distribution, when the shape of the distribution is correctly specified; in the case of time-varying parameters, by construction the instability tests would detect the instabilities with a higher power than our proposed density tests robust to instabilities. Similar results hold in the presence of time-varying mis-specification, as Panel C shows. Overall, our tests are more robust to detecting possibly time-varying mis-specification across all the designs than the other tests which focus only on either time-variation or constant mis-specification.

---

[13] Note that $\left(\eta_{2,t}^2 - 1\right)\sqrt{2}$ is a chi-squared distribution with zero mean and variance one, that is, it has the same mean and variance as $\eta_{1,t}$, although the shape is different.

5 SPF Density Forecast Evaluation

The Survey of Professional Forecasters (SPF) collects information on density forecasts for inflation and output growth made by professional forecasters in the U.S. The forecasters assign a probability value (over pre-defined intervals) to the year-over-year inflation and output growth rates for the current calendar year (nowcast) and the following year (one-year-ahead forecast). The forecasters update their probabilities for the nowcast and the one-year-ahead forecasts on a quarterly basis, and over the course of the year the survey obtains several forecast values for the same target variable.

Diebold et al. (1999) are among the first to evaluate whether the SPF density forecasts of inflation are correctly specified. In particular, they assess whether realized inflation rates are consistent with the empirical distribution of the mean probability forecasts of the survey. Interestingly, they note the presence of time variation in density forecasts, and emphasize that, in their sample, the distribution has shifted from over-estimating a large negative shock to over-estimating large shocks of either sign. While the Diebold et al. (1999) finding of time variation is empirically very interesting, it is based on an ad-hoc choice of sub-sample periods. However, it is not necessarily the case that sub-samples chosen with an ad-hoc procedure identify when the forecast distribution has indeed shifted or become more mis-specified. In addition, if the sub-samples were chosen using information from the data, this should be taken into account when evaluating the presence of mis-specification in sub-samples. Our procedure can detect mis-specification even if it appears only in a sub-sample of the data and the researcher does not need to pre-specify the sub-sample; rather, the data will let the procedure detect it. We therefore proceed to test the correct specification for the SPF density forecasts by using our test. In addition to inflation (the variable considered in Diebold et al., 1999), we also investigate the conditional density forecasts of output growth.

We obtain the mean probability forecasts from the Survey of Professional Forecasters, which are publicly available from the Federal Reserve Bank of Philadelphia. We focus on real GNP/GDP and the GNP/GDP deflator as measures of output and prices. The realized values of inflation and output growth are based on the real-time data set for macroeconomists, also available from the Federal Reserve Bank of Philadelphia.[14] The SPF dataset presents several challenges since the questionnaire has changed over time in various dimensions: there have been changes in the definition of the variables, the intervals over which probabilities have been assigned, as well as the forecast horizon. The most important change in the

---

[14] The data are available at http://www.philadelphiafed.org/research-and-data/real-time-center.

density forecast evaluation process is the switch from GNP to GDP in 1992:I. However, since the change in the definition of the forecasted variable affects both the forecast and the target values simultaneously, we do not consider it as a major impediment. On the other hand, the probability forecasts for output growth are in terms of nominal GNP, as opposed to real GNP, prior to 1981:III. Even though we do have probability forecasts for inflation for the same period of time, it is unclear how to derive an implied probability for the joint dynamics of nominal output and prices in order to estimate the density for real output. To mitigate this problematic issue, we truncate the data set and only consider the mean probability distribution forecasts of GNP/GDP for the period 1981:III-2011:IV. For the GNP/GDP price deflator, however, we use the full sample of data on probability forecasts, available for 1968:IV-2011:IV. We should also note that, in the sample period we consider, both nowcasts and one-year-ahead forecasts are available for the output growth. For inflation, the one-year-ahead forecasts become available only starting 1981:III. Thus, the sample sizes for the various forecast densities that we consider differ depending on the availability of the data, which might affect the empirical results. We deal with the fact that the intervals over which forecasters historically provided probability forecasts has changed over time by considering, for each period of time, a normal approximation to the discrete forecast density distribution provided by the SPF.[15]

We use the year-over-year growth rates of output growth and inflation calculated from the first quarterly vintage of real GNP/GDP and the GNP/GDP deflator in each year to evaluate the density forecasts. For instance, in order to obtain the growth rate of real output for 1981, we take the 1982:I vintage of data and calculate the growth rate of the annual average level of GNP/GDP from 1980 to 1981. We consider the annual-average over annual-average percent change (as opposed to fourth-quarter over forth-quarter percent change) in output and prices to make it comparable with the definition of the variables that SPF forecasters provide probabilistic predictions for.

Table 3 reports the empirical results. The table has three panels. Panel A considers the tests of correct specification robust to instabilities proposed in this paper (the κ_P, C_P tests); Panel B focuses on the Corradi and Swanson (2006a,b) test (κ_P^{CS}, C_P^{CS}). Both tests evaluate whether the Normal density is an appropriate specification for the SPF density nowcasts and one-year-ahead forecasts of output growth and inflation. Panel C reports

---

[15]Note that the use of a normal interpolation would be problematic in case our empirical evidence did not reject the correct specification of the density forecasts, as the latter might be caused by the Normal interpolation. However, we reject the correct specification notwithstanding the Normal interpolation.

results of tests for instability in density forecasts (the κ_P^I, C_P^I tests). The test statistics consistently reject the respective null hypotheses of correct specification at each point in time, correct specification on average, and stability at 5% significance level for both inflation and output growth nowcasts and for one-quarter-ahead inflation forecasts, suggesting both mis-specification and time-variation over the out-of-sample evaluation period. The evidence of time-variation is consistent with Andrade et al. (2012), who also document time-variation in the performance of individual-level SPF density forecasts, with an emphasis on the dynamics of the inter-quantile ranges and skewness measures.

The value of $R + [\tau P]$ associated with the largest value of the κ_P statistic may provide an indication regarding the date of a potential break in the mis-specification of the density forecasts.[16] The date for inflation and output growth nowcasts is around the beginning of 1980s, right at the onset of the Great Moderation. In particular, the date is 1979:II for inflation nowcasts, which coincides with the beginning of Volker's disinflation period. For the one-year ahead inflation and output growth, the date is closer to the late 1990s.[17]

Note that our proposed tests detect whether the correct specification has been violated in at least some part of the sample, but do not directly determine whether it has shifted from being correctly calibrated to mis-specified, or vice versa, or whether the mis-specification has changed over time. We shed further light on the proper calibration of the PITs by plotting the density forecasts before and after the dates suggested by the κ_P test. Figure 1 reports the density forecasts of output growth and Figure 2 those of inflation. In both figures, Panel A reports the density forecast estimated on average over the out-of-sample period for the nowcast (on the left) and the one-year-ahead forecast (on the right). Panel B reports density forecasts for the nowcast and Panel C reports one-year ahead density forecasts, both depicted over the sub-samples determined by our procedure. The solid line plots the expected value of a $U(0,1)$ over 10 bins (i.e. 0.1). Comparing Figures 1 and 2, it is clear that the mis-specification is worse for inflation than for output growth. Interestingly, the density forecast for the current year inflation (in Figure 2, Panel B), suggests a shift towards severe over-estimation of the tail risk after 1979:III. Note that, by looking at the full out-of-sample period only (the left graph in Figure 2, Panel A), one would detect the existence of mis-specification on average but would miss that, before 1979, forecast densities were substantially better calibrated. For the case of one-year-ahead inflation, on the other

---

[16]The date is determined as $R + [\tau^* P]$, where $\tau^* = \arg \sup_\tau \{\sup_r \kappa_P(\tau, r)\}$.

[17]The date is different if one only looks at the instability tests for output growth, which is 2004:I. Thus, structural breaks in density forecasts do not necessarily match potential dates for mis-specification.

hand, the surveys have improved the estimation of the upper tail risk after 1996:I.

Our empirical results are important in the light of the finding that survey forecasts are reportedly providing the best forecasts of inflation. For example, Ang et al. (2007) find that survey forecasts outperform several competing forecasting models (including the Phillips curve, the term structure and ARIMA models) and that, when combining forecasts, the data assign the highest weight on survey information. Our results suggest that surveys are still not providing a correct forecast for the whole distribution of inflation, at least when considering a Normal distribution.

6   Conclusions

This paper proposes new tests for predictive density evaluation. They are designed to be robust to the presence of mis-specification as well as instabilities in the mis-specification. The techniques are based on Kolmogorov-Smirnov and Cramér-von Mises-type test statistics. An empirical application of the proposed methodologies to the SPF uncovers that both output growth and inflation density forecast are mis-specified and finds significant evidence of time-variation in the mis-specification.

Acknowledgments

7 Appendix A. Proofs

Below we prove Theorems 1 and 2.

Proof of Theorem 1. (i) That $\{z_{t+h}\}_{t=R}^{T}$ is $U(0,1)$ under H_0 follows from Corradi and Swanson (2006b, p. 213).

(ii) Note that:

$$\Psi_P(\tau,r) \equiv \frac{1}{\sqrt{P}} \sum_{t=R}^{R+[\tau P]} \left(1\left\{ \Phi\left(y_{t+h}|\Im_t,\widehat{\theta}_{t,R}\right) \le r \right\} - r \right) \tag{18}$$

$$= \frac{1}{\sqrt{P}} \sum_{t=R}^{R+[\tau P]} \left(1\left\{ \Phi\left(y_{t+h}|\Im_t,\theta^{\dagger}\right) \le \Phi\left(\Phi^{-1}\left(r|\Im_t,\widehat{\theta}_{t,R}\right)|\Im_t,\theta^{\dagger}\right) \right\} - r \right) \tag{19}$$

$$= \frac{1}{\sqrt{P}} \sum_{t=R}^{R+[\tau P]} \left(1\left\{ \Phi\left(y_{t+h}|\Im_t,\theta^{\dagger}\right) \le \Phi\left(\Phi^{-1}\left(r|\Im_t,\widehat{\theta}_{t,R}\right)|\Im_t,\theta^{\dagger}\right) \right\} - \Phi\left(\Phi^{-1}\left(r|\Im_t,\widehat{\theta}_{t,R}\right)|\Im_t,\theta^{\dagger}\right) \right) \tag{20}$$

$$+ \frac{1}{\sqrt{P}} \sum_{t=R}^{R+[\tau P]} \left(\Phi\left(\Phi^{-1}\left(r|\Im_t,\widehat{\theta}_{t,R}\right)|\Im_t,\theta^{\dagger}\right) - r \right) \tag{21}$$

$$= \frac{1}{\sqrt{P}} \sum_{t=R}^{R+[\tau P]} \left(1\left\{ \Phi\left(y_{t+h}|\Im_t,\theta^{\dagger}\right) \le \Phi\left(\Phi^{-1}\left(r|\Im_t,\widehat{\theta}_{t,R}\right)|\Im_t,\theta^{\dagger}\right) \right\} - \Phi\left(\Phi^{-1}\left(r|\Im_t,\widehat{\theta}_{t,R}\right)|\Im_t,\theta^{\dagger}\right) \right) \tag{22}$$

$$- \frac{1}{\sqrt{P}} \sum_{t=R}^{R+[\tau P]} \nabla_\theta \Phi\left(\Phi^{-1}\left(r|\Im_t,\widehat{\theta}_{t,R}\right)|\Im_t,\overline{\theta}_{t,R}\right) \left(\widehat{\theta}_{t,R} - \theta^{\dagger}\right) \tag{23}$$

for $\overline{\theta}_{t,R} \in \left(\widehat{\theta}_{t,R}, \theta^{\dagger}\right)$, where (19) follows from Assumption 1a(i),[18] and the last passage follows from a mean value expansion.[19]

We first show that

$$\frac{1}{\sqrt{P}} \sum_{t=R}^{R+[\tau P]} \left(1\left\{ \Phi\left(y_{t+h}|\Im_t,\theta^{\dagger}\right) \le \Phi\left(\Phi^{-1}\left(r|\Im_t,\widehat{\theta}_{t,R}\right)|\Im_t,\theta^{\dagger}\right) \right\} - \Phi\left(\Phi^{-1}\left(r|\Im_t,\widehat{\theta}_{t,R}\right)|\Im_t,\theta^{\dagger}\right) \right)$$

$$= \frac{1}{\sqrt{P}} \sum_{t=R}^{R+[\tau P]} \left(1\left\{ \Phi\left(y_{t+h}|\Im_t,\theta^{\dagger}\right) \le r \right\} - r \right) + o_p(1); \tag{24}$$

---

[18] In fact, note that (19) follows from $\Phi\left(y_{t+1}|\Im_t,\widehat{\theta}_{t,R}\right) \le r$ implies $y_{t+1} \le \Phi^{-1}\left(r|\Im_t,\widehat{\theta}_{t,R}\right)$, which in turn implies that $\Phi\left(y_{t+1}|\Im_t,\theta^{\dagger}\right) \le \Phi\left(\Phi^{-1}\left(r|\Im_t,\widehat{\theta}_{t,R}\right)|\Im_t,\theta^{\dagger}\right)$.

[19] Following Corradi and Swanson (2006b, p. 276), we have: $r \equiv \Phi\left(\Phi^{-1}\left(r|\Im_t,\widehat{\theta}_{t,R}\right)|\Im_t,\widehat{\theta}_{t,R}\right) = \Phi\left(\Phi^{-1}\left(r|\Im_t,\widehat{\theta}_{t,R}\right)|\Im_t,\theta^{\dagger}\right) + \nabla_\theta\Phi\left(\Phi^{-1}\left(r|\Im_t,\widehat{\theta}_{t,R}\right)|\Im_t,\overline{\theta}_{t,R}\right)\left(\widehat{\theta}_{t,R} - \theta^{\dagger}\right)$.

21

we then proceed showing that

$$
\frac{1}{\sqrt{P}} \sum_{t=R}^{R+[\tau P]} \nabla_\theta \Phi \left(\Phi^{-1} \left(r | \Im_t, \widehat{\theta}_{t,R} \right) | \Im_t, \overline{\theta}_{t,R} \right)' \left(\widehat{\theta}_{t,R} - \theta^\dagger \right)
$$

$$
= E \left[\nabla_\theta \Phi \left(\Phi^{-1} \left(r | \Im_t, \theta^\dagger \right) | \Im_t, \theta^\dagger \right) \right]' \frac{1}{\sqrt{P}} \sum_{t=R}^{R+[\tau P]} \left(\widehat{\theta}_{t,R} - \theta^\dagger \right) + o_p(1), \tag{25}
$$

where the $o_p(1)$ term holds uniformly in τ and r. Recall $\xi_{t+h}(r)$ defined in (5).

Regarding (24), under H_0, $\Phi(.)$ is a stationary strong mixing process given the fact that $r \in [0,1]$ and that the indicator function is bounded, as bounded functions of mixing processes are mixing – White, 1984, p. 50) ; thus,

$$
Pr \left(\sup_\tau \sup_{r \in [0,1]} \sup_{r_2 : |r - r_2| < \delta} \left| \frac{1}{\sqrt{P}} \sum_{t=R}^{R+[\tau P]} [\xi_{t+h}(r) - \xi_{t+1}(r_2)] \right| > \varepsilon \right)
$$

$$
= Pr \left(\sup_\tau \sqrt{\tau} \sup_{r \in [0,1]} \sup_{r_2 : |r - r_2| < \delta} \left| \frac{1}{\sqrt{\tau P}} \sum_{j=1}^{[\tau P]} [\xi_{R+j+h}(r) - \xi_{R+j}(r_2)] \right| > \varepsilon \right) \to 0
$$

as $\delta \to 0$ using the fact that τ is bounded between zero and one, and Corradi and Swanson (2006a, p. 796, proof of eq. 12). Note that Assumptions 1b,d ensure that, for $\widetilde{\tau} \in \Upsilon \subset (0,1)$,

$$
\sup_{\widetilde{\tau} \in \Upsilon} \left\| \left[-\frac{1}{T} \sum_{j=1}^{[T\widetilde{\tau}]} \nabla_\theta q_j \left(\overline{\theta}_{t,R} \right) \right]^{-1} - \widetilde{\tau} A \left(\theta^\dagger \right) \right\| \underset{p}{\to} 0,
$$

as in Andrews (1993, eq. A.8, p. 848), which implies that

$$
\sup_{\tau \in \Upsilon} \left\| \left[-\frac{1}{R + [\tau P]} \sum_{j=1}^{R+[\tau P]} \nabla_\theta q_j \left(\overline{\theta}_{t,R} \right) \right]^{-1} - A \left(\theta^\dagger \right) \right\| \underset{p}{\to} 0, \tag{26}
$$

which guarantees that a condition similar to West's (1996) Assumption 2(a) holds. Thus, the latter also implies that $\sup_t |t^a \left(\widehat{\theta}_{t,R} - \theta^\dagger \right)| \underset{p}{\to} 0$ for $0 \le a < 1/2$ by Assumption 1e and Lemma A3 in West (1996), which also implies that $\sup_t |T^a \left(\widehat{\theta}_{t,R} - \theta^\dagger \right)| \underset{p}{\to} 0$. This latter result and Assumptions 1a(ii) ensure that (24) holds following an argument similar to that in Corradi and Swanson (2006a, p. 796, proof of eq. 13).[20]

---

[20]Corradi and Swanson (2006b) note that the argument in Corradi and Swanson (2006a) carries over out-of-sample.

22

Regarding (25):

$$P^{-1/2} \sum_{t=R}^{R+[\tau P]} \nabla_\theta \Phi \left(\Phi^{-1} \left(r | \Im_t, \widehat{\theta}_{t,R} \right) | \Im_t, \overline{\theta}_{t,R} \right)' \left(\widehat{\theta}_{t,R} - \theta^\dagger \right) \tag{27}$$

$$= E \left[\nabla_\theta \Phi \left(\Phi^{-1} \left(r | \Im_t, \theta^\dagger \right) | \Im_t, \theta^\dagger \right) \right]' \frac{1}{\sqrt{P}} \sum_{t=R}^{R+[\tau P]} \left(\widehat{\theta}_{t,R} - \theta^\dagger \right) + \tag{28}$$

$$+ \frac{1}{\sqrt{P}} \sum_{t=R}^{R+[\tau P]} \left\{ \nabla_\theta \Phi \left(\Phi^{-1} \left(r | \Im_t, \widehat{\theta}_{t,R} \right) | \Im_t, \overline{\theta}_{t,R} \right) - E \left[\nabla_\theta \Phi \left(\Phi^{-1} \left(r | \Im_t, \theta^\dagger \right) | \Im_t, \theta^\dagger \right) \right] \right\}' \left(\widehat{\theta}_{t,R} - \theta^\dagger \right) \tag{29}$$

$$= E \left[\nabla_\theta \Phi \left(\kappa(r) | \Im_t, \theta^\dagger \right) \right]' \frac{1}{\sqrt{P}} \sum_{t=R}^{R+[\tau P]} \left(\widehat{\theta}_{t,R} - \theta^\dagger \right) + o_p(1) , \tag{30}$$

where (29) converges in probability to zero uniformly over time by Assumption 1f (cfr. Bai (2003, eq. C5).

It follows from (18), (24) and (25) that:

$$\Psi_P(\tau, r) = \frac{1}{\sqrt{P}} \sum_{t=R}^{R+[\tau P]} \left(1 \left\{ \Phi \left(y_{t+h} | \Im_t, \theta^\dagger \right) \leq r \right\} - r \right) \tag{31}$$

$$- E \left[\nabla_\theta \Phi \left(\kappa(r) | \Im_{t-1}, \theta^\dagger \right) \right]' \frac{1}{\sqrt{P}} \sum_{t=R}^{R+[\tau P]} \left(\widehat{\theta}_{t,R} - \theta^\dagger \right) + o_p(1) .$$

Let $Q_t \equiv \frac{1}{t} \sum_{j=1}^{t} q_j \left(\theta^\dagger \right)$ and, for simplicity, $q_j \equiv q_j \left(\theta^\dagger \right)$. From the first order condition for the estimation of the parameter and a mean value expansion, we have:

$$0 = \frac{1}{t} \sum_{j=1}^{t} q_j \left(\widehat{\theta}_{t,R} \right) = Q_t + \frac{1}{t} \sum_{j=1}^{t} \nabla_\theta q_j \left(\overline{\theta}_{t,R} \right) \left(\widehat{\theta}_{t,R} - \theta^\dagger \right) ; \text{ thus,}$$

$$P^{-1/2} \sum_{t=R}^{R+[\tau P]} \left(\widehat{\theta}_{t,R} - \theta^\dagger \right) = -P^{-1/2} \sum_{t=R}^{R+[\tau P]} \left[\frac{1}{t} \sum_{j=1}^{t} \nabla_\theta q_j \left(\overline{\theta}_{t,R} \right) \right]^{-1} Q_t = P^{-1/2} \sum_{t=R}^{R+[\tau P]} A \left(\theta^\dagger \right) Q_t \tag{32}$$

$$+ P^{-1/2} \sum_{t=R}^{R+[\tau P]} \left\{ \left[-\frac{1}{t} \sum_{j=1}^{t} \nabla_\theta q_j \left(\overline{\theta}_{t,R} \right) \right]^{-1} - A \left(\theta^\dagger \right) \right\} Q_t, \tag{33}$$

23

where $A\left(\theta^{\dagger}\right)$ is defined in Theorem 1. Given (26), the last term in (33) is $o_p\left(1\right)$. By noting that $P^{-1/2}\sum_{t=R}^{R+[\tau P]}\left|\frac{1}{t}\sum_{j=1}^{t}q_j\left(\theta^{\dagger}\right)\right|=O_p\left(1\right)$, it follows from Assumption 1e and West (1996, Lemma A4(c)) that:

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{R+[\tau P]}\left(\widehat{\theta}_{t,R}-\theta^{\dagger}\right)=A\left(\theta^{\dagger}\right)\frac{1}{\sqrt{P}}\sum_{t=R}^{R+[\tau P]}\frac{1}{t}\sum_{j=1}^{t}q_j+o_p\left(1\right).\tag{34}$$

It follows from (31) and (34) that:

$$\Psi_P\left(\tau,r\right)=\frac{1}{\sqrt{P}}\sum_{t=R}^{R+[\tau P]}\xi_{t+h}\left(r\right)-E\left[\nabla_{\theta}\Phi\left(\kappa(r)|\Im_t,\theta^{\dagger}\right)\right]'\,A\left(\theta^{\dagger}\right)\frac{1}{\sqrt{P}}\sum_{t=R}^{R+[\tau P]}Q_t+o_p\left(1\right)\tag{35}$$

$$=\underbrace{\left[1\quad\quad-E\left[\nabla_{\theta}\Phi\left(\kappa(r)|\Im_t,\theta^{\dagger}\right)\right]'A\left(\theta^{\dagger}\right)\right]}_{\equiv\Xi}\underbrace{\frac{1}{\sqrt{P}}\sum_{t=R}^{R+[\tau P]}\left(\begin{array}{c}\xi_{t+h}\left(r\right)\\Q_t\end{array}\right)}_{\equiv K_P(\tau,r)}+o_p\left(1\right).\tag{36}$$

The result will follow by a reasoning similar to Theorem 2.1 in Inoue (2001) if we show that: (a) the sample covariance kernel converges to the specified covariance kernel; and (b) we establish convergence of the finite dimensional distributions of $\Psi_P\left(\tau_1,r_1\right)$ to the finite dimensional distribution of $\Psi\left(.,.\right)$ and tightness holds.

(a) First, we show that $\lim_{T\to\infty}E\left[K_P\left(\tau_1,r_1\right)K_P\left(\tau_2,r_2\right)'\right]$ is absolutely convergent (which implies that the limiting variance of $K_P\left(\tau,r\right)$ is absolutely convergent, since $E\left[K_P\left(\tau,r\right)\right]=0$). From (36):

$$K_P\left(\tau_1,r_1\right)K_P\left(\tau_2,r_2\right)'$$

$$=\left(\frac{1}{\sqrt{P}}\sum_{s=R}^{R+[\tau_1 P]}\left(\begin{array}{c}\xi_{t+h}\left(r_1\right)\\Q_t\end{array}\right)\right)\left(\frac{1}{\sqrt{P}}\sum_{s=R}^{R+[\tau_2 P]}\left(\begin{array}{c}\xi_{s+h}\left(r_2\right)\\Q_s\end{array}\right)'\right)$$

$$=\frac{T}{P}\left(\begin{array}{cc}T^{-1}\sum_{t=R}^{R+[\tau_1 P]}\sum_{s=R}^{R+[\tau_2 P]}\xi_{t+h}\left(r_1\right)\xi_{s+h}\left(r_2\right)&T^{-1}\sum_{t=R}^{R+[\tau_1 P]}\sum_{s=R}^{R+[\tau_2 P]}\xi_{t+h}\left(r_1\right)Q'_s\\T^{-1}\sum_{t=R}^{R+[\tau_1 P]}\sum_{s=R}^{R+[\tau_2 P]}Q_t\xi_{s+h}\left(r_2\right)&T^{-1}\left(\sum_{t=R}^{R+[\tau_1 P]}Q_t\right)\left(\sum_{s=R}^{R+[\tau_2 P]}Q'_s\right)\end{array}\right).$$

Let $a_{R,j,\tau}=\sum_{k=j}^{[\tau P]}\left(R+k\right)^{-1}$ for $j=0,...,[\tau P]$. Note that, since the parameter is estimated via a recursive scheme, $\sum_{t=R}^{R+[\tau P]}Q_t=a_{R,0,\tau}\left(q_1+...+q_R\right)+a_{R,1,\tau}q_{R+1}+...+a_{R,P-1,\tau}q_{R+[\tau P]}.$[21]

---

[21] This follows from a reasoning similar to West (1996, p. 1081).

Thus, $\sum_{t=R}^{R+[\tau P]} Q_t = \sum_{t=1}^{R+[\tau P]} b_{t,\tau} q_t$ where $b_{t,\tau} = a_{R,0,\tau}$ for $t = 1, ..., R$ and $b_{t,\tau} = a_{R,t-R,\tau}$ for $t = R + 1, ..., R + [\tau P]$. Since $\{b_{t,\tau}\}$ is a sequence of deterministic bounded (by unity) constants and q_t is strong mixing, $b_{t,\tau} q_t$ is mixing (White, 1984, Theorem 3.49). Thus,

$$T^{-1} \left(\sum_{t=R}^{R+[\tau_1 P]} Q_t \right) \left(\sum_{s=R}^{R+[\tau_2 P]} Q'_s \right) = T^{-1} \left(\sum_{t=1}^{R+[\tau_1 P]} b_{t,\tau_1} q_t \right) \left(\sum_{s=1}^{R+[\tau_2 P]} b_{s,\tau_2} q'_s \right)$$

$$= T^{-1} \sum_{t=1}^{R+[\tau_1 P]} \sum_{s=1}^{R+[\tau_2 P]} b_{t,\tau_1} b_{s,\tau_2} q_t q'_s.$$

Also, let $\widetilde{\xi}_{i+1} \equiv 0$ for $i < R$ and $\widetilde{\xi}_{i+1} \equiv \xi_{i+1}$ for $i \geq R$; then, from a similar reasoning,

$$T^{-1} \sum_{s=R}^{R+[\tau_2 P]} \left(\sum_{t=R}^{R+[\tau_1 P]} Q_t \xi_{s+h}(r_2) \right) = T^{-1} \sum_{s=1}^{R+[\tau_2 P]} \left(\sum_{t=1}^{R+[\tau_1 P]} b_{t,\tau_1} q_t \widetilde{\xi}_{s+h}(r_2) \right), \text{ and}$$

$$T^{-1} \sum_{t=R}^{R+[\tau_1 P]} \sum_{s=R}^{R+[\tau_2 P]} \xi_{t+h}(r_1) \xi_{s+h}(r_2) = T^{-1} \sum_{t=1}^{R+[\tau_1 P]} \sum_{s=1}^{R+[\tau_2 P]} \widetilde{\xi}_{t+h}(r_1) \widetilde{\xi}_{s+h}(r_2).$$

Thus,

$$K_P(\tau_1, r_1) K_P(\tau_2, r_2)'$$

$$= \frac{T}{P} \begin{pmatrix} T^{-1} \sum_{t=1}^{R+[\tau_1 P]} \sum_{s=1}^{R+[\tau_2 P]} \widetilde{\xi}_{t+h}(r_1) \widetilde{\xi}_{s+h}(r_2) & T^{-1} \sum_{t=1}^{R+[\tau_1 P]} \sum_{s=1}^{R+[\tau_2 P]} \widetilde{\xi}_{t+h}(r_1) b_{s,\tau_2} q'_s \\ T^{-1} \sum_{s=1}^{R+[\tau_2 P]} \sum_{t=1}^{R+[\tau_1 P]} b_{t,\tau_1} q_t \widetilde{\xi}_{s+h}(r_2) & T^{-1} \sum_{t=1}^{R+[\tau_1 P]} \sum_{s=1}^{R+[\tau_2 P]} b_{t,\tau_1} b_{s,\tau_2} q_t q'_s \end{pmatrix}$$

is absolutely convergent since, by Theorem A5 in Hall and Heyde (1980), Assumption 1d and $\lim_{T \to \infty} \left(\frac{T}{P} \right) < \infty$ (which follows from Assumption 1c), each element is finite:

$$T^{-1} \sum_{t=1}^{R+[\tau_1 P]} \sum_{s=1}^{R+[\tau_2 P]} E \left| \widetilde{\xi}_{t+h}(r_1) \widetilde{\xi}_{s+h}(r_2) \right| \leq T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} E \left| \widetilde{\xi}_{t+h}(r_1) \widetilde{\xi}_{s+h}(r_2) \right|$$

$$\leq 4 T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \alpha(t-s) \leq \sum_{j=-\infty}^{\infty} \alpha(j) < \infty,$$

$$T^{-1} \sum_{t=1}^{R+[\tau_1 P]} \sum_{s=1}^{R+[\tau_2 P]} E \left| b_{t,\tau_1} b_{s,\tau_2} q_t q'_s \right| \leq T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} E \left| q_t q'_s \right| \leq 4 T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \alpha(|t-s|) \leq \sum_{j=-\infty}^{\infty} \alpha(j) < \infty,$$

$$T^{-1} \sum_{t=1}^{R+[\tau_1 P]} \sum_{s=1}^{R+[\tau_2 P]} E \left| b_{t,\tau_1} q_t \widetilde{\xi}_{s+h}(r_2) \right| \leq T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} E \left| q_t \widetilde{\xi}_{s+h}(r_2) \right| \leq 4 T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \alpha(|t-s|)$$

$$\leq \sum_{j=-\infty}^{\infty} \alpha(j) < \infty.$$

25

Thus, $\lim_{T\to\infty} K_P(\tau_1, r_1) K_P(\tau_2, r_2)'$ is absolutely convergent.

From Lemma A1(c) in Inoue (2001), we have:

$$E\left[\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{R+[\tau_1 P]}\left(\begin{array}{c}\xi_{t+h}(r_1)\\Q_t\end{array}\right)\right)\left(\frac{1}{\sqrt{P}}\sum_{s=R}^{R+[\tau_2 P]}\left(\begin{array}{c}\xi_{s+h}(r_2)\\Q_s\end{array}\right)\right)'\right]$$

$$\to \min(\tau_1, \tau_2)\lim_{T\to\infty} E\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(\begin{array}{c}\xi_{t+h}(r_1)\\Q_t\end{array}\right)\frac{1}{\sqrt{P}}\sum_{s=R}^{T}\left(\begin{array}{c}\xi_{s+h}(r_2)\\Q_s\end{array}\right)'\right). \qquad (37)$$

Note that $\lim_{T\to\infty} E\left(\frac{1}{P}\sum_{t=R}^{T}Q_t\sum_{s=R}^{T}Q_s'\right) = 2\Pi S_{qq}$ by Assumptions 1c and 1e and Lemma A5 in West (1996). Note also that $\lim_{T\to\infty} E\left(\frac{1}{P}\sum_{t=R}^{T}\sum_{s=R}^{T}Q_t\xi_{s+h}(r)\right) = \Pi S_{q\Phi}(r)$ by Assumptions 1c and 1e and Lemma A6 in West (1996). Also, $\lim_{T\to\infty} E\left(\frac{1}{P}\sum_{t=R}^{T}\sum_{s=R}^{T}\xi_{t+h}(r_1)\xi_{s+h}(r_2)\right) = S_{\Phi\Phi}(r_1, r_2)$, where S_{qq}, $S_{q\Phi}(r)$ and $S_{\Phi\Phi}(r_1, r_2)$ are defined in Theorem 1. Thus,

$$\lim_{T\to\infty} E\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(\begin{array}{c}\xi_{t+h}(r_1)\\Q_t\end{array}\right)\frac{1}{\sqrt{P}}\sum_{s=R}^{T}\left(\begin{array}{c}\xi_{s+h}(r_2)\\Q_s\end{array}\right)'\right) = \left(\begin{array}{cc}S_{\Phi\Phi}(r_1, r_2) & \Pi S_{q\Phi}'(r_1)\\\Pi S_{q\Phi}(r_2) & 2\Pi S_{qq}\end{array}\right). \qquad (38)$$

Hence, by combining (36), (37) and (38), we have:

$$E\left[\Psi_P(\tau_1, r_1)\Psi_P(\tau_2, r_2)\right] = \min(\tau_1, \tau_2)\Xi\cdot\left(\begin{array}{cc}S_{\Phi\Phi}(r_1, r_2) & \Pi S_{q\Phi}'(r_1)\\\Pi S_{q\Phi}(r_2) & 2\Pi S_{qq}\end{array}\right)\cdot\Xi'$$

$$= \min(\tau_1, \tau_2)\cdot\Omega(r_1, r_2), \text{ where}$$

$$\Omega(r_1, r_2) = S_{\Phi\Phi}(r_1, r_2) + 2\Pi E\left[\nabla_\theta\Phi\left(\kappa(r_1)|\Im_t, \theta^\dagger\right)\right]'A(\theta^\dagger)S_{qq}A(\theta^\dagger)E\left[\nabla_\theta\Phi\left(\kappa(r_2)|\Im_t, \theta^\dagger\right)\right]$$

$$- \Pi E\left[\nabla_\theta\Phi\left(\kappa(r_1)|\Im_t, \theta^\dagger\right)\right]'A(\theta^\dagger)S_{q\Phi}(r_1) - \Pi S_{q\Phi}'(r_2)A(\theta^\dagger)E\left[\nabla_\theta\Phi\left(\kappa(r_2)|\Im_t, \theta^\dagger\right)\right].$$

Thus, $\lim_{P\to\infty} E\left[\Psi_P(\tau_1, r_1)\Psi_P(\tau_2, r_2)\right] = \min(\tau_1, \tau_2)\Omega(r_1, r_2)$.

(b) follows from (a), Assumption 1 and Theorem 2.1 in Inoue (2001). ∎

Proof of Theorem 2. Eqs. (9) and (10) together with Theorem 1 imply:

$$R_P(\tau)F_P\left(\tau,r\right) \equiv \begin{bmatrix} \left[(1-\tau)\, P^{-1/2} \sum_{t=R}^{R+[\tau P]} \widehat{\xi}_{t+h}\left(r\right) - \tau P^{-1/2} \sum_{t=R+[\tau P]+1}^{T} \widehat{\xi}_{t+h}\left(r\right)\right] \\ P^{-1/2}\left[\sum_{t=R}^{R+[\tau P]} \widehat{\xi}_{t+h}\left(r\right) + \sum_{t=R+[\tau P]+1}^{T} \widehat{\xi}_{t+h}\left(r\right)\right] \end{bmatrix}$$

$$= \begin{bmatrix} \left[P^{-1/2} \sum_{t=R}^{R+[\tau P]} \widehat{\xi}_{t+h}\left(r\right) - \tau P^{-1/2} \sum_{t=R}^{T} \widehat{\xi}_{t+h}\left(r\right)\right] \\ P^{-1/2} \sum_{t=R}^{T} \widehat{\xi}_{t+h}\left(r\right) \end{bmatrix}$$

$$\Rightarrow \begin{pmatrix} \left[\Psi\left(\tau,r\right) - \tau\Psi\left(1,r\right)\right] \\ \Psi\left(1,r\right) \end{pmatrix} = \begin{pmatrix} \Psi^{\circ}\left(\tau,r\right) \\ \Psi\left(1,r\right) \end{pmatrix}.$$

Thus, by the Continuous Mapping theorem and Theorem 1, $\left[R_P(\tau)F_P\left(\tau,r\right)\right]' \left[R_P(\tau)F_P\left(\tau,r\right)\right] \Rightarrow \Psi^{\circ}\left(\tau,r\right)' \Psi^{\circ}\left(\tau,r\right) + \Psi\left(1,r\right)' \Psi\left(1,r\right).$ ∎

References

Amisano, G. and R. Giacomini, 2007, Comparing density forecasts via weighted likelihood ratio tests. Journal of Business and Economic Statistics 25(2), 177-190.

Andrade, P., E. Ghysels and J. Idier, 2012, Tails of inflation forecasts and tales of monetary policy. Banque de France Working Papers 407.

Andrews, D.W.K., 1993, Tests for parameter instability and structural change with unknown change point. Econometrica 61(4), 821-856.

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.

Bai, J., 2003, Testing parametric conditional distributions of dynamic models. Review of Economics and Statistics 85(3), 531-549.

Corradi, V. and N.R. Swanson, 2006a, Bootstrap conditional distribution tests in the presence of dynamic misspecification. Journal of Econometrics 133(2), 779-806.

Corradi, V. and N.R. Swanson, 2006b, Predictive Density Evaluation, in: G. Elliott, C.W.J. Granger and A. Timmermann, (Eds.), Handbook of Economic Forecasting, Vol. 1, Elsevier, pp. 197-284.

Corradi, V. and N.R. Swanson, 2006c, Predictive density and conditional confidence interval accuracy tests. Journal of Econometrics 135(1–2), 187-228.

Corradi, V. and N.R. Swanson, 2007, Evaluation of dynamic stochastic general equilibrium models based on distributional comparison of simulated and historical data. Journal of Econometrics 136(2), 699-723.

Diks, C., V. Panchenko and D. van Dijk, 2011, Likelihood-based scoring rules for comparing density forecasts in tails. Journal of Econometrics 163(2), 215–230.

Diebold, F.X., T.A. Gunther, and A.S. Tay, 1998, Evaluating density forecasts with applications to financial risk management. International Economic Review 39(4), 863-883.

Diebold F.X., A.S. Tay and K.F. Wallis, 1999, Evaluating density forecasts of inflation: the Survey of Professional Forecasters, in: R.F. Engle and H. White, (Eds.), Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger, Oxford University Press, pp. 76-90.

Hall, P. and P. Heyde, 1980, Martingale limit theory and its application. Academic Press.

Hong, Y.M. and H. Li, 2005, Nonparametric specification testing for continuous time models with applications to term structure of interest rates. Review of Financial Studies 18(1), 37-84.

Inoue, A., 2001, Testing for distributional change in time series. Econometric Theory 17(1), 156-187.

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

Rosenblatt, M., 1952, Remarks on a multivariate transformation. Annals of Mathematical Statistics 23(3), 470-472.

Rossi, B., 2005, Optimal tests for nested model selection with underlying parameter instabilities. Econometric Theory 21(5), 962-990.

van Dijk, D. and P.H. Franses, 2003, Selecting a nonlinear time series model using weighted tests of equal forecast accuracy. Oxford Bulletin of Economics and Statistics 65, 727–744.

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

White, H., 1984, Asymptotic theory for econometricians, Academic Press.

8 Tables and Figures

Table 1. Size Properties

Panel A: Correct Specification Tests Robust to Instabilities

		DGP S1							
		κ_P				C_P			
R	$P:$	1000	500	200	100	1000	500	200	100
1000		0.05	0.04	0.04	0.05	0.04	0.04	0.04	0.05
500		0.04	0.03	0.04	0.04	0.05	0.02	0.02	0.05
200		0.05	0.04	0.04	0.03	0.05	0.04	0.04	0.03
100		0.05	0.05	0.04	0.05	0.05	0.04	0.04	0.04

		DGP S2							
		κ_P				C_P			
R	$P:$	1000	500	200	100	1000	500	200	100
1000		0.09	0.08	0.08	0.10	0.09	0.09	0.08	0.12
500		0.08	0.07	0.08	0.09	0.08	0.07	0.07	0.08
200		0.07	0.08	0.08	0.07	0.08	0.08	0.08	0.08
100		0.09	0.08	0.06	0.09	0.08	0.08	0.08	0.10

Panel B: Correct Specification Tests

		DGP S1							
		κ_P^{CS}				C_P^{CS}			
R	$P:$	1000	500	200	100	1000	500	200	100
1000		0.06	0.05	0.04	0.05	0.05	0.04	0.03	0.06
500		0.06	0.03	0.04	0.04	0.04	0.02	0.03	0.05
200		0.05	0.04	0.04	0.03	0.06	0.04	0.04	0.02
100		0.05	0.05	0.05	0.06	0.04	0.04	0.04	0.04

		DGP S2							
		κ_P^{CS}				C_P^{CS}			
R	$P:$	1000	500	200	100	1000	500	200	100
1000		0.09	0.07	0.07	0.11	0.08	0.08	0.08	0.12
500		0.07	0.05	0.07	0.09	0.08	0.07	0.07	0.08
200		0.07	0.07	0.07	0.07	0.08	0.08	0.08	0.08
100		0.09	0.07	0.07	0.08	0.09	0.08	0.08	0.09

Panel C: Instability Tests

DGP S1

R	$P:$	κ_P^I				C_P^I			
		1000	500	200	100	1000	500	200	100
1000		0.03	0.03	0.05	0.03	0.04	0.03	0.05	0.03
500		0.03	0.03	0.03	0.03	0.02	0.03	0.05	0.04
200		0.04	0.02	0.03	0.02	0.04	0.03	0.04	0.02
100		0.03	0.02	0.02	0.03	0.03	0.02	0.03	0.03

DGP S2

R	$P:$	κ_P^I				C_P^I			
		1000	500	200	100	1000	500	200	100
1000		0.10	0.08	0.10	0.06	0.10	0.08	0.10	0.08
500		0.10	0.09	0.08	0.09	0.08	0.11	0.09	0.11
200		0.10	0.08	0.08	0.08	0.08	0.07	0.07	0.08
100		0.09	0.07	0.07	0.07	0.08	0.09	0.07	0.08

Note. The table reports empirical rejection frequencies for the respective test statistics at the 5% nominal size for various values of P and R. The number of Monte Carlo replications is 500; critical values are simulated with 100 replications; $\tau = [0.15, 0.16, ..., 0.84, 0.85]$ and $r = [0.1, 0.2, ..., 0.8, 0.9]$.

Table 2. Power Properties

	Instability Robust Correct Specification Tests			Correct Specification Tests		Instability Tests	
	κ_P	C_P		κ_P^{CS}	C_P^{CS}	κ_P^I	C_P^I

Panel A: DGP P1

c	κ_P	C_P		κ_P^{CS}	C_P^{CS}	κ_P^I	C_P^I
0	0.04	0.04		0.04	0.04	0.05	0.04
0.5	0.04	0.05		0.04	0.05	0.03	0.04
1	0.13	0.14		0.14	0.15	0.03	0.04
1.5	0.30	0.34		0.33	0.35	0.03	0.04
2	0.50	0.60		0.52	0.61	0.03	0.04
2.5	0.73	0.82		0.77	0.83	0.04	0.04

Panel B: DGP P2

c	κ_P	C_P		κ_P^{CS}	C_P^{CS}	κ_P^I	C_P^I
0	0.04	0.04		0.04	0.04	0.05	0.04
0.25	0.16	0.14		0.12	0.12	0.19	0.24
0.5	0.66	0.47		0.52	0.33	0.88	0.85
0.75	0.97	0.92		0.93	0.75	1.00	1.00
1	1.00	1.00		1.00	0.99	1.00	1.00
1.25	1.00	1.00		1.00	1.00	1.00	1.00

Panel C: DGP P3

T_1/T	κ_P	C_P		κ_P^{CS}	C_P^{CS}	κ_P^I	C_P^I
600	0.04	0.04		0.04	0.04	0.05	0.04
580	0.27	0.29		0.21	0.24	0.33	0.41
560	0.78	0.78		0.63	0.68	0.72	0.79
540	0.96	0.97		0.94	0.96	0.69	0.75

Note. The table reports empirical rejection frequencies for the respective test statistics under the alternatives of DGP P1, DGP P2, and DGP

P3. $R = 500$, $P = 100$. The number of Monte Carlo replications is 500; critical values are simulated with 100 replications;

$\tau = [0.15, 0.16, ..., 0.84, 0.85]$ and $r = [0.1, 0.2, ..., 0.8, 0.9]$.

Table 3: SPF's Mean Probability Forecast Distribution

Series Name:	GDP Growth			GDP Deflator Growth		

Panel A. Forecast Density Tests Robust to Instabilities

Horizon:	κ_P	C_P	Break	κ_P	C_P	Break
0	2.4038*	0.9451*	1985:IV	13.3484*	4.8776*	1979:II
1	0.6177	0.2175	1998:IV	11.3842*	5.0318*	1995:IV

Panel B. Correct Specification Tests

	κ_P^{CS}	C_P^{CS}		κ_P^{CS}	C_P^{CS}	
0	2.3134*	0.8760*	-	12.4986*	4.7087*	-
1	0.3692	0.1281	-	10.2736*	4.8199*	-

Panel C. Instability Tests

|	κ_P^I	C_P^I		κ_P^I	C_P^I	
0	0.7476*	0.0691	2004:I	0.8581*	0.1689*	1982:II
1	0.5838	0.0894	1998:IV	1.7244*	0.2119*	1995:IV

Note. '*' indicates rejections at 5% significance levels. The tests are implemented with $\tau \in [0.15, 0.16, ..., 0.85]$ and $r \in [0.1, 0.2, ..., 0.9]$. The number of Monte Carlo replications used to obtain critical values is 1000. The SPF data samples are as follows: 1968:IV - 2011:IV for inflation nowcast (h=0), 1981:III-2011:IV for output growth nowcast (h=0), 1981:III - 2010:IV for one-year-ahead inflation (h=1), 1981:III-2010:IV for one-year-ahead output growth (h=1).

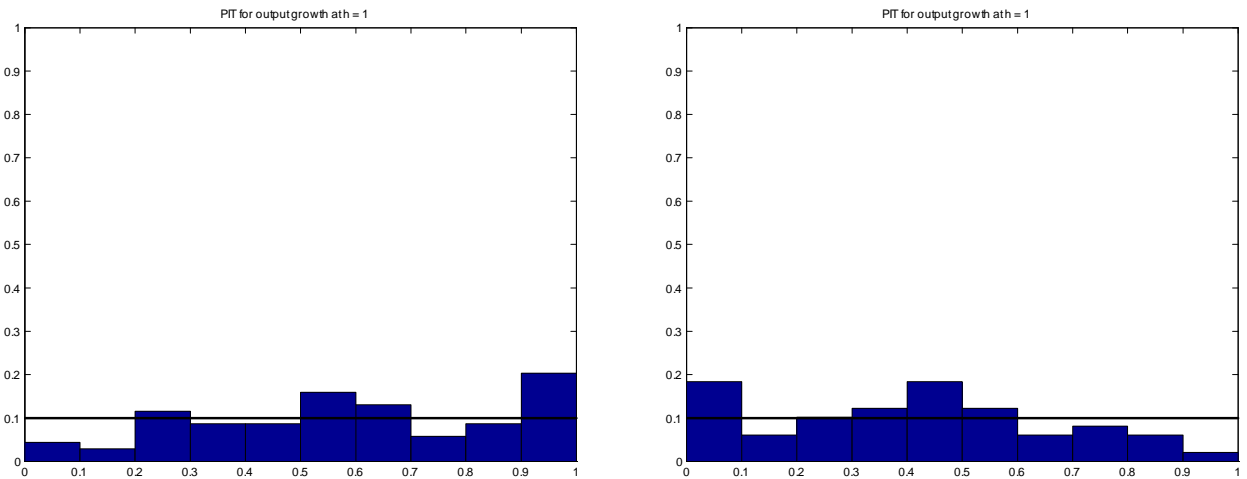## Figure 1: Changes in the Density Forecasts of Output Growth

Panel A: Full Sample (1981:III-2011:IV, 1981:III-2010:IV)
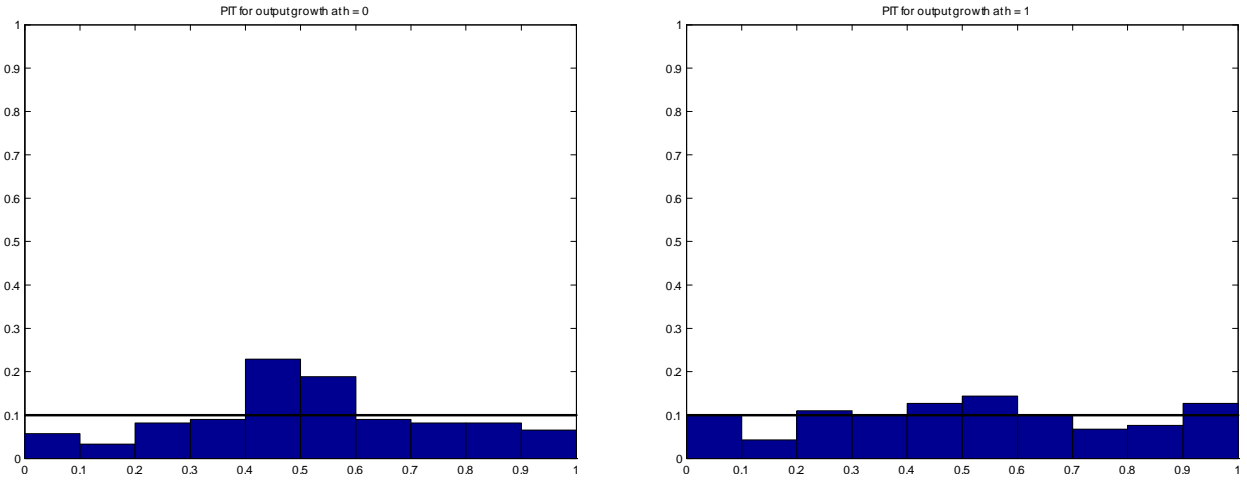
Panel B: Sub-sample Analysis for Nowcast (1981:III-1985:IV, 1986:I-2011:IV)
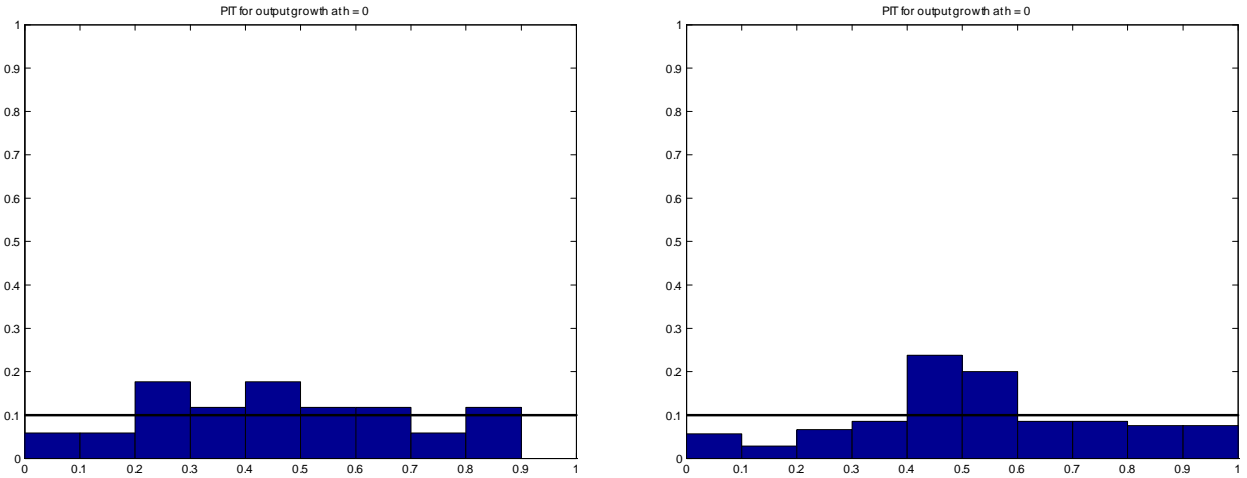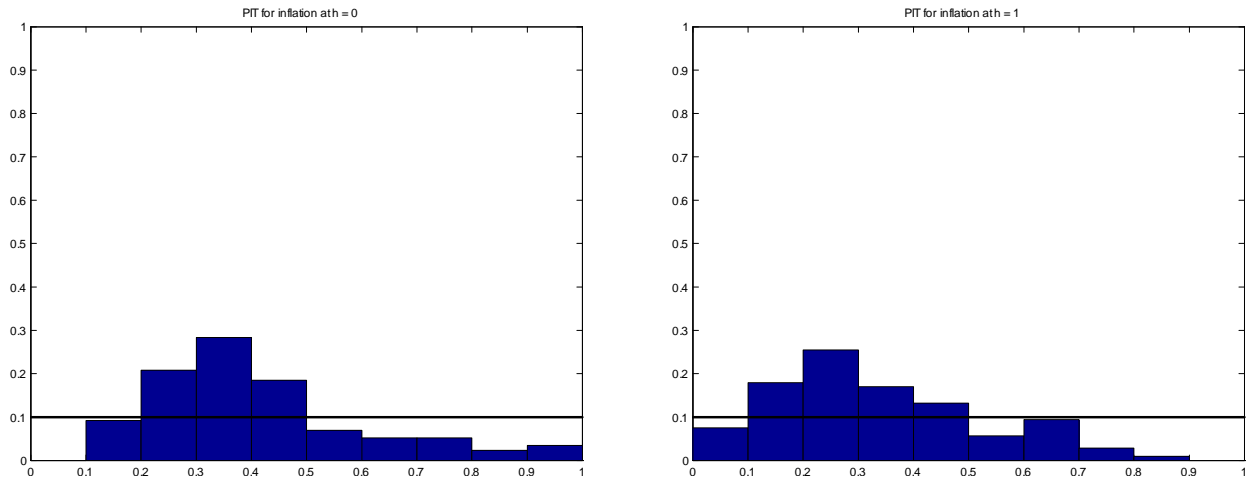

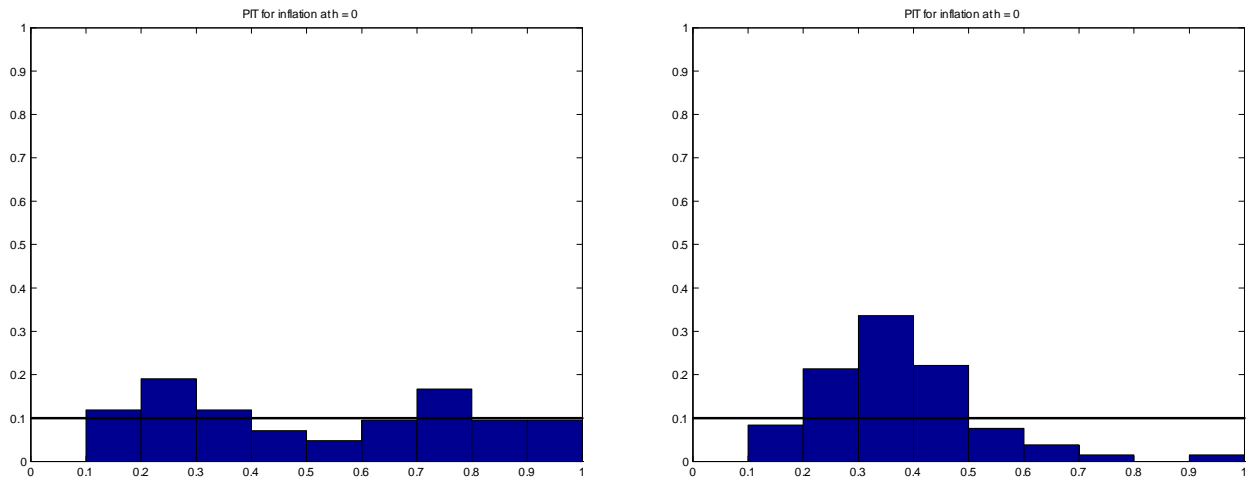### Panel C: Sub-sample Analysis for One-year-ahead Forecasts (1981:III-1998:IV, 1999:I-2010:IV)

Note. The histograms depict the PITS for indicated sample periods. The solid line plots the expected value of a U(0,1) over 10 bins, i.e. 1/10 = 0.1.
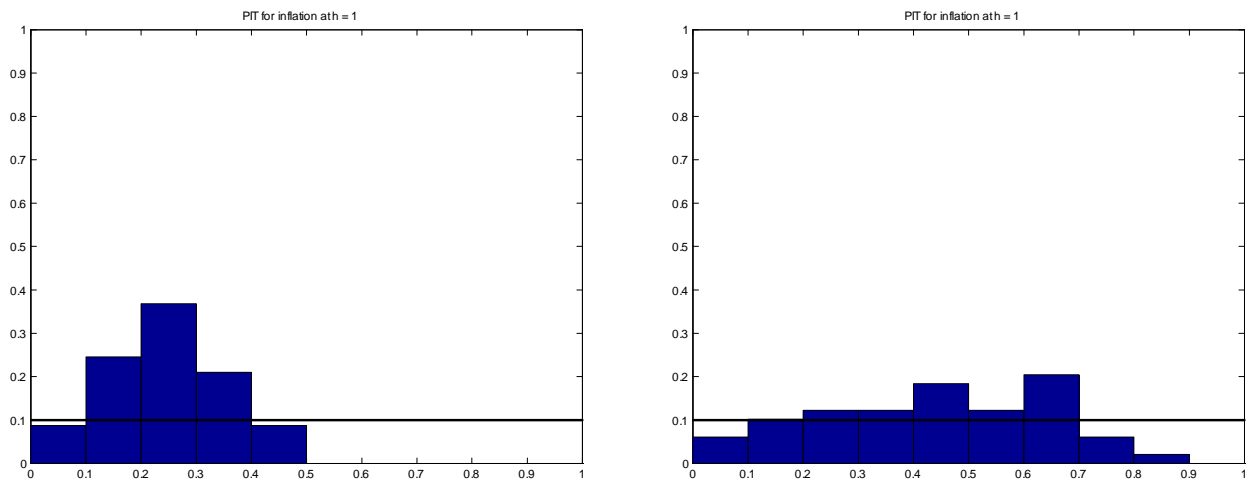
Figure 2: Changes in the Density Forecasts of Inflation

Panel A: Full Sample (1968:IV-2011:IV. 1981:III-2010:IV)

Panel B: Sub-sample Analysis for Nowcast (1968:IV-1979:II, 1979:III-2011:IV)

Panel C: Sub-sample Analysis for One-Year-Ahead Forecast (1981:III-1995:IV, 1996:I-2010:IV)

Note. The histograms depict the PITS for indicated sample periods. The solid line plots the expected value of a U(0,1) over 10 bins, i.e. 1/10 = 0.1.

34