Predictive Density Accuracy Tests*

Valentina Corradi[1] and Norman R. Swanson[2]

[1]Queen Mary, University of London and [2]Rutgers University

February 2004

Abstract

This paper outlines a testing procedure for assessing the relative out-of-sample predictive accuracy of multiple conditional distribution models, and surveys existing related methods in the area of predictive density evaluation, including methods based on the probability integral transform and the Kullback-Leibler Information Criterion. The procedure is closely related to Andrews' (1997) conditional Kolmogorov test and to White's (2000) reality check approach, and involves comparing square (approximation) errors associated with models i, $i = 1, ..., n$, by constructing weighted averages over U of $E\left(\left(F_i(u|Z^t, \theta_i^\dagger) - F_0(u|Z^t, \theta_0)\right)^2\right)$, where $F_0(\cdot|\cdot)$ and $F_i(\cdot|\cdot)$ are true and approximate distributions, $u \in U$, and U is a possibly unbounded set on the real line. Appropriate bootstrap procedures for obtaining critical values for tests constructed using this measure of loss in conjunction with predictions obtained via rolling and recursive estimation schemes are developed. We then apply these bootstrap procedures to the case of obtaining critical values for our predictive accuracy test. A Monte Carlo experiment comparing our bootstrap methods with methods that do not include location bias adjustment terms is provided, and results indicate coverage improvement when our proposed bootstrap procedures are used. Finally, an empirical example comparing alternative predictive densities for U.S. inflation is given.

JEL classification: C22, C51.
Keywords: block bootstrap, rolling and recursive estimation scheme, parameter estimation error, predictive density.

---

* Valentina Corradi, Department of Economics, Queen Mary, University of London, Mile End Road, London E1 4NS, UK, v.corradi@qmul.ac.uk. Norman R. Swanson, Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA, nswanson@econ.rutgers.edu. The authors owe great thanks to Clive W.J. Granger, whose discussions on the matter provided much of the impetus for the research reported in this paper. The authors also wish to thank participants of the conference in honor of Clive W.J. Granger on Predictive Methodology and Application in Economics and Finance for many useful comments. Corradi gratefully acknowledges ESRC grant RES-000-23-0006, and Swanson acknowledges financial support from a Rutgers University Research Council grant.

1 Introduction

In the management of financial risk in the insurance and banking industries, there is often a need for examining confidence intervals or entire conditional distributions. One such case is when value at risk measures are constructed in order to assess the amount of capital at risk from small probability events, such as catastrophes (in insurance markets) or monetary shocks that have large impact on interest rates (see Duffie and Pan (1997) for further discussion). These considerations in part account for the development over the last few years of a new strand of literature addressing the issue of predictive density evaluation. Some of the important recent papers in this area include Diebold, Gunther and Tay (DGT: 1998), Christoffersen (1998), Bai (2003), Diebold, Hahn and Tay (1999), Hong (2001) and Christoffersen, Hahn and Inoue (2001), and Giacomini (2002).[1] This paper has two primary objectives. First, we build on the results of Corradi and Swanson (2003a) by outlining a procedure for assessing the relative out-of-sample predictive accuracy of multiple conditional distribution models that can be used with rolling and recursive estimation schemes. Second, we provide a brief survey of related techniques, such as those based on the use of the probability integral transform and the Kullback-Leibler Information Criterion (KLIC).

The literature on the evaluation of predictive densities is largely concerned with testing the null of correct dynamic specification of an individual conditional distribution model. However, in the literature on the evaluation of point forecast models it is acknowledged that all models in a group that is being evaluated may be misspecified (see e.g. White (2000) and Corradi and Swanson (2002)). In this paper, we draw on elements of these two literatures in order to provide a test for choosing among a group of misspecified out-of-sample predictive density models. Reiterating our above point, the focus of most of the papers cited above is that the density associated with the true conditional distribution is clearly the best predictive density. Therefore, evaluation of predictive densities is usually performed via tests for the correct (dynamic) specification of the conditional distribution. Along these lines, by making use of the probability integral transform, DGT suggest a simple and effective means by which predictive densities can be evaluated. Using the DGT terminology, if $p_t(y_t|\Omega_{t-1})$ is the "true" conditional distribution of $y_t|\Omega_{t-1}$, then $p_t(y_t|\Omega_{t-1})$ is an

[1] Ten years ago, when Clive Granger was asked by one of the authors of this paper in an interview what he thought the most important future areas in time series analysis were, he replied that predictive density construction and evaluation was one of the most critical areas which needed to be developed.

identically and independently distributed uniform random variable on $[0,1]$; so that the difference between an empirical version of $p_t(y_t|\Omega_{t-1})$ constructed using estimated parameters and the 45 degree line can be used as measure of goodness of fit.[2] A feature common to the papers cited above is that the null hypothesis is that of (dynamic) correct specification. Our approach differs from these as we do not assume that any of the competing models (including the benchmark) are correctly specified.[3] Thus, we posit that *all* models should be viewed as approximations of some true unknown underlying data generating process. For this reason, it is our objective in this paper to provide a conditional Kolmogorov test, in the spirit of Andrews (1997), that allows for the joint comparison of multiple misspecified conditional distribution models, for the case of dependent observations. In particular, assume that the object of interest is the conditional distribution of a scalar, Y_{t+1}, given a (possibly vector valued) conditioning set, Z^t, where Z^t contains lags of Y_{t+1} and/or lags other variables. Now, given a group of (possibly) misspecified conditional distributions, $F_1(u|Z^t,\theta_1^\dagger),...,F_m(u|Z^t,\theta_m^\dagger)$, assume that the objective is to compare these models in terms of their closeness to the true conditional distribution, $F_0(u|Z^t,\theta_0)=\Pr(Y_{t+1}\leq u|Z^t)$. If $m>2$, we follow White (2000), in the sense that we choose a particular conditional distribution model as the "benchmark" and test the null hypothesis that no competing model can provide a more accurate approximation of the "true" conditional distribution, against the alternative that at least one competitor outperforms the benchmark model. However, unlike White, we evaluate predictive densities rather than point forecasts. Needless to say, pairwise comparison of alternative models,

[2]Using the same approach, Bai (2003) proposes a Kolmogorov type test based on the comparison of $p_t(y_t|\Omega_{t-1},\widehat{\theta}_T)$ with the CDF of a uniform on $[0,1]$. As a consequence of using estimated parameters, the limiting distribution of his test reflects the contribution of parameter estimation error and is not nuisance parameter free. To overcome this problem, Bai (2003) uses a novel device based on a martingalization argument to construct a modified Kolmogorov test which has a nuisance parameter free limiting distribution. His test has power against violations of uniformity but not against violations of independence. Hong (2001) proposes an interesting test, based on the generalized spectrum, which has power against both uniformity and independence violations, for the case in which the contribution of parameter estimation error vanishes asymptotically. For the case where the null is rejected, Hong (2001) also proposes a test for uniformity that is based on a comparison between a kernel density estimator and the uniform density, and that is robust to non independence (see also Hong and Li (2003)). Diebold, Hahn and Tay (1999) propose a nonparametric correction for improving the density forecast when the uniform (but not the independence) assumption is violated. Finally, Bontemps and Meddahi (2003a,b) suggest a GMM type approach for testing normality and various distributional assumptions.

[3]Corradi and Swanson (2003c) allow for dynamic misspecification under both hypotheses.

in which no benchmark need be specified, follows from our results as a special case. In our context, accuracy is measured using a distributional analog of mean square error. More precisely, the squared (approximation) error associated with model i, $i = 1, ..., m$, is measured in terms of the average over U of $E\left(\left(F_i(u|Z^{t+1}, \theta_i^\dagger) - F_0(u|Z^{t+1}, \theta_0)\right)^2\right)$, where $u \in U$, and U is a possibly unbounded set on the real line.[4] It should be pointed out that one well known measure of distributional accuracy is the Kullback-Leibler Information Criterion (KLIC), in the sense that the "most accurate" model can shown to be that which minimizes the KLIC (see Section 2 for a more precise discussion). Using the KLIC approach, Giacomini (2002) suggests a weighted version of the Vuong (1989) likelihood ratio test for the case of dependent observations, while Kitamura (2002) employs a KLIC based approach to select among misspecified conditional models that satisfy given moment conditions.[5] Furthermore, the KLIC approach has been recently employed for the evaluation of dynamic stochastic general equilibrium models (see e.g. Schörfheide (2000), Fernandez-Villaverde and Rubio-Ramirez (2001), and Chang, Gomes and Schörfheide (2002)). For example, Fernandez-Villaverde and Rubio-Ramirez (2001) show that the KLIC-best model is also the model with the highest posterior probability. In general, there is no reason why our measure of accuracy is more "natural" than the KLIC, or vice-versa. However, in the next section we outline how certain problems (such as comparing conditional confidence intervals) that are difficult to address using the KLIC can be handled quite easily using our measure of distributional accuracy.

The limiting distribution of the suggested statistic turns out to be a functional of a Gaussian process with a covariance kernel reflecting both (dynamic) misspecification and parameter estimation error (PEE). The limiting distribution is not nuisance parameter free and critical values cannot be directly tabulated. Valid asymptotic critical values can be obtained via an empirical version of the block bootstrap which properly takes into account PEE, however. The PEE contribution is summarized by the limiting distribution of $P^{-1/2}\sum_{t=R}^{T-1}\left(\widehat{\theta}_t - \theta^\dagger\right)$, where R denotes the length of the estimation period, P the number of recursively estimated parameters, $\widehat{\theta}_t$ is either a recursive $m-$estimator constructed using the first t observations or a rolling $m-$estimator constructed using observations from $t - R + 1$ to t, and $\theta^\dagger$ is its probability limit. Intuitively, in the recursive case,

[4] To the best of our knowledge, the only other papers in which this measure is considered are Corradi and Swanson (2003a,b).

[5] Of note is that White (1982) shows that quasi maximum likelihood estimators (QMLEs) minimize the KLIC, under mild conditions.

earlier observations are used more frequently than temporally subsequent observations, while in the rolling case, observations in the center of the sample are used more frequently than observations either at the beginning or at the end of the sample. This introduces a location bias to the usual block bootstrap, as under standard resampling with replacement schemes, any block from the original sample has the same probability of being selected.[6] We consider two solutions to this problem. First, we modify the usual resampling scheme and add an adjustment term which corrects for the bootstrap location bias. Second, we retain the usual resampling scheme, but add additional adjustment terms to those needed when our modified resampling scheme is used. Additionally, we consider cases in which all parameters are jointly estimated as well as cases where the conditional mean parameters are first estimated via OLS or NLS, and the error variance is subsequently estimated using the residuals from the conditional mean model.[7] In order to assess the usefulness of our bootstrap procedures, we carry out a series of Monte Carlo experiments evaluating finite sample coverage probabilities of our "PEE" bootstraps for rolling and recursive estimation schemes with analogous bootstrap methods that do not include our "adjustment" terms. Results indicate that the adjustment terms lead to improved coverage probabilities. Thus, our procedures should prove useful for constructing critical values for our predictive density accuracy tests.

The rest of the paper is organized as follows. Section 2 outlines the setup, presents the predictive density accuracy test, and states the asymptotic properties of the test statistic for both the case of recursive and rolling parameter estimation schemes. Section 3 is broken into four subsections. The first two subsections outline bootstrap procedures for mimicking the limiting distribution of parameter estimation error in rolling estimation schemes, while the third subsection summarizes the results of Corradi and Swanson (2003a) for recursive estimation schemes. Finally, the fourth subsection applies the results of the previous two subsections in order to obtain asymptotically valid critical values for the predictive density accuracy test. Section 4 contains the results of a

[6]Note that in the fixed sampling scheme, we just need to take into account the contribution of $\sqrt{R}\left(\widehat{\theta}_R - \theta^{\dagger}\right)$, whose limiting distribution is properly captured by "standard" block bootstrap techniques, using for example the results of Goncalves and White (2003). This case has been considered by Corradi and Swanson (2003b), within the context of in sample evaluation of conditional misspecified distribution models.

[7]From a theoretical perspective, it should be noted that all of our rolling estimation scheme results are new to this paper. Additionally, our recursive estimation scheme results for the case where parameters are estimated sequentially are new, while those for the joint estimation case summarize previous results reported in Corradi and Swanson (2003a).

small Monte Carlo study of the bootstrap procedures developed in the paper, in particular (i) we compare the relative coverage probabilities for recursive and rolling schemes, and (ii) we evaluate the importance of the adjustment term in our bootstrap. In Section 5, an empirical example based on predicting U.S. inflation is presented. Finally, concluding remarks are gathered in Section 6. All proofs are in an appendix. Hereafter, P^* denotes the probability law governing the resampled series, conditional on the sample, E^* and Var^* the mean and variance operators associated with P^*, $o^*_P(1)$ $\Pr -P$ denotes a term converging to zero in P^*-probability, conditional on the sample except a subset of probability measure approaching zero, and finally $O^*_P(1)$ $\Pr -P$ denotes a term which is bounded in P^*-probability, conditional on the sample except a subset of probability measure approaching zero.

2 Predictive Density Evaluation

Our objective is to "choose" a conditional distribution model that provides the most accurate out-of-sample approximation of the true conditional distribution, given multiple predictive densities, and allowing for misspecification under both the null and the alternative hypotheses. One strategy that yields tests of the null of correct specification that are equally as useful as those discussed above is the conditional Kolmogorov test approach of Andrews (1997), which is based on a direct comparison of empirical joint distributions with the product of parametric conditional and nonparametric marginal distributions. Corradi and Swanson (2003c) extend Andrews (1997) in order to allow for the in-sample comparison of multiple misspecified models. As discussed above, one of our main objectives in this paper is the extension of those results to out-of-sample predictive density evaluation in the context of various different estimation schemes. From the perspective of prediction, we assume that the objective is to form parametric conditional distributions for a scalar random variable, y_{t+1}, given Z^t, and to select among these, where $Z^t = (y_t, ..., y_{t-s_1+1}, X_t, ..., X_{t-s_2+1})$, $t = s, ..., \widetilde{T}, ...\widetilde{T} + s$, with $s = \max\{s_1, s_2\}$, and $\widetilde{T} + s = T$, with $T = (s + R) + P$. Assume that $i = 1, ..., n$ different models are estimated. In order to examine *rolling estimation schemes*, define the *rolling* m-estimator for the parameter vector associated with model i as:

$$\widehat{\theta}_{i,t,R} = \arg\max_{\theta_i \in \Theta_i} \frac{1}{R} \sum_{j=t-R+1}^{t} \ln f_i(y_j, Z^{j-1}, \theta_i), \quad R + s \leq t \leq T - 1,\ i = 1, ..., n \tag{1}$$

and

$$\theta_i^{\dagger} = \arg\max_{\theta_i \in \Theta_i} E(\ln f_i(y_j, Z^{j-1}, \theta_i)), \tag{2}$$

where $f_i(\cdot|\cdot,\theta_i)$ is the conditional density associated with $F_i(\cdot|\cdot)$, $i = 1, ..., n$, so that $\theta_i^{\dagger}$ is the probability limit of a quasi maximum likelihood estimator (QMLE). If model i is correctly specified, then $\theta_i^{\dagger} = \theta_0$. We compute a sequence a P estimators, first using observations from $s+1$ to $R+s$, then from to $s+2$ to $R+s+1$, and so on until we use the last R observations, that is from $P+s$ to $T-1$. These estimators are then used to construct sequences of P 1-step ahead forecasts and associated forecast errors, for example. In the context of such rolling estimators, it is necessary to distinguish between the cases of $P \leq R$ and $P > R$, as we shall see below. The rolling and recursive estimation schemes defined above are commonly used in out of sample forecast evaluation (see e.g. West (1996), West and McCracken (1998), Clark and McCracken (2001 and 2003)). Notably exceptions are Giacomini and White (2003), who propose to use a rolling scheme with a fixed window, not increasing with the sample size, so that estimated parameters are treated as mixing variables, and Pesaran and Timmerman (2003), who, in order to take account possible structure breaks, suggest an adaptive manner for choosing the window of observations.

We also consider *recursive estimation schemes*, for which we define the *recursive* m-estimator for the parameter vector associated with model i as:

$$\widehat{\theta}_{i,t} = \arg\max_{\theta_i \in \Theta_i} \frac{1}{t} \sum_{j=s}^{t} \ln f_i(y_j, Z^{j-1}, \theta_i), \quad R+s \leq t \leq T-1,\ i = 1, ..., n \tag{3}$$

and $\theta_i^{\dagger}$ defined as in (2). Again following standard practice, this estimator is first computed using observations from $s+1$ to $R+s$ observations, and then from $s+1$ to $R+1+1$ observations, and so on until the last estimator is constructed using $T-1-s$ observations. As previously, these estimators are then used to construct sequences of P 1-step ahead forecasts and associated forecast errors.

Now, define the group of conditional distribution models from which we want to make a selection as $F_1(u|Z^t, \theta_1^{\dagger}), ..., F_n(u|Z^t, \theta_n^{\dagger})$, and define the true conditional distribution as $F_0(u|Z^t, \theta_0) = \Pr(y_{t+1} \leq u|Z^t)$. In the sequel, $F_1(\cdot|\cdot, \theta_1^{\dagger})$ is taken as the benchmark model, and the objective is to test whether some competitor model can provide a more accurate approximation of $F_0(\cdot|\cdot, \theta_0)$ than the benchmark.[8]

[8] In this test, the competing models are known. This is different than the probability integral transform approach

Following Corradi and Swanson (2003a), we begin by assuming that accuracy is measured using a distributional analog of mean square error. More precisely, the squared (approximation) error associated with model i, $i = 1, ..., n$, is measured in terms of the average over U of $E\left(\left(F_i(u|Z^t, \theta_i^\dagger) - F_0(u|Z^t, \theta_0)\right)^2\right)$, where $u \in U$, and U is a possibly unbounded set on the real line.

In particular, we say that model 1 is more accurate than model 2, if

$$\int_U E\left(\left(F_1(u|Z^t, \theta_1^\dagger) - F_0(u|Z^t, \theta_0)\right)^2 - \left(F_2(u|Z^t, \theta_2^\dagger) - F_0(u|Z^t, \theta_0)\right)^2\right)\phi(u)du < 0,$$

where $\int_U \phi(u)du = 1$ and $\phi(u) \geq 0$, for all $u \in U \subset \Re$. For any given evaluation point, this measure defines a norm and it implies a standard goodness of fit measure.

As mentioned above, another measure of distributional accuracy available in the literature is the KLIC (see e.g. White (1982), Vuong (1989), Giacomini (2002), and Kitamura (2002)), according to which we should choose Model 1 over Model 2 if

$$E(\log f_1(Y_t|Z^t, \theta_1^\dagger) - \log f_2(Y_t|Z^t, \theta_2^\dagger)) > 0.$$

The KLIC is a sensible measure of accuracy, as it chooses the model which on average gives higher probability to events which have actually occurred. Also, it leads to simple likelihood ratio type tests. Interestingly, Fernandez-Villaverde and Rubio-Ramirez (2001) have shown that the best model under the KLIC is also the model with the highest posterior probability. Although our approach and the KLIC approach should perhaps be viewed as alternatives, and as such one might want to implement both tests in some contexts, it should be noted that if we are interested in measuring accuracy over a specific region, or in measuring accuracy for a given conditional confidence interval, say, this cannot be done in a straightforward manner using the KLIC, while it can easily be done using our measure. For example, if we want to evaluate the accuracy of different models for approximating the probability that the rate of inflation tomorrow, given the rate of inflation today, will be between 0.5% and 1.5%, say, we can do so quite easily using the square error criterion, but not using the KLIC.

The hypotheses of interest are:

$$H_0 : \max_{k=2,\ldots,n} \int_U E\left(\left(F_1(u|Z^t, \theta_1^\dagger) - F_0(u|Z^t, \theta_0)\right)^2 - \left(F_k(u|Z^t, \theta_k^\dagger) - F_0(u|Z^t, \theta_0)\right)^2\right)\phi(u)du \leq 0 \tag{4}$$

---

where only the null model is explicitly stated.

versus

$$H_A : \max_{k=2,\ldots,n} \int_U E\left(\left(F_1(u|Z^t,\theta_1^\dagger) - F_0(u|Z^t,\theta_0)\right)^2 - \left(F_k(u|Z^t,\theta_k^\dagger) - F_0(u|Z^t,\theta_0)\right)^2\right)\phi(u)du > 0, \tag{5}$$

where $\phi(u) \geq 0$ and $\int_U \phi(u) = 1$, $u \in U \subset \Re$, U possibly unbounded. Note that for a given u, we compare conditional distributions in terms of their (mean square) distance from the true distribution. We then average over U.[9] The statistic is:

$$Z_{P,j} = \max_{k=2,\ldots,n} \int_U Z_{P,u,j}(1,k)\phi(u)du, \; j = 1,2 \tag{6}$$

where for $j = 1$ (rolling estimation scheme),

$$Z_{P,u,1}(1,k) = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\left(1\{y_{t+1} \leq u\} - F_1(u|Z^t,\widehat{\theta}_{1,t,R})\right)^2 - \left(1\{y_{t+1} \leq u\} - F_k(u|Z^t,\widehat{\theta}_{k,t,R})\right)^2\right) \tag{7}$$

and for $j = 2$ (recursive estimation scheme),

$$Z_{P,u,2}(1,k) = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\left(1\{y_{t+1} \leq u\} - F_1(u|Z^t,\widehat{\theta}_{1,t})\right)^2 - \left(1\{y_{t+1} \leq u\} - F_k(u|Z^t,\widehat{\theta}_{k,t})\right)^2\right), \tag{8}$$

[9] If interest focuses on predictive conditional confidence intervals (see e.g. Christoffersen (1998)), so that the objective is to "approximate" $\Pr(\underline{u} \leq y_{t+1} \leq \overline{u}|Z^t)$, then the null and alternative hypotheses can be stated as:

$$H_0' : \max_{k=2,\ldots,n} E\left(\left(\left(F_1(\overline{u}|Z^t,\theta_1^\dagger) - F_1(\underline{u}|Z^t,\theta_1^\dagger)\right) - \left(F_0(\overline{u}|Z^t,\theta_0) - F_0(\underline{u}|Z^t,\theta_0)\right)\right)^2 - \left(\left(F_k(\overline{u}|Z^t,\theta_k^\dagger) - F_k(\underline{u}|Z^t,\theta_k^\dagger)\right) - \left(F_0(\overline{u}|Z^t,\theta_0) - F_0(\underline{u}|Z^t,\theta_0)\right)\right)^2\right) \leq 0.$$

versus

$$H_A' : \max_{k=2,\ldots,n} E\left(\left(\left(F_1(\overline{u}|Z^t,\theta_1^\dagger) - F_1(\underline{u}|Z^t,\theta_1^\dagger)\right) - \left(F_0(\overline{u}|Z^t,\theta_0) - F_0(\underline{u}|Z^t,\theta_0)\right)\right)^2 - \left(\left(F_k(\overline{u}|Z^t,\theta_k^\dagger) - F_k(\underline{u}|Z^t,\theta_k^\dagger)\right) - \left(F_0(\overline{u}|Z^t,\theta_0) - F_0(\underline{u}|Z^t,\theta_0)\right)\right)^2\right) > 0.$$

Analogously, if interest focuses on testing the null of equal accuracy of only two predictive conditional distribution models, say F_1 and F_k, Diebold-Mariano (1995) type test, we can simply state the hypotheses as:

$$H_0'' : \int_U E\left(\left(F_1(u|Z^t,\theta_1^\dagger) - F_0(u|Z^t,\theta_0)\right)^2 - \left(F_k(u|Z^t,\theta_k^\dagger) - F_0(u|Z^t,\theta_0)\right)^2\right)\phi(u)du = 0$$

versus

$$H_A'' : \int_U E\left(\left(F_1(u|Z^t,\theta_1^\dagger) - F_0(u|Z^t,\theta_0)\right)^2 - \left(F_k(u|Z^t,\theta_k^\dagger) - F_0(u|Z^t,\theta_0)\right)^2\right)\phi(u)du \neq 0.$$

where $\widehat{\theta}_{i,t,R}$ and $\widehat{\theta}_{i,t}$ are defined as in (1) and in (3).

In Corradi and Swanson (2003b), we show how the hypotheses above can be restated as

$$H_0 : \max_{k=2,\ldots,n} \int_U \left(\mu_1^2(u) - \mu_k^2(u)\right) \phi(u)du \leq 0$$

versus

$$H_A : \max_{k=2,\ldots,n} \int_U \left(\mu_1^2(u) - \mu_k^2(u)\right) \phi(u)du > 0,$$

where $\mu_i^2(u) = E\left(\left(1\{y_t \leq u\} - F_i(u|Z^t, \theta_i^\dagger)\right)^2\right)$. In the sequel, we require the following assumptions.

Assumption A1: (y_t, X_t), with y_t scalar and X_t an $R^\zeta-$valued $(0 < \zeta < \infty)$ vector, is a strictly stationary and absolutely regular $\beta-$mixing process with size $-4(4+\psi)/\psi$, $\psi > 0$.

Assumption A2: (i) $\theta_i^\dagger$ is uniquely identified (i.e. $E(\ln f_i(y_t, Z^{t-1}, \theta_i)) < E(\ln f_i(y_t, Z^{t-1}, \theta_i^\dagger))$ for any $\theta_i \neq \theta_i^\dagger$); (ii) $\ln f_i$ is twice continuously differentiable on the interior of Θ_i, for $i = 1, ..., n$, and for Θ_i a compact subset of $R^{\varrho(i)}$; (iii) the elements of $\nabla_{\theta_i} \ln f_i$ and $\nabla^2_{\theta_i} \ln f_i$ are $p-$dominated on Θ_i, with $p > 2(2+\psi)$, where ψ is the same positive constant as defined in Assumption A1; and (iii) $E\left(-\nabla^2_{\theta_i} \ln f_i(\theta_i)\right)$ is positive definite uniformly on Θ_i.[10]

Assumption A3: $T = R + P$, and as $T \to \infty$, $P/R \to \pi$, with $0 < \pi < \infty$.

Assumption A4: (i) $F_i(u|Z^t, \theta_i)$ is continuously differentiable on the interior of Θ_i and $\nabla_{\theta_i} F_i(u|Z^t, \theta_i^\dagger)$ is $2r$-dominated on Θ_i, uniformly in u, $r > 2$, $i = 1, ..., n$;[11] and (ii) let $v_{kk}(u) = \text{plim}_{T\to\infty}$ $Var\left(\frac{1}{\sqrt{T}}\sum_{t=s}^{T}\left(\left(\left(1\{y_{t+1} \leq u\} - F_1(u|Z^t, \theta_1^\dagger)\right)^2 - \mu_1^2(u)\right) - \left(\left(1\{y_{t+1} \leq u\} - F_k(u|Z^t, \theta_k^\dagger)\right)^2 - \mu_k^2(u)\right)\right)\right)$, $k = 2, ..., n$, define analogous covariance terms, $v_{j,k}(u)$, $j, k = 2, ..., n$, and assume that $[v_{j,k}(u)]$ is positive semi-definite, uniformly in u.

Assumptions A1 and A2 are standard memory, moment, smoothness and identifiability conditions. A1 requires (y_t, X_t) to be strictly stationary and absolutely regular. The memory condition is stronger than $\alpha-$mixing, but weaker than (uniform) $\phi-$mixing. Assumption A3 requires that R and P grow at the same rate. Of course, if R grows faster than P, then $\Psi_{R,P,i}$ and $\Theta_{R,P,i}$, $i = 1, 2, 3$ (as defined below) vanish in probability, and there is no need to capture the contribution of parameter estimation error when constructing bootstrap critical values for predictive accuracy tests such

[10] We say that $\nabla_{\theta_i} \ln f_i(y_t, Z^{t-1}, \theta_i)$ is $2r-$dominated on Θ_i if its $j-th$ element, $j = 1, ..., \varrho(i)$, is such that $\left|\nabla_{\theta_i} \ln f_i(y_t, Z^{t-1}, \theta_i)\right|_j \leq D_t$, and $E(|D_t|^{2r}) < \infty$. For more details on domination conditions, see Gallant and White (1988, pp. 33).

[11] We require that for $j = 1, ..., p_i$, $\left(E\left(\nabla_\theta F_i(u|Z^t, \theta_i^\dagger)\right)\right)_j \leq D_t(u)$, with $\sup_t \sup_{u\in\Re} E(D_t(u)^{2r}) < \infty$.

as those discussed in the sequel. Assumptions A4(i) states standard smoothness and domination conditions imposed on the conditional distributions of the models, and assumption A4(ii) states that at least one of the competing models, $F_2(\cdot|\cdot,\theta_1^\dagger),...,F_n(\cdot|\cdot,\theta_n^\dagger)$, has to be nonnested with (and non nesting) the benchmark.

Proposition 1: Let Assumptions A1-A4 hold. Then,

$$\max_{k=2,\dots,n}\int_U \left(Z_{P,u,j}(1,k)-\sqrt{P}\left(\mu_1^2(u)-\mu_k^2(u)\right)\right)\phi_U(u)du \xrightarrow{d} \max_{k=2,\dots,n}\int_U Z_{1,k,j}(u)\phi_U(u)du,$$

where $Z_{1,k,j}(u)$ is a zero mean Gaussian process with covariance $C_{k,j}(u,u')$, $j=1$ corresponding to the rolling and $j=2$ to the recursive estimation scheme, equal to:

$$E\left(\sum_{j=-\infty}^{\infty}\left(\left(1\{y_{s+1}\leq u\}-F_1(u|Z^s,\theta_1^\dagger)\right)^2-\mu_1^2(u)\right)\left(\left(1\{y_{s+j+1}\leq u'\}-F_1(u'|Z^{s+j},\theta_1^\dagger)\right)^2-\mu_1^2(u')\right)\right)$$

$$+E\left(\sum_{j=-\infty}^{\infty}\left(\left(1\{y_{s+1}\leq u\}-F_k(u|Z^s,\theta_k^\dagger)\right)^2-\mu_k^2(u)\right)\left(\left(1\{y_{s+j+1}\leq u'\}-F_k(u'|Z^{s+j},\theta_k^\dagger)\right)^2-\mu_k^2(u')\right)\right)$$

$$-2E\left(\sum_{j=-\infty}^{\infty}\left(\left(1\{y_{s+1}\leq u\}-F_1(u|Z^s,\theta_1^\dagger)\right)^2-\mu_1^2(u)\right)\left(\left(1\{y_{s+j+1}\leq u'\}-F_k(u'|Z^{s+j},\theta_k^\dagger)\right)^2-\mu_k^2(u')\right)\right)$$

$$+4\Pi_j m_{\theta_1^\dagger}(u)'A(\theta_1^\dagger)E\left(\sum_{j=-\infty}^{\infty}\nabla_{\theta_1}\ln f_1(y_{s+1}|Z^s,\theta_1^\dagger)\nabla_{\theta_1}\ln f_1(y_{s+j+1}|Z^{s+j},\theta_1^\dagger)'\right)A(\theta_1^\dagger)m_{\theta_1^\dagger}(u')$$

$$+4\Pi_j m_{\theta_k^\dagger}(u)'A(\theta_k^\dagger)E\left(\sum_{j=-\infty}^{\infty}\nabla_{\theta_k}\ln f_k(y_{s+1}|Z^s,\theta_k^\dagger)\nabla_{\theta_k}\ln f_k(y_{s+j+1}|Z^{s+j},\theta_k^\dagger)'\right)A(\theta_k^\dagger)m_{\theta_k^\dagger}(u')$$

$$-4\Pi_j m_{\theta_1^\dagger}(u,)'A(\theta_1^\dagger)E\left(\sum_{j=-\infty}^{\infty}\nabla_{\theta_1}\ln f_1(y_{s+1}|Z^s,\theta_1^\dagger)\nabla_{\theta_k}\ln f_k(y_{s+j+1}|Z^{s+j},\theta_k^\dagger)'\right)A(\theta_k^\dagger)m_{\theta_k^\dagger}(u')$$

$$-4C\Pi_j m_{\theta_1^\dagger}(u)'A(\theta_1^\dagger)E\left(\sum_{j=-\infty}^{\infty}\nabla_{\theta_1}\ln f_1(y_{s+1}|Z^s,\theta_1^\dagger)\left(\left(1\{y_{s+j+1}\leq u\}-F_1(u|Z^{s+j},\theta_1^\dagger)\right)^2-\mu_1^2(u)\right)\right)$$

$$+4C\Pi_j m_{\theta_1^\dagger}(u)'A(\theta_1^\dagger)E\left(\sum_{j=-\infty}^{\infty}\nabla_{\theta_1}\ln f_1(y_{s+1}|Z^s,\theta_1^\dagger)\left(\left(1\{y_{s+j+1}\leq u\}-F_k(u|Z^{s+j},\theta_k^\dagger)\right)^2-\mu_k^2(u)\right)\right)$$

$$-4C\Pi_j m_{\theta_k^\dagger}(u)' A(\theta_k^\dagger) E\left(\sum_{j=-\infty}^{\infty} \nabla_{\theta_k} \ln f_k(y_{s+1}|Z^s,\theta_k^\dagger)' \left(\left(1\{y_{s+j+1} \leq u\} - F_k(u|Z^{s+j},\theta_k^\dagger)\right)^2 - \mu_k^2(u)\right)\right)$$
$$+4C\Pi_j m_{\theta_k^\dagger}(u)' A(\theta_k^\dagger) E\left(\sum_{j=-\infty}^{\infty} \nabla_{\theta_k} \ln f_k(y_{s+1}|Z^s,\theta_k^\dagger)' \left(\left(1\{y_{s+j+1} \leq u\} - F_1(u|Z^{s+j},\theta_1^\dagger)\right)^2 - \mu_1^2(u)\right)\right) \tag{9}$$

with $m_{\theta_i^\dagger}(u)' = E\left(\nabla_{\theta_i} F_i(u|Z^t,\theta_i^\dagger)'\left(1\{y_{t+1} \leq u\} - F_i(u|Z^t,\theta_i^\dagger)\right)\right)$ and $A(\theta_i^\dagger) = A_i^\dagger = \left(E\left(-\nabla_{\theta_i}^2 \ln f_i(y_{t+1}|Z^t,\theta_i^\dagger)\right)\right)^{-1}$, and for $j=1$ and $P \leq R$, $\Pi_1 = \left(\pi - \frac{\pi^2}{3}\right)$, $C\Pi_1 = \frac{\pi}{2}$, and for $P > R$, $\Pi_1 = \left(1 - \frac{1}{3\pi}\right)$ and $C\Pi_1 = \left(1 - \frac{1}{2\pi}\right)$, finally for $j=2$, $\Pi_2 = 2\left(1 - \pi^{-1}\ln(1+\pi)\right)$ and $C\Pi_2 = 0.5\Pi_2$.

From this proposition, we see that when all competing models provide an approximation to the true conditional distribution that is as (mean square) accurate as that provided by the benchmark (i.e. when $\int_U \left(\mu_1^2(u) - \mu_k^2(u)\right)\phi(u)du = 0, \forall k$), then the limiting distribution is a zero mean Gaussian process with a covariance kernel which is not nuisance parameters free. Additionally, when all competitor models are worse than the benchmark, the statistic diverges to minus infinity at rate $\sqrt{P}$. Finally, when only some competitor models are worse than the benchmark, the limiting distribution provides a conservative test, as Z_P will always be smaller than $\max_{k=2,\ldots,n} \int_U \left(Z_{P,u}(1,k) - \sqrt{P}\left(\mu_1^2(u) - \mu_k^2(u)\right)\right)\phi(u)du$, asymptotically. Of course, when H_A holds, the statistic diverges to plus infinity at rate $\sqrt{P}$.

3 Bootstrap Critical Values

In this section we begin by outlining bootstrap methods for mimicking the limiting distribution of $\frac{1}{\sqrt{P}}\sum_{t=R+s}^{T-1}\left(\widehat{\theta}_{i,t,R} - \theta^\dagger\right)$ and $\frac{1}{\sqrt{P}}\sum_{t=R+s}^{T-1}\left(\widehat{\theta}_{i,t} - \theta^\dagger\right)$ where $\widehat{\theta}_{i,t,R}$ and $\widehat{\theta}_{i,t}$ are the rolling and recursive estimators as defined in (1) and (3). For fixed sampling schemes, the properties of the block bootstrap for $m-$estimators and/or GMM estimators with dependent observations have been studied by several authors. For example, Hall and Horowitz (1996) and Andrews (2002a,b) show that the block bootstrap provides improved critical values, in the sense of asymptotic refinements, for "studentized" GMM estimators and for tests of overidentifying restrictions, in the case where the covariance across moment conditions is zero after a given number of lags. In addition, Inoue and Shintani (2003) show that the block bootstrap provides asymptotic refinements for linear overidentified GMM estimators for general mixing processes. A recent contribution which is useful in our context is that of Goncalves and White (2003), who show that for $m-$estimators, the limiting distribution of $\sqrt{T}(\widehat{\theta}^*_{i,T} - \widehat{\theta}_{i,T})$ provides a valid first order approximation to that of $\sqrt{T}(\widehat{\theta}_{i,T} - \theta^\dagger_i)$ for heterogeneous and near epoch dependent series, where $\widehat{\theta}^*_{i,T}$ is a resampled estimator, and T denotes the length of the entire sample. Based on the results mentioned above, one might expect $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}^*_{t,R} - \widehat{\theta}_{t,R}\right)$ to have the same limiting distribution as $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_{t,R} - \theta^\dagger\right)$ and similarly for the recursive case. However, in the rolling case, observations in the middle of the sample are used more frequently than observation at either the beginning or the end of the sample, while in the recursive case, earlier observations are used more frequently than temporally subsequent observations. This introduces a location bias to the usual block bootstrap, as under standard resampling with replacement, any block from the original sample has the same probability of being selected. Also, the bias term varies across samples and can be either positive or negative, depending on the specific sample. In both the rolling and recursive scheme, we circumvent the problem of bootstrap location bias by first slightly modifying the resampling scheme, and then by adding a proper correction term that offsets the bootstrap bias.

3.1 A Split Sample Block Bootstrap for PEE: Rolling Estimation Scheme

In the rolling estimation scheme, we need to distinguish between the case in which $P \leq R$ and $P > R$. For the time being assume $P \leq R$, we then explain how to modify the resampling procedure for the case of $P > R$. Let $W_t = (y_t, Z^{t-1})$, we first draw b_1 overlapping blocks of length l_1, $b_1 l_1 = P$ from observations $s+1, ..., P+s$, then we draw b_2 overlapping blocks of length l_2, $b_2 l_2 = R + s - P$ from observations $P+s+1, ..., R+s$, and finally b_3 overlapping blocks of length l_3, $b_3 l_3 = (T+s) - (R+s) - 1$ from the last P observations. The first P pseudo observations, $W^*_{s+1}, W^*_{s+2}, ..., W^*_{s+l-1}, ..., W^*_{P+s}$, are equal to $W_{I^1_1}, W_{I^1_1+1}, ..., W_{I^1_1+l_1-1}, ..., W_{I^1_{b_1}+l_1-1}$, where I^1_i, $i = 1, ..., b_1$ are independent uniform random draws on the interval $s+1, ..., P+s-l_1+1$, the following $((R+s)-(P+s))$ observations $W^*_{P+s+1}, W^*_{P+s+2}, ..., W^*_{P+s+l}, ..., W^*_{R+s}$, are equal to $W_{I^2_1}, W_{I^2_1+1}, ..., W_{I^2_1+l_2-1}, ..., W_{I^2_{b_2}+l_2-1}$, where I^2_i, $i = 1, ..., b_2$ are independent uniform random draws from data indexed by $P+s+1, P+s+2, ..., R+s-l_2-1$, and finally the last P observations $W^*_{R+s+1}, W^*_{R+s+2}, ..., W^*_{R+s+l_3}, ..., W^*_{R+s+P-1}$, are equal to $W_{I^3_1}, W_{I^3_1+1}, ..., W_{I^3_1+l_3-1}, ..., W_{I^3_{b_3}+l_3-1}$, where I^3_i, $i = 1, ..., b_3$ are independent uniform random draws from data indexed by $R+s+1, R+s+2, ..., R+s+P-l_3-1$. Thus, conditional on the (entire) sample, the pseudo time series W^*_t,

$t = s, ..., R+s, R+s+1, ..., R+s+P$, consists of $b = b_1 + b_2 + b_3$ asymptotically independent, but non identically distributed blocks of length l_1, l_2 and l_3 respectively. More precisely, each block from $R+s+1, ..., R+s+P-1$ may overlap with any block from say $P+s+1, ..., R+s$ for at most s observations, where s is finite. The case of $P > R$ can be treated in an analogous way, by noting that in this case we first draw b_1 overlapping blocks of length l_1, $b_1 l_1 = R$ from observations $s+1, ..., R+s$, then we draw b_2 overlapping blocks of length l_2, $b_2 l_2 = (P+s)-(R+s)$ from observations $R+s+1, ..., P+s$, and finally b_3 overlapping blocks of length l_3, $b_3 l_3 = (T+s)-(P+s)-1$ from the last R observations. Now, define the rolling bootstrap estimator as,

$$\widehat{\theta}^*_{i,t,R} = \arg\max_{\theta_i \in \Theta_i} \frac{1}{R} \sum_{j=t-R+1}^{t} \ln f_i(y^*_j, Z^{*,j-1}, \theta_i), R+s \leq t \leq T-1,\ i = 1, ..., n. \tag{10}$$

Further, for $P \leq R$, define[12],

$$\begin{aligned}
&\Psi^{*(i)}_{R,P,1} \\
= \;& \frac{1}{\sqrt{P}} \sum_{t=R+s}^{T-1} \left(\widehat{\theta}^*_{i,t,R} - \widehat{\theta}_{i,t,R} \right) + \left(-\frac{1}{T} \sum_{t=s}^{T} \nabla^2_{\theta_i} \ln f_i(y_t, Z^{t-1}, \widehat{\theta}_{i,T}) \right)^{-1} \\
& \times \left[\frac{1}{\sqrt{P}R} \sum_{j=s+1}^{P+s} (j-s) \left(\nabla_{\theta_i} \ln f_i(y_j, Z^{j-1}, \widehat{\theta}_{i,T}) - \frac{1}{P} \sum_{j=s+1}^{P+s} \nabla_{\theta_i} \ln f_i(y_j, Z^{j-1}, \widehat{\theta}_{i,T}) \right) \right. \\
& \left. + \frac{1}{\sqrt{P}R} \sum_{j=R+s+1}^{T-1} (P+s-(j-R)) \left(\nabla_{\theta_i} \ln f_i(y_j, Z^{j-1}, \widehat{\theta}_{i,T}) - \frac{1}{P} \sum_{j=R+s+1}^{T-1} \nabla_{\theta_i} \ln f_i(y_j, Z^{j-1}, \widehat{\theta}_{i,T}) \right) \right]
\end{aligned} \tag{11}$$

[12] Note that in the expression below the average score terms involve using all T observations in constructing $\widehat{\theta}_{i,T}$, but only P observations when forming the average, such as in the terms $\frac{1}{P}\sum_{j=s+1}^{P+s} \nabla_{\theta_i} \ln f_i(y_j, Z^{j-1}, \widehat{\theta}_{i,T})$ and $\frac{1}{P}\sum_{j=R+s+1}^{T-1} \nabla_{\theta_i} \ln f_i(y_j, Z^{j-1}, \widehat{\theta}_{i,T})$. This is done to ensure the terms are not identically zero. Also note that the precise sample period used in these terms is not crucial; it is only crucial that the terms are not identically zero. This is the reason why, here and elsewhere, we sometimes take the sum over the first P observations, sometimes over the last P obervations, etc. Of course, experimentation may ultimately suggest that certain versions of these terms involving particular summands perform better in finite samples than others. This is left to future research, however.

and for $P > R$, define,

$$\begin{aligned}
&\Psi^{*(i)}_{R,P,2} \\
= \; & \frac{1}{\sqrt{P}} \sum_{t=R+s}^{T-1} \left(\widehat{\theta}^{*}_{i,t,R} - \widehat{\theta}_{i,t,R} \right) + \left(-\frac{1}{T} \sum_{t=s}^{T} \nabla^2_{\theta_i} \ln f_i(y_t, Z^{t-1}, \widehat{\theta}_{i,T}) \right)^{-1} \\
& \times \left[\frac{1}{\sqrt{P}R} \sum_{j=s+1}^{R+s} (j-s) \left(\nabla_{\theta_i} \ln f_i(y_j, Z^{j-1}, \widehat{\theta}_{i,T}) - \frac{1}{R} \sum_{j=s+1}^{R+s} \nabla_{\theta_i} \ln f_i(y_j, Z^{j-1}, \widehat{\theta}_{i,T}) \right) \right. \\
& \left. + \frac{1}{\sqrt{P}R} \sum_{j=P+s+1}^{T-1} (R+s-(j-P)) \left(\nabla_{\theta_i} \ln f_i(y_j, Z^{j-1}, \widehat{\theta}_{i,T}) - \frac{1}{R} \sum_{j=P+s+1}^{T-1} \nabla_{\theta_i} \ln f_i(y_j, Z^{j-1}, \widehat{\theta}_{i,T}) \right) \right]
\end{aligned} \tag{12}$$

Proposition 2: Let A1-A3 hold.

(i) Assume that as $P \to \infty$ and $l_1 \to \infty$, $l_1/P^{1/4} \to 0$, and as $R \to \infty$ and $l_3 \to \infty$, $l_3/P^{1/4} \to 0$, and finally as $R - P \to \infty$ and $l_2 \to \infty$, $l_2/(R-P)^{1/4} \to 0$. Then, as $P \to \infty$ and $R \to \infty$, for $P \leq R$,

$$P\left(\omega : \sup_{v \in \Re^{\varrho(i)}} \left| P^*_{R,P}\left(\Psi^{*(i)}_{R,P,1} \leq v \right) - P\left(\frac{1}{\sqrt{P}} \sum_{t=R+s}^{T-1} \left(\widehat{\theta}_{i,t,R} - \theta^{\dagger}_{i} \right) \leq v \right) \right| > \varepsilon \right) \to 0.$$

(ii) Assume that as $R \to \infty$ and $l_1 \to \infty$, that $l_1/R^{1/4} \to 0$, and as $P \to \infty$ and $l_3 \to \infty$, $l_3/R^{1/4} \to 0$, and finally as $P - R \to \infty$ and $l_2 \to \infty$, $l_2/(P-R)^{1/4} \to 0$. Then, as P and $R \to \infty$, for $P > R$,

$$P\left(\omega : \sup_{v \in \Re^{\varrho(i)}} \left| P^*_{R,P}\left(\Psi^{*(i)}_{R,P,2} \leq v \right) - P\left(\frac{1}{\sqrt{P}} \sum_{t=R+s}^{T-1} \left(\widehat{\theta}_{i,t,R} - \theta^{\dagger}_{i} \right) \leq v \right) \right| > \varepsilon \right) \to 0,$$

where $P^*_{R,P}$ denotes the probability law of the resampled series, conditional on the (entire) sample.

Broadly speaking, Proposition 2 states that for $P \leq R$, $\Psi^{*(i)}_{R,P,1}$ and for $P > R$, $\Psi^{*(i)}_{R,P,2}$ has the same limiting distribution as $\frac{1}{\sqrt{P}} \sum_{t=R+s}^{T-1} \left(\widehat{\theta}_{i,t,R} - \theta^{\dagger}_{i} \right)$, conditional on sample, and for all samples except a set with probability measure approaching zero. Note that given A3, both R and P grow with the sample size at the same rate as T. As can be clearly seen in the proof of the proposition, if $|R-P| = o(T)$, then the contribution of the observations in that range is asymptotically negligible. Also, note that we do not need any adjustment term for the observations between P and R, or between R and P, depending whether P is larger or smaller than R. The intuitive reason is that all observations in that range carry the same weight (i.e. are used the same number of times), and therefore the standard block bootstrap, when "applied" to the observations in that range, works properly.

Though a detailed proof of Proposition 2 is given in the appendix, it is worthwhile to give an intuitive explanation of why there is an adjustment term in $\Psi^{*(i)}_{R,P,1}$ (and in $\Psi^{*(i)}_{R,P,2}$) as one might expect that $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}^{*}_{i,t,R}-\widehat{\theta}_{i,t,R}\right)$ has the same limiting distribution as $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_{i,t,R}-\theta^{\dagger}_{i,R}\right)$. For notational simplicity in the current discussion, let $h_{i,t}=\nabla_{\theta_i}\ln f_i(y_t,Z^{t-1},\theta^{\dagger}_i)$ and $h^{*}_{i,t}=\nabla_{\theta_i}\ln f_i(y^{*}_t,Z^{*,t-1},\theta^{\dagger}_i)$. Via a mean value expansion around $\theta^{\dagger}$, using arguments similar to those used in Lemma 4.1 of West and McCracken (1998), for the case of $P\leq R$ we have,

$$\begin{aligned}&\frac{1}{\sqrt{P}}\sum_{t=R+s}^{T-1}\left(\widehat{\theta}_{i,t,R}-\theta^{\dagger}_i\right)\\ =\;& A^{\dagger}_i\frac{1}{\sqrt{PR}}\left(\sum_{j=s+1}^{P+s}(j-s)h_{i,j}+P\sum_{j=P+s+1}^{R+s}h_{i,j}+\sum_{j=R+s+1}^{T-1}(P+s-(j-R))h_{i,j}\right)+o_P(1)\end{aligned}\tag{13}$$

where it should be recalled that $A^{\dagger}_i=\left(E\left(-\nabla^2_{\theta_i}\ln f_i(y_t,Z^{t-1},\theta^{\dagger}_i)\right)\right)^{-1}$. Also,

$$\begin{aligned}&\frac{1}{\sqrt{P}}\sum_{t=R+s}^{T-1}\left(\widehat{\theta}^{*}_{i,t,R}-\widehat{\theta}_{i,t,R}\right)\\ =\;& A^{\dagger}_i\frac{1}{\sqrt{PR}}\left(\sum_{j=s+1}^{P+s}(j-s)(h^{*}_{i,j}-h_{i,j})+P\sum_{j=P+s+1}^{R+s}(h^{*}_{i,j}-h_{i,j})+\sum_{j=R+s+1}^{T-1}(P+s-(j-R))(h^{*}_{i,j}-h_{i,j})\right)\\ &+o^{*}_P(1),\;\; \Pr-P\end{aligned}\tag{14}$$

Now, up to a term of order $O^{*}_P\left(l/\sqrt{P}\right)$,

$$E^{*}\left(\frac{1}{\sqrt{PR}}\sum_{j=s+1}^{P+s}(j-s)h^{*}_{i,j}\right)=\frac{1}{\sqrt{PR}}\sum_{j=s+1}^{P+s}(j-s)\frac{1}{P}\sum_{j=s+1}^{P+s}h_{i,j}\neq\frac{1}{\sqrt{PR}}\sum_{j=s+1}^{P+s}(j-s)h_{i,j},$$

and similarly,

$$\begin{aligned}&E^{*}\left(\frac{1}{\sqrt{PR}}\sum_{j=P+s+1}^{R+s}(P+s-(j+R))h^{*}_{i,j}\right)\;=\\ &\frac{1}{\sqrt{PR}}\sum_{j=R+s+1}^{T-1}(P+s-(j-R))\frac{1}{P}\sum_{j=R+s+1}^{T-1}h_{i,j}\;\neq\;\frac{1}{\sqrt{PR}}\sum_{j=R+s+1}^{T-1}(P+s-(j-R))h_{i,j},\end{aligned}$$

Therefore, the expectation of the RHS in (14), computed under the bootstrap law, $P^*_{R,P}$, is not zero, so that we cannot expect $\frac{1}{\sqrt{P}}\sum_{t=R+s}^{T-1}\left(\widehat{\theta}^*_{i,t,R}-\widehat{\theta}_{i,t,R}\right)$ to converge in $P^*_{R,P}-$distribution to a zero mean normal. Now, rewrite (13) as,

$$
\begin{aligned}
&\frac{1}{\sqrt{P}}\sum_{t=R+s}^{T-1}\left(\widehat{\theta}^*_{i,t}-\widehat{\theta}_{i,t}\right)\\
=\ & A_i^{\dagger}\left[\frac{1}{\sqrt{P}R}\sum_{j=s+1}^{P}(j-s)\left(h^*_{i,j}-\overline{h}_{i,P}\right)+\frac{\sqrt{P}}{R}\sum_{j=P+1}^{R}\left(h^*_{i,R+j}-\overline{h}_{i,R-P}\right)\right.\\
&\left.+\frac{1}{\sqrt{P}R}\sum_{j=R+s+1}^{T-1}(P+s-(j-R))(h^*_{i,j}-\overline{h}_{i,T-R})\right]\\
&-A_i^{\dagger}\left[\frac{1}{\sqrt{P}R}\sum_{j=s+1}^{P}(j-s)\left(h_{i,j}-\overline{h}_{i,P}\right)+\frac{1}{\sqrt{P}R}\sum_{j=R+s+1}^{T-1}(P+s-(j-R))(h_{i,j}-\overline{h}_{i,T-R})\right]\\
&+o^*_P(1),\ \ \Pr-P,
\end{aligned}
\tag{15}
$$

where $\overline{h}_P$, $\overline{h}_{R-P}$, and $\overline{h}_{T-R}$ are the sample means constructed observations from $s+1$ to $P+s$, observations between $P+s$ and $R+s$ and from the last P observations. As shown in the proof of the proposition, the first term on the RHS of (15) mimics the limiting distribution of $\frac{1}{\sqrt{P}}\sum_{t=R+s}^{T-1}\left(\widehat{\theta}_{i,t,R}-\theta^*_i\right)$, conditional on sample. On the other hand, the second term on the RHS is $O(1)$, conditional on sample, and for all samples except a set with probability measure approaching zero. Therefore, the second term in (15) can be interpreted as a location bias term of the standard block bootstrap. Such bias can be either positive or negative across different samples. Also, the difference between the second term on the RHS of (12) and the second term on the RHS of (15) vanishes asymptotically. Therefore, the adjustment term completely offsets the second term on the RHS of (15), as R and P go to infinity.

So far we have considered the case in which all parameters are jointly estimated. However, it is quite customary to first estimate conditional mean parameters via OLS or NLS and subsequently estimate the error variance using residuals. Along these lines, let $\theta_i=(\beta_i,\sigma^2)$, where β_i is $\Re^{p_i-1}$ valued and σ^2 is a scalar. Additionally, let $\ln f_i(y_j,Z^{j-1},\beta_i)=-(y_j-g_i(Z^{j-1},\beta_i))^2$,

$$\widehat{\beta}_{i,t,R}=\arg\min_{\beta_i\in B_i}\frac{1}{R}\sum_{j=t-R+1}^{t}(y_j-g(Z^{j-1},\beta_i))^2=,\quad R+s\le t\le T-1,\ i=1,...,n$$

where g is twice differentiable and $2r-$dominated on B, and $\widehat{\sigma}^2_{i,t,R}=\frac{1}{R}\sum_{j=t-R+1}^{t}(y_j-g_i(Z^{j-1},\widehat{\beta}_{i,t,R}))^2$.

The bootstrap analogs are

$$\widehat{\beta}^*_{i,t,R} = \arg\min_{\beta_i \in B_i} \frac{1}{R} \sum_{j=t-R+1}^{t} (y^*_j - g_i(Z^{*,j-1}, \beta_i))^2 =, \quad R+s \leq t \leq T-1, \; i = 1, ..., n$$

and $\widehat{\sigma}^{2,*}_{i,t,R} = \frac{1}{R}\sum_{j=t-R+1}^{t}(y^*_j - g_i(Z^{*,j-1}, \widehat{\beta}^*_{i,t,R}))^2$.

Furthermore, let $h_{i,j} = 2\epsilon_j \nabla_{\beta_i} g_i(Z^{j-1}, \beta^\dagger_i)$, where $\epsilon_j = (y_j - g(Z^{j-1}, \beta^\dagger_i))$, and $h^*_{i,j} = 2\epsilon^*_j \nabla_{\beta_i} g_i(Z^{*,j-1}, \beta^\dagger_i)$, for $t-R < j \leq t$, where $\epsilon^*_j = (y^*_j - g(Z^{*,j-1}, \widehat{\beta}_{i,t,R}))$, and finally let $\widehat{h}_{i,j} = 2\widehat{\epsilon}_j \nabla_{\beta_i} g_i(Z^{j-1}, \widehat{\beta}_{i,T})$, with $\widehat{\beta}_{i,T}$ be the estimator based on the full sample, and $\widehat{\epsilon}_j = (y_j - g(Z^{j-1}, \widehat{\beta}_{i,T}))$. For $P \leq R$, define:

$$\Phi^{*(i)}_{R,P,1} = \frac{1}{\sqrt{P}} \sum_{t=R+s}^{T-1} \begin{pmatrix} \widehat{\beta}^*_{i,t,R} - \widehat{\beta}_{i,t,R} \\ \widehat{\sigma}^{2*}_{i,t,R} - \widehat{\sigma}^2_{i,t,R} \end{pmatrix} + \begin{pmatrix} -\frac{1}{T}\sum_{t=s}^{T} \nabla^2_{\beta_i} g_i(Z^{t-1}, \widehat{\theta}_{i,T}) & 0 \\ 0 & 1 \end{pmatrix}^{-1}$$
$$\begin{pmatrix} \frac{1}{\sqrt{PR}} \left(\sum_{j=s+1}^{P+s}(j-s)\left(\widehat{h}_{i,j} - \widehat{\overline{h}}_{i,P}\right) + \sum_{j=R+s+1}^{T-1}(P+s-(j-R))(\widehat{h}_{i,j} - \widehat{\overline{h}}_{i,R-P})\right) \\ \frac{1}{\sqrt{PR}} \left(\sum_{j=s+1}^{P+s}(j-s)\left(\widehat{\epsilon}^2_{i,j} - \widehat{\overline{\epsilon}}^2_{i,P}\right) + \sum_{j=R+s+1}^{T-1}(P+s-(j-R))(\widehat{\epsilon}^2_{i,j} - \widehat{\overline{\epsilon}}^2_{i,T-R})\right) \end{pmatrix}$$

and for $P > R$ define:

$$\Phi^{*(i)}_{R,P,2} = \frac{1}{\sqrt{P}} \sum_{t=R+s}^{T-1} \begin{pmatrix} \widehat{\beta}^*_{i,t,R} - \widehat{\beta}_{i,t,R} \\ \widehat{\sigma}^{2*}_{i,t,R} - \widehat{\sigma}^2_{i,t,R} \end{pmatrix} + \begin{pmatrix} -\frac{1}{T}\sum_{t=s}^{T} \nabla^2_{\beta_i} g_i(Z^{t-1}, \widehat{\theta}_{i,T}) & 0 \\ 0 & 1 \end{pmatrix}^{-1}$$
$$\begin{pmatrix} \frac{1}{\sqrt{PR}} \left(\sum_{j=s+1}^{R+s}(j-s)\left(\widehat{h}_{i,j} - \widehat{\overline{h}}_{i,R}\right) + \sum_{j=P+s+1}^{T-1}(R+s-(j-P))(\widehat{h}_{i,j} - \widehat{\overline{h}}_{i,T-R})\right) \\ \frac{1}{\sqrt{PR}} \left(\sum_{j=s+1}^{R+s}(j-s)\left(\widehat{\epsilon}^2_{i,j} - \widehat{\overline{\epsilon}}^2_{i,R}\right) + \sum_{j=P+s+1}^{T-1}(R+s-(j-P))(\widehat{\epsilon}^2_{i,j} - \widehat{\overline{\epsilon}}^2_{i,T-R})\right) \end{pmatrix},$$

where $\widehat{\overline{h}}_{i,P}$, $\widehat{\overline{h}}_{i,R-P}$, $\widehat{\overline{h}}_{i,T-R}$ are defined as $\overline{h}_{i,P}$, $\overline{h}_{i,R-P}$, $\overline{h}_{i,T-R}$ but with $\theta^\dagger_i$ replaced by $\widehat{\theta}_{i,T}$, and $\widehat{\overline{\epsilon}}^2_{i,P} = P^{-1}\sum_{t=s+1}^{P+s} \widehat{\epsilon}^2_{i,t}$, and $\widehat{\overline{\epsilon}}^2_{i,R} = R^{-1}\sum_{t=s+1}^{R+s} \widehat{\epsilon}^2_{i,t}$.

Proposition 3: Let A1-A3 hold.

(i) Assume that as $P \to \infty$ and $l_1 \to \infty$, $l_1/P^{1/4} \to 0$, and as $R \to \infty$ and $l_3 \to \infty$, $l_3/R^{1/4} \to 0$, and finally as $R - P \to \infty$ and $l_2 \to \infty$, $l_2/(R-P)^{1/4} \to 0$. Then, as P and $R \to \infty$, for $P \leq R$,

$$P\left(\omega : \sup_{v \in \Re^{\varrho(i)}} \left| P^*_{R,P}\left(\Phi^{*(i)}_{R,P,1} \leq v\right) - P\left(\frac{1}{\sqrt{P}} \sum_{t=R+s}^{T-1} \left(\widehat{\theta}_{i,t,R} - \theta^\dagger_i\right) \leq v\right)\right| > \varepsilon\right) \to 0.$$

(ii) Assume that as $R \to \infty$ and $l_1 \to \infty$, $l_1/R^{1/4} \to 0$, and as $P \to \infty$ and $l_3 \to \infty$, $l_3/P^{1/4} \to 0$, and finally as $P - R \to \infty$ and $l_2 \to \infty$, $l_2/(P-R)^{1/4} \to 0$. Then, as P and $R \to \infty$, for $P > R$,

$$P\left(\omega : \sup_{v \in \Re^{\varrho(i)}} \left| P^*_{R,P}\left(\Phi^{*(i)}_{R,P,2} \leq v\right) - P\left(\frac{1}{\sqrt{P}} \sum_{t=R+s}^{T-1} \left(\widehat{\theta}_{i,t,R} - \theta^\dagger_i\right) \leq v\right)\right| > \varepsilon\right) \to 0,$$

where $P^*_{R,P}$ denotes the probability law of the resampled series, conditional on the (entire) sample.

3.2 A Full Sample Block Bootstrap for PEE: Rolling Estimation Scheme

Suppose we instead resample $P+R$ observations from the entire sample. Let let $W_t = (y_t, Z^{t-1})$, and draw b overlapping blocks of length l, where $bl = T-s$. The resampled observations, $W_s^{**}, W_{s+1}^{**}, ..., W_{s+l-1}^{**}, ..., W_T^{**}$, are equal to $W_{I_1}, W_{I_1+1}, ..., W_{I_1+l-1}, ..., W_{I_b+l-1}$, where I_i, $i = 1, ..., b$ are independent uniform random draws on the interval $s, ..., T-l+1$. Let $\widehat{\theta}_{i,t,R}^{**}$ be defined as in (1), but using W_t^{**} instead of W_t^*. Also, let $h_{i,t}^{**} = \nabla_{\theta_i} q_i(y_t^{**}, Z^{**,t-1}, \theta_i^\dagger)$. Now, from (14), we have

$$
\begin{aligned}
&\frac{1}{\sqrt{P}} \sum_{t=R+s}^{T-1} \left(\widehat{\theta}_{i,t,R}^{**} - \widehat{\theta}_{i,t,R} \right) \\
= \; & A_i^\dagger \frac{1}{\sqrt{P}R} \left(\sum_{j=s+1}^{P+s} (j-s)(h_{i,j}^* - h_{i,j}) + P \sum_{j=P+s+1}^{R+s} (h_{i,j}^* - h_{i,j}) + \sum_{j=R+s+1}^{T-1} (P+s-(j-R))(h_{i,j}^* - h_{i,j}) \right) \\
&+ o_P^*(1), \;\; \Pr -P
\end{aligned} \tag{16}
$$

Now, up to a term of order $O_P^*\left(l/\sqrt{P}\right)$,

$$
E^* \left(\frac{1}{\sqrt{P}R} \sum_{j=s+1}^{P+s} (j-s) h_{i,j}^* \right) = \frac{1}{\sqrt{P}R} \sum_{j=s+1}^{P+s} (j-s) \frac{1}{T} \sum_{j=s+1}^{T} h_{i,j} \neq \frac{1}{\sqrt{P}R} \sum_{j=s+1}^{P+s} (j-s) h_{i,j},
$$

$$
E^* \left(\frac{\sqrt{P}}{R} \sum_{j=s+1}^{P+s} h_{i,j}^* \right) = \frac{P^{3/2}}{TR} \sum_{j=s+1}^{T} h_{i,j} \neq \frac{1}{\sqrt{P}R} \sum_{j=s+1}^{P+s} (j-s) h_{i,j},
$$

and similarly,

$$
\begin{aligned}
&E^* \left(\frac{1}{\sqrt{P}R} \sum_{j=R+s+1}^{T-1} (P+s-(j+R)) h_{i,j}^* \right) \quad = \\
&\frac{1}{\sqrt{P}R} \sum_{j=R+s+1}^{T-1} (P+s-(j-R)) \frac{1}{T} \sum_{j=s+1}^{T-1} h_{i,j} \quad \neq \quad \frac{1}{\sqrt{P}R} \sum_{j=R+s+1}^{T-1} (P+s-(j-R)) h_{i,j},
\end{aligned}
$$

Hereafter, let $\overline{h}_{i,T} = \frac{1}{T} \sum_{j=s+1}^{T} h_{i,j}$. Now,

$$
\begin{aligned}
&\frac{1}{\sqrt{P}} \sum_{t=R+s}^{T-1} \left(\widehat{\theta}_{i,t,R}^{**} - \widehat{\theta}_{i,t,R} \right) \\
= \; & A_i^\dagger \frac{1}{\sqrt{P}R} \left(\sum_{j=s+1}^{P+s} (j-s)(h_{i,j}^* - \overline{h}_{i,T}) + P \sum_{j=P+s+1}^{R+s} (h_{i,j}^* - \overline{h}_{i,T}) + \sum_{j=R+s+1}^{T-1} (P+s-(j-R))(h_{i,j}^* - \overline{h}_{i,T}) \right) \\
&- A_i^\dagger \frac{1}{\sqrt{P}R} \left(\sum_{j=s+1}^{P+s} (j-s)(h_{i,j} - \overline{h}_{i,T}) + P \sum_{j=P+s+1}^{R+s} (h_{i,j} - \overline{h}_{i,T}) + \sum_{j=R+s+1}^{T-1} (P+s-(j-R))(h_{i,j} - \overline{h}_{i,T}) \right) \\
&+ o_P^*(1), \;\; \Pr -P
\end{aligned} \tag{17}
$$

Note that the first line on the RHS of (17) has the same limiting distribution as $\frac{1}{\sqrt{P}}\sum_{t=R+s}^{T-1}\left(\widehat{\theta}_{i,t,R}-\theta_i^{\dagger}\right)$, conditional on the sample and for all sample but a set of probability measure approaching zero. On the other hand, the last line on the RHS of (17) is a location bias term, which is either positive or negative across different samples. For convenience, define $\widehat{A}_{i,T}=\left(-\frac{1}{T}\sum_{t=s}^{T}\nabla_{\theta_i}^2\ln f_i(y_t,Z^{t-1},\widehat{\theta}_{i,P})\right)^{-1}$, $\widehat{h}_{i,t}=\nabla_{\theta_i}\ln f_i(y_t,Z^{t-1},\widehat{\theta}_{i,P})$ and $\widehat{\overline{h}}_i=\frac{1}{T}\sum_{t=s+1}^{T}\nabla_{\theta_i}\ln f_i(y_t,Z^{t-1},\widehat{\theta}_{i,P})$. Consider,

$$\begin{aligned}\Psi_{R,P}^{(i)**} &= \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_{i,t}^{**}-\widehat{\theta}_{i,t}\right)+\\ &\quad +\widehat{A}_{i.T}\frac{1}{\sqrt{P}R}\Big(\sum_{j=s+1}^{P+s}(j-s)(\widehat{h}_{i,j}-\widehat{\overline{h}}_{i,T})+P\sum_{j=P+s+1}^{R+s}(\widehat{h}_{i,j}-\widehat{\overline{h}}_{i,T})+\\ &\quad +\sum_{j=R+s+1}^{T-1}(P+s-(j-R))(\widehat{h}_{i,j}-\widehat{\overline{h}}_{i,T})\Big) \end{aligned} \tag{18}$$

Now, $\Psi_{R,P}^{(i)**}-\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_{i,t}^{**}-\widehat{\theta}_{i,t}\right)$ offsets the location bias term, and thus $\Psi_{R,P}^{(i)**}$ has the same limiting distribution as $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_{i,t}-\theta_i^{*}\right)$, conditional on sample.

It follows immediately that $\Psi_{R,P}^{(i)*}$ only contains a correction term for the first and the last P observations, while $\Psi_{R,P}^{(i)**}$ contains an extra correction term, also for the observations between P and R. In this sense, one may prefer $\Psi_{R,P}^{(i)*}$ to $\Psi_{R,P}^{(ii)**}$. However, a comparison of the two statistics is left to future research, as the Monte Carlo experiments reported in Section 4 focus on the finite sample behavior of $\Psi_{R,P}^{(i)*}$, although our empirical findings suggest there may be little to choose between split and full bootstrap sampling approaches (see Section 5).

3.3 A Split Sample Block Bootstrap for PEE: Recursive Estimation Scheme

This bootstrap procedure is discussed in detail in Corradi and Swanson (2003a). Here, we recap their results for the split sample version of the block bootstrap. Results for the full sample version of the block bootstrap are analogous to those given in the previous subsection for the case of rolling estimation schemes.

Form bootstrap samples by first resampling from observations $s+1,...,R+s$, and then concatenating onto this an additional P observations resampled from the P remaining sample observations. More specifically, let $b_1l_1+b_2l_2=T$, with $b_1l_1=R$ and $b_2l_2=P$. Also, let $W_t=(y_t,Z^{t-1})$. First, draw b_1 overlapping blocks, of length l_1, from $s+1,...,R+s$ and then draw b_2 overlapping blocks, of length l_2, from data indexed by $R+s+1,...,R+s+P$, with replacement. The first R pseudo observations, $W_{s+1}^*, W_{s+2}^*, ..., W_{s+l-1}^*, ..., W_{R+s}^*$, are equal to $W_{I_1^R}, W_{I_1^R+1}, ..., W_{I_1^R+l_1-1}, ..., W_{I_{b_1}^R+l_1-1}$, where I_i^R, $i = 1, ..., b_1$ are independent uniform random draws on the interval $s, ..., R+s-l_1+1$; and the remaining P pseudo observations, $W_{R+s+1}^*, W_{R+s+2}^*, ..., W_{R+s+l_2}^*, ..., W_{R+s+P}^*$, are equal to $W_{I_1^P}, W_{I_1^P+1}, ..., W_{I_1^P+l_2-1}, ..., W_{I_{b_2}^P+l_2-1}$, where I_i^P, $i = 1, ..., b_2$ are independent uniform random draws from data indexed by $R+s, R+2, ..., R+s+P-l_2-1$. Thus, conditional on the (entire) sample, the pseudo time series W_t^*, $t = s, ..., R+s, R+s+1, ..., R+s+P$, consists of $b = b_1 + b_2$ asymptotically independent, but non identically distributed blocks of length l_1 and l_2 respectively.[13] Now, define the recursive PEE bootstrap m-estimator as,

$$\widehat{\theta}_{i,t}^* = \arg\max_{\theta_i \in \Theta_i} \frac{1}{t} \sum_{j=s}^{t} \ln f_i(y_j^*, Z^{*,j-1}, \theta_i), \ \ R+s \leq t \leq T-1, \ \ i = 1, ..., n.$$

Finally, define

$$\begin{aligned} & \Psi_{R,P,3}^* \\ = \ & \frac{1}{\sqrt{P}} \sum_{t=R+s}^{T-1} \left(\widehat{\theta}_{i,t}^* - \widehat{\theta}_{i,t} \right) + \left(-\frac{1}{T} \sum_{t=s}^{T} \nabla_{\theta_i}^2 \ln f_i(y_t, Z^{t-1}, \widehat{\theta}_{i,T}) \right)^{-1} \\ & \times \frac{1}{\sqrt{P}} \sum_{j=s+1}^{P+s-1} a_{R,j} \left(\nabla_{\theta_i} \ln f_i(y_{R+j}, Z^{R+j-1}, \widehat{\theta}_{i,T}) - \frac{1}{P} \sum_{j=1}^{P} \nabla_{\theta_i} \ln f_i(y_{R+j}, Z^{R+j-1}, \widehat{\theta}_{i,T}) \right), \end{aligned} \tag{19}$$

where $a_{R,j} = \frac{1}{R+j} + \frac{1}{R+j+1} + ... + \frac{1}{R+P-1}$, $j = 0, 1, ..., P-1$.

Proposition 4: Let A1-A3 hold. Also, assume that as $P, R \to \infty$, $l_1, l_2 \to \infty$, and that $\frac{l_2}{P^{1/4}} \to 0$ and $\frac{l_1}{R^{1/4}} \to 0$. Then, as P and $R \to \infty$,

$$P\left(\omega : \sup_{v \in \Re^{\varrho(i)}} \left| P_{R,P}^* \left(\Psi_{R,P,3}^* \leq v \right) - P\left(\frac{1}{\sqrt{P}} \sum_{t=R+s}^{T-1} \left(\widehat{\theta}_{i,t} - \theta_i^\dagger \right) \leq v \right) \right| > \varepsilon \right) \to 0,$$

where $P_{R,P}^*$ denotes the probability law of the resampled series, conditional on the (entire) sample.

Now let $\widehat{\beta}_{i,t}$, $\widehat{\beta}_{i,t}^*$, $\sigma_{i,t}^2$, $\widehat{\sigma}_{i,t}^{2,*}$ be defined as $\widehat{\beta}_{i,t,R}$, $\widehat{\beta}_{i,t,R}^*$, $\sigma_{i,t,R}^2$, $\widehat{\sigma}_{i,t,R}^{2,*}$ above but using a recursive

[13] More precisely, each block from $R+s+1, ..., R+s+P-l_2-1$ may overlap with any block from $s+1, ..., R+s$ for at most s observations, where s is finite.

instead of a rolling scheme, and define

$$\Phi^*_{R,P,3} = \frac{1}{\sqrt{P}}\sum_{t=R+s}^{T-1}\begin{pmatrix}\widehat{\beta}^*_{i,t}-\widehat{\beta}_{i,t}\\ \widehat{\sigma}^{2*}_{i,t}-\widehat{\sigma}^2_{i,t}\end{pmatrix}+\begin{pmatrix}-\frac{1}{T}\sum_{t=s}^{T}\nabla^2_{\beta_i}g_i(Z^{t-1},\widehat{\theta}_{i,T}) & 0\\ 0 & 1\end{pmatrix}^{-1}$$
$$\begin{pmatrix}\frac{1}{\sqrt{P}}\sum_{j=s+1}^{P+s-1}a_{R,j}\left(\widehat{h}_{i,R+j}-\frac{1}{P}\sum_{j=s+1}^{P+s-1}\widehat{h}_{R+j}\right)\\ \frac{1}{\sqrt{P}}\sum_{j=s+1}^{P+s-1}a_{R,j}\left(\widehat{\epsilon}^2_{i,R+j}-\frac{1}{P}\sum_{j=s+1}^{P+s-1}\widehat{\epsilon}^2_{R+j}\right)\end{pmatrix},$$

we have

Proposition 5: Let A1-A3 hold. Also assume that as $P,R\to\infty$, $l_1,l_2\to\infty$, and that $\frac{l_2}{P^{1/4}}\to 0$ and $\frac{l_1}{R^{1/4}}\to 0$. Then, as P and $R\to\infty$,

$$P\left(\omega:\sup_{v\in\Re^{\varrho(i)}}\left|P^*_{R,P}\left(\Phi^*_{R,P,3}\leq v\right)-P\left(\frac{1}{\sqrt{P}}\sum_{t=R+s}^{T-1}\left(\widehat{\theta}_{i,t}-\theta_i^{\dagger}\right)\leq v\right)\right|>\varepsilon\right)\to 0,$$

where $P^*_{R,P}$ denotes the probability law of the resampled series, conditional on the (entire) sample.

3.4 Bootstrap Critical Values for the Predictive Density Accuracy Test

Turning again to our predictive density accuracy test, we are now in a position to construct an appropriate bootstrap statistic, from whence bootstrap critical values can be constructed. Using the bootstrap sampling procedures defined in the previous section, one first constructs appropriate bootstrap samples. Thereafter, form bootstrap statistics as follows,

$$Z^*_{P,j}=\max_{k=2,\ldots,n}\int_U Z^*_{P,u,j}(1,k)\phi(u)du,$$

where for $j=1$ (rolling estimation scheme) and $P\leq R$,

$$\begin{aligned}
Z^*_{P,u,1}(1,k) \;=\; & \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\left(\left(1\{y^*_{t+1}\leq u\}-F_1(u|Z^{*,t},\widehat{\theta}^*_{1,t,R})\right)^2-\left(1\{y_{t+1}\leq u\}-F_1(u|Z^t,\widehat{\theta}_{1,t,R})\right)^2\right)\right.\\
& \left.-\left(\left(1\{y^*_{t+1}\leq u\}-F_k(u|Z^{*,t},\widehat{\theta}^*_{k,t,R})\right)^2-\left(1\{y_{t+1}\leq u\}-F_k(u|Z^t,\widehat{\theta}_{k,t,R})\right)^2\right)\right)\\
& -\frac{2}{T}\sum_{t=s}^{T-1}\left(\nabla_{\theta_1}F_1(u|Z^t,\widehat{\theta}_{1,T})\left(1\{y^*_{t+1}\leq u\}-F_1(u|Z^t,\widehat{\theta}_{1,T})\right)\right)'\\
& \times\left(\Psi^{*(1)}_{R,P,1}-\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}^*_{1,t,R}-\widehat{\theta}_{1,t,R}\right)\right)\\
& +\frac{2}{T}\sum_{t=s}^{T-1}\left(\nabla_{\theta_k}F_k(u|Z^t,\widehat{\theta}_{k,T})'\left(1\{y^*_{t+1}\leq u\}-F_k(u|Z^t,\widehat{\theta}_{k,T})\right)\right)\\
& \times\left(\Psi^{*(k)}_{R,P,1}-\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}^*_{k,t,R}-\widehat{\theta}_{k,t,R}\right)\right).
\end{aligned} \tag{20}$$

For $j=1$ and $P>R$, $Z^{*}_{P,u,1}(1,k)$ is defined as above, but with $\Psi^{*(1)}_{R,P,1},\Psi^{*(k)}_{R,P,1}$ replaced by $\Psi^{*(1)}_{R,P,2},\Psi^{*(k)}_{R,P,2}$.

For $j=2$ (recursive estimation scheme),

$$
\begin{aligned}
Z^{*}_{P,u,2}(1,k) \quad = \quad & \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\Bigg(\Bigg(\left(1\{y^{*}_{t+1}\le u\}-F_1(u|Z^{*,t},\widehat{\theta}^{*}_{1,t})\right)^2-\left(1\{y_{t+1}\le u\}-F_1(u|Z^{t},\widehat{\theta}_{1,t})\right)^2\Bigg) \\
& -\left(\left(1\{y^{*}_{t+1}\le u\}-F_k(u|Z^{*,t},\widehat{\theta}^{*}_{k,t})\right)^2-\left(1\{y_{t+1}\le u\}-F_k(u|Z^{t},\widehat{\theta}_{k,t})\right)^2\right)\Bigg) \\
& -\frac{2}{T}\sum_{t=s}^{T-1}\left(\nabla_{\theta_1}F_1(u|Z^{t},\widehat{\theta}_{1,T})'\left(1\{y^{*}_{t+1}\le u\}-F_1(u|Z^{t},\widehat{\theta}_{1,T})\right)\right)' \\
& \times\left(\Psi^{*(1)}_{R,P,3}-\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}^{*}_{1,t}-\widehat{\theta}_{1,t}\right)\right) \\
& +\frac{2}{T}\sum_{t=s}^{T-1}\left(\nabla_{\theta_k}F_k(u|Z^{t},\widehat{\theta}_{k,T})'\left(1\{y^{*}_{t+1}\le u\}-F_k(u|Z^{t},\widehat{\theta}_{k,T})\right)\right)' \\
& \times\left(\Psi^{*(k)}_{R,P,3}-\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}^{*}_{k,t}-\widehat{\theta}_{k,t}\right)\right).
\end{aligned}
$$

Finally, when the conditional mean parameters are estimated by (N)LS and the variance is subsequently estimated using residuals, replace $\Psi^{*(l)}_{R,P,i}$, with $\Phi^{*(l)}_{R,P,i}$, $i=1,2,3$, $l=1,k$.

Proposition 6: Let Assumptions A1-A4 hold.. Also, assume that: (i) for the rolling estimation scheme and $P\le R$, as $P\to\infty$ and $l_1\to\infty$, $l_1/P^{1/4}\to 0$, and as $R\to\infty$ and $l_3\to\infty$, $l_3/P^{1/4}\to 0$, and finally as $R-P\to\infty$ and $l_2\to\infty$, $l_2/(R-P)^{1/4}\to 0$; or (ii) for the rolling estimation scheme and $P>R$, as $R\to\infty$ and $l_1\to\infty$, $l_1/R^{1/4}\to 0$, and as $P\to\infty$ and $l_3\to\infty$, $l_3/R^{1/4}\to 0$, and finally as $P-R\to\infty$ and $l_2\to\infty$, $l_2/(P-R)^{1/4}\to 0$, or (iii) for the recursive estimation scheme, as $P,R\to\infty$ and $l_1,l_2\to\infty$, then $\frac{l_2}{P^{1/4}}\to 0$ and $\frac{l_1}{R^{1/4}}\to 0$. Then, as P and $R\to\infty$, for $j=1,2$

$$
P\left(\omega:\sup_{v\in\Re}\left|P^{*}_{R,P}\left(\max_{k=2,\dots,n}\int_U Z^{*}_{P,u,j}(1,k)\phi(u)du\le v\right)-P\left(\max_{k=2,\dots,n}\int_U Z^{\mu}_{P,u,j}(1,k)\phi(u)du\le v\right)\right|>\varepsilon\right)\to 0,
$$

where $Z^{\mu}_{P,u,j}(1,k)=Z_{P,u,j}(1,k)-\sqrt{P}\left(\mu_1^2(u)-\mu_k^2(u)\right)$.

The above result suggests proceeding in the following manner. For any bootstrap replication, compute the bootstrap statistic, $Z^{*}_{P,j}$. Perform B bootstrap replications (B large) and compute the quantiles of the empirical distribution of the B bootstrap statistics. Reject H_0, if $Z_{P,j}$ is greater than the $(1-\alpha)th$-percentile. Otherwise, do not reject. Now, for all samples except a set with probability measure approaching zero, $Z_{P,j}$ has the same limiting distribution as the corresponding bootstrapped statistic when $E\left(\mu_1^2(u)-\mu_k^2(u)\right)=0$, $\forall\ k$, ensuring asymptotic size equal to α. On the other hand, when one or more competitor models are strictly dominated by the benchmark, the rule provides a test with asymptotic size between 0 and α. Under the alternative, $Z_{P,j}$ diverges to (plus) infinity, while the corresponding bootstrap statistic has a well defined limiting distribution, ensuring unit asymptotic power. From the above discussion, we see that the bootstrap distribution provides correct asymptotic critical values only for the least favorable case under the null hypothesis; that is, when all competitor models are as good as the benchmark model. When $\max_{k=2,\ldots,m}\int_U\left(\mu_1^2(u)-\mu_k^2(u)\right)\phi(u)du=0$, but $\int_U\left(\mu_1^2(u)-\mu_k^2(u)\right)\phi(u)du<0$ for some k, then the bootstrap critical values lead to conservative inference. An alternative to our bootstrap critical values in this case is the construction of critical values based on subsampling (see e.g. Politis, Romano and Wolf (1999), Ch. 3). Heuristically, construct $T-2b_T$ statistics using subsamples of length b_T, where $b_T/T\to 0$. The empirical distribution of these statistics computed over the various subsamples properly mimics the distribution of the statistic. Thus, subsampling provides valid critical values even for the case where $\max_{k=2,\ldots,m}\int_U\left(\mu_1^2(u)-\mu_k^2(u)\right)\phi(u)du=0$, but $\int_U\left(\mu_1^2(u)-\mu_k^2(u)\right)\phi(u)du<0$ for some k. This is the approach used by Linton, Maasoumi and Whang (2003), for example, in the context of testing for stochastic dominance. Needless to say, one problem with subsampling is that unless the sample is very large, the empirical distribution of the subsampled statistics may yield a poor approximation of the limiting distribution of the statistic. An alternative approach for addressing the conservative nature of our bootstrap critical values is suggested in Hansen (2001). Hansen's idea is to recenter the bootstrap statistics using the sample mean, whenever the latter is larger than (minus) a bound of order $\sqrt{2T\log\log T}$. Otherwise, do not recenter the bootstrap statistics. In the current context, his approach leads to correctly sized inference when $\max_{k=2,\ldots,m}\int_U\left(\mu_1^2(u)-\mu_k^2(u)\right)\phi(u)du=0$, but $\int_U\left(\mu_1^2(u)-\mu_k^2(u)\right)\phi(u)du<0$ for some k. Additionally, his approach has the feature that if all models are characterized by a sample mean below the bound, the null is "accepted" and no bootstrap statistic is constructed.

4 Monte Carlo Results

In this section we build on the Monte Carlo results of Corradi and Swanson (2003a), where the bootstrap for PEE in recursive estimation schemes is analyzed via experimentation using $\Psi^*_{R,P,3}$, as defined in (19). In particular, in this section we compare $\Psi^*_{R,P,1}$ and $\Psi^*_{R,P,3}$ (where the superscript or subscript i is suppressed for simplicity) with analogous bootstrap PEE statistics where no bias adjustment is made.[14]

As in Corradi and Swanson (2003a), two data generating processes are specified, namely $y_t = c + \rho y_{t-1} + \varepsilon_t$ and $y_t = c + \rho_1 y_{t-1} + \rho_2 y_{t-1} + \varepsilon_t$, with $\varepsilon_t \sim IN(0,1)$, $c = 0.1$, $\rho = \{0.2, 0.4, 0.6, 0.8\}$ and $\rho_1 = \rho_2 = \{0.1, 0.2, 0.3, 0.4\}$. Given this setup, we proceed to estimate both AR(1) and AR(2) models for each of the two alternative DGPs. Thus, when we estimate (via OLS) an AR(1) (or an AR(2)) model, $\widehat{\theta}_{l,t} = (\widehat{c}_{l,t}, \hat{\rho}_{l,t})'$ (or $\widehat{\theta}_{l,t} = (\widehat{c}_{l,t}, \hat{\rho}_{1,l,t}, \hat{\rho}_{2,l,t})'$), with $l = 1,2$ denoting the estimate models (AR(1) and AR(2), respectively), and $\theta_l^\dagger = (c_l^\dagger, \rho_l^\dagger)'$ (or $\theta_l^\dagger = (c_l^\dagger, \rho_{1,l}^\dagger, \rho_{2,l}^\dagger)'$), where $\theta_l^\dagger$ denotes the probability limit of $\widehat{\theta}_{l,t}$. Needless to say, in the case of correct dynamic specification, $\theta_l^\dagger$ represents the parameters characterizing the conditional expectation, while in the case of dynamic misspecification (e.g. the DGP is AR(2) and we estimate an AR(1)), $\theta_l^\dagger$ represents pseudo true values, which can be explicitly computed.

We confine our attention to the slope parameters in the above regression models. For notational simplicity, consider the case in which we estimate a AR(1) and the DGP is also AR(1), so that we compute a $P-$sequence of estimators $\widehat{\rho}_t$, bootstrap estimators $\widehat{\rho}_t^*$, and we know that $\rho^\dagger = \{0.2, 0.4, 0.6, 0.8\}$. Now, the rolling estimation scheme bootstrap is thus given by:[15]

$$\begin{aligned}
\Psi^*_{R,P,1} \quad = \quad & \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T-1} (\widehat{\rho}_t^* - \widehat{\rho}_t) + \left(\frac{1}{T} \sum_{t=2}^{T} (y_{t-1} - \overline{y})^2 \right)^{-1} \\
& \times [\frac{1}{\sqrt{P}R} \sum_{j=2}^{P+1} (j-1) \left(\widehat{e}_{R+j-1} (y_{R+j-1} - \overline{y}) - \frac{1}{P} \sum_{j=2}^{P+1} \widehat{e}_{R+j-1} (y_{R+j-1} - \overline{y}) \right) \\
& + \frac{1}{\sqrt{P}R} \sum_{j=R+2}^{T-1} (P + 1 - (j-2)) \left(\widehat{e}_{j+1} (y_{j+1} - \overline{y}) - \frac{1}{P} \sum_{j=2}^{P+1} \widehat{e}_{j+1} (y_{j+1} - \overline{y}) \right)],
\end{aligned}$$

where $\widehat{e}_{R+j} = (y_{R+j} - \overline{y}) - \widehat{\rho}_T (y_{R+j-1} - \overline{y})$, $\overline{y} = T^{-1} \sum_{t=s}^T y_t$. Furthermore, the recursive estima-

[14]Subsequent analysis of finite sample properties of predictive density tests constructed as outlined above using our bootstrap results is the subject of ongoing research and will be reported in a later paper.

[15]In our experiments, $\widehat{\rho}_t^*$ is computed using the pseudo time series obtained by first resampling b_1 blocks from the first R observations and then concatenating b_2 blocks resampled from the last P observations, as described in Section 2. Examination of the alternative PEE bootstrap methods developed in this paper, including the method for the case where the entire sample is used, and extra adjustment terms are added to the bootstrap statistic, is left to future research.

tion scheme bootstrap is given by:

$$\Psi^*_{R,P,3} = \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T-1} (\widehat{\rho}^*_t - \widehat{\rho}_t) + \left(\frac{1}{T} \sum_{t=2}^{T} (y_{t-1} - \overline{y})^2 \right)^{-1} \times \frac{1}{\sqrt{P}} \sum_{j=2}^{P} a_{R,j} \left(\widehat{e}_{R+j} (y_{R+j} - \overline{y}) - \frac{1}{P} \sum_{j=1}^{P} \widehat{e}_{R+j} (y_{R+j} - \overline{y}) \right).$$

Furthermore, define analogous bootstrap statistics without adjustment as

$$\widetilde{\Psi}^*_{R,P,1} = \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T-1} (\widehat{\rho}^*_t - \widehat{\rho}_t)$$

and

$$\widetilde{\Psi}^*_{R,P,3} = \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T-1} (\widehat{\rho}^*_t - \widehat{\rho}_t)$$

Finally, let $z^*_{j,\alpha}$ be the $(1-\alpha)$ quantile of the distribution of $\Psi^*_{R,P,j}$, $j = 1,3$ and let $\widetilde{z}^*_{j,\alpha}$ be the $(1-\alpha)$ quantile of the distribution of $\widetilde{\Psi}^*_{R,P,j}$, $j = 1,3$. Recall that the adjusted and non-adjusted bootstrap statistics are characterized by the same asymptotic variance; the only difference is that the latter is biased. Thus, we can directly compare the coverage probabilities of the different bootstraps with and without adjustment terms. Thus, we define $100(1-\alpha)\%$, equal-tailed, two-sided confidence intervals corresponding to the rolling bootstrap with adjustment and the rolling bootstrap without adjustment, respectively: $CI^*_1 : \left\{ \frac{1}{P} \sum_{t=R}^{P-1} \widehat{\rho}_t - \frac{z^*_{1,\alpha/2}}{\sqrt{P}}, \frac{1}{P} \sum_{t=R}^{P-1} \widehat{\rho}_t + \frac{z^*_{1,(1-\alpha/2)}}{\sqrt{P}} \right\}$ and $CI^*_2 : \left\{ \frac{1}{P} \sum_{t=R}^{P-1} \widehat{\rho}_t - \frac{z^*_{3,\alpha/2}}{\sqrt{P}}, \frac{1}{P} \sum_{t=R}^{P-1} \widehat{\rho}_t + \frac{z^*_{3,(1-\alpha/2)}}{\sqrt{P}} \right\}$. Similarly, for the recursive bootstrap we have: $\widetilde{CI}^*_1 : \left\{ \frac{1}{P} \sum_{t=R}^{P-1} \widehat{\rho}_t - \frac{\widetilde{z}^*_{1,\alpha/2}}{\sqrt{P}}, \frac{1}{P} \sum_{t=R}^{P-1} \widehat{\rho}_t + \frac{\widetilde{z}^*_{1,(1-\alpha/2)}}{\sqrt{P}} \right\}$ and $\widetilde{CI}^*_2 : \left\{ \frac{1}{P} \sum_{t=R}^{P-1} \widehat{\rho}_t - \frac{\widetilde{z}^*_{3,\alpha/2}}{\sqrt{P}}, \frac{1}{P} \sum_{t=R}^{P-1} \widehat{\rho}_t + \frac{\widetilde{z}^*_{3,(1-\alpha/2)}}{\sqrt{P}} \right\}$.

The coverage probabilities for CI^*_1 and CI^*_2, for example, are then obtained by computing the proportion of times, across simulation replications, for which $\rho^\dagger$ falls into the respective interval. By comparing these coverage probabilities we have a direct measure of the impact of the adjustment term. Broadly speaking, if the difference between the actual and nominal coverage is smaller for CI^*_1 than for $\widetilde{CI}^*_1$, then it is definitely worthwhile to construct bootstrap critical values based on the bootstrap with adjustment. Furthermore, direct inspection of the coverage probabilities for CI^*_1 will yield evidence concerning block length selection and overall performance of the PEE bootstrap methods. All bootstrap empirical distributions are based on 200 bootstrap replications, and all tabulated results are based on 500 Monte Carlo simulations. In addition, samples of $T = \{800{,}1600\}$ observations are used, and the number of estimators constructed in the context of the PEE recursive scheme bootstrap is $P = 0.5T$, with the first estimator constructed using $T - P$ observations, the second with $T - P + 1$ observations, etc. The number of estimators constructed in the context of the PEE rolling scheme bootstrap is also $P = 0.5T$ (and hence our use of $\Psi^*_{R,P,1}$ instead of $\Psi^*_{R,P,2}$), with all estimators constructed using R observations. The nominal coverage probability, across all experiments, is set equal to 0.90. We have tried a variety of values of α in the construction of the confidence intervals. However, as the results are qualitatively the same, we report results only for $\alpha = 0.10$.

Our findings are reported in Tables 1-4, and are organized as follows. The second column lists the bootstrap used to mimic the distribution of PEE associated with either the AR(1) autoregressive parameter (denoted $\hat{\rho}$ in the tables) or the autoregressive parameters from the AR(2) model (denoted $\hat{\rho}_1$ and $\hat{\rho}_2$ in the table). Entries given under the heading *roll*1 correspond to coverage probabilities associated with CI_1^*, while those given under the heading *roll*2 correspond to coverage probabilities associated with $\widetilde{CI}_1^*$. Similarly, entries given under the heading *rec*1 correspond to coverage probabilities associated with CI_2^*, while those given under the heading *rec*2 correspond to coverage probabilities associated with $\widetilde{CI}_2^*$.Tables 1-4 is broken into two panels, depending upon whether data were generated according to an AR(1) process (Panel A) or an AR(2) process (Panel B), and the autoregressive parameters of the DGPs are given in the header line for each panel. In addition, block lengths used are denoted by the various values of $l_1 = l_2$. (The same block length when resampling from the first R observations and from the last P observations.)

A number of clear-cut findings emerge upon inspection of the tables. First, the adjustment terms in the rolling and recursive bootstrap PEE statistics are required in order to improve coverage. Probabilities associated with the respective versions of the bootstrap statistics that do not contain adjustment terms ($\widetilde{\Psi}^*_{R,P,1}$ and $\widetilde{\Psi}^*_{R,P,3}$) are generally poor, relative to the properly adjusted versions. Second, and as expected, coverage is best when the autoregressive parameters in the models are smaller, with performance worsening as these parameters increase from 0.2 to 0.8 in the AR(1) case (see Panel A of Tables 1-4) and from 0.1 to 0.4 in the AR(2) case (see Panel B of the same tables). This is particularly true, again as expected, for the smaller block lengths. Finally, misspecification does not play a great roll in coverage probability accuracy. For example, whether an AR(2) is estimated when the true DGP is an AR(1) (as is the case for the $\hat{\rho}_1$ and $\hat{\rho}_2$ rows of entries in Panel A of each table) is of secondary importance. Of primary importance appears to be block length and the magnitude of the autoregressive component of the model. This is a promising finding, in the sense that the bootstrap methods discussed here are in this sense robust to model misspecification - a good property given our assumption in our predictive density test that all models may be misspecified. Although much further research will need to be undertaken before all of the properties of the bootstraps discussed in this paper are known, and before the related properties of tests (such as the predictive density test) based on the use of our bootstrap techniques become clear, we take the results of this paper to be a positive step in that direction.

5 Empirical Illustration - Forecasting Inflation

In this section we use a simple stylized macroeconomic example to illustrate how to apply the predictive density accuracy test discussed in Section 2. In particular, assume that the objective is to select amongst 4 different predictive density models for inflation, including an linear AR model and an ARX model, where the ARX model differs from the AR model only through the inclusion of unemployment as an additional explanatory variable. Assume also that 2 versions of each of these models are used, one assuming normality, and one assuming that the conditional distribution being evaluated follows a Student's t distribution with 5 degrees of freedom. Further, assume that the number of lags used in these models is selected via use of either the SIC or the AIC. This example can thus be thought of as an out-of-sample evaluation of simplified Phillips curve type models of inflation.

The data used were obtained from the St. Louis Federal Reserve website. For unemployment, we use the seasonally adjusted civilian unemployment rate. For inflation, we use the 12th difference of the log of the seasonally adjusted CPI for all urban consumers, all items. Both data series were found to be I(0), based on application of standard augmented Dickey-Fuller unit root tests. All data are monthly, and the sample period is 1954:1-2003:12. This 600 observation sample was broken into two equal parts for test construction, so that $R = P = 300$. Additionally, all predictions were 1-step ahead, and were constructed using the recursive estimation scheme discussed above. Bootstrap percentiles were calculated based on 100 bootstrap replications, and we set $u \in U \subset [Inf_{\min}, Inf_{\max}]$, where Inf_t is the inflation variable being examined, and 100 equally spaced values for u across this range were used (i.e. $\phi(u)$ is the uniform density). Lags were selected as follows. First, and using only the initial R sample observations, autoregressive lags were selected according to both the SIC and the AIC. Thereafter, fixing the number of autoregressive lags, the number of lags of unemployment ($Unem_t$) was chosen, again using each of the SIC and the AIC. This framework enabled us to compare various permutations of 4 different models using the $Z_{P,2}$ statistic, where

$$Z_{P,2} = \max_{k=2,\dots,4} \int_U Z_{P,u,2}(1,k)\phi(u)du$$

and

$$Z_{P,u,2}(1,k) = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\left(1\{Inf_{t+1} \leq u\} - F_1(u|Z^t, \widehat{\theta}_{1,t})\right)^2 - \left(1\{Inf_{t+1} \leq u\} - F_k(u|Z^t, \widehat{\theta}_{k,t})\right)^2 \right),$$

as discussed in Section 2. In particular, we consider (i) a comparison of AR and ARX models, with lags selected using the SIC; (ii) a comparison of AR and ARX models , with lags selected using the AIC; (iii) a comparison of AR models, with lags selected using either the SIC or the AIC; and (iv) a comparison of ARX models, with lags selected using either the SIC or the AIC. Recalling that each model is specified with either a Gaussian or Student's t error density,we thus have 4 applications, each of which involves the comparison of 4 different predictive density models. Results are gathered in Tables 5-8. The tables contain: mean square forecast errors - MSFE (so that our density accuracy results can be compared with model rankings based on conditional mean evaluation); lags used; $\int_U \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(1\{Inf_{t+1} \leq u\} - F_1(u|Z^t, \widehat{\theta}_{1,t})\right)^2 \phi(u)du = DMSFE$ (for "ranking" based on our density type mean square error measures), and {50,60,70,80,90} split and full sample bootstrap percentiles for block lengths of {3,5,10,15,20} observations (for conducting inference using $Z_{P,2}$).

Although this empirical application is presented only for illustrative purposes, we claim that the results presented in Tables 5-8 are indicative of the types of results that may generally be obtained upon application of the tools developed in this paper. For example, notice that lower MSFEs are uniformly associated with models that have lags selected via the AIC (see MSFE values in Tables 1-4). This rather surprising result suggests that parsimony is not always the best "rule of thumb" for selecting models for predicting conditional mean, and is a finding in agreement with one of the main conclusions of Marcellino, Stock and Watson (2004). Interestingly, though, the density based mean square forecast error measure that we consider (i.e. $DMSFE$) is not generally lower when the AIC is used. This suggests that the choice of lag selection criterion is sensitive to whether individual moments or entire distributions are being evaluated. Of further note is that $\max_{k=2,\dots,4} \int_U Z_{P,u,2}(1,k)\phi(u)du$ in Table 1 is -0.046, which fails to reject the null hypothesis that the benchmark AR(1)-normal density model is at least as "good" as any other SIC selected model. Furthermore, when only AR models are evaluated (see Table 3), there is nothing gained by using the AIC instead of the SIC, and the normality assumption is again not "bested" by assuming fatter predictive density tails (notice that in this case, failure to reject occurs even when 50th percentiles of either the split or full sample recursive block bootstrap distributions are used to form critical values). In contrast to the above results, when either the AIC is used for all competitor models (Table 2), or when only ARX models are considered with lags selected by either SIC or AIC (Table 4), the null hypothesis of normality is rejected using 90th percentile critical values. Further, in both of these cases, the "preferred model", based on ranking according to $DMSFE$, is (i) an ARX model with Student's t errors (when only the AIC is used to select lags) or (ii) an ARX model with Gaussian errors and lags selected via the SIC (when only ARX models are compared). This result indicates the importance of comparing a wide variety of models. If we were only to compare AR and ARX models using the AIC, as in Table 2, then we would conclude that ARX models beat AR models, and that fatter tails should replace Gaussian tails in error density specification. However, inspection of the density based MSFE measures across all models considered in the tables makes clear that the lowest $DMSFE$ values are always associated with more parsimonious models (with lags selected using the SIC) that assume Gaussianity.

6 Concluding Remarks

In this paper we discuss a test for predictive density accuracy. In addition, we provide a survey of related predictive density evaluation methods, and stress that our method differs from many of these in the sense that we allow all competing models to be misspecified. From a theoretical perspective, we outline 3 block bootstrap procedures applicable to a wide class of test statistics (those for which the limit distribution is a functional of Gaussian processes) constructed based on estimators obtained via rolling estimation schemes. Additionally, we survey 2 other block bootstrap procedures for recursive estimators due to Corradi and Swanson (2003a). The paper also contains a small Monte Carlo investigation that illustrates the sorts of coverage probabilities that might be expected upon use of the bootstrap procedures. Finally, an empirical example based on forecasting models of inflation is used to illustrate the predictive density accuracy test, and it is found that density evaluation based on AR models leaves nothing to choose between $AR(1)$ models under normality and models under alternative Student's t distributional assumptions and those with lags selected using the AIC instead of the SIC. On the other hand, when the lag selection device is fixed to be the AIC, then ARX predictive density models "win", and the Student's t distribution better mimics the actual distribution of the predictive density than the Gaussian distribution.

This paper is meant as a starting point. Much further research is needed, both theoretical and empirical, before the full impact of the bootstrap procedures and predictive density accuracy tests that we have outlined will become clear. For example, alternative bootstrap procedures such as the full sample procedure with additional adjustment terms discussed here need to be further developed and examined, both theoretically, and via Monte Carlo experimentation. Additionally, empirical and Monte Carlo investigation comparing and contrasting the various predictive density accuracy tests discussed in this paper remains to be done.

7 Appendix

The main theoretical contributions of this paper are contained in the proofs of Propositions 2 and 3, as the other propositions follow in a fairly straightforward manner, given the results of Corradi and Swanson (2003a,b).

Proof of Proposition 1: This proof requires a simple modification to the proof of Theorem 1 in Corradi and Swanson (2003b). In fact, the only difference is that in the current context parameters are estimated either recursively (see Corradi and Swanson (2003a) for further discussion of the recursive case), or using a rolling estimation scheme. Let $\mu_i^2(u) = E\left(\left(1\{y_{t+1} \leq u\} - F_i(u|Z^t, \theta_i^\dagger)\right)^2\right)$ $= E\left(\left(1\{y_{t+1} \leq u\} - F_0(u|Z^t, \theta_0)\right)^2\right) + E\left(\left(F_0(u|Z^t, \theta_0) - F_i(u|Z^t, \theta_i^\dagger)\right)^2\right)$. We begin by considering the rolling case. For any given u,

$$
\begin{aligned}
Z_{P,u,1}(1,k) = \frac{1}{\sqrt{P}} \sum_{t=R+s}^{T-1} \left(\left(1\{y_{t+1} \leq u\} - F_1(u|Z^t, \widehat{\theta}_{1,t,R})\right)^2 - \left(1\{y_{t+1} \leq u\} - F_k(u|Z^t, \widehat{\theta}_{k,t,R})\right)^2 \right)
\end{aligned}
$$

$$
\begin{aligned}
= \quad & \frac{1}{\sqrt{P}} \sum_{t=R+s}^{T-1} \left(\left(1\{y_{t+1} \leq u\} - F_1(u|Z^t, \widehat{\theta}_{1,t,R})\right)^2 - \mu_1^2(u) \right) \\
& - \frac{1}{\sqrt{P}} \sum_{t=R+s}^{T-1} \left(\left(1\{y_{t+1} \leq u\} - F_k(u|Z^t, \widehat{\theta}_{k,t,R})\right)^2 - \mu_k^2(u) \right) + \sqrt{P}(\mu_1^2(u) - \mu_k^2(u)) \\
= \quad & \frac{1}{\sqrt{P}} \sum_{t=R+s}^{T-1} \left(\left(1\{y_{t+1} \leq u\} - F_1(u|Z^t, \theta_1^\dagger)\right)^2 - \mu_1^2(u) \right) \\
& - \frac{1}{\sqrt{P}} \sum_{t=s}^{T-1} \left(\left(1\{y_{t+1} \leq u\} - F_k(u|Z^t, \theta_k^\dagger)\right)^2 - \mu_k^2(u) \right) \\
& - \frac{2}{P} \sum_{t=R+s}^{T-1} \nabla_{\theta_1} F_1(u|Z^t, \overline{\theta}_{1,t,R})' \left(1\{y_{t+1} \leq u\} - F_1(u|Z^t, \theta_1^\dagger)\right) \sqrt{P} \left(\widehat{\theta}_{1,t,R} - \theta_1^\dagger\right) \\
& + \frac{2}{P} \sum_{t=R+s}^{T-1} \nabla_{\theta_k} F_k(u|Z^t, \overline{\theta}_{k,t,R})' \left(1\{y_{t+1} \leq u\} - F_k(u|Z^t, \theta_k^\dagger)\right) \sqrt{P} \left(\widehat{\theta}_{k,t,R} - \theta_k^\dagger\right) \\
& + \sqrt{P}(\mu_1^2(u) - \mu_k^2(u)) + o_P(1)
\end{aligned}
$$

$$
\begin{aligned}
= & \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\left(1\{y_{t+1}\leq u\}-F_1(u|Z^t,\theta_1^\dagger)\right)^2-\mu_1^2(u)\right)\\
& -\frac{1}{\sqrt{P}}\sum_{t=s}^{T-1}\left(\left(1\{y_{t+1}\leq u\}-F_k(u|Z^t,\theta_k^\dagger)\right)^2-\mu_k^2(u)\right)\\
& -2m_{\theta_1^\dagger}(u)'A(\theta_1^\dagger)\frac{1}{\sqrt{P}}\sum_{t=R+s}^{T-1}\frac{1}{R}\sum_{j=t-R+1}^{t}\ln f_1(y_j,Z^{j-1},\theta_1)\\
& +2m_{\theta_k^\dagger}(u)'A(\theta_k^\dagger)\frac{1}{\sqrt{P}}\sum_{t=R+s}^{T-1}\frac{1}{R}\sum_{j=t-R+1}^{t}\ln f_k(y_j,Z^{j-1},\theta_k)\\
& +\sqrt{P}(\mu_1^2(u)-\mu_k^2(u))+o_P(1)
\end{aligned}
\tag{21}
$$

where $\overline{\theta}_{i,t,R}\in(\widehat{\theta}_{i,t,R},\theta_i^\dagger)$, $i=1,...,n$, and $m_{\theta_i^\dagger}(u)'=E\left(\nabla_{\theta_i}F_i(u|Z^t,\theta_i^\dagger)'\left(1\{y_{t+1}\leq u\}-F_i(u|Z^t,\theta_i^\dagger)\right)\right)$ and $A(\theta_i^\dagger)=\left(E\left(-\nabla^2_{\theta_i}\ln f_i(y_{t+1}|Z^t,\theta_i^\dagger)\right)\right)^{-1}$ and where the $o_P(1)$ term holds uniformly in $u\in U$. We need to distinguish between the case of $P\leq R$ and $P>R$. In the former case, by Lemma 4.1 in West and McCracken (1998, WM), $\frac{1}{\sqrt{P}}\sum_{t=R+s}^{T-1}\frac{1}{R}\sum_{j=t-R+1}^{t}\ln f_k(y_j,Z^{j-1},\theta_1^\dagger)$ is asymptotically normal with variance $\left(\pi-\frac{\pi^2}{3}\right)E\left(\sum_{j=-\infty}^{\infty}\nabla_{\theta_1}\ln f_1(y_{s+1}|Z^s,\theta_1^\dagger)\nabla_{\theta_1}\ln f_1(y_{s+j+1}|Z^{s+j},\theta_1^\dagger)'\right)$, while the long run covariance between
$\frac{1}{\sqrt{P}}\sum_{t=R+s}^{T-1}\frac{1}{R}\sum_{j=t-R+1}^{t}\ln f_k(y_j,Z^{j-1},\theta_1^\dagger)$ and
$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\left(1\{y_{t+1}\leq u\}-F_1(u|Z^t,\theta_1^\dagger)\right)^2-\mu_1^2(u)\right)$ is given by
$\frac{\pi}{2}E\left(\sum_{j=-\infty}^{\infty}\nabla_{\theta_1}\ln f_1(y_{s+1}|Z^s,\theta_1^\dagger)\left(\left(1\{y_{s+j+1}\leq u\}-F_k(u|Z^{s+j},\theta_k^\dagger)\right)^2-\mu_k^2(u)\right)\right)$. Again from Lemma 4.1 in WM, for the case of $P>R$, $\left(\pi-\frac{\pi^2}{3}\right)$ and $\frac{\pi}{2}$ are replaced by $\left(1-\frac{1}{3\pi}\right)$ and $\left(1-\frac{1}{2\pi}\right)$.
In the recursive case, the second last line in (21) becomes,

$$
-2m_{\theta_1^\dagger}(u)'A(\theta_1^\dagger)\frac{1}{\sqrt{P}}\sum_{t=R+s}^{T-1}\frac{1}{t}\sum_{j=s+1}^{t}\ln f_1(y_j,Z^{j-1},\theta_1)+2m_{\theta_k^\dagger}(u)'A(\theta_k^\dagger)\frac{1}{\sqrt{P}}\sum_{t=R+s}^{T-1}\frac{1}{t}\sum_{j=s+1}^{t}\ln f_k(y_j,Z^{j-1},\theta_k)
$$

and the asymptotic variance of the parameter estimation error component as well as the covariance term follow from Lemma A5 in West (1996). Finally, convergence of finite dimensional distributions and stochastic equicontinuity follows by the same argument as in the proof of Theorem 1 in Corradi and Swanson (2003b).
The proofs of Propositions 2 and 3 require three Lemmas, which are given below.

As the statement of Proposition 2 holds for $i=1,...,n$, and the proof is the same regardless which model we consider, for notational simplicity we drop the subscript i. Also, we only consider the case where $P\leq R$, as the case where $P>R$ follows straightforwardly using the same arguments.

Lemma A1: Let A1-A3 hold. Assume that for $P \leq R$, as $P \to \infty$ and $l_1 \to \infty$, $l_1/P \to 0$, and as $R \to \infty$ and $l_3 \to \infty$, $l_3/R \to 0$, and finally as $R-P \to \infty$ and $l_2 \to \infty$, $l_2/(R-P) \to 0$, then (i) $\sup_{t\geq R}\left|\widehat{\theta}^*_{t,R} - \widehat{\theta}_{t,R}\right| = o_{P^*}(1)$, $\Pr-P$, and (ii) $\sup_{t\geq R}\left|\widehat{\theta}^*_{t,R} - \theta^\dagger\right| = o_{P^*}(1)$, $\Pr-P$.[16]

Lemma A2: Let A1-A3 hold. If as $R \to \infty$ and $P \to \infty$, $l_1, l_3 \to \infty$, $l_1/P^{1/4} \to 0$ and $l_3/R^{1/4} \to 0$, then $\sup_{t\geq R} t^{\vartheta}\left|(\widehat{\theta}^*_{t,R} - \theta^\dagger)\right| = o_{P^*}(1)$, $\Pr-P$, for all $\vartheta < 0.5$.

Lemma A3: Let A1-A3 hold. If as $R \to \infty$ and $P \to \infty$, $l_1, l_3 \to \infty$, $l_1/P^{1/4} \to 0$ and $l_3/R^{1/4} \to 0$, then if $P/R \to \pi > 0$,

$$Var^*\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\frac{1}{R}\sum_{j=t-P+1}^{t+R}\left(\nabla_\theta q(y^*_j, Z^{*,j-1}, \theta^\dagger)\right)\right) = \Pi C_{00}, \;\; \Pr-P, \;\; t$$

where $C_{00} = \sum_{j=-\infty}^{\infty} E\left(\left(\nabla_\theta q(y_{1+s}, Z^s, \theta^\dagger)\right)\left(\nabla_\theta q(y_{1+s+j}, Z^{s+j}, \theta^\dagger)\right)'\right)$ and $\Pi = \pi - \pi^2/3$ for $P \leq R$ and $1 - \pi^2/3$ for $P > R$.

Proof of Lemma A1: (i) We need to extend the consistency results for bootstrap $m-$estimators of Goncalves and White (2003, Theorem 2.1), to the case of rolling $m-$estimators. Recalling that for $t \geq R+s$,

$$\widehat{\theta}_{t,R} = \arg\max_{\theta\in\Theta}\frac{1}{R}\sum_{j=t-R+1}^{t}\ln f(y_j, Z^{j-1}, \theta_i) \text{ and } \widehat{\theta}^*_{t,R} = \arg\max_{\theta\in\Theta}\frac{1}{R}\sum_{j=t-R+1}^{t}\ln f(y^*_j, Z^{*,j-1}, \theta_i)$$

and given that the argmax is a measurable function, and because of the unique identifiability conditions in A2(ii), it suffices to show that

$$\sup_{t\geq R+s}\sup_{\theta\in\Theta}\left|\frac{1}{R}\sum_{j=t-R+1}^{t}\left(\ln f(y^*_j, Z^{*,j-1}, \theta) - \ln f(y_j, Z^{j-1}, \theta)\right)\right| = o_{P^*}(1), \;\; \Pr-P.$$

Hereafter, for notational simplicity let $\ln f(y^*_j, Z^{*,j-1}, \theta) = q^*_j(\theta)$ and $\ln f(y_j, Z^{j-1}, \theta) = q_j(\theta)$, and let $\mu_\theta = E(q_j(\theta))$. Now,

$$\sup_{t\geq R+s}\sup_{\theta\in\Theta}\left|\frac{1}{R}\sum_{j=t-R+1}^{t}\left(q^*_j(\theta) - q_j(\theta)\right)\right| \leq \sup_{t\geq R+s}\sup_{\theta\in\Theta}\left|\frac{1}{R}\sum_{j=t-R+1}^{t}\left(q^*_j(\theta) - E^*\left(q^*_j(\theta)\right)\right)\right| \quad (22)$$

$$+\sup_{t\geq R+s}\sup_{\theta\in\Theta}\left|\frac{1}{R}\sum_{j=t-R+1}^{t}\left(q_j(\theta) - \mu_\theta\right)\right| + \sup_{t\geq R+s}\sup_{\theta\in\Theta}\left|\frac{1}{R}\sum_{j=t-R+1}^{t}\left(E^*\left(q^*_j(\theta)\right) - \mu_\theta\right)\right|. \quad (23)$$

[16] Recall that when $|P-R| = o(T)$, then the contribution of the observation in the range $|P-R|$ is negligible, whichever values we choose for l_2.

Now, assuming without loss of generality, $P \leq R$,

$$\begin{aligned}\frac{1}{R}\sum_{j=t-R+1}^{t} E^*(q_j^*(\theta)) &= \overline{h}_P(\theta)\frac{1\{j(t)\leq P\}}{R}+\overline{h}_{R-P}\frac{1\{P<j(t)\leq R\}}{R}+\overline{h}_{T-R}\frac{1\{R<j(t)\leq T-1\}}{R}\\ &\quad +O\left(\frac{l}{R}\right),\ \Pr-P\\ &= \overline{h}_P\alpha_1(t)+\overline{h}_{R-P}\alpha_2(t)+\overline{h}_{T-R}\alpha_3(t)+O\left(\frac{l}{T}\right),\ \ \Pr-P \qquad (24)\end{aligned}$$

uniformly in θ, as under A3, P and R grow at the same rate, as the sample size increases, and for $i=1,2,3$ $0\leq\alpha_i(t)\leq 1$ and $\sum_{i=1}^{3}\alpha_i(t)=1$. Also the $O(l/T)$ term holds uniformly in t. Therefore, the last term on the RHS of (23) writes as:

$$\sup_{t\geq R+s}\sup_{\theta\in\Theta}\left|\left(\overline{h}_P(\theta)-\mu_\theta\right)\alpha_1(t)\right|+\sup_{t\geq R+s}\sup_{\theta\in\Theta}\left|\left(\overline{h}_{R-P}(\theta)-\mu_\theta\right)\alpha_2(t)\right|+\sup_{t\geq R+s}\sup_{\theta\in\Theta}\left|\left(\overline{h}_P(\theta)-\mu_\theta\right)\alpha_2(t)\right| \qquad (25)$$

where given the mixing and moment conditions in A1 and A2, the first and third term on the RHS of (25) are $O(T^{-1/2})O(1)$, $\Pr-P$, because of the uniform law of large numbers, the second term is also $O(T^{-1/2})O(1)$, if $R-P=O(T)$, otherwise if $R-P=o(T)$, then is $O(1)o(1)$. Therefore the sum is (25) is $o(1)-\Pr-P$.

With regard to the first term on the RHS of (23), note that

$$\sup_{t\geq R+s}\sup_{\theta\in\Theta}\left|\frac{1}{R}\sum_{j=t-R+1}^{t}(q_j(\theta)-\mu_\theta)\right|\leq\sup_{t\geq R+s}\sup_{\theta\in\Theta}\left|\frac{1}{R}\sum_{j=s}^{t}(q_j(\theta)-\mu_\theta)\right|+\sup_{\theta\in\Theta}\left|\frac{1}{R}\sum_{j=s+1}^{R}(q_j(\theta)-\mu_\theta)\right|=o_p(1)$$

by the same argument as in the proof of Lemma A1 in CS (2003a). Finally, with regard to the RHS of (22), note that

$$\begin{aligned}&\sup_{t\geq R+s}\sup_{\theta\in\Theta}\left|\frac{1}{R}\sum_{j=t-R+1}^{t}\left(q_j^*(\theta)-E^*\left(q_j^*(\theta)\right)\right)\right|\\ \leq\ &\sup_{t\geq R+s}\sup_{\theta\in\Theta}\left|\frac{1}{R}\sum_{j=s+1}^{t}\left(q_j^*(\theta)-E^*\left(q_j^*(\theta)\right)\right)\right|+\sup_{t\geq R+s}\sup_{\theta\in\Theta}\left|\frac{1}{R}\sum_{j=s+1}^{R}\left(q_j^*(\theta)-E^*\left(q_j^*(\theta)\right)\right)\right|=o(1),\ \ \Pr-P,\end{aligned}$$

because of the uniform law of large number for heterogeneous, independent observations.

Proof of Lemma A2: Without loss of generality we consider the case of $P\leq R$. First note that,

$$t^{\vartheta}\left(\widehat{\theta}^*_{t,R}-\theta^\dagger\right)=\left(\frac{1}{t}\sum_{j=t-R+1}^{t}\nabla^2_\theta\ln f(y_j^*,Z^{*,j-1},\overline{\theta}^*_{t,R})\right)^{-1}\left(\frac{1}{t^{1-\vartheta}}\sum_{j=t-R+1}^{t}\nabla_\theta\ln f(y_j^*,Z^{*,j-1},\theta^\dagger)\right),$$

with $\overline{\theta}^*_{t,R} \in (\widehat{\theta}^*_{t,R}, \theta^\dagger)$. Hereafter, for notational simplicity let $\nabla_\theta \ln f(y^*_j, Z^{*,j-1}, \theta) = \nabla_\theta q^*_j(\theta)$ and $\nabla^2_\theta \ln f(y_j, Z^{j-1}, \theta) = \nabla^2_\theta q_j(\theta)$, and $A^\dagger = \left(E\left(-\nabla^2_\theta q_t(\theta^\dagger)\right)\right)^{-1}$,

$$\sup_{t\geq R+s}\left|\frac{1}{t}\sum_{j=t-R+1}^{t}\left(\nabla^2_\theta q^*_j(\overline{\theta}^*_{t,R}) - A^{\dagger-1}\right)\right| \leq \sup_{t\geq R+s}\left|\frac{1}{t}\sum_{j=t-R+1}^{t}\left(\nabla^2_\theta q^*_j(\overline{\theta}^*_{t,R}) - E^*\left(\nabla^2_\theta q^*_j(\overline{\theta}^*_{t,R})\right)\right)\right| \tag{26}$$

$$+\sup_{t\geq R+s}\left|\frac{1}{t}\sum_{j=t-R+1}^{t}\left(\nabla^2_\theta q_j(\overline{\theta}_{t,R}) - A^{\dagger-1}\right)\right| + \sup_{t\geq R+s}\left|\frac{1}{t}\sum_{j=s}^{t}\left(\nabla^2_\theta q_j(\overline{\theta}_{t,R}) - E^*\left(\nabla^2_\theta q^*_j(\overline{\theta}^*_{t,R})\right)\right)\right|, \tag{27}$$

with as $\overline{\theta}^*_{t,R} \in (\widehat{\theta}^*_{t,R}, \theta^\dagger)$ and $\overline{\theta}_{t,R} \in (\widehat{\theta}_{t,R}, \theta^\dagger)$. As for the RHS of (26),

$$\sup_{t\geq R+s}\left|\frac{1}{t}\sum_{j=t-R+1}^{t}\left(\nabla^2_\theta q^*_j(\overline{\theta}^*_{t,R}) - E^*\left(\nabla^2_\theta q^*_j(\overline{\theta}^*_{t,R})\right)\right)\right| \leq \sup_{t\geq R}\sup_{\theta\in\Theta}\left|\frac{1}{t}\sum_{j=t-R+1}^{t}\left(\nabla^2_\theta q^*_j(\theta) - E^*\left(\nabla^2_\theta q^*_j(\theta)\right)\right)\right|$$

First note that,

$$E^*\left(\nabla^2_\theta q^*_j(\theta)\right) = \frac{1}{P}\sum_{j=1}^{P}\nabla^2_\theta q_j(\theta)\alpha_1(t) + \frac{1}{R-P}\sum_{j=P+1}^{R}\nabla^2_\theta q_j(\theta)\alpha_2(t) + \frac{1}{T-R-1}\sum_{j=R+1}^{T-1}\nabla^2_\theta q_j(\theta)\alpha_3(t)$$

thus RHS of (26) is $o(1)$, $\Pr - P$, by the same argument as in Lemma A1. Given Lemma A1, $\sup_{t\geq R}\left|\overline{\theta}^*_{t,R} - \overline{\theta}_{t,R}\right| = o_{P^*}(1)$ $\Pr - P$, and $\sup_{t\geq R}\left|\overline{\theta}_{t,R} - \theta^\dagger\right| = o_P(1)$, thus the sum of the two terms in (27) is $o_{P^*}(1)$ $\Pr - P$, by the same argument used in the proof of Lemma A1.

Let $n_t = (2t\log\log t)^{1/2}$, and let $\nabla_\theta \ln f(y^*_j, Z^{*,j-1}, \theta) = h^*_j(\theta)$, and $\nabla_\theta \ln f(y_j, Z^{j-1}, \theta) = h_j(\theta)$,

$$\begin{aligned}\sup_{t\geq R+s}\left|\frac{1}{n_t}\sum_{j=t-R+1}^{t}h^*_j(\theta^\dagger)\right| &\leq \sup_{t\geq R+s}\left|\frac{1}{n_t}\sum_{j=t-R+1}^{t}\left(h^*_j(\theta^\dagger) - E^*\left(h^*_j(\theta^\dagger)\right)\right)\right| \\ &\quad + \sup_{t\geq R+s}\left|\frac{1}{n_t}\sum_{j=t-R+1}^{t}E^*\left(h^*_j(\theta^\dagger)\right)\right|,\end{aligned} \tag{28}$$

and noting that, by the same argument as in the proof of Lemma A1, up to a term of order $O(l/P^{1/2})$, $\Pr - P$, recalling that $\alpha_i \in (0,1]$ for $i = 1, 2.3$, there are constants C_1, C_2, C_3 such that,

$$\sup_{t\geq R+s}\left|\frac{1}{n_t}\sum_{j=t-R+1}^{t}E^*\left(h^*_j(\theta^\dagger)\right)\right| \leq C_1\left|\frac{1}{\sqrt{2R\log\log R}}\sum_{j=1}^{P}h_j(\theta^\dagger)\right|$$

$$+C_2\left|\frac{1}{\sqrt{2R\log\log R}}\sum_{j=P+1}^{R}h_j(\theta^\dagger)\right| + \sup_{t\geq R+s}\left|\frac{1}{\sqrt{2R\log\log R}}\sum_{j=R+1}^{t}h_j(\theta^\dagger)\right|, \tag{29}$$

and all the terms on the RHS of (29) are $O(1)$, $a.s. - P$, as, given A1 and A3 each terms satisfies the conditions for the functional law of the iterated logarithm (e.g. Theorem 2 in Eberlain (1986)).

It remains to show that the first term on the RHS of (28) is $O_{P^*}(1)$, $\Pr - P$. To further simplify the notation, we denote $h_j^*(\theta^\dagger)$ and $h_j(\theta^\dagger)$ as h_j^* and h_j, respectively. By a similar argument as in the proof of Lemma A2 in CS (2003a), it can be shown that $V^* = \lim_{T\to\infty} Var^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h_j^*(\theta^\dagger)\right)$ is $O(1)$, $\Pr - P$. The desired result then follows from Eberlain's (1986) law of iterated logarithm for dependent and heterogeneous process, given A1 and A2.

Proof of Lemma A3: As in the proof of Lemma A2, let $\nabla_\theta \ln f(y_j^*, Z^{*,j-1}, \theta) = h_j^*(\theta)$, and $\nabla_\theta \ln f(y_j, Z^{j-1}, \theta) = h_j(\theta)$, also let $\nabla_\theta \ln f(y_j^*, Z^{*,j-1}, \theta^\dagger) = h_j^*$, and $\nabla_\theta \ln f(y_j, Z^{j-1}, \theta^\dagger) = h_j$. Along the lines of West and McCracken (1998, proof of Lemma 4.1), for the case of $P \leq R$,

$$\frac{1}{\sqrt{P}}\sum_{t=R+s}^{T-1}\frac{1}{R}\sum_{j=t-R+1}^{t} h_j^* = \frac{1}{\sqrt{PR}}\sum_{j=s+1}^{P+s}(j-s)h_j^* + \frac{\sqrt{P}}{R}\sum_{j=P+s+1}^{R+s} h_j^* + \frac{1}{\sqrt{PR}}\sum_{j=R+s+1}^{T-1}\left(P-s-(j-R)\right)h_j^* \tag{30}$$

Thus,

$$Var^*\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\frac{1}{R}\sum_{j=t-R+1}^{t} h_j^*\right) = Var^*\left(\frac{1}{\sqrt{PR}}\sum_{j=s+1}^{P+s}(j-s)h_j^*\right)$$

$$+Var^*\left(\frac{\sqrt{P}}{R}\sum_{j=P+s+1}^{R+s} h_j^*\right) + Var^*\left(\frac{1}{\sqrt{PR}}\sum_{j=R+s+1}^{T-1}\left(P-s-(j-R)\right)h_j^*\right) + o(1)\ \Pr - P \tag{31}$$

where the $o(1)$ $\Pr - P$ term comes from the fact that the covariance term are $o(1)$ $\Pr - P$. In fact, given the resampling scheme outlined in Section 3.1.1 any block from the first b_i $i = 1, 2$ blocks can overlap with any of the following b_{i+1} blocks for at most s observations. We begin by analyzing the first term on the RHS of (31), Now, for $j \leq P$, $E^*\left(h_j^*\right) = \overline{h}_P + O(l/P)$, thus, given that s is finite, up to a term of order $O(l/P^{1/2})$,[17]

$$Var^*\left(\frac{1}{\sqrt{PR}}\sum_{j=s+1}^{P}(j-s)h_j^*\right) = Var^*\left(\frac{1}{\sqrt{PE}}\sum_{k=1}^{b_1}\sum_{i=1}^{l}((k-1)l+i)h_{I_k^1+i}\right)$$

[17] For notational simplicity, we start summation from 1 instead than from s. As s is finite, this has no consequence on the asymptotic behavior.

$$= E^*\left(\frac{1}{\sqrt{P}R}\sum_{k=1}^{b_1}\sum_{i=1}^{l}\sum_{j=1}^{l}((k-1)l+i)((k-1)l+j)(h_{I_k+i}-\overline{h}_P)(h_{I_k+j}-\overline{h}_P)'\right)$$

$$= \frac{1}{P}\frac{1}{R^2}\sum_{k=1}^{b_1}\sum_{i=1}^{l}\sum_{j=1}^{l}((k-1)l+i)((k-1)l+j)E^*\left((h_{I_k+i}-\overline{h}_P)(h_{I_k+j}-\overline{h}_P)'\right)$$

$$\begin{aligned}
&= \frac{1}{P}\frac{1}{R^2}\sum_{k=1}^{b_1}\sum_{i=1}^{l}\sum_{j=1}^{l}((k-1)l+i)((k-1)l+j)\left(\frac{1}{P}\sum_{t=l}^{P-l}(h_{t+i}-\overline{h}_P)(h_{t+j}-\overline{h}_P)'\right)+O(l/P^{1/2})\ \Pr-P\\
&= \frac{1}{P}\frac{1}{R^2}\sum_{k=1}^{b_1}\sum_{i=1}^{l}\sum_{j=1}^{l}((k-1)l+i)((k-1)l+j)\gamma_{|i-j|}\\
&\quad+\frac{1}{P}\frac{1}{R^2}\sum_{k=1}^{b_1}\sum_{i=1}^{l}\sum_{j=1}^{l}((k-1)l+i)((k-1)l+j)\left(\frac{1}{P}\sum_{t=l}^{P-l}\left((h_{t+i}-\overline{h}_P)(h_{t+j}-\overline{h}_P)'-\gamma_{i-j}\right)\right)\\
&\quad+O(l/P^{1/2})\ \Pr-P
\end{aligned}\tag{32}$$

We need to show that the last term on the last equality in (32) is $o(1)\ \Pr-P$. First, as for all k,i,j $\frac{((k-1)l+i)((k-1)l+j)}{R^2}\leq 1$, it is majorized by

$$\begin{aligned}
&\left|\frac{b_1}{P}\sum_{i=1}^{l}\sum_{j=1}^{l}\left(\frac{1}{P}\sum_{t=l}^{P-l}\left((h_{t+i}-\overline{h}_P)(h_{t+j}-\overline{h}_P)'-\gamma_{i-j}\right)\right)\right|\\
=&\left|\frac{1}{P}\sum_{t=l}^{P-l}\sum_{j=-l}^{l}\left((h_t-\overline{h}_P)(h_{t+j}-\overline{h}_P)'-\gamma_j\right)\right|+O(l/P^{1/2})\ \Pr-P
\end{aligned}\tag{33}$$

The first term on the RHS of (33) goes to zero in probability, by the same argument as in Lemma 2 in Corradi (1999)[18]. For the first term on the RHS of the last equality in (32), note that

$$\frac{1}{P}\frac{1}{R^2}\sum_{k=1}^{b_2}\sum_{i=1}^{l}\sum_{j=1}^{l}((k-1)l+i)((k-1)l+j)\gamma_{|i-j|}=\frac{1}{P}\sum_{t=l}^{P-l}\sum_{j=-l}^{l}t(t+j)\gamma_j+O(l/P^{1/2})\ \Pr-P$$

$$=\frac{1}{P}\frac{1}{R^2}\sum_{t=l}^{P-l}t^2\sum_{j=-l}^{l}\gamma_j+\frac{1}{P}\sum_{t=l}^{P-l}\sum_{j=-l}^{l}(t(t+j)-t^2)\gamma_j+O(l/P^{1/2})\ \Pr-P$$

By the same argument as in Lemma 4.1 in West and McCracken (1998), the second term on the RHS above approaches zero, while

$$\frac{1}{P}\sum_{t=l}^{P-l}t^2\sum_{j=-l}^{l}\gamma_j\to\frac{\pi^2}{3}C_{00}.$$

[18]The domination condition here are weaker than those in Lemma 2 in Corradi (1999) as we require only convergence to zero in probability and not almot surely.

By a similar argument, and following the proof of Lemma 4.1 in West and McCracken (1998), it can be shown that

$$\begin{aligned} Var^*\left(\frac{\sqrt{P}}{R}\sum_{j=P+s+1}^{R+s} h_j^*\right) &= (\pi-\pi^2)C_{00}+o_P(1) \\ Var^*\left(\frac{1}{\sqrt{P}R}\sum_{j=R+s+1}^{T-1}\left(P-s-(j-R)\right)h_j^*\right) &= \frac{\pi^2}{3}C_{00}+o_P(1). \end{aligned}$$

Finally, the case of $P>R$ can be treated along the same lines.

Proof of Proposition 2:

$$\frac{1}{P^{1/2}}\sum_{t=R+s}^{T-1}\left(\widehat{\theta}^*_{t,R}-\widehat{\theta}_{t,R}\right)=\frac{1}{P^{1/2}}\sum_{t=R+s}^{T-1}\left(\widehat{\theta}^*_{t,R}-\theta^\dagger\right)-\frac{1}{P^{1/2}}\sum_{t=R+s}^{T-1}\left(\widehat{\theta}_{t,R}-\theta^\dagger\right)$$

$$\begin{aligned} &= \frac{1}{P^{1/2}}\sum_{t=R+s}^{T-1}\left(-\frac{1}{R}\sum_{j=t-R+1}^{t}\nabla^2_\theta \ln f(y_j^*,Z^{*,j-1},\overline{\theta}^*_{t,R})\right)^{-1}\left(\frac{1}{R}\sum_{j=t-R+1}^{t}\nabla_\theta \ln f(y_j^*,Z^{*,j-1},\theta^\dagger)\right) \\ &\quad -\frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(-\frac{1}{R}\sum_{j=t-R+1}^{t}\nabla^2_\theta \ln f(y_j,Z^{j-1},\overline{\theta}_{t,R})\right)^{-1}\left(\frac{1}{R}\sum_{j=t-R+1}^{t}\nabla_\theta \ln f(y_j,Z^{j-1},\theta^\dagger)\right), \end{aligned} \quad (34)$$

where $\overline{\theta}^*_{t,R}\in\left(\widehat{\theta}^*_{t,R},\theta^\dagger\right)$ and $\overline{\theta}_{t,R}\in\left(\widehat{\theta}_{t,R},\theta^\dagger\right)$.

Given Lemma A1 and A2 and given A1-A3,

$$\begin{aligned} &\sup_{t\geq R+s}\left(\left(\frac{1}{R}\sum_{j=t-R+1}^{t}\nabla_\theta \ln f(y_j^*,Z^{*,j-1},\overline{\theta}^*_{t,R})\right)^{-1}-\left(\frac{1}{R}\sum_{j=t-R+1}^{t}\nabla_\theta \ln f(y_j,Z^{j-1},\overline{\theta}_{t,R})\right)^{-1}\right) \\ &= o_P^*(1),\ \ \Pr-P, \end{aligned}$$

and also

$$\sup_{t\geq R+s}\left(\left(-\frac{1}{R}\sum_{j=t-R+1}^{t}\nabla^2_\theta \ln f(y_j^*,Z^{*,j-1},\overline{\theta}^*_{t,R})\right)^{-1}-A^\dagger\right)=o_P^*(1),\ \ \Pr-P, \quad (35)$$

so the RHS of (34) can be written as:

$$\begin{aligned} &\frac{1}{P^{1/2}}\sum_{t=R+s}^{T-1}A^\dagger\left(\frac{1}{R}\sum_{j=t-R+1}^{t}\nabla_\theta \ln f(y_j^*,Z^{*,j-1},\theta^\dagger)-\frac{1}{R}\sum_{j=t-R+1}^{t}\nabla_\theta \ln f(y_j,Z^{j-1},\theta^\dagger)\right)+o_P^*(1),\ \Pr-P \\ &= A^\dagger\frac{1}{P^{1/2}}\sum_{t=R+s}^{T-1}\left(\frac{1}{R}\sum_{j=t-R+1}^{t}h_j^*-\frac{1}{R}\sum_{j=t-R+1}^{t}h_t\right)+o_P^*(1),\ \ \Pr-P, \end{aligned} \quad (36)$$

by letting $\nabla_\theta \ln f(y_j^*, Z^{*,j-1}, \theta^\dagger) = h_t^*$, $\nabla_\theta \ln f(y_j, Z^{j-1}, \theta^\dagger) = h_t$. Recalling (24), the RHS (36) for $P \leq R$, can be written as,

$$\begin{aligned}
& A^\dagger \frac{1}{\sqrt{PR}} \sum_{t=s+1}^{P+s} (t-s) \left(h_t^* - \overline{h}_P\right) + A^\dagger \frac{\sqrt{P}}{R} \sum_{t=P+s+1}^{R+s} (h_t^* - \overline{h}_{R-P}) \\
& + A^\dagger \frac{1}{\sqrt{PR}} \sum_{t=R+s+1}^{T-1} (P+s-(t-R)) \left(h_t^* - \overline{h}_{T-R}\right) \\
& - A^\dagger \frac{1}{\sqrt{PR}} \sum_{i=s+1}^{P+s-1} (i-s)(h_i - \overline{h}_P) - A^\dagger \frac{1}{\sqrt{PR}} \sum_{t=R+s+1}^{T-1} (P+s-(t-R)) \left(h_t - \overline{h}_{T-R}\right) \\
& + o_P^*(1), \ \ \Pr -P. \qquad (37)
\end{aligned}$$

The sum of the first three terms in (37) satisfies a central limit theorem for mixing triangular arrays (Wooldridge and White (1988)) and, by Lemma A3, has asymptotic variance equal to ΠC_{00}, which is the same as the asymptotic variance of $P^{-1/2} \sum_{t=R+s}^{T-1} (\widehat{\theta}_{t,R} - \theta_t^\dagger)$ (see Lemma 4.1, in West and McCracken (1998)), conditionally on the samples and for all samples but a subset of measure approaching zero. Therefore, it suffices to show that the last term on the RHS of (19), i.e. the adjustment term, is equal to $A^\dagger \frac{1}{\sqrt{P}} \sum_{i=s+1}^{P+s-1} (i-s)(h_i - \overline{h}_P) - A^\dagger \frac{1}{\sqrt{PR}} \sum_{t=R+s+1}^{T-1} (P+s-(t-R)) \left(h_t - h_t\right)$, up to a term vanishing asymptotically. Given A1 and A2, $\left(-\frac{1}{T} \sum_{t=s}^{T-1} \nabla_\theta^2 \ln f(y_t, Z^{t-1}, \widehat{\theta}_T)\right)^{-1} - A^\dagger = o(1)$ $\Pr -P$ (i.e. $o_P(1)$), where $\widehat{\theta}_T$ is the estimator constructed using all T observations.

Now let $h_t(\widehat{\theta}_T) = \nabla_\theta \ln f(y_t, Z^{t-1}, \widehat{\theta}_T)$, and $\overline{h}_P(\widehat{\theta})$, $\overline{h}_{R-P}(\widehat{\theta}_T)$, $\overline{h}_{T-R}(\widehat{\theta}_T)$ be defined as $\overline{h}_P$, $\overline{h}_{R-P}$, $\overline{h}_{T-R}$ with $\theta^\dagger$ replaced by $\widehat{\theta}_T$, and let $\nabla^2 h_t(\widehat{\theta}_T) = \nabla_\theta^2 \ln f(y_t, Z^{t-1}, \widehat{\theta}_T)$. Now,

$$\begin{aligned}
& A^\dagger \frac{1}{\sqrt{PR}} \sum_{t=s+1}^{P+s-1} (t-s) \left(\left(h_t(\widehat{\theta}_T) - \overline{h}_P(\widehat{\theta}_T)\right) - \left(h_t - \overline{h}_P\right)\right) \\
= \ & A^\dagger \frac{1}{PR} \sum_{t=s+1}^{P+s-1} (t-s) \left(\nabla^2 h_t(\overline{\theta}_T) - \overline{\nabla^2 h}_P(\overline{\theta}_T)\right) \sqrt{P} \left(\widehat{\theta}_T - \theta^\dagger\right) = o(1), \ \ \Pr -P, \qquad (38)
\end{aligned}$$

as $\sqrt{P}\left(\widehat{\theta}_T - \theta_0^\dagger\right) = O(1)$ $\Pr -P$, and by the uniform law of large numbers for mixing triangular arrays, $\frac{1}{PR} \sum_{t=s+1}^{P+s-1} (t-s) \left(\nabla^2 h_{R+i}(\overline{\theta}_T) - \overline{\nabla^2 h}_P(\overline{\theta}_T)\right) = o(1)$ $\Pr -P$. The term concerning observations from $P+s+1$ to $R+s$ does not require adjustment, as all observations carry the same weights, the term concerning observations from $R+s+1$ to $T-1$ can be treated as above.

Proof of Proposition 3: We consider only the case of $P \leq R$. The fact that $\frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \left(\widehat{\beta}_{t,R}^* - \widehat{\beta}_{t,R}\right)$ has the same limiting distribution as in Proposition 1, follows by exactly the same arguments as in the proofs of Lemmas A1-A3 and Proposition 2 above. Now,

$$\widehat{\sigma}^2_{t,R} = \frac{1}{R}\sum_{j=t-R+1}^{t} \widehat{\epsilon}^2_j = \frac{1}{R}\sum_{j=t-R+1}^{t}\left(y_j - g(Z^{j-1}, \widehat{\beta}_{t,R})\right)^2 = \frac{1}{R}\sum_{j=t-R+1}^{t} \epsilon^2_j + O_P(R^{-1/2}),$$

as for $\vartheta < 1/2$, $\sup_{t\geq R} t^{\vartheta}\left(\widehat{\beta}_{t,R} - \beta^{\dagger}\right) = o_p(1)$ (see Lemma A1 in West and McCracken (1998)) and $\sup_{t\geq R}\sqrt{R}\left(\widehat{\beta}_{t,R} - \beta^{\dagger}\right)\frac{1}{R}\sum_{j=t-R+1}^{t}\epsilon_j = o_P(1)$. Thus,

$$\sqrt{R}\left(\widehat{\sigma}^2_{t,R} - \sigma^{2\dagger}\right) = \frac{1}{\sqrt{R}}\sum_{j=t-R+1}^{t}\left(\widehat{\epsilon}^2_j - \sigma^{2\dagger}\right) + o_P(1).$$

Now,

$$\begin{aligned}
\widehat{\sigma}^{2*}_{t,R} &= \frac{1}{R}\sum_{j=t-R+1}^{t}\widehat{\epsilon}^{*2}_j = \frac{1}{R}\sum_{j=t-R+1}^{t}\left(y^*_j - g(Z^{*,j-1}, \widehat{\beta}^*_{t,R})\right)^2 \\
&= \frac{1}{R}\sum_{j=t-R+1}^{t}\left(y^*_j - g(Z^{*,j-1}, \widehat{\beta}_{t,R})\right)^2 = \frac{1}{R}\sum_{j=t-R+1}^{t}\epsilon^{*2}_j + O_{P^*}(R^{-1/2}), \ \ \Pr -P,
\end{aligned}$$

where the last equality on the RHS of the above equation follows given Lemma A1 in West and McCracken (1998) and Lemma A3 (since these results in turn ensure that $\sup_{t\geq R} t^{\vartheta}\left(\widehat{\beta}^*_{t,R} - \widehat{\beta}_{t,R}\right) = o_{P^*}(1)$, $\Pr -P$ and $\sup_{t\geq R}\sqrt{R}\left(\widehat{\beta}^*_{t,R} - \widehat{\beta}_{t,R}\right)\frac{1}{R}\sum_{j=t-R+1}^{t}\epsilon^*_j = o_{P^*}(1)$ $\Pr -P$). Thus,

$$\sqrt{R}\left(\widehat{\sigma}^{2*}_{t,R} - \widehat{\sigma}^2_{t,R}\right) = \frac{1}{\sqrt{R}}\sum_{j=t-R+1}^{t}\left(\widehat{\epsilon}^{*2}_j - \widehat{\epsilon}^2_j\right)$$

and up to a $o_P(1)$ term,

$$\begin{aligned}
\frac{1}{P^{1/2}}\sum_{t=R+s}^{T-1}\left(\widehat{\sigma}^{2*}_{t,R} - \widehat{\sigma}^2_{t,R}\right) &= \frac{1}{\sqrt{P}R}\sum_{t=R+s}^{T-1}\sum_{j=t-R+1}^{t}\left(\widehat{\epsilon}^{*2}_j - \widehat{\epsilon}^2_j\right) = \frac{1}{\sqrt{P}R}\sum_{j=s+1}^{P+s}(j-s)\left(\widehat{\epsilon}^{*2}_j - \widehat{\epsilon}^2_j\right) \\
&+ \frac{\sqrt{P}}{R}\sum_{j=P+s+1}^{R+s}\left(\widehat{\epsilon}^{*2}_j - \widehat{\epsilon}^2_j\right) + \frac{1}{\sqrt{P}R}\sum_{j=R+s+1}^{T-1}(P+s-(j-R))\left(\widehat{\epsilon}^{*2}_j - \widehat{\epsilon}^2_j\right). \qquad (39)
\end{aligned}$$

Letting $\overline{\widehat{\epsilon}}^2_P = \frac{1}{P}\sum_{i=1}^{P}\widehat{\epsilon}^2_i$ and $\overline{\widehat{\epsilon}}^2_{R-P}$, $\overline{\widehat{\epsilon}}^2_{T-R}$ defined in an analogous way. So, the LHS of (39) can be written as:

$$\begin{aligned}
&\left[\frac{1}{\sqrt{P}R}\sum_{j=s+1}^{P+s}(j-s)\left(\widehat{\epsilon}^{*2}_j - \overline{\widehat{\epsilon}}^2_P\right) + \frac{\sqrt{P}}{R}\sum_{j=P+s+1}^{R+s}\left(\widehat{\epsilon}^{*2}_j - \overline{\widehat{\epsilon}}^2_{R-P}\right)\right. \\
&\left.\frac{1}{\sqrt{P}R}\sum_{j=R+s+1}^{T-1}(P+s-(j-R))\left(\widehat{\epsilon}^{*2}_j - \overline{\widehat{\epsilon}}^2_{T-R}\right)\right] \\
&-\left[\frac{1}{\sqrt{P}R}\sum_{j=s+1}^{P+s}(j-s)\left(\widehat{\epsilon}^2_j - \overline{\widehat{\epsilon}}^2_P\right) + \frac{1}{\sqrt{P}R}\sum_{j=R+s+1}^{T-1}(P+s-(j-R))\left(\widehat{\epsilon}^2_j - \overline{\widehat{\epsilon}}^2_{T-R}\right)\right]. \qquad (40)
\end{aligned}$$

Note that for $j = s+1, ..., P+s$ $E^*(\widehat{\epsilon}_j^{*2}) = \overline{\overline{\epsilon}}_P^2 + O(l/P)$, by the same argument use in the proofs of Lemmas A1-A3 and Theorem 1, the term in the first square bracket in (40) has the same limiting distribution as $\sqrt{R}\left(\widehat{\sigma}_{t,R}^2 - \sigma^{2\dagger}\right)$, conditional on the sample and for all samples but a set of probability measure approaching zero. The term in the second square bracket is the adjustment term used in the construction of $\Phi_{R,P,1}^*$. The case of $P > R$ can be treated in an analogous fashion.

Proof of Proposition 4: This proof follows from Theorem 1 in Corradi and Swanson (2003a).

Proof of Proposition 5: This proof follows using arguments similar to those used in the proof of Proposition 3.

Proof of Proposition 6: The proof to this proposition follows as a straightforward modification of Proposition 7 in Corradi and Swanson (2003a).

8 References

Andrews, D.W.K., (1997), A Conditional Kolmogorov Test, *Econometrica,* 65, 1097-1128.

Andrews, D.W.K., (2002a), Higher-Order Improvements of a Computationally Attractive $k-$step Bootstrap for Extremum Estimators, *Econometrica,* 70, 119-162.

Andrews, D.W.K., (2002b), The Block-Block Bootstrap: Improved Asymptotic Refinements, Manuscript, Cowles Foundation in Economics, Yale University.

Bai, J., (2003), Testing Parametric Conditional Distributions of Dynamic Models, *Review of Economics and Statistics,* 85, 531-549.

Bontemps, C., and N. Meddahi, (2003a), Testing Normality: a GMM Approach, *Journal of Econometrics*, forthcoming.

Bontemps, C., and N. Meddahi, (2003b), Testing Distributional Assumptions: a GMM Approach, Working Paper, University of Montreal.

Chang, Y.S., J.F. Gomes, and F. Schorfheide, (2002), Learning-by-Doing as a Propagation Mechanism, *American Economic Review*, 92, 1498-1520.

Christoffersen, P.F., (1998), Evaluating Interval Forecasts, *International Economic Review*, 39, 841-862.

Christoffersen, P.F., J. Hahn and A. Inoue, (2001), Testing and Comparing Value-at-Risk Measures, *Journal of Empirical Finance,* 8, 325-342.

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

Clark, T.E. and M.W. McCracken, (2003), Forecast Based Model Selection in the Presence of Structural Breaks, *Journal of Econometrics,* forthcoming.

Corradi, V., (1999), Deciding between $I(0)$ and $I(1)$ via FLIL-Based Bounds, *Econometric Theory,* 15, 643-663.

Corradi, V. and N.R. Swanson, (2002), A Consistent Test for Out of Sample Nonlinear Predictive Ability, *Journal of Econometrics,* 110, 353-381.

Corradi, V. and N.R. Swanson, (2003a), The Block Bootstrap for Parameter Estimation Error In Recursive Estimation Schemes, With Applications to Predictive Evaluation, Working Paper, Queen Mary, University of London and Rutgers University.

Corradi, V., N.R. Swanson, (2003b), A Test For Comparing Multiple Misspecified Conditional Distribution Models, Working Paper, Queen Mary, University of London and Rutgers University.

Corradi, V. and N.R. Swanson, (2003c), Bootstrap Conditional Distribution Tests in the Presence of Dynamic Misspecification, *Journal of Econometrics,* forthcoming.

Diebold, F.X., T. Gunther and A.S. Tay, (1998), Evaluating Density Forecasts with Applications to Finance and Management, *International Economic Review,* 39, 863-883.

Diebold, F.X., J. Hahn and A.S. Tay, (1999), Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High Frequency Returns on Foreign Exchange, *Review of Economics and Statistics*, 81, 661-673.

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

Duffie, D., and J. Pan, (1997), An Overview of Value at Risk, *Journal of Derivatives*, 4, 7-49.

Eberlain, E., (1986), On Strong Invariance Principles under Dependence Assumptions, *Annals of Probability,* 14, 260-270.

Fernandez-Villaverde, J., and J.F. Rubio-Ramirez, (2001), Comparing Dynamic Equilibrium Models to Data, Working Paper, University of Pennsylvania.

Gallant, A.R. and H. White, (1988), *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models,* Blackwell, Oxford.

Giacomini, R. (2002), Comparing Density Forecasts via Weighted Likelihood Ratio Tests: Asymptotic and Bootstrap Methods, Working Paper, University of California, San Diego.

Giacomini, R., and H. White, (2003), Conditional Tests for Predictive Ability, Working Paper, University of California, San Diego.

Goncalves, S., and H. White, (2003), Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models, *Journal of Econometrics,* forthcoming.

Hall, P., and J.L. Horowitz, (1996), Bootstrap Critical Values for Tests Based on Generalized Method of Moments Estimators, *Econometrica,* 64, 891-916.

Hansen, P.R., (2001), An Unbiased and Powerful Test for Superior Predictive Ability, Working Paper, Brown University.

Hong, Y., (2001), Evaluation of Out of Sample Probability Density Forecasts with Applications to S&P 500 Stock Prices, Working Paper, Cornell University.

Hong, Y.M., and H. Li, (2003) Nonparametric Specification Testing for Continuous Time Models with Applications to Term Structure Interest Rates, *Review of Financial Studies,* forthcoming.

Inoue, A., and M. Shintani, (2003), Bootstrapping GMM Estimators for Time Series, Mimeo, *Journal of Econometrics,* forthcoming.

Kitamura, Y., (2002), Econometric Comparisons of Conditional Models, Working Paper, University of Pennsylvania.

Künsch H.R., (1989), The Jackknife and the Bootstrap for General Stationary Observations, *Annals of Statistics,* 17, 1217-1241.

Linton, O., E. Maasoumi and Y.J. Whang, (2003), Consistent Testing for Stochastic Dominance Under General Sampling Schemes, Manuscript, LSE, Southern Methodist University and Ewha University.

Marcellino, M., J. Stock and M. Watson, (2004), A Comparison of Direct and Iterated AR Methods for Forecasting Macroeconomic Series h-Steps Ahead, Working Paper, Princeton University.

Pesaran M. H., and A. Timmerman, (2003), How Costly is to Ignore Breaks when Forecasting the Direction of a Time Series? *International Journal of Forecasting,* forthcoming.

Politis, D.N., J.P. Romano and M. Wolf, (1999), *Subsampling,* Springer and Verlag, New York.

Schörfheide, F., (2000), Loss Function Based Evaluation of DSGE Models, *Journal of Applied Econometrics,* 15, 645-670.

West, K.D., (1996), Asymptotic Inference About Predictive Ability, *Econometrica*, 64, 1067-1084.

West, K.D., and M.W. McCracken, (1998), Regression-Based Tests for Predictive Ability, *International Economic Review*, 39, 817-840.

White, H., (2000), A Reality Check for Data Snooping, *Econometrica*, 68, 1097-1126.

Wooldridge, J.M. and H. White, (1988), Some Invariance Principles and Central Limit Theorems for Dependent and Heterogeneous Processes, Econometric Theory, 4, 210-230.

Table 1: Finite Sample Properties: Rolling and Recursive PEE Bootstrap: Part I$^{(*)}$

smpl	*boot*	*coeff*	$l=4$	$l=6$	$l=10$	$l=12$	$l=15$	$l=20$	$l=25$	$l=30$	$l=50$	$l=60$
Panel A: DGP is an AR(1) Process - $\rho=0.2$												
800	*roll*1	$\hat{\rho}$	1.000	1.000	0.270	0.585	0.800	0.890	0.870	0.835	0.825	0.845
		$\hat{\rho}_1$	0.840	0.800	0.295	0.620	0.805	0.895	0.885	0.830	0.825	0.850
		$\hat{\rho}_2$	0.795	0.790	0.760	0.800	0.885	0.865	0.840	0.880	0.845	0.875
	*roll*2	$\hat{\rho}$	0.790	0.800	0.310	0.510	0.720	0.760	0.785	0.730	0.690	0.755
		$\hat{\rho}_1$	0.715	0.720	0.300	0.565	0.705	0.765	0.795	0.745	0.720	0.755
		$\hat{\rho}_2$	0.730	0.695	0.700	0.720	0.735	0.780	0.720	0.780	0.705	0.715
	*rec*1	$\hat{\rho}$	0.675	0.665	0.260	0.730	0.850	0.870	0.935	0.925	0.860	0.885
		$\hat{\rho}_1$	0.820	0.855	0.290	0.745	0.860	0.920	0.930	0.910	0.900	0.890
		$\hat{\rho}_2$	0.825	0.880	0.775	0.880	0.930	0.915	0.910	0.915	0.885	0.880
	*rec*2	$\hat{\rho}$	0.810	0.780	0.260	0.730	0.850	0.855	0.905	0.900	0.855	0.865
		$\hat{\rho}_1$	0.795	0.820	0.305	0.735	0.845	0.880	0.910	0.895	0.890	0.865
		$\hat{\rho}_2$	0.780	0.835	0.750	0.865	0.905	0.910	0.870	0.865	0.885	0.850
1600	*roll*1	$\hat{\rho}$	0.875	0.850	0.765	1.000	0.070	0.410	0.680	0.755	0.915	0.860
		$\hat{\rho}_1$	0.880	0.900	0.875	0.835	0.085	0.435	0.710	0.740	0.910	0.850
		$\hat{\rho}_2$	0.880	0.900	0.875	0.830	0.750	0.800	0.870	0.855	0.835	0.855
	*roll*2	$\hat{\rho}$	0.850	0.845	0.870	0.810	0.120	0.425	0.605	0.675	0.755	0.760
		$\hat{\rho}_1$	0.755	0.770	0.755	0.755	0.120	0.435	0.640	0.685	0.785	0.795
		$\hat{\rho}_2$	0.770	0.780	0.730	0.700	0.705	0.690	0.785	0.720	0.705	0.775
	*rec*1	$\hat{\rho}$	0.725	0.725	0.760	0.690	0.115	0.555	0.765	0.885	0.890	0.925
		$\hat{\rho}_1$	0.895	0.910	0.910	0.900	0.115	0.555	0.780	0.890	0.905	0.930
		$\hat{\rho}_2$	0.920	0.935	0.915	0.880	0.810	0.895	0.915	0.960	0.940	0.940
	*rec*2	$\hat{\rho}$	0.970	0.945	0.895	0.895	0.120	0.500	0.740	0.855	0.885	0.920
		$\hat{\rho}_1$	0.865	0.890	0.875	0.880	0.120	0.530	0.770	0.870	0.890	0.925
		$\hat{\rho}_2$	0.890	0.890	0.850	0.890	0.820	0.870	0.895	0.925	0.915	0.880
Panel B: DGP is an AR(2) Process - $\rho=0.1$												
800	*roll*1	$\hat{\rho}$	0.865	0.895	0.825	0.865	0.845	1.000	1.000	0.565	0.725	0.855
		$\hat{\rho}_1$	0.845	0.820	0.830	0.805	0.830	0.820	0.730	0.615	0.745	0.860
		$\hat{\rho}_2$	0.840	0.830	0.835	0.830	0.830	0.825	0.755	0.660	0.765	0.830
	*roll*2	$\hat{\rho}$	0.900	0.885	0.870	0.875	0.840	0.790	0.765	0.540	0.635	0.780
		$\hat{\rho}_1$	0.705	0.710	0.730	0.680	0.720	0.705	0.580	0.570	0.660	0.790
		$\hat{\rho}_2$	0.700	0.725	0.720	0.710	0.725	0.725	0.600	0.605	0.645	0.745
	*rec*1	$\hat{\rho}$	0.775	0.765	0.725	0.760	0.725	0.695	0.645	0.620	0.820	0.875
		$\hat{\rho}_1$	0.905	0.920	0.860	0.920	0.910	0.860	0.805	0.640	0.860	0.915
		$\hat{\rho}_2$	0.910	0.925	0.865	0.930	0.925	0.870	0.835	0.705	0.830	0.890
	*rec*2	$\hat{\rho}$	0.900	0.920	0.950	0.920	0.925	0.830	0.855	0.610	0.815	0.850
		$\hat{\rho}_1$	0.890	0.880	0.830	0.895	0.870	0.840	0.760	0.630	0.845	0.880
		$\hat{\rho}_2$	0.880	0.900	0.840	0.890	0.890	0.825	0.780	0.705	0.805	0.875
1600	*roll*1	$\hat{\rho}$	0.775	0.810	0.870	0.870	0.840	0.845	0.800	0.820	1.000	0.395
		$\hat{\rho}_1$	0.685	0.855	0.850	0.875	0.825	0.835	0.895	0.795	0.805	0.450
		$\hat{\rho}_2$	0.695	0.885	0.850	0.890	0.835	0.835	0.915	0.790	0.800	0.515
	*roll*2	$\hat{\rho}$	0.745	0.870	0.875	0.900	0.855	0.875	0.855	0.855	0.880	0.440
		$\hat{\rho}_1$	0.595	0.740	0.710	0.755	0.725	0.720	0.775	0.675	0.685	0.490
		$\hat{\rho}_2$	0.620	0.790	0.695	0.750	0.740	0.730	0.775	0.680	0.675	0.505
	*rec*1	$\hat{\rho}$	0.695	0.735	0.795	0.820	0.730	0.775	0.730	0.765	0.780	0.505
		$\hat{\rho}_1$	0.745	0.920	0.900	0.905	0.925	0.930	0.890	0.890	0.925	0.545
		$\hat{\rho}_2$	0.775	0.920	0.905	0.915	0.930	0.930	0.905	0.885	0.920	0.550
	*rec*2	$\hat{\rho}$	0.800	0.910	0.910	0.930	0.915	0.920	0.885	0.910	0.915	0.495
		$\hat{\rho}_1$	0.735	0.890	0.870	0.880	0.905	0.920	0.875	0.860	0.890	0.550
		$\hat{\rho}_2$	0.760	0.905	0.880	0.890	0.905	0.925	0.880	0.855	0.890	0.565

$^{(*)}$ Notes:

$DMSFE{=}\int_U \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(1\{Inf_{t+1}\leq u\}-F_1(u|Z^t,\widehat{\theta}_{1,t})\right)^2\phi(u)du.$

The second column lists the bootstrap used to examine parameter estimation error (PEE) associated with either an AR(1) autoregressive parameter ($\hat{\rho}$) or two autoregressive parameters from an AR(2) model ($\hat{\rho}_1$ and $\hat{\rho}_2$). Additionally, two different DGPs are used to generate data; and AR(1) DGP slope parameter equal to 0.2 and an AR(2) with slope parameters both equal to 0.1. (See Tables 2-4 for alternative parameterizations.) Bootstrap mnemonics ending with a "1" denote methods that account for PEE, while those ending with a "2" indicate the the same rolling (or recursive) bootstrap procedure was used, but with no adjustment terms. Numerical entires are 90% coverage probabilities. In all experiments, 500 Monte Carlo iterations were carried out (see above for further details).

Table 2: Finite Sample Properties: Rolling and Recursive PEE Bootstrap: Part II$^{(*)}$

$smpl$	$boot$	$coeff$	$l=4$	$l=6$	$l=10$	$l=12$	$l=15$	$l=20$	$l=25$	$l=30$	$l=50$	$l=60$
Panel A: DGP is an AR(1) Process - $\rho=0.4$												
800	$roll1$	$\hat{\rho}$	0.015	0.200	0.590	0.725	0.770	0.805	0.810	0.800	0.720	0.770
		$\hat{\rho}_1$	0.020	0.265	0.630	0.720	0.805	0.825	0.830	0.805	0.775	0.760
		$\hat{\rho}_2$	0.790	0.835	0.825	0.870	0.865	0.845	0.850	0.840	0.750	0.780
	$roll2$	$\hat{\rho}$	0.025	0.240	0.560	0.680	0.645	0.670	0.760	0.695	0.605	0.680
		$\hat{\rho}_1$	0.035	0.315	0.580	0.660	0.680	0.680	0.755	0.745	0.695	0.695
		$\hat{\rho}_2$	0.720	0.715	0.735	0.785	0.745	0.745	0.750	0.730	0.645	0.685
	$rec1$	$\hat{\rho}$	0.015	0.190	0.575	0.700	0.780	0.810	0.800	0.795	0.785	0.780
		$\hat{\rho}_1$	0.015	0.255	0.605	0.745	0.855	0.840	0.780	0.790	0.820	0.805
		$\hat{\rho}_2$	0.710	0.835	0.835	0.865	0.875	0.830	0.830	0.845	0.780	0.805
	$rec2$	$\hat{\rho}$	0.010	0.225	0.560	0.680	0.785	0.760	0.785	0.775	0.760	0.775
		$\hat{\rho}_1$	0.025	0.270	0.600	0.730	0.820	0.800	0.745	0.780	0.770	0.785
		$\hat{\rho}_2$	0.695	0.815	0.780	0.855	0.830	0.795	0.790	0.825	0.745	0.765
1600	$roll1$	$\hat{\rho}$	0.000	0.070	0.390	0.565	0.815	0.815	0.805	0.870	0.870	0.830
		$\hat{\rho}_1$	0.000	0.115	0.460	0.630	0.835	0.835	0.845	0.885	0.870	0.855
		$\hat{\rho}_2$	0.780	0.855	0.910	0.855	0.895	0.890	0.845	0.880	0.880	0.835
	$roll2$	$\hat{\rho}$	0.000	0.090	0.420	0.510	0.720	0.700	0.700	0.770	0.750	0.710
		$\hat{\rho}_1$	0.005	0.110	0.475	0.615	0.735	0.720	0.750	0.805	0.745	0.705
		$\hat{\rho}_2$	0.745	0.745	0.795	0.750	0.750	0.800	0.760	0.775	0.750	0.690
	$rec1$	$\hat{\rho}$	0.000	0.045	0.360	0.535	0.720	0.775	0.815	0.900	0.805	0.830
		$\hat{\rho}_1$	0.000	0.070	0.430	0.605	0.760	0.775	0.870	0.910	0.850	0.830
		$\hat{\rho}_2$	0.815	0.855	0.875	0.850	0.885	0.855	0.875	0.820	0.810	0.885
	$rec2$	$\hat{\rho}$	0.000	0.060	0.360	0.540	0.715	0.755	0.775	0.850	0.765	0.775
		$\hat{\rho}_1$	0.000	0.085	0.440	0.605	0.735	0.750	0.840	0.870	0.815	0.780
		$\hat{\rho}_2$	0.805	0.855	0.830	0.820	0.840	0.820	0.830	0.800	0.800	0.835
Panel B: DGP is an AR(2) Process - $\rho=0.2$												
800	$roll1$	$\hat{\rho}$	0.195	0.545	0.750	0.825	0.795	0.790	0.865	0.835	0.815	0.865
		$\hat{\rho}_1$	0.260	0.595	0.780	0.885	0.800	0.835	0.855	0.855	0.820	0.860
		$\hat{\rho}_2$	0.330	0.670	0.855	0.870	0.870	0.825	0.865	0.870	0.795	0.805
	$roll2$	$\hat{\rho}$	0.235	0.500	0.605	0.720	0.705	0.675	0.775	0.735	0.690	0.740
		$\hat{\rho}_1$	0.305	0.565	0.665	0.760	0.725	0.740	0.750	0.730	0.710	0.730
		$\hat{\rho}_2$	0.330	0.595	0.735	0.765	0.790	0.735	0.735	0.765	0.690	0.720
	$rec1$	$\hat{\rho}$	0.190	0.525	0.715	0.790	0.840	0.825	0.800	0.825	0.790	0.720
		$\hat{\rho}_1$	0.285	0.610	0.730	0.815	0.840	0.870	0.795	0.860	0.780	0.735
		$\hat{\rho}_2$	0.300	0.610	0.815	0.825	0.845	0.840	0.845	0.865	0.765	0.815
	$rec2$	$\hat{\rho}$	0.205	0.525	0.655	0.780	0.800	0.785	0.775	0.790	0.780	0.670
		$\hat{\rho}_1$	0.280	0.575	0.705	0.800	0.800	0.835	0.775	0.835	0.770	0.705
		$\hat{\rho}_2$	0.305	0.590	0.770	0.790	0.795	0.830	0.805	0.840	0.755	0.765
1600	$roll1$	$\hat{\rho}$	0.065	0.375	0.665	0.765	0.780	0.870	0.880	0.855	0.845	0.835
		$\hat{\rho}_1$	0.100	0.415	0.690	0.780	0.815	0.855	0.885	0.875	0.840	0.835
		$\hat{\rho}_2$	0.090	0.515	0.775	0.755	0.855	0.905	0.900	0.845	0.815	0.870
	$roll2$	$\hat{\rho}$	0.125	0.395	0.625	0.710	0.690	0.730	0.805	0.710	0.780	0.725
		$\hat{\rho}_1$	0.160	0.430	0.665	0.730	0.720	0.720	0.820	0.705	0.770	0.730
		$\hat{\rho}_2$	0.120	0.520	0.730	0.680	0.760	0.785	0.770	0.740	0.750	0.750
	$rec1$	$\hat{\rho}$	0.055	0.335	0.645	0.685	0.850	0.860	0.845	0.835	0.820	0.840
		$\hat{\rho}_1$	0.095	0.405	0.695	0.725	0.855	0.890	0.855	0.845	0.835	0.850
		$\hat{\rho}_2$	0.105	0.475	0.745	0.740	0.805	0.840	0.845	0.855	0.830	0.820
	$rec2$	$\hat{\rho}$	0.060	0.355	0.615	0.650	0.810	0.855	0.805	0.815	0.795	0.805
		$\hat{\rho}_1$	0.100	0.425	0.640	0.740	0.785	0.860	0.800	0.815	0.805	0.805
		$\hat{\rho}_2$	0.115	0.465	0.760	0.725	0.800	0.805	0.840	0.805	0.795	0.795

$^{(*)}$ Notes: See notes to Table 1.

Table 3: Finite Sample Properties: Rolling and Recursive PEE Bootstrap: Part III(*)

smpl	*boot*	*coeff*	$l=4$	$l=6$	$l=10$	$l=12$	$l=15$	$l=20$	$l=25$	$l=30$	$l=50$	$l=60$
Panel A: DGP is an AR(1) Process - $\rho=0.6$												
800	*roll*1	$\hat{\rho}$	0.000	0.020	0.300	0.470	0.685	0.735	0.825	0.845	0.830	0.805
		$\hat{\rho}_1$	0.000	0.045	0.410	0.600	0.780	0.780	0.870	0.855	0.830	0.815
		$\hat{\rho}_2$	0.850	0.885	0.870	0.870	0.845	0.880	0.880	0.855	0.825	0.790
	*roll*2	$\hat{\rho}$	0.000	0.040	0.325	0.450	0.620	0.635	0.730	0.730	0.705	0.710
		$\hat{\rho}_1$	0.000	0.090	0.440	0.535	0.700	0.640	0.740	0.705	0.775	0.690
		$\hat{\rho}_2$	0.775	0.780	0.760	0.765	0.705	0.785	0.720	0.755	0.665	0.660
	*rec*1	$\hat{\rho}$	0.000	0.010	0.275	0.420	0.715	0.765	0.815	0.815	0.815	0.760
		$\hat{\rho}_1$	0.000	0.055	0.440	0.575	0.815	0.850	0.820	0.830	0.800	0.765
		$\hat{\rho}_2$	0.850	0.865	0.900	0.855	0.875	0.860	0.830	0.855	0.780	0.755
	*rec*2	$\hat{\rho}$	0.000	0.010	0.285	0.420	0.715	0.735	0.765	0.800	0.795	0.685
		$\hat{\rho}_1$	0.000	0.045	0.440	0.585	0.795	0.820	0.775	0.800	0.750	0.700
		$\hat{\rho}_2$	0.840	0.845	0.845	0.815	0.830	0.795	0.810	0.825	0.750	0.700
1600	*roll*1	$\hat{\rho}$	0.000	0.000	0.090	0.215	0.635	0.710	0.800	0.770	0.870	0.825
		$\hat{\rho}_1$	0.000	0.010	0.215	0.430	0.715	0.785	0.830	0.810	0.870	0.810
		$\hat{\rho}_2$	0.835	0.895	0.865	0.890	0.890	0.885	0.805	0.875	0.870	0.790
	*roll*2	$\hat{\rho}$	0.000	0.005	0.125	0.235	0.585	0.645	0.715	0.655	0.785	0.685
		$\hat{\rho}_1$	0.000	0.010	0.245	0.450	0.670	0.710	0.700	0.695	0.805	0.690
		$\hat{\rho}_2$	0.760	0.765	0.775	0.855	0.790	0.760	0.710	0.810	0.790	0.740
	*rec*1	$\hat{\rho}$	0.000	0.000	0.085	0.130	0.605	0.630	0.740	0.760	0.815	0.775
		$\hat{\rho}_1$	0.000	0.000	0.165	0.375	0.710	0.780	0.795	0.820	0.860	0.800
		$\hat{\rho}_2$	0.805	0.870	0.925	0.865	0.880	0.855	0.870	0.845	0.820	0.870
	*rec*2	$\hat{\rho}$	0.000	0.000	0.090	0.140	0.575	0.630	0.700	0.740	0.760	0.750
		$\hat{\rho}_1$	0.000	0.000	0.200	0.365	0.685	0.740	0.755	0.785	0.815	0.770
		$\hat{\rho}_2$	0.765	0.845	0.875	0.825	0.870	0.820	0.830	0.805	0.805	0.840
Panel B: DGP is an AR(2) Process - $\rho=0.3$												
800	*roll*1	$\hat{\rho}$	0.025	0.285	0.625	0.720	0.810	0.820	0.830	0.810	0.750	0.785
		$\hat{\rho}_1$	0.090	0.485	0.715	0.830	0.830	0.850	0.840	0.820	0.810	0.765
		$\hat{\rho}_2$	0.135	0.455	0.710	0.820	0.830	0.835	0.895	0.830	0.805	0.780
	*roll*2	$\hat{\rho}$	0.045	0.300	0.540	0.635	0.715	0.735	0.710	0.700	0.635	0.665
		$\hat{\rho}_1$	0.140	0.435	0.660	0.705	0.725	0.755	0.685	0.675	0.690	0.640
		$\hat{\rho}_2$	0.190	0.425	0.660	0.695	0.730	0.710	0.725	0.680	0.695	0.660
	*rec*1	$\hat{\rho}$	0.025	0.255	0.605	0.640	0.780	0.835	0.880	0.770	0.735	0.735
		$\hat{\rho}_1$	0.055	0.415	0.730	0.730	0.840	0.835	0.880	0.805	0.790	0.745
		$\hat{\rho}_2$	0.105	0.470	0.750	0.790	0.835	0.850	0.830	0.830	0.725	0.705
	*rec*2	$\hat{\rho}$	0.030	0.270	0.565	0.625	0.755	0.820	0.865	0.745	0.705	0.715
		$\hat{\rho}_1$	0.065	0.430	0.675	0.725	0.805	0.825	0.865	0.760	0.765	0.710
		$\hat{\rho}_2$	0.115	0.480	0.725	0.775	0.765	0.830	0.790	0.785	0.710	0.705
1600	*roll*1	$\hat{\rho}$	0.000	0.095	0.520	0.605	0.750	0.830	0.815	0.830	0.895	0.820
		$\hat{\rho}_1$	0.005	0.180	0.700	0.720	0.830	0.835	0.830	0.830	0.905	0.860
		$\hat{\rho}_2$	0.010	0.245	0.540	0.670	0.840	0.855	0.855	0.800	0.860	0.850
	*roll*2	$\hat{\rho}$	0.005	0.150	0.510	0.560	0.690	0.740	0.735	0.730	0.750	0.705
		$\hat{\rho}_1$	0.015	0.235	0.660	0.655	0.730	0.740	0.755	0.740	0.780	0.720
		$\hat{\rho}_2$	0.020	0.295	0.510	0.615	0.700	0.735	0.705	0.725	0.745	0.750
	*rec*1	$\hat{\rho}$	0.000	0.070	0.440	0.615	0.785	0.725	0.875	0.845	0.850	0.830
		$\hat{\rho}_1$	0.010	0.145	0.585	0.675	0.830	0.795	0.885	0.870	0.875	0.825
		$\hat{\rho}_2$	0.005	0.205	0.585	0.760	0.775	0.835	0.830	0.845	0.850	0.800
	*rec*2	$\hat{\rho}$	0.005	0.085	0.460	0.580	0.745	0.700	0.825	0.795	0.790	0.770
		$\hat{\rho}_1$	0.015	0.160	0.565	0.680	0.825	0.795	0.850	0.815	0.825	0.820
		$\hat{\rho}_2$	0.010	0.225	0.555	0.730	0.750	0.820	0.795	0.785	0.820	0.760

(*) Notes: See notes to Table 1.

Table 4: Finite Sample Properties: Rolling and Recursive PEE Bootstrap: Part IV[(*)]

$smpl$	$boot$	$coeff$	$l=4$	$l=6$	$l=10$	$l=12$	$l=15$	$l=20$	$l=25$	$l=30$	$l=50$	$l=60$
Panel A: DGP is an AR(1) Process - $\rho = 0.8$												
800	$roll1$	$\hat{\rho}$	0.000	0.000	0.015	0.150	0.560	0.645	0.730	0.795	0.745	0.745
		$\hat{\rho}_1$	0.000	0.035	0.285	0.555	0.725	0.845	0.840	0.855	0.785	0.810
		$\hat{\rho}_2$	0.940	0.895	0.935	0.885	0.915	0.860	0.895	0.885	0.820	0.765
	$roll2$	$\hat{\rho}$	0.000	0.000	0.060	0.185	0.500	0.575	0.675	0.650	0.670	0.650
		$\hat{\rho}_1$	0.000	0.055	0.325	0.530	0.670	0.730	0.770	0.750	0.645	0.650
		$\hat{\rho}_2$	0.865	0.820	0.830	0.810	0.790	0.770	0.800	0.770	0.745	0.665
	$rec1$	$\hat{\rho}$	0.000	0.000	0.060	0.090	0.540	0.660	0.685	0.750	0.690	0.750
		$\hat{\rho}_1$	0.000	0.000	0.300	0.440	0.765	0.825	0.800	0.835	0.800	0.770
		$\hat{\rho}_2$	0.865	0.925	0.910	0.910	0.870	0.880	0.860	0.840	0.775	0.785
	$rec2$	$\hat{\rho}$	0.000	0.000	0.065	0.105	0.510	0.620	0.680	0.710	0.655	0.720
		$\hat{\rho}_1$	0.000	0.005	0.330	0.435	0.765	0.805	0.785	0.790	0.750	0.690
		$\hat{\rho}_2$	0.855	0.905	0.875	0.895	0.830	0.860	0.840	0.820	0.725	0.785
1600	$roll1$	$\hat{\rho}$	0.000	0.000	0.000	0.010	0.260	0.425	0.730	0.740	0.790	0.800
		$\hat{\rho}_1$	0.000	0.005	0.115	0.250	0.605	0.775	0.820	0.865	0.815	0.850
		$\hat{\rho}_2$	0.920	0.895	0.900	0.900	0.890	0.870	0.895	0.900	0.850	0.865
	$roll2$	$\hat{\rho}$	0.000	0.000	0.010	0.025	0.300	0.430	0.660	0.675	0.680	0.720
		$\hat{\rho}_1$	0.000	0.005	0.175	0.295	0.590	0.700	0.730	0.790	0.730	0.750
		$\hat{\rho}_2$	0.900	0.845	0.815	0.775	0.825	0.740	0.775	0.795	0.730	0.715
	$rec1$	$\hat{\rho}$	0.000	0.000	0.000	0.010	0.305	0.450	0.670	0.695	0.795	0.760
		$\hat{\rho}_1$	0.000	0.000	0.055	0.180	0.580	0.680	0.830	0.820	0.810	0.835
		$\hat{\rho}_2$	0.895	0.945	0.930	0.875	0.875	0.900	0.855	0.850	0.820	0.820
	$rec2$	$\hat{\rho}$	0.000	0.000	0.000	0.010	0.320	0.445	0.700	0.670	0.755	0.750
		$\hat{\rho}_1$	0.000	0.000	0.085	0.200	0.575	0.710	0.805	0.790	0.790	0.825
		$\hat{\rho}_2$	0.885	0.940	0.900	0.850	0.870	0.870	0.835	0.815	0.780	0.820
Panel B: DGP is an AR(2) Process - $\rho = 0.4$												
800	$roll1$	$\hat{\rho}$	0.000	0.055	0.465	0.580	0.765	0.760	0.865	0.770	0.795	0.780
		$\hat{\rho}_1$	0.035	0.285	0.700	0.775	0.830	0.870	0.855	0.860	0.780	0.760
		$\hat{\rho}_2$	0.040	0.345	0.655	0.775	0.860	0.865	0.815	0.845	0.800	0.760
	$roll2$	$\hat{\rho}$	0.005	0.080	0.415	0.495	0.650	0.665	0.810	0.660	0.630	0.665
		$\hat{\rho}_1$	0.055	0.315	0.650	0.655	0.725	0.765	0.795	0.740	0.675	0.655
		$\hat{\rho}_2$	0.055	0.415	0.620	0.715	0.785	0.730	0.725	0.720	0.705	0.665
	$rec1$	$\hat{\rho}$	0.000	0.050	0.405	0.505	0.755	0.755	0.800	0.725	0.770	0.715
		$\hat{\rho}_1$	0.005	0.290	0.650	0.790	0.845	0.820	0.895	0.845	0.780	0.720
		$\hat{\rho}_2$	0.045	0.325	0.715	0.675	0.825	0.840	0.870	0.815	0.765	0.765
	$rec2$	$\hat{\rho}$	0.000	0.065	0.395	0.495	0.725	0.735	0.790	0.690	0.720	0.690
		$\hat{\rho}_1$	0.010	0.310	0.650	0.790	0.820	0.790	0.855	0.795	0.730	0.670
		$\hat{\rho}_2$	0.045	0.330	0.670	0.675	0.805	0.830	0.830	0.815	0.735	0.780
1600	$roll1$	$\hat{\rho}$	0.000	0.000	0.225	0.280	0.715	0.750	0.810	0.810	0.810	0.825
		$\hat{\rho}_1$	0.000	0.050	0.465	0.595	0.830	0.860	0.855	0.855	0.810	0.835
		$\hat{\rho}_2$	0.000	0.085	0.505	0.605	0.865	0.825	0.865	0.845	0.830	0.860
	$roll2$	$\hat{\rho}$	0.000	0.010	0.265	0.315	0.600	0.670	0.725	0.725	0.650	0.720
		$\hat{\rho}_1$	0.000	0.090	0.470	0.595	0.765	0.765	0.810	0.725	0.715	0.740
		$\hat{\rho}_2$	0.000	0.165	0.520	0.520	0.740	0.745	0.765	0.700	0.760	0.705
	$rec1$	$\hat{\rho}$	0.000	0.000	0.170	0.330	0.580	0.610	0.815	0.800	0.780	0.790
		$\hat{\rho}_1$	0.000	0.020	0.455	0.585	0.780	0.755	0.865	0.790	0.870	0.855
		$\hat{\rho}_2$	0.000	0.090	0.465	0.545	0.820	0.795	0.865	0.830	0.900	0.820
	$rec2$	$\hat{\rho}$	0.000	0.005	0.175	0.345	0.560	0.600	0.780	0.780	0.730	0.770
		$\hat{\rho}_1$	0.000	0.030	0.450	0.560	0.750	0.745	0.805	0.785	0.815	0.790
		$\hat{\rho}_2$	0.000	0.110	0.485	0.540	0.780	0.775	0.820	0.800	0.850	0.795

[(*)] Notes: See notes to Table 1.

Table 5: Comparison of Autoregressive Inflation Models with and Without Unemployment Using SIC[(*)]

	Model 1 - Normal	Model 2 - Normal	Model 3 - Student's t	Model 4 - Student's t
Specification	AR	ARX	AR	ARX
lag Selection	SIC (1)	SIC (1,1)	SIC (1)	SIC (1,1)
$MSFE$	0.00083352	0.00004763	0.00083352	0.00004763
$DMSFE$	1.80129635	2.01137942	1.84758927	1.93272971
$Z_{P,u,2}(1,k)$	benchmark	-0.21008307	-0.04629293	-0.13143336

Critical Values

	Split Sample Bootstrap					Full Sample Bootstrap				
Percentile	3	5	10	15	20	3	5	10	15	20
50	0.021162	0.024060	0.029225	0.032261	0.035047	0.024781	0.028650	0.031658	0.033059	0.039597
60	0.025038	0.029217	0.035260	0.042024	0.048347	0.030310	0.033776	0.038414	0.041436	0.049562
70	0.029217	0.033260	0.046050	0.062857	0.085990	0.037022	0.039206	0.047596	0.051924	0.065609
80	0.037753	0.044869	0.104205	0.116851	0.146838	0.047352	0.048774	0.060000	0.067258	0.093197
90	0.049772	0.112000	0.169281	0.197268	0.239285	0.071591	0.067820	0.096591	0.104021	0.170241

[(*)] Notes: Entires in the table are given in two parts (i) summary statistics, and (ii) bootstrap percentiles. In (i): "specification" lists the model used. For each specification, lags may be chosen either with the SIC or the AIC, and the predictive density may be either Gaussian or Student's t, as denoted in the various columns of the table. The bracketed entires beside SIC and AIC denote the number of lags chosen for the autoregressive part of the model and the number of lags of unemployment used, respectively. $MSFE$ is the out-of-sample mean square forecast error based on evaluation of P=300 1-step ahead predictions using recursively estimated models, and $DMSFE = \int_U \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(1\{Inf_{t+1} \leq u\} - F_1(u|Z^t, \widehat{\theta}_{1,t})\right)^2 \phi(u)du$, where $R = 300$, corresponding to the sample period from 1954:1-1978:12, is our analogous density based square error loss measure. Finally, $Z_{P,u,2}(1,k)$ is the accuracy test statistic, for each benchmark/alternative model comparison. The density accuracy test is the maximum across the $Z_{P,u,2}(1,k)$ values. In (ii) percentiles of split and full sample bootstrap empirical distributions under different block length sampling regimes are given. Testing is carried out using 90th percentiles (ee above for further details).

Table 6: Comparison of Autoregressive Inflation Models with and Without Unemployment Using AIC[(*)]

	Model 1 - Normal	Model 2 - Normal	Model 3 - Student's t	Model 4 - Student's t
Specification	AR	ARX	AR	ARX
lag Selection	AIC (3)	AIC (3,1)	AIC (3)	AIC (3,1)
$MSFE$	0.00000841	0.00000865	0.00000841	0.00000865
$DMSFE$	2.17718449	2.17189485	2.11242940	2.10813786
$Z_{P,u,2}(1,k)$	benchmark	0.00528965	0.06475509	0.06904664

Critical Values

	Split Sample Bootstrap					Full Sample Bootstrap				
Percentile	3	5	10	15	20	3	5	10	15	20
50	-0.002736	-0.002844	-0.002719	-0.002855	-0.002866	-0.000348	0.000541	0.000745	0.000517	0.000301
60	-0.001674	-0.001489	-0.000748	-0.001035	-0.001230	0.001530	0.002071	0.002289	0.002202	0.002289
70	0.000745	0.000937	0.001086	0.001088	0.001086	0.002865	0.003447	0.004013	0.004036	0.004267
80	0.002635	0.002842	0.003430	0.004151	0.004440	0.004446	0.004919	0.005859	0.006300	0.007121
90	0.005140	0.005883	0.006333	0.006879	0.008406	0.007112	0.007578	0.008466	0.009770	0.010420

[(*)] Notes: See notes to Table 5.

Table 7: Comparison of Autoregressive Inflation Models Using SIC and AIC$^{(*)}$

	Model 1 - Normal	Model 2 - Normal	Model 3 - Student's t	Model 4 - Student's t
Specification	AR	AR	AR	AR
lag Selection	SIC (1)	AIC (3)	SIC (1)	AIC (3)
$MSFE$	0.00083352	0.00000841	0.00083352	0.00000841
$DMSFE$	1.80129635	2.17718449	1.84758927	2.11242940
$Z_{P,u,2}(1,k)$	benchmark	-0.37588815	-0.04629293	-0.31113305

Critical Values

	Split Sample Bootstrap					Full Sample Bootstrap				
Percentile	3	5	10	15	20	3	5	10	15	20
50	0.030910	0.034325	0.044692	0.049984	0.056810	0.029054	0.028849	0.033184	0.037521	0.041006
60	0.038460	0.046735	0.062769	0.083326	0.098715	0.034439	0.034439	0.039774	0.046804	0.053350
70	0.049358	0.080635	0.108230	0.132668	0.143861	0.037828	0.039183	0.051636	0.060352	0.071022
80	0.123676	0.134695	0.162630	0.184618	0.202485	0.048282	0.055104	0.078584	0.098374	0.110733
90	0.164544	0.177596	0.238318	0.265352	0.289242	0.098374	0.117644	0.138269	0.167334	0.207614

$^{(*)}$ Notes: See notes to Table 5.

Table 8: Comparison of Autoregressive Inflation Models with Unemployment Using SIC and AIC$^{(*)}$

	Model 1 - Normal	Model 2 - Normal	Model 3 - Student's t	Model 4 - Student's t
Specification	ARX	ARX	ARX	ARX
lag Selection	SIC (1,1)	AIC (3,1)	SIC (1,1)	AIC (3,1)
$MSFE$	0.00004763	0.00000865	0.00004763	0.00000865
$DMSFE$	2.01137942	2.17189485	1.93272971	2.10813786
$Z_{P,u,2}(1,k)$	benchmark	-0.16051543	0.07864972	-0.09675844

Critical Values

	Split Sample Bootstrap					Full Sample Bootstrap				
Percentile	3	5	10	15	20	3	5	10	15	20
50	0.034626	0.034213	0.036984	0.038688	0.040339	0.009987	0.011698	0.013288	0.014661	0.016318
60	0.037691	0.037691	0.040489	0.044661	0.046770	0.012823	0.014629	0.016761	0.018989	0.020060
70	0.041699	0.044492	0.048036	0.051974	0.054885	0.014879	0.018140	0.020144	0.024000	0.025739
80	0.050567	0.051521	0.055278	0.059678	0.065561	0.019038	0.022255	0.025917	0.030161	0.033054
90	0.059443	0.059906	0.066012	0.073324	0.079340	0.025917	0.026484	0.034474	0.038606	0.041419

$^{(*)}$ Notes: See notes to Table 5.