The Block Bootstrap for Parameter Estimation Error In Recursive Estimation Schemes, With Applications to Predictive Evaluation*

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

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

December 2003

Abstract

We consider the comparison of multiple (possibly all misspecified) models in terms of their out of sample predictive ability. Typically, candidate models compared contain parameters estimated using recursive (or related rolling) estimation schemes. In some cases, such as the comparison of predictive densities considered here, we show that the limiting distribution of appropriate test statistics is a functional over a Gaussian process, with a covariance kernel that reflects the contribution of parameter uncertainty. The limiting distributions are not nuisance parameter free and valid critical values are thus generally obtained via the bootstrap. Given these considerations, our approach in this paper is to develop a bootstrap procedure that properly captures the contribution of parameter estimation error in recursive estimation schemes. Intuitively, when parameters are estimated recursively, as in done in our framework, earlier observations in the sample are used more heavily than subsequent observations. However, in the standard block bootstrap, all blocks have equal chance of being drawn. This induces a location bias in the bootstrap distribution, which can be either positive or negative across different samples. Within the context of tests of predictive accuracy, we suggest how to offset the location bias via the construction of properly adjusted bootstrap statistics. The usefulness of our approach is illustrated via two applications: one is an out of sample version of integrated conditional moment tests of Bierens (1982, 1990) and Bierens and Ploberger (1997), the other concerns predictive density evaluation, and is an extension of the Andrews (1997) conditional Kolmogorov test, which allows for comparison of multiple misspecified conditional distribution models. The main findings from a small Monte Carlo experiment indicate that the adjustment term used in the suggested bootstrap substantially improves coverage rates relative to a bootstrap without adjustment.

JEL classification: C22, C51.
Keywords: block bootstrap, recursive estimation scheme, nonlinear causality, 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. We thank Jean-Marie Dufour, Silvia Goncalves, Stephen Gordon, Clive Granger, Oliver Linton, Brendan McCabe, Antonio Mele, Andrew Patton, Rodney Strachan, Christian Schluter, Allan Timmerman, and seminar participants at Cornell University, the London School of Economics, Laval University, Queen Mary, University of London, CIREQ-Universite' de Montreal, University of Liverpool, University of Southampton, University of Virginia, and the Bank of Canada for useful comments and suggestions. Corradi gratefully acknowledges ESRC grant RES-000-23-0006.

1 Introduction

In the predictive evaluation literature, when multiple misspecified models are compared in terms of their out of sample predictive ability interest may focus on pointwise prediction, confidence intervals or prediction of the entire distribution. In such contexts, one often compares parametric models containing estimated parameters. Hence, it is important to take into account the contribution of parameter estimation error, particularly in (small) finite samples. Furthermore, it is common practice to split the sample T into $T = R + P$ observations, where the last P observations are used for predictive evaluation. We consider such a setup, and assume that parameters are estimated in a recursive fashion, such that R observations are used to construct a first parameter estimator, say $\widehat{\theta}_R$ and a first prediction error, taken for simplicity to be a 1-step ahead prediction error. Then, $R+1$ observations are used to construct $\widehat{\theta}_{R+1}$ and a second prediction error. This is continued until a final estimator is constructed using $T-1$ observations, resulting in a sequence of $P = T-R$ estimators and prediction errors. If R and P grow at the same rate as the sample size increases, the limiting distributions of predictive accuracy tests using this setup generally reflect the contribution of parameters uncertainty (i.e. the contribution of $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_t - \theta^{\dagger}\right)$, where $\widehat{\theta}_t$ is a recursive $m-$estimator constructed using the first t observations, and $\theta^{\dagger}$ is its probability limit). Our objectives in this paper are twofold. First, we will introduce block bootstrap techniques that are (first order) valid in recursive estimation frameworks. Thereafter, we will outline predictive accuracy and predictive density tests that can be made operational using our new bootstrap procedures.

In some circumstances, such as when constructing Diebold and Mariano (1995) tests for equal (pointwise) predictive accuracy of two models, the limiting distribution is a normal random variable. In this case, the contribution of parameter estimation error can be addressed using the framework of West (1996), and essentially involves estimating an "extra" covariance term. However, in other circumstances, such as when comparing predictive densities, test statistic limiting distributions can be shown to be functionals of Gaussian processes with covariance kernels that reflect both (dynamic) misspecification as well as the contribution of (recursive) parameter estimation error. Such limiting distributions are not nuisance parameters free, and critical values cannot be tabulated. However, valid asymptotic critical values can be obtained via appropriate application of the (block) bootstrap. This requires the formulation of a bootstrap procedures that allows us to form statistics which properly mimic the contribution of $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_t-\theta^{\dagger}\right)$.[1] The main objective of this paper is thus to suggest a new block bootstrap procedure which is valid for recursive m-estimators, in the sense that its use suffices to mimic the limiting distribution of $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_t-\theta^{\dagger}\right)$, where R denotes the length of the estimation period, P the number of recursively estimated parameters, $\widehat{\theta}_t$ is a recursive $m-$estimator constructed using the first t observations, and $\theta^{\dagger}$ is its probability limit.

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 schemes, any block from the original sample has the same probability of being selected. Such a location bias can be either positive or negative, depending on the sample that we observe. One way to circumvent this problem by first forming bootstrap samples as follows: *Resample R observations from the initial R sample observations, and then concatenate onto this vector an additional P resampled observations from the remaining sample.* Thereafter, construct $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_t^*-\widehat{\theta}_t\right)$ and add an adjustment term in order to ensure that the distribution of the sum of both components is the same as the distribution of $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_t-\theta^{\dagger}\right)$, conditional on the sample, and for all samples except a set of probability measure approaching zero. The adjustment term compensates for the fact that, when resampling from the last P observations, each block (of length l) has the same chance of being drawn, while in the construction of the actual $m-$estimator, earlier observations are more heavily used. We call this bootstrap the parameter estimation error (PEE) bootstrap for recursive estimation schemes (or alternatively, the recursive PEE bootstrap).[2] In addition to the above procedure, we also discuss another block bootstrap procedure that is also appropriate for the predictive density and accuracy tests that we introduce, and which is based on

[1]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, 2002b) 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 (2001) 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 (2002a and 2002b), who show that for $m-$estimators, the limiting distribution of $\sqrt{T}(\widehat{\theta}_T^*-\widehat{\theta}_T)$ provides a valid first order approximation to that of $\sqrt{T}(\widehat{\theta}_T-\theta^{\dagger})$ for heterogeneous and near epoch dependent series, where $\widehat{\theta}_T^*$ is a resampled estimator, and T denotes the length of the entire sample.

[2]Bootstrap statistics for the case of rolling estimation scheme are considered in Corradi and Swanson (2004).

resampling from the whole sample. However, in this case we need more "complicated" adjustment terms in order to offset the bias term.[3]

As alluded to above, the recursive PEE bootstrap can be used to provide valid critical values in a variety of interesting predictive testing contexts, and two such leading applications are developed in this paper. The first is an out-of-sample version of the integrated conditional moment (ICM) test of Bierens (1982,1990) and Bierens and Ploberger (1997) which provides out of sample tests consistent against generic (nonlinear) alternatives. The second is a procedure assessing the relative out-of-sample predictive accuracy of multiple conditional distribution models. This procedure is based on an extension of the Andrews (1997) conditional Kolmogorov test. There are two key links between these applications. First, both applications are made operational via use of the recursive PEE bootstrap. Second, all applications allow for misspecification among all models being estimated and compared, as opposed to the usual practice of assuming correct (dynamic) specification under the null hypothesis.[4],[5] In both cases, the limiting distribution is a functional over a Gaussian process with a covariance kernel reflecting the contribution of parameters estimation error in the recursive estimation scheme. Hence the need of a block bootstrap procedure able to mimic the limiting distribution of $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_t-\theta^{\dagger}\right)$.

In order to assess the finite sample performance of the PEE recursive bootstrap, we carry out a set of Monte Carlo experiments in which we compare the recursive PEE bootstrap with a version thereof that does not contain adjustment terms. Interestingly, there is clear evidence that the adjustment terms yield substantive improvements relative to the version of the recursive PEE bootstrap that does not contain the adjustment terms.

The rest of the paper is organized as follows. Section 2 outlines block bootstrap procedures

[3]In principle, one could also devise a resampling scheme in which blocks are more heavily drawn from the beginning of the sample, and within each block, "earlier" observations are more heavily weighted. However, in practice, it is in general not feasible to implement a weighted resampling scheme which exactly mimics the long run covariance of $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_t-\theta^{\dagger}\right)$.

[4]This second feature is important when constructing predictive models, for example, as it is natural not to impose correct specification of any of the competing models; and if one were to impose correct specification under the null hypothesis, then inference would not be valid in general (e.g. when the assumption of martingale difference sequence errors is violated - see Corradi and Swanson (2003a)).

[5]Motivation for comparing misspecified models can be taken from the finance literature. For example, Hansen, Heaton and Luttmer (1995) and Hansen and Jeganathan (1997) consider the problem of comparing multiple misspecified models from the perspective of asset pricing and stochastic discount factor models.

for recursive $m-$estimators and establishes the first order validity of the procedures. Sections 3 and 4 outline the two applications of the recursive block bootstrap: out of sample integrated conditional moment tests and predictive density evaluation. Monte Carlo findings are discussed in Section 5. Finally, concluding remarks are given in Section 6. All proofs are collected 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 Bootstrap Methods

In the following subsections, we establish the first order validity of block bootstrap procedures that allow us to properly capture the effect of PEE in contexts where parameters are estimated using the *recursive* m-estimator defined as follows. Let $Z^t = (y_t, ..., y_{t-s_1+1}, X_t, ..., X_{t-s_2+1})$, $t = 1, ..., T$, and let $s = \max\{s_1, s_2\}$. Additionally, we henceforth assume that $i = 1, ..., n$ models are estimated, as in the applications outlined in Section 3 and 4 below. Now, define the *recursive* m-estimator for the parameter vector associated with model i as:

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

and

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

where q_i denotes the objective function for model i. Following standard practice (such as in the real-time forecasting literature), this estimator is first computed using R observations. In our applications we focus on 1-step ahead prediction, and the recursive estimators are thus computed first using $R+1$ observations, and then $R+2$ observations, and so on until the last estimator is constructed using $T-1$ observations; resulting in a sequence of $P = T - R$ estimators. These estimators are then used to construct sequences of P 1-step ahead forecasts and associated forecast errors, for example.

2.1 A Block Bootstrap Procedure for Parameter Estimation Error in Recursive Schemes

In order to properly capture the contribution of parameter estimation error given the above recursive sampling scheme, it suffices to form bootstrap samples by first resampling R observations from the initial R sample observations, and then concatenating onto this an additional P observations resampled from the P remaining sample observations. More specifically, let $b_1 l_1 + b_2 l_2 = T - s + 1$, with $b_1 l_1 = R - s + 1$ and $b_2 l_2 = P$. Also, let $W_t = (y_t, Z^{t-1})$. First, draw b_1 overlapping blocks, of length l_1, from $s, ..., R$ and then draw b_2 overlapping blocks, of length l_2, from data indexed by $R+1, ..., R+P$, with replacement. The first $R-s+1$ pseudo observations, $W_s^*, W_{s+1}^*, ..., W_{s+l-1}^*, ..., W_R^*$, 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 $s, ..., R - l_1 + 1$; and the remaining P pseudo observations, $W_{R+1}^*, W_{R+2}^*, ..., W_{R+l}^*, ..., W_{R+P}^*$, 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 $R+1, R+2, ..., R+P-l-1$. Thus, conditional on the (entire) sample, the pseudo time series W_t^*, $t = s, ..., R, R+1, ..., R+P$, consists of $b = b_1 + b_2$ asymptotically independent, but non identically distributed blocks of length l_1 and l_2 respectively.[6] Also, conditionally on the sample, all of the moments of each block are asymptotically homogeneous, under the assumption that the underlying sample is strictly stationary. Therefore, the rescaled (partial) sums of the blocks satisfy the law of large numbers and the central limit theorem for asymptotically independent and homogeneous random variables, conditional on the (entire) sample. Now, define the recursive PEE bootstrap m-estimator as,

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

In order to establish the asymptotic validity of this version of the block bootstrap, which we have called the recursive PEE bootstrap, we require the following assumptions.

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

Assumption A2: (i) $\theta_i^\dagger$ is uniquely identified (i.e. $E(q_i(y_t, Z^{t-1}, \theta_i)) > E(q_i(y_t, Z^{t-1}, \theta_i^\dagger))$ for any $\theta_i \neq \theta_i^\dagger$); (ii) q_i is twice continuously differentiable on the interior of Θ_i, for $i = 1, ..., n$, and

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

for Θ_i a compact subset of $R^{\varrho(i)}$; (iii) the elements of $\nabla_{\theta_i} q_i$ and $\nabla^2_{\theta_i} q_i$ 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) $E\left(-\nabla^2_{\theta_i} q_i(\theta_i)\right)$ is negative definite uniformly on Θ_i.[7]

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

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 $\Psi^*_{R,P}$ (as defined below) vanishes 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 as those discussed in the sequel.

Define:

$$\begin{aligned}\Psi^*_{R,P} &= \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}^*_{i,t} - \widehat{\theta}_{i,t}\right) + \left(-\frac{1}{T}\sum_{t=s}^{T}\nabla^2_{\theta_i} q_i(y_t, Z^{t-1}, \widehat{\theta}_{i,T-1})\right)^{-1} \\ &\quad \times \frac{1}{\sqrt{P}}\sum_{j=1}^{P-1} a_{R,j}\left(\nabla_{\theta_i} q_i(y_{R+j}, Z^{R+j-1}, \widehat{\theta}_{i,T-1}) - \frac{1}{P}\sum_{j=1}^{P}\nabla_{\theta_i} q_i(y_{R+j}, Z^{R+j-1}, \widehat{\theta}_{i,T-1})\right), \qquad (4)\end{aligned}$$

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

Note that the adjustment term can be seen as a recentering term. In fixed sample estimation, recentering is necessary for the first order validity of overidentified GMM (see e.g. Hall and Horowitz (1996), Andrews (2002a), Inoue and Shintani (2003)). On the other hand, in the case of $m-$estimator, recentering is necessary only to obtain higher order refinements (e.g. Andrews (2002a, 2002b)). In the recursive estimation case, the adjustment term is necessary also for the first order validity of the bootstrap $m-$estimators.

Theorem 1: 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 $R \to \infty$,

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

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

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

Broadly speaking, Theorem 1 states that $\Psi^*_{R,P}$ has the same limiting distribution as $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_{i,t}-\theta_i^{\dagger}\right)$, conditional on sample, and for all samples except a set with probability measure approaching zero. As outlined in the following sections, application of Theorem 1 allows us to capture the contribution of (recursive) parameter estimation error to the covariance kernel of the limiting distribution of various statistics.

Though a detailed proof of Theorem 1 is given in the appendix, it is worthwhile giving an intuitive explanation of why there is an adjustment term in $\Psi^*_{R,P}$. For notational simplicity, let $h_{i,t}=\nabla_{\theta_i}q_i(y_t,Z^{t-1},\theta_i^{\dagger})$ and $h^*_{i,t}=\nabla_{\theta_i}q_i(y^*_t,Z^{*,t-1},\theta_i^{\dagger})$. Via a mean value expansion around $\theta^{\dagger}$, using a similar argument as in Lemma A5 in West (1996), we have

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_{i,t}-\theta_i^{\dagger}\right)=B_i^{\dagger}\frac{a_{R,0}}{\sqrt{P}}\sum_{t=s}^{R}h_{i,t}+B_i^{\dagger}\frac{1}{\sqrt{P}}\sum_{j=1}^{P-1}a_{R,j}h_{i,R+j}+o_P(1), \tag{5}$$

where $B_i^{\dagger}=\left(E\left(-\nabla^2_{\theta i}q_i(y_t,Z^{t-1},\theta_i^{\dagger})\right)\right)^{-1}$. Also,

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}^*_{i,t}-\widehat{\theta}_{i,t}\right)=B_i^{\dagger}\frac{a_{R,0}}{\sqrt{P}}\sum_{t=s}^{R}\left(h^*_{i,t}-h_{i,t}\right)+B_i^{\dagger}\frac{1}{\sqrt{P}}\sum_{j=1}^{P-1}a_{R,j}\left(h^*_{i,R+j}-h_{i,R+j}\right)+o^*_P(1),\ \ \mathrm{Pr}-P. \tag{6}$$

Now, the first term on the RHS of (6) has the same limiting distribution as the first term on the RHS of (5), conditional on sample. However, the second term on the RHS of (6) does not have the same limiting distribution as the second term on the RHS of (5), conditional on sample. The reason for this is that, up to a term of order $O^*_P\left(l/\sqrt{P}\right)$, $E^*\left(\frac{1}{\sqrt{P}}\sum_{j=1}^{P-1}a_{R,j}h^*_{i,R+j}\right)=\frac{1}{\sqrt{P}}\sum_{j=1}^{P-1}a_{R,j}\frac{1}{P}\sum_{j=1}^{P-1}h_{i,R+j}\neq\frac{1}{\sqrt{P}}\sum_{j=1}^{P-1}a_{R,j}h_{i,R+j}$. Now, rewrite (6) as,

$$\begin{aligned}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}^*_{i,t}-\widehat{\theta}_{i,t}\right) &= \left[B_i^{\dagger}\frac{a_{R,0}}{\sqrt{P}}\sum_{t=s}^{R}\left(h^*_{i,t}-h_{i,t}\right)+B_i^{\dagger}\frac{1}{\sqrt{P}}\sum_{j=1}^{P-1}a_{R,j}\left(h^*_{i,R+j}-\frac{1}{P}\sum_{j=1}^{P-1}h_{i,j}\right)\right]\\ &\quad -B_i^{\dagger}\frac{1}{\sqrt{P}}\sum_{j=1}^{P-1}a_{R,j}\left(h_{i,R+j}-\frac{1}{P}\sum_{j=1}^{P-1}h_{i,j}\right)+o^*_P(1),\ \ \mathrm{Pr}-P.\end{aligned} \tag{7}$$

As shown in the proof of the theorem, the term in square brackets in (7) mimics the limiting distribution of $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_{i,t}-\theta_i^{*}\right)$, conditional on sample. Also, the difference between the second term on the RHS of (4) and $B_i^{\dagger}\frac{1}{\sqrt{P}}\sum_{j=1}^{P-1}a_{R,j}\left(h_{i,R+j}-\frac{1}{P}\sum_{j=1}^{P-1}h_{i,j}\right)$, vanishes asymptotically. Therefore, the adjustment term completely offsets the second term on the RHS of (7), as R,P go to infinity.

2.2 An Alternative Block Bootstrap Procedure for Parameter Estimation Error in Recursive Schemes

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, $W_s^{**}, W_{s+1}^{**}, ..., W_{s+l-1}^{**}, ...$ 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 $\widehat{\theta}_t^{**}$ be defined as in (3), 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,

$$\frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\widehat{\theta}_{i,t}^{**} - \widehat{\theta}_{i,t}\right) = B_i^{\dagger} \frac{a_{R,0}}{\sqrt{P}} \sum_{t=s}^{R} \left(h_{i,t}^{**} - h_{i,t}\right) + B_i^{\dagger} \frac{1}{\sqrt{P}} \sum_{j=1}^{P-1} a_{R,j} \left(h_{i,R+j}^{**} - h_{i,R+j}\right) + o_P^*(1), \ \Pr -P, \tag{8}$$

Note that up to an error of order $O\left(\frac{l}{\sqrt{T}}\right)$, $\Pr -P$, $E^{**}\left(\frac{a_{R,0}}{\sqrt{P}} \sum_{t=s}^{R} h_{i,t}^{**}\right) = \frac{a_{R,0}(R-s)}{\sqrt{P}} \overline{h}_T$, and $E^{**}\left(\frac{1}{\sqrt{P}} \sum_{j=1}^{P-1} a_{R,j} h_{i,R+j}^{**}\right) = \frac{1}{\sqrt{P}} \sum_{j=1}^{P-1} a_{R,j} \overline{h}_T$, where $\overline{h}_T = \frac{1}{T} \sum_{t=s}^{T} h_t$. The LHS of (8) can then be written as:

$$\frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\widehat{\theta}_{i,t}^{**} - \widehat{\theta}_{i,t}\right) = \left[B_i^{\dagger} \frac{a_{R,0}}{\sqrt{P}} \sum_{t=s}^{R} \left(h_{i,t}^{**} - \overline{h}_T\right) + B_i^{\dagger} \frac{1}{\sqrt{P}} \sum_{j=1}^{P-1} a_{R,j} \left(h_{i,R+j}^{**} - \overline{h}_T\right)\right]$$

$$-B_i^{\dagger} \frac{a_{R,0}}{\sqrt{P}} \sum_{t=s}^{R} \left(h_t - \overline{h}_T\right) - B_i^{\dagger} \frac{1}{\sqrt{P}} \sum_{j=1}^{P-1} a_{R,j} \left(h_{i,R+j} - \overline{h}_T\right) + o_P^*(1), \ \Pr -P. \tag{9}$$

The first term in square brackets on the RHS of (9) 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, while the remaining two terms on the RHS of (9) constitute the location bias term. It is then possible to construct a bootstrap a statistic that offsets the bias term. Consider,

$$\begin{aligned}\Psi_{R,P}^{**} &= \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\widehat{\theta}_{i,t}^{**} - \widehat{\theta}_{i,t}\right) + \left(-\frac{1}{P} \sum_{t=s}^{P} \nabla_{\theta_i}^2 q_i(y_t, Z^{t-1}, \widehat{\theta}_{i,P-1})\right)^{-1} \\ &\quad \times \left(B_i^{\dagger} \frac{a_{R,0}}{\sqrt{P}} \sum_{t=s}^{R} \left(\nabla_{\theta_i} q_i(y_t, Z^{t-1}, \widehat{\theta}_{i,P-1}) - \frac{1}{T} \sum_{j=1}^{T} \nabla_{\theta_i} q_i(y_j, Z^{j-1}, \widehat{\theta}_{i,P-1})\right)\right. \\ &\quad \left. + \frac{1}{\sqrt{P}} \sum_{j=1}^{P-1} a_{R,j} \left(\nabla_{\theta_i} q_i(y_{R+j}, Z^{R+j-1}, \widehat{\theta}_{i,P-1}) - \frac{1}{T} \sum_{j=1}^{T} \nabla_{\theta_i} q_i(y_j, Z^{j-1}, \widehat{\theta}_{i,P-1})\right)\right)\end{aligned} \tag{10}$$

Now, $\Psi_{R,P}^{**} - \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\widehat{\theta}_{i,t}^{**} - \widehat{\theta}_{i,t}\right)$ offsets the location bias term, and thus $\Psi_{R,P}^{**}$ 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.[8]

[8]Note that we use only the last P observation to construct the estimator to be used in the correction term. In fact, if we used $\widehat{\theta}_{i,T-1}$ instead of $\widehat{\theta}_{i,P-1}$, $\frac{1}{T} \sum_{j=1}^{T} \nabla_{\theta_i} q_i(y_j, Z^{j-1}, \widehat{\theta}_{i,T-1}) = 0$ and so we cannot offset the bias term.

It follows immediately that $\Psi^*_{R,P}$ only contains a correction term for the last P observations, while $\Psi^{**}_{R,P}$ contains two correction terms, one for each of the first R and last P observations. In this sense, $\Psi^*_{R,P}$ may be considered "better" than $\Psi^{**}_{R,P}$. However, a comparison of the two statistics is left to future research, as the Monte Carlo experiments reported in Section 5 focus on the finite sample behavior of $\Psi^*_{R,P}$.

The next two sections outline two applications of the bootstrap procedure suggested above, and although the applications are based on the use of $\Psi^*_{R,P}$, analogous versions that use $\Psi^{**}_{R,P}$ follow directly.

3 Out-of-Sample Integrated Conditional Moment Tests

Corradi and Swanson (CS: 2002) draw on both the consistent specification and predictive ability testing literatures and propose a test for predictive accuracy which is consistent against generic nonlinear alternatives, and which is designed for comparing nested models. The test is based on an out-of-sample version of the integrated conditional moment (ICM) test of Bierens (1982,1990) and Bierens and Ploberger (1997). One reason why using an ICM type test is more intuitively appealing in the framework that we are interested in, than Diebold and Mariano (DM: 1995), West (1996) or reality check type tests is that in addition to comparing nested models, we also use the same loss function for estimation and for predictive evaluation.[9] This is important because if the same loss function is used throughout, and the null model is nested, DM (and related) tests vanish in probability under the null.[10] It turns out that the limiting distribution of the ICM type test statistic proposed by CS is a functional of a Gaussian process with a covariance kernel that reflects both the time series structure of the data as well as the contribution of parameter estimation error. As a consequence, critical values are data dependent and cannot be directly tabulated.

[9]For a compelling recent discussion outlining the importance of loss functions in economic forecasting, please see Elliott and Timmerman (2002a,b) and the references cited therein.

[10]Note that McCracken (1999) shows that a particular version of the DM test in which a null model is compared against a fixed alternative and the numerator is multiplied by $\sqrt{P}$ has a nonstandard limiting distribution. However, his approach does not apply in the current context, as we consider generic alternatives, and as we allow for misspecification of all models being compared rather than assuming correct specification under the null. On the other hand, Giacomini and White (2003) provides a test for conditional predictive ability valid for both nested and nonnested models. The key ingredient of her test is the fact that parameters are estimated using a fixed rolling window.

CS establish the validity of the conditional p-value method for constructing critical values in this context, thus extending earlier work by Hansen (1996) and Inoue (2001). However, the conditional p-value approach suffers from the fact that under the alternative, the simulated statistics diverge (at rate as high as $\sqrt{\widetilde{l}}$), conditional on the sample and for all samples except a set of measure zero, where $\widetilde{l}$ plays a role analogous to l in the block bootstrap. As this feature may lead to reduced power in finite samples, we establish in this application that the recursive PEE bootstrap can also be used.

Summarizing the testing approach used in this application, assume that the objective is to test whether there exists any unknown alternative model that has better predictive accuracy than a given benchmark model, for a given loss function. A typical example is the case in which the benchmark model is a simple autoregressive model and we want to check whether a more accurate forecasting model can be constructed by including possibly unknown (non)linear functions of the past of the process or of the past of some other process(es).[11] Although this is the case that we focus on, the benchmark model can in general be any (non)linear model. As mentioned above, one important feature of this application is that the same loss function is used for in-sample estimation and out-of-sample prediction (see Granger (1993) and Weiss (1996)). In contrast to the previous application, however, this does not ensure that parameter estimation error vanishes asymptotically.

Let the benchmark model be:

$$y_t = \theta_{1,1}^{\dagger} + \theta_{1,2}^{\dagger} y_{t-1} + u_{1,t}, \tag{11}$$

where $\theta_1^{\dagger} = (\theta_{1,1}^{\dagger}, \theta_{1,2}^{\dagger})' = \arg\min_{\theta_1 \in \Theta_1} E(q_1(y_t - \theta_{1,1} - \theta_{1,2} y_{t-1}))$, $\theta_1 = (\theta_{1,1}, \theta_{1,2})'$, y_t is a scalar, $q_1 = g$, as the same loss function is used both for in-sample estimation and out-of-sample predictive evaluation, and everything else is defined above. The generic alternative model is:

$$y_t = \theta_{2,1}^{\dagger}(\gamma) + \theta_{2,2}^{\dagger}(\gamma) y_{t-1} + \theta_{2,3}^{\dagger}(\gamma) w(Z^{t-1}, \gamma) + u_{2,t}(\gamma), \tag{12}$$

where $\theta_2^{\dagger}(\gamma) = (\theta_{2,1}^{\dagger}(\gamma), \theta_{2,2}^{\dagger}(\gamma), \theta_{2,3}^{\dagger}(\gamma))' = \arg\min_{\theta_2 \in \Theta_2} E(q_1(y_t - \theta_{2,1} - \theta_{2,2} y_{t-1} - \theta_{2,3} w(Z^{t-1}, \gamma)))$, $\theta_2(\gamma) = (\theta_{2,1}(\gamma), \theta_{2,2}(\gamma), \theta_{2,3}(\gamma))'$, $\theta_2 \in \Theta_2$, where Γ is a compact subset of $\Re^d$, for some finite d. The alternative model is called "generic" because of the presence of $w(Z^{t-1}, \gamma)$, which is a generically comprehensive function, such as Bierens' exponential, a logistic, or a cumulative distribution

[11] For example, Swanson and White (1997) compare the predictive accuracy of various linear models against neural network models using both in-sample and out-of-sample model selection criteria.

function (see e.g. Stinchcombe and White (1998) for a detailed explanation of generic comprehensiveness). One example has $w(Z^{t-1},\gamma) = \exp(\sum_{i=1}^{s_2}\gamma_i\Phi(X_{t-i}))$, where Φ is a measurable one to one mapping from $\Re$ to a bounded subset of $\Re$, so that here $Z^t = (X_t, ..., X_{t-s_2+1})$, and we are thus testing for nonlinear Granger causality. The hypotheses of interest are:

$$H_0 : E(g(u_{1,t+1}) - g(u_{2,t+1}(\gamma))) = 0 \text{ versus } H_A : E(g(u_{1,t+1}) - g(u_{2,t+1}(\gamma))) > 0. \qquad (13)$$

Clearly, the reference model is nested within the alternative model, and given the definitions of $\theta_1^\dagger$ and $\theta_2^\dagger(\gamma)$, the null model can never outperform the alternative. For this reason, H_0 corresponds to equal predictive accuracy, while H_A corresponds to the case where the alternative model outperforms the reference model, as long as the errors above are loss function specific forecast errors. It follows that H_0 and H_A can be restated as:

$$H_0 : \theta_{2,3}^\dagger(\gamma) = 0 \text{ versus } H_A : \theta_{2,3}^\dagger(\gamma) \neq 0,$$

for $\forall\gamma \in \Gamma$, except for a subset with zero Lebesgue measure. Now, given the definition of $\theta_2^\dagger(\gamma)$, note that

$$E\left(g'(y_{t+1} - \theta_{2,1}^\dagger(\gamma) - \theta_{2,2}^\dagger(\gamma)y_t - \theta_{2,3}^\dagger(\gamma)w(Z^t,\gamma)) \times \begin{pmatrix} -1 \\ -y_t \\ -w(Z^t,\gamma) \end{pmatrix}\right) = 0,$$

where g' is defined as above. Thus, under H_0 we have that $\theta_{2,3}^\dagger(\gamma) = 0$, $\theta_{2,1}^\dagger(\gamma) = \theta_{1,1}^\dagger$, $\theta_{2,2}^\dagger(\gamma) = \theta_{1,2}^\dagger$, and $E(g'(u_{1,t+1})w(Z^t,\gamma)) = 0$. Thus, we can once again restate H_0 and H_A as:

$$H_0 : E(g'(u_{1,t+1})w(Z^t,\gamma)) = 0 \text{ versus } H_A : E(g'(u_{1,t+1})w(Z^t,\gamma)) \neq 0, \qquad (14)$$

for $\forall\gamma \in \Gamma$, except for a subset with zero Lebesgue measure. Finally, define $\widehat{u}_{1,t+1} = y_{t+1} - \begin{pmatrix} 1 & y_t \end{pmatrix}\widehat{\theta}_{1,t}$. Following CS, the test statistic is:

$$M_P = \int_\Gamma m_P(\gamma)^2\phi(\gamma)d\gamma, \qquad (15)$$

and

$$m_P(\gamma) = \frac{1}{P^{1/2}}\sum_{t=R}^{T-1} g'(\widehat{u}_{1,t+1})w(Z^t,\gamma), \qquad (16)$$

where $\int_\Gamma \phi(\gamma)d\gamma = 1$, $\phi(\gamma) \geq 0$, and $\phi(\gamma)$ is absolutely continuous with respect to Lebesgue measure. In the sequel, we need:

Assumption A4: (i) w is a bounded, twice continuously differentiable function on the interior of Γ and $\nabla_\gamma w(Z^t,\gamma)$ is bounded uniformly in Γ; and (ii) $\nabla_\gamma \nabla_{\theta_1} q'_{1,t}(\theta_1) w(Z^{t-1},\gamma)$ is continuous on $\Theta_1 \times \Gamma$, where $q'_{1,t}(\theta_1) = q'_1(y_t - \theta_{1,1} - \theta_{1,2}y_{t-1})$, Γ a compact subset of R^d and is $2r-$dominated uniformly in $\Theta_1 \times \Gamma$, with $r \geq 2(2+\psi)$ where ψ is the same positive constant as that defined in Assumption A1.

Assumption A5 requires the function w to be bounded and twice continuously differentiable; such a requirement is satisfied by logistic or exponential functions, for example.

Proposition 2: Let assumptions A1-A4 hold. Then, the following results hold: (i) Under H_0,

$$M_P = \int_\Gamma m_P(\gamma)^2 \phi(\gamma) d\gamma \xrightarrow{d} \int_\Gamma Z(\gamma)^2 \phi(\gamma) d\gamma,$$

where $m_P(\gamma)$ is defined in equation (16) and Z is a Gaussian process with covariance kernel given by:

$$\begin{aligned} K(\gamma_1,\gamma_2) &= S_{gg}(\gamma_1,\gamma_2) + 2\Pi \mu'_{\gamma_1} B^\dagger S_{hh} B^\dagger \mu_{\gamma_2} + \Pi B^\dagger \mu'_{\gamma_1} S_{gh}(\gamma_2) \\ &\quad + \Pi \mu'_{\gamma_2} B^\dagger S_{gh}(\gamma_1), \end{aligned}$$

with $\mu_{\gamma_1} = E(\nabla_{\theta_1}(g'_{t+1}(u_{1,t+1})w(Z^t,\gamma_1)))$, $B^\dagger = (-E(\nabla^2_{\theta_1} q_1(u_{1,t})))^{-1}$,
$S_{gg}(\gamma_1,\gamma_2) = \sum_{j=-\infty}^{\infty} E(g'(u_{1,s+1})w(Z^s,\gamma_1)g'(u_{1,s+j+1})w(Z^{s+j},\gamma_2))$,
$S_{hh} = \sum_{j=-\infty}^{\infty} E(\nabla_{\theta_1} q_1(u_{1,s}) \nabla_{\theta_1} q_1(u_{1,s+j})')$,
$S_{gh}(\gamma_1) = \sum_{j=-\infty}^{\infty} E(g'(u_{1,s+1})w(Z^s,\gamma_1)\nabla_{\theta_1} q_1(u_{1,s+j})')$, and γ, γ_1, and γ_2 are generic elements of Γ.

(ii) Under H_A, for $\varepsilon > 0$, $\lim_{P\to\infty} \Pr\left(\frac{1}{P}\int_\Gamma m_P(\gamma)^2\phi(\gamma)d\gamma > \varepsilon\right) = 1$.

As in the previous application, the limiting distribution under H_0 is a Gaussian process with a covariance kernel that reflects both the dependence structure of the data and, for $\pi > 0$, the effect of parameter estimation error. Hence, critical values are data dependent and cannot be tabulated.

In order to implement this statistic using the recursive PEE bootstrap, define[12]

$$\widehat{\theta}^*_{1,t} = (\widehat{\theta}^*_{1,1,t}, \widehat{\theta}^*_{1,2,t})' = \arg\min_{\theta_1\in\Theta_1} \frac{1}{t}\sum_{j=2}^{t} q_1(y^*_j - \theta_{1,1} - \theta_{1,2} y^*_{j-1}). \tag{17}$$

Also, define $\widehat{u}^*_{1,t+1} = y^*_{t+1} - \begin{pmatrix} 1 & y^*_t \end{pmatrix} \widehat{\theta}^*_{1,t}$. The bootstrap test statistic is:

$$M^*_P = \int_\Gamma m^*_P(\gamma)^2 \phi(\gamma) d\gamma,$$

[12]Recall that $y^*_t, Z^{*,t}$ has been obtained via the resampling procedure described in Section 2

where, recalling that $g = q_1$,

$$\begin{aligned} m_P^*(\gamma) &= \frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \left(g'(\widehat{u}^*_{1,t+1}) w(Z^{*,t},\gamma) - g'(\widehat{u}_{1,t+1}) w(Z^t,\gamma) \right) \\ &+ \frac{1}{T} \sum_{t=2}^{T} \nabla_{\theta_1} q_1 (y_t - \widehat{\theta}_{1,1,T} - \widehat{\theta}_{1,2,T} y_{t-1}) w(Z^t,\gamma) \left(-\frac{1}{T} \sum_{t=2}^{T} \nabla^2_{\theta_1} q_1 (y_t - \widehat{\theta}_{1,1,T} - \widehat{\theta}_{1,2,T} y_{t-1}) \right)^{-1} \\ &\times \frac{1}{\sqrt{P}} \sum_{i=1}^{P-1} a_{R,i} \left(\nabla_{\theta_1} q_1 (y_{R+i} - \widehat{\theta}_{1,1,T} - \widehat{\theta}_{1,2,T} y_{R+i-1}) \right. \\ &\left. - \frac{1}{P} \sum_{i=1}^{P-1} \nabla_{\theta_1} q_1 (y_{R+i} - \widehat{\theta}_{1,1,T} - \widehat{\theta}_{1,2,T} y_{R+i-1}) \right) \end{aligned}$$

Proposition 3: Let assumptions A1-A4 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 $R \to \infty$,

$$P\left(\omega : \sup_{v \in \Re} \left| P^*_{R,P}\left(\int_\Gamma m_P^*(\gamma)^2 \phi(\gamma) d\gamma \leq v \right) - P\left(\int_\Gamma m_P^\mu(\gamma)^2 \phi(\gamma) d\gamma \leq v \right) \right| > \varepsilon \right) \to 0,$$

where $a_{R,i} = \frac{1}{R+i} + \frac{1}{R+i+1} + ... + \frac{1}{R+P-1}$ and $m_P^\mu(\gamma) = m_P(\gamma) - \sqrt{P} E\left(g'(u_{1,t+1}) w(Z^t,\gamma) \right).$

The above result suggests proceeding the same way as in the first application. For any bootstrap replication, compute the bootstrap statistic, M_P^*. Perform B bootstrap replications (B large) and compute the percentiles of the empirical distribution of the B bootstrap statistics. Reject H_0 if M_P is greater than the $(1-\alpha)th$-percentile. Otherwise, do not reject. Now, for all samples except a set with probability measure approaching zero, M_P has the same limiting distribution as the corresponding bootstrap statistic under H_0, thus ensuring asymptotic size equal to α. Under the alternative, M_P diverges to (plus) infinity, while the corresponding bootstrap statistic has a well defined limiting distribution, ensuring unit asymptotic power.

4 Predictive Density Evaluation

In several instances, such as financial risk management, for example, one is interested in predicting either a particular confidence interval (e.g. Value at Risk) or the entire conditional distribution of a variable of interest. Hence, over the last few years, a new strand of literature addressing the issue of predictive density evaluation has arisen (see e.g. Diebold, Gunther and Tay (DGT: 1998), Christoffersen (1998), Bai (2003), Clements and Smith (2000,2002) Diebold, Hahn and Tay (1999), Hong (2001) and Christoffersen, Hahn and Inoue (2001)). 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. On the other hand, 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), Corradi, Swanson and Olivetti (2001) and Corradi and Swanson (2002)). In this application, 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 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.[13] 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. Thus, we posit that *all* models should be viewed as approximations of some true unknown underlying data generating process.

In this application, 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

[13]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. If the null is rejected, Hong (2001) also proposes a test for uniformity robust to non independence, which is based on the comparison between a kernel density estimator and the uniform density. Diebold, Hahn and Tay (1999) propose a nonparametric correction for improving the density forecast when the uniform (but not the independence) assumption is violated.

predictive densities, and allowing for misspecification under both the null and the alternative hypotheses. At first blush, it may seem that the probability integral transform approach will yield a solution to this problem. However, it is difficult to map the degree of deviation from uniformity and independence into a meaningful measure of the degree of misspecification of a model, so that the probability integral transform method does not easily extend to the evaluation of multiple misspecified models.

Another 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 (2003b) extend Andrews (1997) in order to allow for the in-sample comparison of multiple misspecified models. Our focus in this application is to extend their results to out-of-sample predictive density evaluation via use of the block recursive bootstrap. More specifically, and using the notation outlined above, our objective is to form parametric conditional distributions for a scalar random variable, y_{t+1}, given Z^t, and to select among these. 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 $F_0(u|Z^t, \theta_0) = \Pr(y_{t+1} \leq u|Z^t)$. Hereafter, assume that $q_i(y_t, Z^{t-1}, \theta_i) = -\ln f_i(y_t|Z^{t-1}, \theta_i)$, where $f_i(\cdot|\cdot, \theta_i)$ is the conditional density associated with F_i, $i = 1, ..., n$, so that in this application, $\theta_i^\dagger$ is the probability limit of a quasi maximum likelihood estimator (QMLE). If model i is correctly specified, then $\theta_i^\dagger = \theta_0$. 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.[14]

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. The hypotheses of interest are:

$$H_0: \max_{k=2,\dots,n} \int_U E\left(\left(F_1(u|Z^t, \theta_1^\dagger) - F_0(u|Z^t, \theta_0)\right)^2 - \left(F_k(u|Z^t, \theta_k^\dagger) - F_0(u|Z^t, \theta_0)\right)^2\right) \phi(u) du \leq 0 \tag{18}$$

[14]In this test, the competing models are known. This is different than the probability integral transform approach where only the null model is explicitly stated.

versus

$$H_A : \max_{k=2,\dots,n} \int_U E\left(\left(F_1(u|Z^t,\theta_1^\dagger) - F_0(u|Z^t,\theta_0)\right)^2 - \left(F_k(u|Z^t,\theta_k^\dagger) - F_0(u|Z^t,\theta_0)\right)^2\right)\phi(u)du > 0, \tag{19}$$

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.[15,16] The statistic is:

$$Z_P = \max_{k=2,\dots,n} \int_U Z_{P,u}(1,k)\phi(u)du, \tag{20}$$

where

$$Z_{P,u}(1,k) = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\left(1\{y_{t+1} \leq u\} - F_1(u|Z^t,\widehat{\theta}_{1,t})\right)^2 - \left(1\{y_{t+1} \leq u\} - F_k(u|Z^t,\widehat{\theta}_{k,t})\right)^2\right). \tag{21}$$

[15] Kitamura (2002) proposes a comparison among misspecified conditional models, subject to given moment restrictions, in terms of the conditional entropy. Giacomini (2002) proposed an evaluation method for predictive densities based on a weighted likelihood ratio, accuracy is measured in terms of the Kullback Leibler Information Criterion (KLIC). The relation between our approach and the KLIC is outlined in Corradi and Swanson (2003b).

[16] 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:

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

versus

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

Analogously, if interest focuses on testing the null of equal accuracy of only two predictive conditional distribution models, say F_1 and F_k, 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

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

In this application we find more "natural" to estimate each model by QMLE, so that in terms of equation (??) $q_i = -\ln f_i$, where f_i is the conditional density associated with model i, and $\widehat{\theta}_{i,t}$ is defined as $\widehat{\theta}_{i,t} = \arg\max_{\theta_i \in \Theta_i} \frac{1}{t}\sum_{j=s}^{t} \ln f_i(y_j, Z^{j-1}, \theta_i), \quad R \leq t \leq T-1,\ i = 1, ..., n.$

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 additional assumption.

Assumption A5: (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, $r > 2$, $i = 1, ..., n$;[17] and (ii) let $v_{kk}(u) =\mathrm{plim}_{T\to\infty}$ $Var\left(\frac{1}{\sqrt{T}}\sum_{t=s}^{T}\left(\left(\left(1\{y_{t+1} \leq u\} - F_1(u|Z^t, \theta_1^\dagger)\right)^2 - \mu_1^2(u)\right) - \left(\left(1\{y_{t+1} \leq u\} - F_k(u|Z^t, \theta_k^\dagger)\right)^2 - \mu_k^2(u)\right)\right)\right)$ $k = 2, ..., n$, 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.

Analogous to assumption A4, assumptions A6(i)-(ii) are standard smoothness and domination conditions imposed on the conditional distributions of the models, and assumption A6(iii) 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 4: Let assumptions A1-A3 and A6 hold.[18] Then:

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

where $Z_{1,k}(u)$ is a zero mean Gaussian process with covariance $C_k(u, u')$ equal to:

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

[17]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 $\sup_t \sup_{u\in\Re} E(D_t(u)^{2r}) < \infty$.

[18]Note that A2 should hold with $q_i = -\ln f_i$, and A2(ii) should read as $E(f_i(y_{t+1}|Z^t, \theta_i)) < (f_i(y_{t+1}|Z^t, \theta_i^\dagger))$, for all $\theta_i \neq \theta_i^\dagger$.

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

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

$$+8\Pi m_{\theta_1^\dagger}(u)'A(\theta_1^\dagger)E\left(\sum_{j=-\infty}^{\infty}\nabla_{\theta_1}\ln f_1(y_{s+1}|Z^s,\theta_1^\dagger)\nabla_{\theta_1}\ln f_1(y_{s+j+1}|Z^{s+j},\theta_1^\dagger)'\right)A(\theta_1^\dagger)m_{\theta_1^\dagger}(u')$$

$$+8\Pi m_{\theta_k^\dagger}(u)'A(\theta_k^\dagger)E\left(\sum_{j=-\infty}^{\infty}\nabla_{\theta_k}\ln f_k(y_{s+1}|Z^s,\theta_k^\dagger)\nabla_{\theta_k}\ln f_k(y_{s+j+1}|Z^{s+j},\theta_k^\dagger)'\right)A(\theta_k^\dagger)m_{\theta_k^\dagger}(u')$$

$$-8\Pi m_{\theta_1^\dagger}(u,)'A(\theta_1^\dagger)E\left(\sum_{j=-\infty}^{\infty}\nabla_{\theta_1}\ln f_1(y_{s+1}|Z^s,\theta_1^\dagger)\nabla_{\theta_k}\ln f_k(y_{s+j+1}|Z^{s+j},\theta_k^\dagger)'\right)A(\theta_k^\dagger)m_{\theta_k^\dagger}(u')$$

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

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

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

$$+4\Pi m_{\theta_k^\dagger}(u)'A(\theta_k^\dagger)E\left(\sum_{j=-\infty}^{\infty}\nabla_{\theta_k}\ln f_k(y_{s+1}|Z^s,\theta_k^\dagger)'\left(\left(1\{y_{s+j+1}\leq u\}-F_1(u|Z^{s+j},\theta_1^\dagger)\right)^2-\mu_1^2(u)\right)\right) \quad (22)$$

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 $A(\theta_i^\dagger)=\left(E\left(-\nabla^2_{\theta_i}\ln f_i(y_{t+1}|Z^t,\theta_i^\dagger)\right)\right.$ 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 $\sqrt{P}$.

In a recent paper, Hansen (2001) explores the point made by White (2000) that the reality check test can have level going to zero at the same time that power goes to unity (i.e. that the test is conservative unless $\int_U \left(\mu_1^2(u)-\mu_k^2(u)\right)\phi(u)du=0,\forall k$), and suggests a mean correction for the statistic in order to address this feature of the test. Our predictive density test has the same features and can also be modified using the method proposed by Hansen.[19]

Now, define the bootstrap statistic as:[20]

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

where

$$\begin{aligned}
Z_{P,u}^*(1,k) \;=\; & \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\left(\left(1\{y_{t+1}^*\le u\}-F_1(u|Z^{*,t},\widehat{\theta}_{1,t}^*)\right)^2-\left(1\{y_{t+1}\le u\}-F_1(u|Z^t,\widehat{\theta}_{1,t})\right)^2\right)\right.\\
& \left.-\left(\left(1\{y_{t+1}^*\le u\}-F_k(u|Z^{*,t},\widehat{\theta}_{k,t}^*)\right)^2-\left(1\{y_{t+1}\le u\}-F_k(u|Z^t,\widehat{\theta}_{k,t})\right)^2\right)\right)\\
& -\frac{2}{T}\sum_{t=s}^{T-1}\left(\nabla_{\theta_1}F_1(u|Z^t,\widehat{\theta}_{1,T})'\left(1\{y_{t+1}^*\le u\}-F_1(u|Z^t,\widehat{\theta}_{1,T})\right)\right)'\left(-\frac{1}{T}\sum_{t=s}^{T-1}\nabla_{\theta_1}^2 f_1(y_t|Z^{t-1},\widehat{\theta}_{1,T})\right)^{-1}\\
& \times\frac{1}{\sqrt{P}}\sum_{i=1}^{P-1}a_{R,i}\left(\nabla_{\theta_1}f_1(y_{R+i}|Z^{R+i-1},\widehat{\theta}_{1,T})-\frac{1}{P}\sum_{i=1}^{P}\nabla_{\theta_1}f_1(y_{R+i}|Z^{R+i-1},\widehat{\theta}_{1,T})\right)\\
& +\frac{2}{T}\sum_{t=s}^{T}\nabla_{\theta_k}F_k(u|Z^t,\widehat{\theta}_{k,T})\left(1\{y_{t+1}\le u\}-F_k(u|Z^t,\widehat{\theta}_{k,T})\right)'\left(-\frac{1}{T}\sum_{t=s}^{T-1}\nabla_{\theta_k}^2 f_k(y_t|Z^{t-1},\widehat{\theta}_{k,T})\right)^{-1}\\
& \times\frac{1}{\sqrt{P}}\sum_{i=1}^{P-1}a_{R,i}\left(\nabla_{\theta_k}f_k(y_{R+i}|Z^{R+i-1},\widehat{\theta}_{k,T})-\frac{1}{P}\sum_{i=1}^{P}\nabla_{\theta_k}f_k(y_{R+i}|Z^{R+i-1},\widehat{\theta}_{k,T})\right).
\end{aligned}\tag{23}$$

[19] An alternative to the bootstrap critical values, may the construction of critical values based on subsampling (e.g. Politis, Romano and Wolf (1999), Ch.3). Heuristically, we construct $T-2b_T$ statistics using subsamples of length b_T, where $b_T/T\to 0$; the empirical distribution of the statistics computed over the various subsamples, properly mimics the distribution of the statistic. Thus, it provides valid critical values even for the case of $\max_{k=2,\ldots,m}E\left(g(u_{1,t+1})-g(u_{k,t+1})\right)=0$, but $E\left(g(u_{1,t+1})-g(u_{k,t+1})\right)<0$ for some k.Needless to say, the problem is that unless the sample is very large, the empirical distribution of the subsampled statistics provides a poor approximation to the limiting distribution of the statistic. The subsampling approach has been followed by Linton, Maasoumi and Whang (2003), in the context of testing for stochastic dominance.

[20] As in the previous two applications, y_t^* and $Z^{*,t}$ have been obtained via the resampling procedure described in Section 2

Proposition 5: Let assumptions A1-A3 and A6 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 $R \to \infty$,

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

where $a_{R,i} = \frac{1}{R+i} + \frac{1}{R+i+1} + \ldots + \frac{1}{R+P-1}$ and $Z^{\mu}_{P,u}(1,k) = Z_{P,u}(1,k) - \sqrt{P}\left(\mu_1^2(u) - \mu_k^2(u)\right)$.

The above result suggests proceeding in the following manner. For any bootstrap replication, compute the bootstrap statistic, Z^*_P. Perform B bootstrap replications (B large) and compute the quantiles of the empirical distribution of the B bootstrap statistics. Reject H_0, if S_P 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 has the same limiting distribution as the corresponding bootstrapped statistic when $\int_U \left(\mu_1^2(u) - \mu_k^2(u)\right)\phi(u)du = 0, \forall k$, 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, S_P diverges to (plus) infinity, while the corresponding bootstrap statistic has a well defined limiting distribution, ensuring unit asymptotic power.

5 Monte Carlo Results

Proposition 1 establishes the first order validity of the recursive PEE bootstrap. In this section we study its finite sample behavior via a small Monte Carlo experiment. In particular our objective is to answer the following two questions: (i) Does the inclusion of the adjustment term lead to (substantially) improved coverage probabilities? (ii) What sorts of coverage probabilities can we expect for different block lengths in finite sample applications?

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, in the present context, when we estimate (via OLS) an AR(1) (or an AR(2)) model, $\widehat{\theta}_{i,t} = (\widehat{c}_{i,t}, \hat{\rho}_{i,t})'$ (or $\widehat{\theta}_{i,t} = (\widehat{c}_{i,t}, \hat{\rho}_{1,i,t}, \hat{\rho}_{2,i,t})'$), with $i = 1,2$ denoting the estimate models (AR(1) and AR(2), respectively), and $\theta_i^\dagger = (c_i^\dagger, \rho_i^\dagger)'$ (or $\theta_i^\dagger = (c_i^\dagger, \rho_{1,i}^\dagger, \rho_{2,i}^\dagger)'$), where $\theta_i^\dagger$ denotes the probability limit of $\widehat{\theta}_{i,t}$. Needless to say, in the case of correct dynamic specification, $\theta_i^\dagger$ 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)), $\theta_i^\dagger$ represents pseudo true values, which can be explicitly computed. We confine our attention on the slope parameters. 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 $\rho^\dagger = \{0.2, 0.4, 0.6, 0.8\}$. In this case the bootstrap statistic with adjustment is given by:[21]

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

where $\widehat{e}_{R+j} = \left(y_{R+j}-\overline{y}\right) - \widehat{\rho}_T\left(y_{R+j-1}-\overline{y}\right)$. Also, define the bootstrap statistic without adjustment as $\Psi_{2,R,P}^* = \frac{1}{\sqrt{P}}\sum_{t=R}^{P-1}\left(\widehat{\rho}_t^* - \widehat{\rho}_t\right)$. Hereafter, let $z_{1,\alpha}^*$ be the $(1-\alpha)$ quantile of the distribution of $\Psi_{1,R,P}^*$ and $z_{2,\alpha}^*$ the $(1-\alpha)$ quantile of the distribution of $\Psi_{2,R,P}^*$. We now define $100(1-\alpha)\%$, equal-tailed, two-sided confidence intervals corresponding to the recursive bootstrap with adjustment and the recursive bootstrap without adjustment, respectively:

$$CI_1^* : \left\{\frac{1}{P}\sum_{t=R}^{P-1}\widehat{\rho}_t - \frac{z_{1,\alpha/2}^*}{\sqrt{P}}, \frac{1}{P}\sum_{t=R}^{P-1}\widehat{\rho}_t + \frac{z_{1,(1-\alpha/2)}^*}{\sqrt{P}}\right\} \tag{24}$$

$$CI_2^* : \left\{\frac{1}{P}\sum_{t=R}^{P-1}\widehat{\rho}_t - \frac{z_{2,\alpha/2}^*}{\sqrt{P}}, \frac{1}{P}\sum_{t=R}^{P-1}\widehat{\rho}_t + \frac{z_{2,(1-\alpha/2)}^*}{\sqrt{P}}\right\} \tag{25}$$

The coverage probabilities for CI_1^* and CI_2^* 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 CI_2^*, then it is definitely worthwhile to construct bootstrap critical values based on the recursive 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 recursive PEE bootstrap.

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 $T = \{600,1200,2400\}$

[21]Note that $\widehat{\rho}_t^*$ 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.

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 nominal coverage probability, across all experiments, is set equal to 0.95. 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 $\alpha = 0.05$.

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 corresponding to *rec* correspond to coverage probabilities associated with CI_1^*. Entries corresponding to *no adj* are the rejection probabilities associated with 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 $l_1 = l_2$.[22]

Turning now to the results, two clear-cut conclusions emerge. First, inspection of Tables 1-4 suggests that the recursive PEE bootstrap (*rec*) performs better than the version without adjustment (*no adj*). This is true for almost all block lengths, parameters, and regardless of DGP. 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). In addition, performance improvement is rather similar across the two bootstraps when sample size and block length are increased. These findings suggest that as long as blocks of moderate to large length are used, the recursive PEE bootstrap performs adequately, and should be useful when constructing critical values for tests such as the predictive density test discussed above. Second, when $\rho_2^\dagger = 0$, which is the case in Panel A of Tables 1-4, as the true DGP in these cases is an AR(1) process, coverage is often better than in the case in which we estimate an AR(2) than when we estimate an AR(1) (see, for example, the first 2 rows of Table 2, Panel A). Interestingly, the converse does not always hold. In particular, when the true DGP is an AR(2) and an AR(1) is estimated then coverage associated with $\hat{\rho}$ is often as good as that associated with $\hat{\rho}_1$ and $\hat{\rho}_2$, and is often better

[22] As $P = R = 0.5T$, we use the same block length when resampling from the first R observations and from the last P observations.

than that associated with $\hat{\rho}_2$ (see, for example, the first 2 rows of Table 2, Panel B).This is perhaps surprising, given that we always set $\rho_1 = \rho_2$, and suggests a complicated interaction between parsimony and model specification.

6 Concluding Remarks

In this paper we introduce a parameter estimation error (PEE) bootstrap for recursive estimation schemes (named the recursive PEE bootstrap) and show its first order validity (i.e. the ability of appropriately constructed bootstrap statistics to mimic the limiting distribution of $\frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\widehat{\theta}_t - \theta^{\dagger} \right)$, where R denotes the length of the estimation period, P the number of recursively estimated parameters, $\widehat{\theta}_t$ is a recursive $m-$estimator constructed using the first t observations, and $\theta^{\dagger}$ is its probability limit). We also establish the validity of a block bootstrap based on resampling from the entire original sample of data, as is done in non-recursive setups, and show that such a bootstrap resampling scheme essentially requires the use of twice as many "bias adjustment" terms when statistical tests are constructed using recursively estimated models. Two applications of the recursive PEE bootstrap are developed. The first is an out-of-sample version of the integrated conditional moment (ICM) test of Bierens (1982,1990) and Bierens and Ploberger (1997) which provides out of sample tests consistent against generic (nonlinear) alternatives. The second is a procedure assessing the relative out-of-sample predictive accuracy of multiple conditional distribution models. This procedure is based on an extension of the Andrews (1997) conditional Kolmogorov test and is an alternative to the predictive density test of Diebold, Gunther and Tay (1998) and others. Finally, we examine the finite sample behavior of the recursive PEE bootstrap via a small Monte Carlo study, and show that the recursive PEE bootstrap has much better (finite sample) coverage than a related bootstrap procedure which neglects to include appropriate (bias) adjustment terms.

7 Appendix

The proof of Theorem 1 requires the following three lemmas. As the statement below 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, for notational simplicity, in the sequel we set $l_1 = l_2 = l$.

Lemma A1: Let A1-A3 hold. If as $R \to \infty$ and $P \to \infty$, $l \to \infty$, $l/R \to 0$ and $l/P \to 0$, then (i) $\sup_{t \geq R} \left| \widehat{\theta}_t^* - \widehat{\theta}_t \right| = o_{P^*}(1)$, $\Pr -P$, and (ii) $\sup_{t \geq R} \left| \widehat{\theta}_t^* - \theta^\dagger \right| = o_{P^*}(1)$, $\Pr -P$.

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

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

$$Var^* \left(\frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \frac{1}{t} \sum_{j=s}^{t} \left(\nabla_\theta q(y_j^*, Z^{*,j-1}, \theta^\dagger) \right) \right) = 2\Pi C_{00}, \ \ \Pr -P,$$

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 $\Pi = 1 - \pi^{-1} \ln(1+\pi)$.

Proof of Lemma A1: (i) We need to extend the consistency results for bootstrap $m-$estimators of Goncalves and White (2002b, Theorem 2.1), to the case of recursive $m-$estimators. Recalling that for $t \geq R$,

$$\widehat{\theta}_t = \arg\min_{\theta \in \Theta} \frac{1}{t} \sum_{j=s}^{t} q(y_j, Z^{j-1}, \theta) \text{ and } \widehat{\theta}_t^* = \arg\min_{\theta \in \Theta} \frac{1}{t} \sum_{j=s}^{t} q(y_j^*, Z^{*,j-1}, \theta),$$

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

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

Hereafter, for notational simplicity let $q(y_j^*, Z^{*,j-1}, \theta) = q_j^*(\theta)$ and $q(y_j, Z^{j-1}, \theta) = q_j(\theta)$, and let $\mu = E(q_j(\theta))$, $\forall \theta \in \Theta$. Now,

$$\sup_{t \geq R} \sup_{\theta \in \Theta} \left| \frac{1}{t} \sum_{j=s}^{t} \left(q_j^*(\theta) - q_j(\theta) \right) \right| \leq \sup_{t \geq R} \sup_{\theta \in \Theta} \left| \frac{1}{t} \sum_{j=s}^{t} \left(q_j^*(\theta) - E^* \left(q_j^*(\theta) \right) \right) \right| \tag{26}$$

$$+ \sup_{t \geq R} \sup_{\theta \in \Theta} \left| \frac{1}{t} \sum_{j=s}^{t} \left(q_j(\theta) - \mu \right) \right| + \sup_{t \geq R} \sup_{\theta \in \Theta} \left| \frac{1}{t} \sum_{j=s}^{t} \left(E^* \left(q_j^*(\theta) \right) - \mu \right) \right|. \tag{27}$$

Now, recalling that $R = b_1 l$ and $P = b_2 l$, let $t = (b_1 + k)l$ for any generic $k = 1, ..., b_2$,

$$\frac{1}{t}\sum_{j=s}^{t} E^*(q_j^*(\theta)) = \frac{1}{(b_1+k)l}\sum_{j=s}^{R} E^*(q_j^*(\theta)) + \frac{1}{(b_1+k)l}\sum_{j=R+1}^{R+kl} E^*(q_j^*(\theta)) + O\left(\frac{l}{R}\right), \ \Pr - P,$$

uniformly in θ, as under A3, P and R grow at the same rate, as the sample size increases. Now,

$$\begin{aligned}
\frac{1}{t}\sum_{j=s}^{R} E^*(q_j^*(\theta)) &= \frac{b_1}{t}\left(\frac{q_{s+1}(\theta)+...+q_{l+s}(\theta)}{R-l+1} + \frac{q_{2+s}(\theta)+...+q_{l+1+s}(\theta)}{R-l+1} + ... + \frac{q_{R-l+1}(\theta)+...+q_R(\theta)}{R-l+1}\right) \\
&= \frac{b_1}{b_1+k}\frac{1}{R}\sum_{j=1}^{R} q_j(\theta) + O\left(\frac{l}{R}\right), \ \Pr - P,
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{t}\sum_{j=R+1}^{R+kl} E^*(q_j^*(\theta)) &= \frac{k}{t}\left(\frac{q_{R+1}(\theta)+...+q_{R+l}(\theta)}{P-l+1} + ... + \frac{q_{R+P-l+1}(\theta)+...+q_{R+P-1}(\theta)}{P-l+1}\right) \\
&= \frac{k}{k+b_1}\frac{1}{P}\sum_{j=R+1}^{R+P} q_j(\theta) + O\left(\frac{l}{R}\right), \ \Pr - P.
\end{aligned}$$

Thus

$$\frac{1}{t}\sum_{j=1}^{t} E^*(q_j^*(\theta)) = \frac{b_1}{(b_1+k)}\frac{1}{R}\sum_{j=1}^{R} q_j(\theta) + \frac{k}{(b_1+k)}\frac{1}{P}\sum_{j=R+1}^{R+P} q_j(\theta) + O\left(\frac{l}{R}\right), \ \Pr - P,$$

and so the second term on the right hand side of (27), by the uniform strong law of large number approaches zero in probability. Analogously, the first term on the right hand side of (27), by the uniform strong law of large number is $o_P(1)$. As for the first term on the RHS of (26), it is majorized by:

$$\sup_{\theta\in\Theta}\left|\frac{1}{b_1}\sum_{j=s}^{b_1}\left(\frac{U_j(\theta) - E^*(U_j(\theta))}{l}\right)\right| + \sup_{k\geq 1}\sup_{\theta\in\Theta}\left|\frac{k}{b_1+k}\frac{1}{k}\sum_{j=s}^{k}\left(\frac{U_j(\theta) - E^*(U_j(\theta))}{l}\right)\right|, \qquad (28)$$

where for $j = s-1, ..., R-l$, U_i are independent discrete uniform taking value $(q_{j+1}(\theta)+...+q_{j+l}(\theta))$ with probability $1/(R-l+1)$, and for $j = R, ..., R+P-l$, U_i are independent discrete uniform taking value $(q_{j+1+R}(\theta) + ... + q_{j+l+R}(\theta))$ with probability $1/(P-l+1)$. By the uniform law of large numbers for asymptotically independent and homogeneous observations, the sum of the two term in (28) approaches zero in P^*-probability.

(ii) Immediate as,

$$\sup_{t\geq R}\left|\widehat{\theta}_t^* - \theta^\dagger\right| \leq \sup_{t\geq R}\left|\widehat{\theta}_t^* - \widehat{\theta}_t\right| + \sup_{t\geq R}\left|\widehat{\theta}_t - \theta^\dagger\right|,$$

and the first term is $o_P^*(1)$, $\Pr-P$ by part (i), while the second term is $o(1)$ $\Pr-P$.

Proof of Lemma A2: First note that,

$$t^{\vartheta}\left(\widehat{\theta}_t^* - \theta^\dagger\right) = \left(\frac{1}{t}\sum_{j=s}^{t}\nabla_\theta^2 q(y_j^*, Z^{*,j-1}, \overline{\theta}_t^*)\right)^{-1}\left(\frac{1}{t^{1-\vartheta}}\sum_{j=s}^{t}\nabla_\theta q(y_j^*, Z^{*,j-1}, \theta^\dagger)\right),$$

with $\overline{\theta}_t^* \in (\widehat{\theta}_t^*, \theta^\dagger)$. Hereafter, for notational simplicity let $\nabla_\theta^2 q(y_j^*, Z^{*,j-1}, \theta) = \nabla^2 q_j^*(\theta)$, $\nabla_\theta^2 q(y_j, Z^{j-1}, \theta) = \nabla^2 q_j(\theta)$, and let $B^\dagger = \left(E\left(-\nabla_\theta^2 q_t(\theta^\dagger)\right)\right)^{-1}$,

$$\sup_{t\geq R}\left|\frac{1}{t}\sum_{j=s}^{t}\left(\nabla^2 q_j^*(\overline{\theta}_t^*) - B^{\dagger-1}\right)\right| \leq \sup_{t\geq R}\left|\frac{1}{t}\sum_{j=s}^{t}\left(\nabla^2 q_j^*(\overline{\theta}_t^*) - E^*\left(\nabla^2 q_j^*(\overline{\theta}_t^*)\right)\right)\right| \tag{29}$$

$$+\sup_{t\geq R}\left|\frac{1}{t}\sum_{j=s}^{t}\left(\nabla^2 q_j(\overline{\theta}_t) - B^{\dagger-1}\right)\right| + \sup_{t\geq R}\left|\frac{1}{t}\sum_{j=s}^{t}\left(\nabla^2 q_j(\overline{\theta}_t) - E^*\left(\nabla^2 q_j^*(\overline{\theta}_t^*)\right)\right)\right|, \tag{30}$$

as $\overline{\theta}_t^* \in (\widehat{\theta}_t^*, \theta^\dagger)$ and $\overline{\theta}_t \in (\widehat{\theta}_t, \theta^\dagger)$, given Lemma A1, $\sup_{t\geq R}\left|\overline{\theta}_t^* - \overline{\theta}_t\right| = o_{P^*}(1)$ $\Pr-P$, thus the right hand side of (29) and the sum of the two term in (30) are $o_{P^*}(1)$ $\Pr-P$, by the same argument used in the proof of Lemma A1. Given A3(ii), it follows immediately that

$$\sup_{t\geq R}\left|\left(\frac{1}{t}\sum_{j=s}^{t}\nabla_\theta^2 q(y_j^*, Z^{*,j-1}, \overline{\theta}_t^*)\right)^{-1} - B^\dagger\right| = o_{P^*}(1),\ \ \Pr-P. \tag{31}$$

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

$$\sup_{t\geq R}\left|\frac{1}{n_t}\sum_{j=s}^{t}h_j^*(\theta^\dagger)\right| \leq \sup_{t\geq R}\left|\frac{1}{n_t}\sum_{j=s}^{t}\left(h_j^*(\theta^\dagger) - E^*\left(h_j^*(\theta^\dagger)\right)\right)\right| + \sup_{t\geq R}\left|\frac{1}{n_t}\sum_{j=2}^{t}E^*\left(h_j^*(\theta^\dagger)\right)\right|, \tag{32}$$

and noting that, by the same argument as in the proof of Lemma A1, up to a term of order $O(l/P^{1/2})$, $\Pr-P$,

$$\begin{aligned}\sup_{t\geq R}\left|\frac{1}{n_t}\sum_{j=s}^{t}E^*\left(h_j^*(\theta^\dagger)\right)\right| \quad &\leq \quad \left|\frac{1}{\sqrt{2R\log\log R}}\sum_{j=s}^{R}h_j(\theta^\dagger)\right| \\ &+\sup_{k\geq 1}\left|\frac{1}{\sqrt{2kl\log\log(kl)}}\sum_{j=R}^{R+kl}h_j(\theta^\dagger)\right|,\end{aligned} \tag{33}$$

and both the terms on the RHS of (33) are $O(1)$, $a.s.-P$, as, given A1 and A3 each component of both 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 (32) 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. First, recalling that, except for an overlapping of at most s observations, all blocks are independent each other, conditionally on the sample,

$$Var^*\left(\frac{1}{\sqrt{T}}\sum_{t=s}^{T-1} h_t^*\right) = Var^*\left(\frac{1}{\sqrt{T}}\sum_{t=s}^{R} h_t^*\right) + Var^*\left(\frac{1}{\sqrt{T}}\sum_{t=R+1}^{R+P-1} h_t^*\right) + O\left(\frac{l}{T^{1/2}}\right),\ \Pr -P$$

$$\begin{aligned} &= Var^*\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{b_1}\sum_{j=1}^{l} h_{I_i^R+j}\right) + Var^*\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{b_2}\sum_{j=1}^{l} h_{I_i^P+j}\right) + O\left(\frac{l}{T^{1/2}}\right),\ \Pr -P, \\ &= E^*\left(\frac{R}{T}\frac{1}{R}\sum_{i=1}^{b_1}\sum_{j=1}^{l}\sum_{k=1}^{l} h_{I_i^R+j}h_{I_i^R+k}\right) + E^*\left(\frac{P}{T}\frac{1}{P}\sum_{i=1}^{b_2}\sum_{j=1}^{l}\sum_{k=1}^{l} h_{I_i^P+j}h_{I_i^P+k}\right) + O\left(\frac{l}{T^{1/2}}\right),\ \Pr -P \end{aligned} \quad (34)$$

where for $i = 1, ..., b_1$ I_i^R are independent discrete uniform on $s, ..., R - l$, while for $i = 1, ..., b_2$ I_i^P are independent discrete uniform on $R, R + 1, ..., R + P - l$, thus after few simple manipulations, the equality on (34) can be rewritten as

$$\frac{R}{T}\frac{1}{R}\sum_{k=l}^{R-l}\sum_{j=-l}^{l} h_k h'_{k+j} + \frac{P}{T}\frac{1}{P}\sum_{k=R+l}^{R+P-l}\sum_{j=-l}^{l} h_k h'_{k+j} + O\left(\frac{l}{T^{1/2}}\right),\ \Pr -P, \quad (35)$$

also given A(1) and A(3) and the growth conditions on the parameter l, the sum of the first to two terms in (35), up to term approaching zero $\Pr -P$, is equal to C_{00}, and so
$V^* = \lim_{T\to\infty} Var^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h_j^*(\theta^\dagger)\right)$ is $O(1)$, $\Pr -P$. Now,

$$\begin{aligned} \sup_{t\geq R}\left\|V^{*-1/2}\frac{1}{n_t}\sum_{j=s}^{t}(h_j^* - E^*(h_j^*))\right\| \leq & \left\|V^{*-1/2}\frac{1}{\sqrt{2b_1\log\log b_1}}\sum_{i=1}^{b_1}\left(\frac{U_i - E^*(U_i)}{l}\right)\right\| \\ & + \sup_{k\geq 1}\left\|V^{*-1/2}\frac{1}{\sqrt{2k\log\log b_1}}\sum_{i=1}^{b_1}\left(\frac{U_i - E^*(U_i)}{l}\right)\right\|, \quad (36) \end{aligned}$$

where for $i = s, ..., R-l$, U_i is independent discrete uniform taking value $(h_{i+1}(\theta) + ... + h_{i+l}(\theta))$ with probability $1/(R - l + 1)$, and for $i = R, ..., R + P - l$, U_i is independent discrete uniform taking value $(h_{i+1+R}(\theta) + ... + h_{i+l+R}(\theta))$ with probability $1/(P - l + 1)$. Therefore the assumptions of Theorem 1 in Eberlain (1986) are satisfied and so the right hand side of (36) is $O^*_{a.s.}(1)$ $\Pr -P$, thus $\sup_{t\geq R}\left|\frac{1}{b_t}\sum_{j=s}^{t}\left(h_j^*(\theta^\dagger) - E^*\left(h_j^*(\theta^\dagger)\right)\right)\right|$ is also $O^*_{a.s.}(1)$ $\Pr -P$. Recalling (31), the desired statement then follows.

Proof of Lemma A3:

As in the proof of Lemma A2, let $\nabla_\theta q(y_j^*, Z^{*,j-1}, \theta) = h_j^*(\theta)$, and $\nabla_\theta q(y_j, Z^{j-1}, \theta) = h_j(\theta)$, also let $\nabla_\theta q(y_j^*, Z^{*,j-1}, \theta^\dagger) = h_j^*$, and $\nabla_\theta q(y_j, Z^{j-1}, \theta^\dagger) = h_j$. Along the lines of West (1996, proof of Lemma A5),

$$\frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \frac{1}{t} \sum_{j=s}^{t} h_j^* = \frac{a_{R,0}}{\sqrt{P}} \sum_{j=s}^{R} h_j^* + \frac{1}{\sqrt{P}} (a_{R,1} h_{R+1}^* + ... + a_{R,P-1} h_{R+P-1}^*) + o_{P^*}(1), \ \Pr -P \tag{37}$$

where $a_{R,i} = (R+i)^{-1} + ... + (R+P-1)^{-1}$, for $0 \leq i < P-1$.[23] Thus,

$$Var^* \left(\frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \frac{1}{t} \sum_{j=1}^{t} h_j^* \right) = \frac{R}{P} Var^* \left(a_{R,0} \frac{1}{\sqrt{R}} \sum_{j=1}^{R} h_j^* \right)$$

$$+ \frac{1}{P} Var^* \left(\sum_{j=1}^{P-1} a_{R,j} h_{R+j}^* \right) + \frac{1}{P} Cov^* \left(a_{R,0} \sum_{j=1}^{R} h_j^*, \sum_{j=1}^{P-1} a_{R,j} h_{R+j}^* \right)$$

Given that any of the last b_2 blocks can be correlated with any of the first b_1 blocks for at most s observations, s finite, up to a term approaching zero conditionally on the sample, the covariance term is equal to zero. Now, for $t = s, ...R$, $E^*\left(h_t^*\right) = R^{-1} \sum_{t=s}^{R} h_t + O(l/P) = \overline{h}_R + O(l/P)$, thus up to a term of order $O(l/R^{1/2})$,

$$\begin{aligned} Var^* \left(a_{R,0} \frac{1}{\sqrt{R}} \sum_{j=1}^{R} h_j^* \right) &= a_{R,0}^2 Var^* \left(\frac{1}{\sqrt{R}} \sum_{k=1}^{b_1} \sum_{i=1}^{l} h_{I_k+i} \right) \\ &= a_{R,0}^2 E^* \left(\frac{1}{R} \sum_{k=1}^{b_1} \sum_{i=1}^{l} \sum_{k=1}^{l} (h_{I_k+i} - \overline{h}_R)(h_{I_k+j} - \overline{h}_R)' \right) \end{aligned}$$

$$= a_{R,0}^2 \left(\frac{1}{R} \sum_{t=l}^{R-l} \sum_{j=-l}^{l} (h_t - \overline{h}_R)(h_{t+j} - \overline{h}_R)' \right) + O(l/R^{1/2}) \ \Pr -P.$$

So,

$$\begin{aligned} &\frac{R}{P} Var^* \left(a_{R,0} \frac{1}{\sqrt{R}} \sum_{j=1}^{R} h_j^* \right) \\ &= \frac{R a_{R,0}^2}{P} \sum_{j=-l}^{l} \gamma_j + \frac{R a_{R,0}^2}{P} \left(\frac{1}{R} \sum_{t=l}^{R-l} \sum_{j=-l}^{l} ((h_t - \overline{h}_R)(h_{t+j} - \overline{h}_R)' - \gamma_j) \right) + O\left(\frac{l^2}{R} \right), \end{aligned} \tag{38}$$

[23] The $o_{P^*}(1)$, $\Pr -P$ term on the RHS of (37) comes from the fact that the summation run from $j = s$ instead of $j = 1$.

where $\gamma_j = Cov(h_1, h_{1+j})$. By West (1996, proof of Lemma A5), it follows that $\frac{Ra^2_{R,0}}{P}\sum_{j=-l}^{l}\gamma_j \rightarrow \pi^{-1}\ln^2(1+\pi)C_{00}$, while the second term on the RHS above goes to zero $\Pr -P$ (see e.g. Theorem 2 in Newey and West (1987)). Now, for $j = 1, ..., P$, $E^*(a_{R,j}h^*_{R+j}) = a_{R,j}P^{-1}\sum_{j=1}^{P-1} h_{R+j} + O(l/P) = a_{R,j}\overline{h}_P + O(l/P)$, thus up to a term of order $O(l/P^{1/2})$ $\Pr -P$,

$$Var^*\left(\frac{1}{\sqrt{P}}\sum_{j=1}^{P-1} a_{R,j}h^*_{R+j}\right) = Var^*\left(\frac{1}{\sqrt{P}}\sum_{k=1}^{b_2}\sum_{i=1}^{l} a_{R,((k-1)l+i)}h_{R+I_k+i}\right)$$

$$= \frac{1}{P}E^*\left(\sum_{k=1}^{b_2}\sum_{i=1}^{l}\sum_{j=1}^{l} a_{R,((k-1)l+i)}a_{R,((k-1)l+j)}(h_{R+I_k+i} - \overline{h}_P)(h_{R+I_k+j} - \overline{h}_P)'\right)$$

$$= \frac{1}{P}\sum_{k=1}^{b_2}\sum_{i=1}^{l}\sum_{j=1}^{l} a_{R,((k-1)l+i)}a_{R,((k-1)l+j)}E^*\left((h_{R+I_k+i} - \overline{h}_P)(h_{R+I_k+j} - \overline{h}_P)'\right)$$

$$= \frac{1}{P}\sum_{k=1}^{b_2}\sum_{i=1}^{l}\sum_{j=1}^{l} a_{R,((k-1)l+i)}a_{R,((k-1)l+j)}\left(\frac{1}{P}\sum_{t=l}^{P-l}(h_{R+t+i} - \overline{h}_P)(h_{R+t+j} - \overline{h}_P)'\right) + O(l/P^{1/2}) \ \Pr -P$$

$$= \frac{1}{P}\sum_{k=1}^{b_2}\sum_{i=1}^{l}\sum_{j=1}^{l} a_{R,((k-1)l+i)}a_{R,((k-1)l+j)}\gamma_{i-j}$$
$$+\frac{1}{P}\sum_{k=1}^{b_2}\sum_{i=1}^{l}\sum_{j=1}^{l} a_{R,((k-1)l+i)}a_{R,((k-1)l+j)}\left(\frac{1}{P}\sum_{t=l}^{P-l}\left((h_{R+t+i} - \overline{h}_P)(h_{R+t+j} - \overline{h}_P)' - \gamma_{i-j}\right)\right)$$
$$+O(l/P^{1/2}) \ \Pr -P \tag{39}$$

We need to show that the last term on the last equality in (39) is $o(1)$ $\Pr -P$. First note that it is majorized by

$$\left|\frac{b_2}{P}\sum_{i=1}^{l}\sum_{j=1}^{l}\left(\frac{1}{P}\sum_{t=l}^{P-l}\left((h_{R+t+i} - \overline{h}_P)(h_{R+t+j} - \overline{h}_P)' - \gamma_{i-j}\right)\right)\right|$$
$$= \left|\frac{1}{P}\sum_{t=l}^{P-l}\sum_{j=-l}^{l}\left((h_{R+t} - \overline{h}_P)(h_{R+t+j} - \overline{h}_P)' - \gamma_j\right)\right| + O(l/P^{1/2}) \ \Pr -P \tag{40}$$

The first term on the RHS of (40) goes to zero in probability, by the same argument as in Lemma 2 in Corradi (1999)[24]. As for the first term on the RHS of the last equality in (39),

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

[24]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.

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

By the same argument as in Lemma A5 in West (1996), the second term on the RHS above approaches zero, while

$$\frac{1}{P}\sum_{t=l}^{P-l} a_{R,t}^2 \sum_{j=-l}^{l}\gamma_j \to \left(2[1-\pi^{-1}\ln(1+\pi)] - \pi^{-1}\ln^2(1+\pi)\right) C_{00}.$$

As the first term on the RHS of (38) converges to $\pi^{-1}\ln^2(1+\pi)C_{00}$ (see West (1996), p.1082), the desired outcome then follows.

Proof of Theorem 1:

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

$$\begin{aligned} = \ & \frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(-\frac{1}{t}\sum_{j=s}^{t}\nabla_\theta^2 q(y_j^*, Z^{*,j-1}, \overline{\theta}_t^*)\right)^{-1}\left(\frac{1}{t}\sum_{j=s}^{t}\nabla_\theta q(y_j^*, Z^{*,j-1}, \theta^\dagger)\right) \\ & -\frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(-\frac{1}{t}\sum_{j=s}^{t}\nabla_\theta^2 q(y_j, Z^{j-1}, \overline{\theta}_t)\right)^{-1}\left(\frac{1}{t}\sum_{j=s}^{t}\nabla_\theta q(y_j, Z^{j-1}, \theta^\dagger)\right), \end{aligned} \tag{41}$$

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

Given Lemma A1 and A2 and given A1-A3,

$$\sup_{t\geq R}\left(\left(\frac{1}{t}\sum_{j=s}^{t}\nabla_\theta q(y_j^*, Z^{*,j-1}, \overline{\theta}_t^*)\right)^{-1} - \left(\frac{1}{t}\sum_{j=s}^{t}\nabla_\theta q(y_j, Z^{j-1}, \overline{\theta}_t)\right)^{-1}\right) = o_P^*(1), \ \Pr -P,$$

and also

$$\sup_{t\geq R}\left(\left(-\frac{1}{t}\sum_{j=s}^{t}\nabla_\theta^2 q(y_j^*, Z^{*,j-1}, \overline{\theta}_t^*)\right)^{-1} - B^\dagger\right) = o_P^*(1), \ \Pr -P, \tag{42}$$

so the RHS of (41) can be written as:

$$\begin{aligned} & \frac{1}{P^{1/2}}\sum_{t=R}^{T-1} B^\dagger\left(\frac{1}{t}\sum_{j=s}^{t}\nabla_\theta q(y_j^*, Z^{*,j-1}, \theta^\dagger) - \frac{1}{t}\sum_{j=s}^{t}\nabla_\theta q(y_j, Z^{j-1}, \theta^\dagger)\right) + o_P^*(1), \ \Pr -P \\ = \ & B^\dagger\frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(\frac{1}{t}\sum_{j=s}^{t} h_j^* - \frac{1}{t}\sum_{j=s}^{t} h_t\right) + o_P^*(1), \ \Pr -P, \end{aligned} \tag{43}$$

by letting $\nabla_\theta q(y_j^*, Z^{*,j-1}, \theta^\dagger) = h_t^*$, $\nabla_\theta q(y_j, Z^{j-1}, \theta^\dagger) = h_t$. Recalling that $a_{R,s} = (R+s)^{-1} + ... + (R+P-1)$, $0 \leq s \leq P-1$, and that for $t = 1, ...R$, $E^*\left(h_t^*\right) = R^{-1}\sum_{t=1}^{R} h_t + O(l/P) = \overline{h}_R + O(l/P)$, and for $j = 1, ..., P$, $E^*(a_{R,j}h_{R+j}^*) = a_{R,j}P^{-1}\sum_{j=1}^{P-1} h_{R+j} + O(l/P) = a_{R,j}\overline{h}_P + O(l/P)$, the RHS of (43) writes as,

$$B^\dagger a_{R,0}\frac{1}{\sqrt{P}}\sum_{t=1}^{R}(h_t^* - h_t) + B^\dagger\frac{1}{\sqrt{P}}\sum_{i=1}^{P-1} a_{R,i}(h_{R+i}^* - \overline{h}_P)$$
$$-B^\dagger\frac{1}{\sqrt{P}}\sum_{i=1}^{P-1} a_{R,i}(h_{R+i} - \overline{h}_P) + o_P^*(1), \ \ \Pr -P. \qquad (44)$$

The sum of the first two terms in (44) satisfies a central limit theorem for mixing triangular arrays (Wooldridge and White (1988)) and, by Lemma A3, has asymptotic variance equal to $2\Pi C_{00}$, which is the same as the asymptotic variance of $P^{-1/2}\sum_{t=R}^{T-1}(\widehat{\theta}_t - \theta_t^\dagger)$ (see Lemma A5, in West (1996)), conditionally on the samples and for all samples but a subset of measure approaching zero. Therefore, it suffices to show that the the last term on the RHS of (4), i.e. the adjustment term, is equal to $B^\dagger\frac{1}{\sqrt{P}}\sum_{i=1}^{P-1} a_{R,i}(h_{R+i} - \overline{h}_P)$, up to a term vanishing asymptotically. Given A1 and A2, $\left(-\frac{1}{T}\sum_{t=s}^{T-1}\nabla_{\theta_i}^2 q_i(y_t, Z^{t-1}, \widehat{\theta}_T)\right)^{-1} - B^\dagger = o(1)$ $\Pr -P$ (i.e. $o_P(1)$), where $\widehat{\theta}_T$ is the estimator constructed using all T observations.
Now let $h_{R+i}(\widehat{\theta}_T) = \nabla_\theta q(y_{R+i}, Z^{R+i-1}, \widehat{\theta}_T)$, and $\overline{h}_P(\widehat{\theta}_T) = P^{-1}\sum_{i=1}^{P}\nabla_\theta q(y_{R+i}, Z^{R+i-1}, \widehat{\theta}_T)$, and let $\nabla^2 h_{R+i}(\widehat{\theta}_T) = \nabla_\theta^2 q(y_{R+i}, Z^{R+i-1}, \widehat{\theta}_T)$. Now,

$$B^\dagger\frac{1}{\sqrt{P}}\sum_{i=1}^{P-1} a_{R,i}\left(\left(h_{R+i}(\widehat{\theta}_T) - \overline{h}_P(\widehat{\theta}_T)\right) - \left(h_{R+i} - \overline{h}_P\right)\right)$$
$$= \ B^\dagger\frac{1}{P}\sum_{i=1}^{P-1} a_{R,i}\left(\nabla^2 h_{R+i}(\overline{\theta}_T) - \overline{\nabla^2 h}_P(\overline{\theta}_T)\right)\sqrt{P}\left(\widehat{\theta}_T - \theta^\dagger\right) = o(1), \ \ \Pr -P, \qquad (45)$$

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, $\frac{1}{P}\sum_{i=1}^{P-1} a_{R,i}\left(\nabla^2 h_{R+i}(\overline{\theta}_T) - \overline{\nabla^2 h}_P(\overline{\theta}_T)\right) = o(1)$ $\Pr -P$. The desired result then follows.

Proof of Proposition 2: From Theorem 1 in Corradi and Swanson (2002).

Proof of Proposition 3: Recall that $g = q_1$, also let $\overline{u}_{1,t+1}^* = y_{t+1}^* - \begin{pmatrix} 1 & y_t^* \end{pmatrix}\overline{\theta}_{1,t}^*$, where $\overline{\theta}_{1,t}^* \in \left(\widehat{\theta}_{1,t}^*, \theta_1^\dagger\right)$, and $\widehat{\theta}_{1,t}^*$ is defined in (17). Then,

$$\frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(g'(\widehat{u}_{1,t+1}^*)w(Z^{*,t},\gamma) - g'(\widehat{u}_{1,t+1})w(Z^t,\gamma)\right) = \frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(g'(u_{1,t+1}^*)w(Z^{*,t},\gamma) - g'(u_{1,t+1})g(Z^t,\gamma)\right)$$

$$+\frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(\nabla_{\theta_1}g'(\overline{u}^*_{1,t+1})\right)'w(Z^{*,t},\gamma)\left(\widehat{\theta}^*_{1,t}-\theta_1^\dagger\right)-\frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(\nabla_{\theta_1}g'(\overline{u}_{1,t+1})\right)'w(Z^t,\gamma)\left(\widehat{\theta}_{1,t}-\theta_1^\dagger\right)$$

$$= \frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(g'(u^*_{1,t+1})w(Z^{*,t},\gamma)-g'(u_{1,t+1})w(Z^t,\gamma)\right)+\frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(\nabla_{\theta_1}g'(\overline{u}^*_{1,t+1})\right)'w(Z^{*,t},\gamma)\left(\widehat{\theta}^*_{1,t}-\widehat{\theta}_{1,t}\right)$$

$$+\frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(\left(\nabla_{\theta_1}g'(\overline{u}^*_{1,t+1})\right)'w(Z^{*,t},\gamma)-\left(\nabla_{\theta_1}g'(\overline{u}_{1,t+1})\right)'w(Z^t,\gamma)\right)\left(\widehat{\theta}_{1,t}-\theta_1^\dagger\right), \tag{46}$$

the last term in (46) is $o^*_P(1)$ $\Pr -P$, as given A3, $\sup_{t\geq R}P^{1/2}\left(\widehat{\theta}_{1,t}-\theta_1^\dagger\right)=O_P(1)$, and

$$\sup_{\theta_1\in\Theta_1}\frac{1}{P}\sum_{t=R}^{T-1}\left(\left(\nabla_{\theta_1}g'(u^*_{1,t+1}(\theta_1))\right)'w(Z^{*,t},\gamma)-\left(\nabla_{\theta_1}g'(u_{1,t+1}(\theta_1))\right)'w(Z^t,\gamma)\right)=o_{P^*}(1),\ \ \Pr -P,$$

where $u^*_{1,t+1}(\theta_1)=y^*_{t+1}-\begin{pmatrix}1 & y^*_t\end{pmatrix}\theta_1$, and $u_{1,t+1}(\theta_1)=y_{t+1}-\begin{pmatrix}1 & y_t\end{pmatrix}\theta_1$. Now, pointwise in γ, the first term on the RHS of the last equality in (46), has the same limiting distribution as $\frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(g'(u_{1,t+1})w(Z^t,\gamma)-E\left(g'(u_{1,t+1})w(Z^t,\gamma)\right)\right)$. Stochastic equicontinuity on Γ can be shown along the lines of Theorem 2 in Corradi and Swanson (2002). Therefore, under H_0, any continuous functional over Γ of $\frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(g'(u^*_{1,t+1})w(Z^{*,t},\gamma)-g'(u_{1,t+1})w(Z^t,\gamma)\right)$ has the same limiting distribution of the same functional of $\frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(g'(u_{1,t+1})w(Z^t,\gamma)-E\left(g'(u_{1,t+1})w(Z^t,\gamma)\right)\right)$, conditional on the sample and for all samples but a subset of measure approaching zero. The last term on the RHS of the last equality in (46) is $o^*_P(1)$ $\Pr -P$. It now remains to consider the second term on the RHS of the last equality in (46). Hereafter, let $\overline{\nabla_{\theta_1}g_P}=P^{-1}\sum_{i=1}^{P-1}\nabla_{\theta_1}g(u_{1,j})$, and $\mu_\gamma=E\left(\nabla_{\theta_1}g'(u_{1,t+1})w(Z^t,\gamma)\right)$. Recalling that $q_1=g$, by the same argument used above, it writes as,

$$\mu'_\gamma\frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(\left(-\frac{1}{t}\sum_{j=s}^{t}\nabla^2_{\theta_1}g(\overline{u}^*_{1,j})\right)^{-1}\left(\frac{1}{t}\sum_{j=s}^{t}\nabla_{\theta_1}g(u^*_{1,j})\right)\right.$$
$$\left.-\left(-\frac{1}{t}\sum_{j=s}^{t}\nabla^2_{\theta_1}g(\overline{u}_{1,j})\right)^{-1}\left(\frac{1}{t}\sum_{j=s}^{t}\nabla_{\theta_1}g(u_{1,j})\right)\right)+o^*_P(1),\ \ \Pr -P$$

$$= \mu'_\gamma B^\dagger\frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(\frac{1}{t}\sum_{j=s}^{t}\left(\nabla_{\theta_1}g(u^*_{1,j})-\nabla_{\theta_1}g(u_{1,j})\right)\right)+o^*_P(1),\ \ \Pr -P$$

$$= \mu'_\gamma B^\dagger\frac{a_{R,0}}{P^{1/2}}\sum_{t=s}^{R}\left(\nabla_{\theta_1}g(u^*_{1,j})-\nabla_{\theta_1}g(u_{1,j})\right)+\mu'_\gamma B^\dagger\frac{1}{P^{1/2}}\sum_{i=1}^{P-1}a_{R,i}\left(\nabla_{\theta_1}g(u^*_{1,R+i})-\overline{\nabla_{\theta_1}g_P}\right)$$
$$-\mu'_\gamma B^\dagger\frac{1}{P^{1/2}}\sum_{i=1}^{P-1}a_{R,i}\left(\nabla_{\theta_1}g(u_{1,R+i})-\overline{\nabla_{\theta_1}g_P}\right)+o^*_P(1),\ \ \Pr -P.$$

Now, the sum of the first two terms on the RHS of the last equality, mimic the contribution of parameter estimation error, while the last term is offset by the adjustment term. The desired outcome then follows.

Proof of Proposition 4: This proof requires just a small 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 recursively. Also, recall that parameters are estimated by QMLE (using the density associated with the candidate conditional model), so that $q_i = -\ln f_i$, $i = 1, ..., n$ with f_i being the density associated with F_i. Let $\mu_i^2(u) = E\left(\left(1\{y_{t+1} \leq u\} - F_i(u|Z^t, \theta_i^\dagger)\right)^2\right)$ $= E\left(\left(1\{y_{t+1} \leq u\} - F_0(u|Z^t, \theta_0)\right)^2\right) + E\left(\left(F_0(u|Z^t, \theta_0) - F_i(u|Z^t, \theta_i^\dagger)\right)^2\right)$. Thus,

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

where $\overline{\theta}_{i,t} \in (\widehat{\theta}_{i,t}, \theta_i^\dagger)$, $i = 1, ..., n$, and where the $o_P(1)$ term holds uniformly in $u \in U$. Now, given A1,A2 and A6, by the uniform law of large numbers for $\beta-$mixing processes,

$$
\sup_{\theta_i \in \Theta_i} \left| \frac{1}{P} \sum_{t=R}^{T-1} \nabla_{\theta_i} F_i(u|Z^t, \theta_i)' \left(1\{y_{t+1} \leq u\} - F_i(u|Z^t, \theta_i)\right) - \mu_{\theta_i} \right| = o_P(1).
$$

Now,

$$\begin{aligned}
\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_{i,t}-\theta_i^\dagger\right) &= \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(-\frac{1}{t}\sum_{j=s}^{t}\nabla^2_{\theta_i}\ln f_i(y_j|Z^{j-1},\theta_i^\dagger)\right)^{-1}\left(-\frac{1}{t}\sum_{j=s}^{t}\nabla_{\theta_i}\ln f_i(y_j|Z^{j-1},\theta_i^\dagger)\right) \\
&= B_i^\dagger\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(-\frac{1}{t}\sum_{j=s}^{t}\nabla_{\theta_i}\ln f_i(y_j|Z^{j-1},\theta_i^\dagger)\right)+o_P(1),
\end{aligned}$$

where $B_i^\dagger=\left(E\left(-\nabla^2_{\theta_i}\ln f_i(y_{t+1}|Z^t,\theta_i^\dagger)\right)\right)^{-1}$ and $\mu_{\theta_i}=E\left(\nabla_{\theta_i}F_i(u|Z^t,\theta_i)'\left(1\{y_{t+1}\leq u\}-F_i(u|Z^t,\theta_i)\right)\right)$. For any given u,u' convergence in distribution to a normal with the same covariance as in the statement of the theorem follows by the same argument as in Proposition 6 and by the Cramer Wold device. Finally stochastic equicontinuity in u follows by the same argument as in the proof of Theorem 1 in Corradi and Swanson (2003b). Convergence of finite dimension distribution and stochastic equicontinuity ensure weak convergence to a process on U.

Proof of Proposition 5: As for the first two lines on the RHS of the bootstrap statistic, equation (23):

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

$$\begin{aligned}
=\ &\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\left(\left(\left(1\{y^*_{t+1}\leq u\}-F_1(u|Z^{*,t},\theta_1^\dagger)\right)-\nabla_{\theta_1}F_1(u|Z^{*,t},\overline{\theta}^*_{1,t})\left(\widehat{\theta}^*_{1,t}-\theta_1^\dagger\right)\right)^2\right.\right. \\
&-\left.\left(\left(1\{y_{t+1}\leq u\}-F_1(u|Z^t,\theta_1^\dagger)\right)-\nabla_{\theta_1}F_1(u|Z^t,\overline{\theta}_{1,t})\left(\widehat{\theta}_{1,t}-\theta_1^\dagger\right)\right)^2\right) \\
&-\left(\left(\left(1\{y^*_{t+1}\leq u\}-F_k(u|Z^{*,t},\theta_k^\dagger)\right)-\nabla_{\theta_k}F_k(u|Z^{*,t},\overline{\theta}^*_{k,t})\left(\widehat{\theta}^*_{k,t}-\theta_k^\dagger\right)\right)^2\right. \\
&-\left.\left.\left(\left(1\{y_{t+1}\leq u\}-F_k(u|Z^t,\theta_k^\dagger)\right)-\nabla_{\theta_k}F_k(u|Z^t,\overline{\theta}_{k,t})\left(\widehat{\theta}_{k,t}-\theta_k^\dagger\right)\right)^2\right)\right), \qquad (48)
\end{aligned}$$

where $\overline{\theta}^*_{i,t}\in\left(\widehat{\theta}^*_{i,t},\theta_i^\dagger\right)$, $\overline{\theta}_{i,t}\in\left(\widehat{\theta}_{i,t},\theta_i^\dagger\right)$. Now, by a similar argument as that used in the proof of

Theorem 2 in Corradi and Swanson (2003b), (48) writes as:

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

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

$$
\begin{aligned}
&+\frac{2}{P}\sum_{t=R}^{T-1}\left(\left(1\{y_{t+1}^*\leq u\}-F_k(u|Z^{*,t},\theta_k^\dagger)\right)\nabla_{\theta_k}F_k(u|Z^{*,t},\overline{\theta}_{k,t}^*)'\right)\sqrt{P}\left(\widehat{\theta}_{k,t}^*-\theta_k^\dagger\right)\\
&-\frac{2}{P}\sum_{t=R}^{T-1}\left(\left(1\{y_{t+1}\leq u\}-F_k(u|Z^t,\theta_k^\dagger)\right)\nabla_{\theta_k}F_k(u|Z^t,\overline{\theta}_{k,t})'\right)\sqrt{P}\left(\widehat{\theta}_{k,t}-\theta_k^\dagger\right)+o_{P^*}(1),\Pr-P, \qquad (49)
\end{aligned}
$$

As shown in the proof of Theorem 2 in Corradi and Swanson (2003b), $\frac{1}{\sqrt{T}}\sum_{t=s}^{T}\left(F_i^2(u|Z^{*t},\theta_i^\dagger)-F_i^2(u|Z^t,\theta_i^\dagger)\right)$ has the same limiting distribution as $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(F_i^2(u|Z^t,\theta_i^\dagger)-E\left(F_i^2(u|Z^t,\theta_1^\dagger)\right)\right)$, as a process over U, and $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(F_i(u|Z^{*,t},\theta_i^\dagger)1\{y_{t+1}^*\leq u\}-F_i(u|Z^t,\theta_i^\dagger)1\{y_{t+1}\leq u\}\right)$ has the same limiting distribution as $\frac{1}{\sqrt{P}}\sum_{t=s}^{T-1}\left(F_i(u|Z^t,\theta_i^\dagger)1\{y_{t+1}\leq u\}-E\left(F_i(u|Z^t,\theta_i^\dagger)1\{y_{t+1}\leq u\}\right)\right)$, as a process over U.

For sake of simplicity we just analyze the parameter estimation error component of model 1.

Now, by a similar argument as that used in the proof of Proposition 3,

$$
\begin{aligned}
& -\frac{2}{P}\sum_{t=R}^{T-1}\left(\left(1\{y_{t+1}^{*}\leq u\}-F_1(u|Z^{*,t},\theta_1^{\dagger})\right)\nabla_{\theta_1}F_1(u|Z^{*,t},\overline{\theta}_{1,t}^{*})'\right)\sqrt{P}\left(\widehat{\theta}_{1,t}^{*}-\theta_1^{\dagger}\right) \\
& +\frac{2}{P}\sum_{t=R}^{T-1}\left(\left(1\{y_{t+1}\leq u\}-F_1(u|Z^{t},\theta_1^{\dagger})\right)\nabla_{\theta_1}F_1(u|Z^{t},\overline{\theta}_{1,t})'\right)\sqrt{P}\left(\widehat{\theta}_{1,t}-\theta_1^{\dagger}\right) \\
= & -2\mu_{\theta_1^{\dagger}}A(\theta_1^{\dagger})\sqrt{P}\left(\widehat{\theta}_{1,t}^{*}-\widehat{\theta}_{1,t}\right)+o_p^{*}(1),\ \ \Pr-P \\
= & -2\mu_{\theta_1^{\dagger}}A(\theta_1^{\dagger})\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\frac{1}{t}\sum_{j=s}^{t}\left(\ln f_1(y_j^{*}|Z^{*,j-1},\theta_1^{\dagger})-\ln f_1(y_j|Z^{j-1},\theta_1^{\dagger})\right)\right)+o_p^{*}(1),\ \ \Pr-P, \qquad (50)
\end{aligned}
$$

where $m_{\theta_1^{\dagger}}(u)'=E\left(\nabla_{\theta_1}F_1(u|Z^{t},\theta_1^{\dagger})'\left(1\{y_{t+1}\leq u\}-F_1(u|Z^{t},\theta_1^{\dagger})\right)\right)$ and $A(\theta_1^{\dagger})=\left(E\left(-\nabla_{\theta_1}^{2}f_1(y_{t+1}|Z^{t},\theta_1^{\dagger})\right)\right.$

Let $h_{1,t+1}^{*}=\nabla_{\theta_1}\ln f_1(y_{t+1}^{*}|Z^{*,t},\theta_1^{\dagger})$, $h_{1,t+1}=\nabla_{\theta_1}\ln f_1(y_{t+1}|Z^{t},\theta_1^{\dagger})$. Also, the last line in (50) writes as:

$$
\begin{aligned}
& -2\mu_{\theta_1^{\dagger}}A(\theta_1^{\dagger})\left(a_{R,0}^{2}\frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(h_{1,t}^{*}-h_{1,t}\right)+\frac{1}{P^{1/2}}\sum_{i=1}^{P-1}a_{R,i}\left(h_{1,R+i}^{*}-\overline{h}_{1,P}\right)\right) \\
& +2\mu_{\theta_1^{\dagger}}A(\theta_1^{\dagger})\frac{1}{P^{1/2}}\sum_{i=1}^{P-1}a_{R,i}\left(h_{1,R+i}-\overline{h}_{1,P}\right)+o_P^{*}(1),\ \ \Pr-P.
\end{aligned}
$$

The desired outcome then follows by the same argument as in the proof of Theorem 1.

8 References

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

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, Working Paper, Cowles Foundation in Economics, Yale University.

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

Bierens, H.B., (1982): Consistent model specification tests, *Journal of Econometrics*, 20, 105-134.

Bierens, H.B., (1990): A Conditional Moment Test of Functional Form, *Econometrica* , 58, 1443-1458.

Bierens, H.J. and W. Ploberger, (1997): Asymptotic theory of integrated conditional moment tests, *Econometrica*, 65, 1129-1152.

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

Christoffersen, P. and F.X. Diebold, (1996), Further Results on Forecasting and Model Selection under Asymmetric Loss, *Journal of Applied Econometrics, 11, 561-572.*

Christoffersen, P. and F.X. Diebold, (1997), Optimal Prediction Under Asymmetric Loss, *Econometric Theory, 13, 808-817.*

Christoffersen, P. and F.X. Diebold, (2000), How Relevant is Volatility Forecasting For Financial Risk Management?, *Review of Economics and Statistics,* 82, 12-22.

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

Clements, M.P. and J. Smith, (2000), Evaluating the Forecast Densities of Linear and Nonlinear Models: Applications to Output Growth and Unemployment, *Journal of Forecasting,* 19, 255-276.

Clements, M.P. and J. Smith, (2002), Evaluating Multivariate Forecast Densities: A Comparison of Two Approaches, *International Journal of Forecasting,* 18, 397-407.

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

Corradi, V., N.R. Swanson and C. Olivetti, (2001), Predictive Ability with Cointegrated Variables, *Journal of Econometrics,* 104, 315-358.

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), Bootstrap Conditional Distribution Tests in the Presence of Dynamic Misspecification, *Journal of Econometrics,* forthcoming.

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., N.R. Swanson, (2004), Predictive Density Evaluation in the Presence of Generic Misspecification, Working Paper, Rutgers University and Queen Mary-University of London.

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.

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

Elliott, G. and A. Timmerman, (2002a), Optimal Forecast Combinations Under General Loss Functions and Forecast Error Distributions, Working Paper, University of California, San Diego.

Elliott, G. and A. Timmerman, (2002b), Optimal Forecast Combination Under Regime Switching, Working Paper, University of California, San Diego.

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, manuscript, University of California, San Diego.

Goncalves, S., and H. White, (2002a), The Bootstrap of the Mean for Dependent and Heterogeneous Arrays, *Econometric Theory*, 18, 1367-1384.

Goncalves, S., and H. White, (2002b), Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models, *Journal of Econometrics,* forthcoming.

Granger, C.W.J., (1993), On the Limitations of Comparing Mean Squared Forecast Errors: Comment, *Journal of Forecasting,* 12, 651-652.

Granger, C.W.J., (1999), Outline of Forecast Theory Using Generalized Cost Functions, *Spanish Economic Review,* 1, 161-173.

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

Hansen, B.E., (1996), Inference When a Nuisance Parameter is Not Identified Under the Null Hypothesis, *Econometrica,* 64, 413-430.

Hansen, L.P., J. Heaton and E.G.J, Luttmer (1995), Econometric Evaluation of Asset Pricing Models, *Review of Financial Studies,* 8, 237-279.

Hansen, L.P. and R. Jeganathan (1997), Assessing Specification Errors in Stochastic Discount Factor Models, *Journal of Finance*, 52, 557-590.

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

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

Inoue, A., (2001), Testing for Distributional Change in Time Series, *Econometric Theory*, 17, 156-187.

Inoue, A., and M. Shintani, (2001), Bootstrapping GMM Estimators for Time Series, Mimeo, North Carolina State University and Vanderbilt University.

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

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.

McCracken, M.W., (1999), Asymptotics for Out of Sample Tests of Causality, Working Paper, University of Missouri-Columbia.

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

Politis, D.N. and J.P. Romano, (1994), The Stationary Bootstrap, *Journal of the American Statistical Association,* 89, 1303-1313.

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

Stinchcombe, M.B. and H. White, (1998), Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative, *Econometric Theory,* 14, 3, 295-325.

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

Weiss, A., (1996) Estimating Time Series Models Using the Relevant Cost Function, *Journal of Applied Econometrics,* 11, 539-560.

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

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.

Zellner, A., (1986), Bayesian Estimation and Prediction Using Asymmetric Loss Function, *Journal of the American Statistical Association,* 81, 446-451.

Table 1: Finite Sample Properties of the Bootstrap for Parameter Estimation Error: Recursive Scheme with and without Adjustment Terms: Part I(*)

Panel A: DGP is an AR(1) Process - $\rho = 0.2$												
smpl	*boot*	*coeff*	$l = 4$	$l = 6$	$l = 10$	$l = 12$	$l = 15$	$l = 20$	$l = 25$	$l = 30$	$l = 50$	$l = 60$
600	*rec*	$\hat{\rho}$	0.732	0.800	0.896	0.870	0.912	0.904	0.894	0.870	0.864	0.844
		$\hat{\rho}_1$	0.730	0.812	0.904	0.890	0.910	0.898	0.900	0.884	0.878	0.850
		$\hat{\rho}_2$	0.898	0.898	0.908	0.916	0.938	0.910	0.896	0.908	0.862	0.856
	no adj	$\hat{\rho}$	0.710	0.766	0.876	0.842	0.882	0.868	0.874	0.844	0.822	0.816
		$\hat{\rho}_1$	0.710	0.784	0.894	0.860	0.882	0.882	0.872	0.852	0.832	0.818
		$\hat{\rho}_2$	0.870	0.882	0.880	0.888	0.902	0.894	0.868	0.888	0.838	0.814
			$l = 4$	$l = 10$	$l = 15$	$l = 20$	$l = 25$	$l = 30$	$l = 40$	$l = 50$	$l = 60$	$l = 100$
1200	*rec*	$\hat{\rho}$	0.612	0.868	0.902	0.900	0.924	0.926	0.908	0.880	0.904	0.882
		$\hat{\rho}_1$	0.624	0.874	0.900	0.904	0.934	0.918	0.900	0.896	0.914	0.890
		$\hat{\rho}_2$	0.874	0.924	0.924	0.918	0.918	0.888	0.940	0.912	0.884	0.866
	no adj	$\hat{\rho}$	0.602	0.846	0.874	0.862	0.912	0.892	0.888	0.854	0.876	0.842
		$\hat{\rho}_1$	0.636	0.860	0.880	0.862	0.918	0.896	0.886	0.880	0.876	0.854
		$\hat{\rho}_2$	0.848	0.898	0.894	0.886	0.902	0.878	0.912	0.892	0.870	0.842
			$l = 4$	$l = 10$	$l = 20$	$l = 30$	$l = 40$	$l = 50$	$l = 60$	$l = 80$	$l = 100$	$l = 120$
2400	*rec*	$\hat{\rho}$	0.436	0.780	0.898	0.934	0.928	0.920	0.914	0.910	0.934	0.888
		$\hat{\rho}_1$	0.452	0.796	0.908	0.934	0.924	0.920	0.918	0.920	0.928	0.888
		$\hat{\rho}_2$	0.890	0.904	0.930	0.932	0.914	0.924	0.930	0.938	0.892	0.894
	no adj	$\hat{\rho}$	0.446	0.770	0.874	0.916	0.896	0.886	0.882	0.880	0.892	0.868
		$\hat{\rho}_1$	0.456	0.768	0.884	0.916	0.906	0.890	0.890	0.890	0.898	0.854
		$\hat{\rho}_2$	0.856	0.878	0.898	0.912	0.898	0.902	0.914	0.898	0.876	0.866
Panel B: DGP is an AR(2) Process - $\rho_1 = \rho_2 = 0.1$												
			$l = 4$	$l = 6$	$l = 10$	$l = 12$	$l = 15$	$l = 20$	$l = 25$	$l = 30$	$l = 50$	$l = 60$
600	*rec*	$\hat{\rho}$	0.828	0.850	0.898	0.892	0.910	0.894	0.894	0.864	0.868	0.834
		$\hat{\rho}_1$	0.862	0.872	0.914	0.902	0.924	0.908	0.912	0.878	0.880	0.834
		$\hat{\rho}_2$	0.832	0.886	0.900	0.920	0.912	0.896	0.904	0.890	0.886	0.840
	no adj	$\hat{\rho}$	0.350	0.580	0.825	0.795	0.795	0.860	0.845	0.830	0.850	0.800
		$\hat{\rho}_1$	0.435	0.610	0.830	0.815	0.835	0.885	0.870	0.835	0.880	0.815
		$\hat{\rho}_2$	0.920	0.910	0.850	0.930	0.895	0.895	0.855	0.880	0.835	0.860
			$l = 4$	$l = 10$	$l = 15$	$l = 20$	$l = 25$	$l = 30$	$l = 40$	$l = 50$	$l = 60$	$l = 100$
1200	*rec*	$\hat{\rho}$	0.794	0.892	0.916	0.918	0.914	0.906	0.914	0.898	0.886	0.872
		$\hat{\rho}_1$	0.816	0.906	0.928	0.926	0.912	0.912	0.916	0.900	0.896	0.882
		$\hat{\rho}_2$	0.786	0.912	0.920	0.940	0.932	0.902	0.930	0.892	0.912	0.856
	no adj	$\hat{\rho}$	0.135	0.670	0.820	0.865	0.865	0.840	0.880	0.870	0.860	0.815
		$\hat{\rho}_1$	0.200	0.720	0.845	0.900	0.875	0.880	0.895	0.885	0.885	0.840
		$\hat{\rho}_2$	0.875	0.880	0.925	0.905	0.895	0.920	0.890	0.900	0.920	0.860
			$l = 4$	$l = 10$	$l = 20$	$l = 30$	$l = 40$	$l = 50$	$l = 60$	$l = 80$	$l = 100$	$l = 120$
2400	*rec*	$\hat{\rho}$	0.702	0.882	0.894	0.938	0.914	0.936	0.910	0.920	0.922	0.896
		$\hat{\rho}_1$	0.734	0.892	0.898	0.936	0.912	0.942	0.912	0.920	0.926	0.902
		$\hat{\rho}_2$	0.738	0.890	0.916	0.940	0.912	0.910	0.938	0.906	0.904	0.916
	no adj	$\hat{\rho}$	0.015	0.530	0.790	0.830	0.870	0.885	0.865	0.920	0.875	0.860
		$\hat{\rho}_1$	0.040	0.615	0.810	0.830	0.910	0.875	0.875	0.890	0.875	0.860
		$\hat{\rho}_2$	0.885	0.895	0.965	0.875	0.875	0.880	0.900	0.885	0.900	0.875

(*) Notes: 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 $\hat{\rho}_2$). Entries corresponding to *rec* correspond to coverage probabilities based on the recursive PEE bootstrap with adjustment terms, in the context of a recursive estimation scheme, so that in our framework they denote the proportion of times that the (pseudo) true parameter ρ falls into CI^{**}, with α equal to 0.05. Analogous values for the same bootstrap procedure, but without adjustment terms are given in rows denoted by *no adj*. Results based on different values of α are available upon request. In the context of the recursive scheme, P recursive estimators are constructed, where P=0.5T and T is the *full* sample, set at either 600, 1200, or 2400 observations. Data are generated according to either 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. Block lengths used are denoted by the various values of l. In all experiments, 500 Monte Carlo iterations were carried out (see above for further details).

Table 2: Finite Sample Properties of the Bootstrap for Parameter Estimation Error: Recursive Scheme with and without Adjustment Terms: Part II(*)

Panel A: DGP is an AR(1) Process - $\rho = 0.4$												
smpl	*boot*	*coeff*	$l=4$	$l=6$	$l=10$	$l=12$	$l=15$	$l=20$	$l=25$	$l=30$	$l=50$	$l=60$
600	*rec*	$\hat{\rho}$	0.340	0.590	0.830	0.815	0.815	0.880	0.855	0.840	0.870	0.840
		$\hat{\rho}_1$	0.420	0.635	0.850	0.825	0.860	0.905	0.890	0.870	0.905	0.870
		$\hat{\rho}_2$	0.930	0.920	0.880	0.945	0.925	0.910	0.905	0.885	0.865	0.885
	no adj	$\hat{\rho}$	0.804	0.794	0.876	0.856	0.890	0.854	0.890	0.846	0.834	0.802
		$\hat{\rho}_1$	0.838	0.838	0.886	0.868	0.894	0.876	0.890	0.864	0.842	0.806
		$\hat{\rho}_2$	0.802	0.864	0.866	0.896	0.878	0.874	0.878	0.868	0.852	0.814
			$l=4$	$l=10$	$l=15$	$l=20$	$l=25$	$l=30$	$l=40$	$l=50$	$l=60$	$l=100$
1200	*rec*	$\hat{\rho}$	0.120	0.660	0.835	0.900	0.890	0.890	0.890	0.895	0.880	0.845
		$\hat{\rho}_1$	0.195	0.720	0.845	0.910	0.895	0.890	0.910	0.910	0.900	0.860
		$\hat{\rho}_2$	0.910	0.915	0.930	0.940	0.935	0.945	0.930	0.905	0.945	0.900
	no adj	$\hat{\rho}$	0.770	0.864	0.884	0.880	0.876	0.894	0.894	0.872	0.860	0.854
		$\hat{\rho}_1$	0.802	0.882	0.894	0.900	0.888	0.898	0.894	0.868	0.866	0.866
		$\hat{\rho}_2$	0.780	0.884	0.890	0.906	0.884	0.882	0.906	0.874	0.894	0.828
			$l=4$	$l=10$	$l=20$	$l=30$	$l=40$	$l=50$	$l=60$	$l=80$	$l=100$	$l=120$
2400	*rec*	$\hat{\rho}$	0.015	0.495	0.810	0.840	0.900	0.910	0.890	0.910	0.895	0.895
		$\hat{\rho}_1$	0.030	0.605	0.825	0.860	0.925	0.915	0.880	0.900	0.925	0.900
		$\hat{\rho}_2$	0.905	0.920	0.960	0.925	0.920	0.910	0.920	0.895	0.925	0.875
	no adj	$\hat{\rho}$	0.706	0.866	0.876	0.904	0.888	0.910	0.884	0.892	0.898	0.862
		$\hat{\rho}_1$	0.726	0.870	0.872	0.898	0.890	0.910	0.892	0.878	0.904	0.868
		$\hat{\rho}_2$	0.732	0.862	0.890	0.908	0.892	0.898	0.912	0.880	0.872	0.886
Panel B: DGP is an AR(2) Process $\rho_1 = \rho_2 = 0.2$												
600	*rec*	$\hat{\rho}$	0.710	0.825	0.895	0.885	0.915	0.870	0.885	0.845	0.885	0.845
		$\hat{\rho}_1$	0.790	0.880	0.925	0.885	0.905	0.875	0.895	0.885	0.900	0.865
		$\hat{\rho}_2$	0.745	0.815	0.880	0.845	0.885	0.895	0.890	0.880	0.870	0.845
	no adj	$\hat{\rho}$	0.680	0.780	0.855	0.845	0.860	0.865	0.845	0.840	0.865	0.800
		$\hat{\rho}_1$	0.765	0.850	0.890	0.890	0.875	0.880	0.880	0.845	0.875	0.825
		$\hat{\rho}_2$	0.710	0.825	0.845	0.840	0.850	0.860	0.875	0.835	0.855	0.805
			$l=4$	$l=10$	$l=15$	$l=20$	$l=25$	$l=30$	$l=40$	$l=50$	$l=60$	$l=100$
1200	*rec*	$\hat{\rho}$	0.530	0.790	0.880	0.905	0.900	0.880	0.935	0.895	0.940	0.900
		$\hat{\rho}_1$	0.605	0.855	0.910	0.940	0.930	0.885	0.945	0.895	0.960	0.895
		$\hat{\rho}_2$	0.660	0.860	0.880	0.910	0.910	0.915	0.905	0.875	0.925	0.870
	no adj	$\hat{\rho}$	0.515	0.765	0.850	0.905	0.900	0.840	0.900	0.880	0.900	0.880
		$\hat{\rho}_1$	0.605	0.820	0.900	0.910	0.925	0.870	0.915	0.900	0.935	0.850
		$\hat{\rho}_2$	0.640	0.855	0.850	0.885	0.885	0.885	0.885	0.865	0.915	0.840
			$l=4$	$l=10$	$l=20$	$l=30$	$l=40$	$l=50$	$l=60$	$l=80$	$l=100$	$l=120$
2400	*rec*	$\hat{\rho}$	0.300	0.775	0.875	0.915	0.920	0.925	0.920	0.925	0.890	0.890
		$\hat{\rho}_1$	0.395	0.790	0.905	0.945	0.945	0.940	0.915	0.930	0.890	0.880
		$\hat{\rho}_2$	0.430	0.810	0.925	0.860	0.925	0.920	0.905	0.895	0.930	0.915
	no adj	$\hat{\rho}$	0.315	0.755	0.870	0.920	0.895	0.900	0.885	0.910	0.875	0.850
		$\hat{\rho}_1$	0.400	0.780	0.895	0.905	0.885	0.925	0.875	0.925	0.880	0.840
		$\hat{\rho}_2$	0.445	0.775	0.915	0.860	0.905	0.880	0.895	0.850	0.865	0.880

(*) Notes: See notes to Table 1.

Table 3: Finite Sample Properties of the Bootstrap for Parameter Estimation Error: Recursive Scheme with and without Adjustment Terms: Part III(*)

Panel A: DGP is an AR(1) Process - $\rho = 0.6$												
smpl	*boot*	*coeff*	$l=4$	$l=6$	$l=10$	$l=12$	$l=15$	$l=20$	$l=25$	$l=30$	$l=50$	$l=60$
600	*rec*	$\hat{\rho}$	0.945	0.730	0.385	0.360	0.290	0.150	0.180	0.145	0.140	0.150
		$\hat{\rho}_1$	0.805	0.510	0.240	0.250	0.175	0.095	0.110	0.120	0.120	0.115
		$\hat{\rho}_2$	0.130	0.070	0.110	0.065	0.115	0.085	0.075	0.110	0.110	0.130
	no adj	$\hat{\rho}$	0.055	0.310	0.605	0.645	0.690	0.800	0.805	0.825	0.830	0.830
		$\hat{\rho}_1$	0.190	0.505	0.720	0.730	0.800	0.875	0.860	0.840	0.850	0.830
		$\hat{\rho}_2$	0.845	0.905	0.880	0.910	0.840	0.905	0.915	0.850	0.840	0.845
			$l=4$	$l=10$	$l=15$	$l=20$	$l=25$	$l=30$	$l=40$	$l=50$	$l=60$	$l=100$
1200	*rec*	$\hat{\rho}$	1.000	0.630	0.350	0.290	0.195	0.235	0.150	0.120	0.130	0.095
		$\hat{\rho}_1$	0.975	0.515	0.235	0.170	0.140	0.100	0.090	0.105	0.105	0.145
		$\hat{\rho}_2$	0.085	0.055	0.075	0.050	0.075	0.085	0.065	0.075	0.085	0.125
	no adj	$\hat{\rho}$	0.000	0.400	0.630	0.675	0.795	0.765	0.815	0.825	0.860	0.870
		$\hat{\rho}_1$	0.045	0.505	0.750	0.795	0.855	0.840	0.890	0.870	0.865	0.840
		$\hat{\rho}_2$	0.895	0.930	0.900	0.910	0.920	0.890	0.925	0.880	0.895	0.845
			$l=4$	$l=10$	$l=20$	$l=30$	$l=40$	$l=50$	$l=60$	$l=80$	$l=100$	$l=120$
2400	*rec*	$\hat{\rho}$	1.000	0.880	0.440	0.245	0.185	0.115	0.145	0.105	0.090	0.095
		$\hat{\rho}_1$	1.000	0.735	0.300	0.210	0.130	0.110	0.110	0.080	0.100	0.100
		$\hat{\rho}_2$	0.070	0.035	0.060	0.105	0.095	0.070	0.090	0.075	0.105	0.105
	no adj	$\hat{\rho}$	0.000	0.155	0.540	0.750	0.810	0.870	0.815	0.855	0.875	0.865
		$\hat{\rho}_1$	0.000	0.290	0.670	0.770	0.875	0.855	0.865	0.865	0.890	0.860
		$\hat{\rho}_2$	0.925	0.950	0.935	0.880	0.885	0.920	0.890	0.890	0.865	0.875
Panel B: DGP is an AR(2) Process - $\rho_1 = \rho_2 = 0.3$												
			$l=4$	$l=6$	$l=10$	$l=12$	$l=15$	$l=20$	$l=25$	$l=30$	$l=50$	$l=60$
600	*rec*	$\hat{\rho}$	0.580	0.350	0.225	0.165	0.170	0.140	0.155	0.160	0.130	0.165
		$\hat{\rho}_1$	0.460	0.205	0.095	0.105	0.070	0.095	0.120	0.125	0.125	0.125
		$\hat{\rho}_2$	0.310	0.220	0.150	0.105	0.105	0.100	0.080	0.115	0.095	0.145
	no adj	$\hat{\rho}$	0.000	0.035	0.280	0.465	0.540	0.640	0.660	0.765	0.760	0.800
		$\hat{\rho}_1$	0.100	0.365	0.675	0.755	0.815	0.890	0.850	0.840	0.850	0.780
		$\hat{\rho}_2$	0.925	0.975	0.920	0.920	0.910	0.920	0.885	0.885	0.905	0.800
			$l=4$	$l=10$	$l=15$	$l=20$	$l=25$	$l=30$	$l=40$	$l=50$	$l=60$	$l=100$
1200	*rec*	$\hat{\rho}$	0.785	0.345	0.110	0.145	0.145	0.150	0.115	0.130	0.130	0.120
		$\hat{\rho}_1$	0.645	0.235	0.075	0.095	0.110	0.090	0.100	0.090	0.135	0.135
		$\hat{\rho}_2$	0.605	0.195	0.090	0.100	0.100	0.130	0.130	0.090	0.095	0.125
	no adj	$\hat{\rho}$	0.000	0.045	0.315	0.455	0.575	0.680	0.725	0.790	0.820	0.830
		$\hat{\rho}_1$	0.000	0.405	0.620	0.755	0.775	0.860	0.860	0.870	0.870	0.870
		$\hat{\rho}_2$	0.975	0.915	0.895	0.905	0.945	0.895	0.895	0.870	0.875	0.835
			$l=4$	$l=10$	$l=20$	$l=30$	$l=40$	$l=50$	$l=60$	$l=80$	$l=100$	$l=120$
2400	*rec*	$\hat{\rho}$	0.950	0.370	0.225	0.175	0.095	0.130	0.130	0.075	0.080	0.145
		$\hat{\rho}_1$	0.875	0.280	0.155	0.115	0.065	0.085	0.090	0.065	0.055	0.095
		$\hat{\rho}_2$	0.880	0.310	0.110	0.065	0.085	0.125	0.020	0.070	0.095	0.115
	no adj	$\hat{\rho}$	0.000	0.000	0.235	0.545	0.580	0.700	0.790	0.815	0.815	0.815
		$\hat{\rho}_1$	0.000	0.190	0.610	0.775	0.805	0.865	0.905	0.855	0.890	0.900
		$\hat{\rho}_2$	0.950	0.935	0.925	0.940	0.935	0.935	0.935	0.870	0.915	0.915

(*) Notes: See notes to Table 1.

Table 4: Finite Sample Properties of the Bootstrap for Parameter Estimation Error: Recursive Scheme with and without Adjustment Terms: Part IV$^{(*)}$

Panel A: DGP is an AR(1) Process - $\rho = 0.8$												
smpl	*boot*	*coeff*	$l=4$	$l=6$	$l=10$	$l=12$	$l=15$	$l=20$	$l=25$	$l=30$	$l=50$	$l=60$
600	*rec*	$\hat{\rho}$	0.000	0.025	0.285	0.420	0.545	0.675	0.695	0.775	0.780	0.830
		$\hat{\rho}_1$	0.095	0.345	0.710	0.765	0.800	0.895	0.860	0.850	0.870	0.855
		$\hat{\rho}_2$	0.940	0.975	0.960	0.945	0.925	0.960	0.925	0.905	0.920	0.850
	no adj	$\hat{\rho}$	0.405	0.635	0.775	0.800	0.790	0.850	0.825	0.815	0.850	0.825
		$\hat{\rho}_1$	0.535	0.760	0.885	0.855	0.885	0.895	0.855	0.870	0.840	0.840
		$\hat{\rho}_2$	0.665	0.755	0.855	0.880	0.870	0.885	0.880	0.865	0.850	0.825
			$l=4$	$l=10$	$l=15$	$l=20$	$l=25$	$l=30$	$l=40$	$l=50$	$l=60$	$l=100$
1200	*rec*	$\hat{\rho}$	0.000	0.040	0.305	0.460	0.595	0.695	0.730	0.785	0.815	0.860
		$\hat{\rho}_1$	0.000	0.405	0.635	0.770	0.770	0.875	0.885	0.905	0.885	0.885
		$\hat{\rho}_2$	0.960	0.910	0.920	0.935	0.955	0.950	0.930	0.905	0.880	0.875
	no adj	$\hat{\rho}$	0.195	0.630	0.855	0.830	0.825	0.865	0.850	0.835	0.850	0.845
		$\hat{\rho}_1$	0.345	0.750	0.885	0.895	0.855	0.880	0.865	0.905	0.835	0.835
		$\hat{\rho}_2$	0.395	0.800	0.875	0.870	0.880	0.865	0.850	0.865	0.895	0.855
			$l=4$	$l=10$	$l=20$	$l=30$	$l=40$	$l=50$	$l=60$	$l=80$	$l=100$	$l=120$
2400	*rec*	$\hat{\rho}$	0.000	0.000	0.220	0.545	0.635	0.710	0.805	0.835	0.840	0.845
		$\hat{\rho}_1$	0.000	0.175	0.615	0.800	0.825	0.885	0.935	0.905	0.900	0.910
		$\hat{\rho}_2$	0.950	0.955	0.945	0.960	0.950	0.940	0.945	0.920	0.950	0.930
	no adj	$\hat{\rho}$	0.055	0.630	0.785	0.810	0.875	0.865	0.840	0.880	0.865	0.835
		$\hat{\rho}_1$	0.145	0.720	0.820	0.835	0.915	0.885	0.855	0.890	0.920	0.855
		$\hat{\rho}_2$	0.140	0.685	0.870	0.920	0.890	0.870	0.935	0.905	0.870	0.870
Panel B: DGP is an AR(2) Process - $\rho_1 = \rho_2 = 0.4$												
			$l=4$	$l=6$	$l=10$	$l=12$	$l=15$	$l=20$	$l=25$	$l=30$	$l=50$	$l=60$
600	*rec*	$\hat{\rho}$	0.155	0.360	0.650	0.665	0.755	0.830	0.785	0.795	0.840	0.830
		$\hat{\rho}_1$	0.580	0.695	0.855	0.840	0.895	0.920	0.865	0.865	0.875	0.890
		$\hat{\rho}_2$	0.515	0.690	0.855	0.885	0.890	0.905	0.900	0.865	0.840	0.855
	no adj	$\hat{\rho}$	0.165	0.350	0.660	0.620	0.750	0.780	0.770	0.775	0.800	0.820
		$\hat{\rho}_1$	0.570	0.695	0.840	0.830	0.870	0.885	0.845	0.835	0.835	0.865
		$\hat{\rho}_2$	0.510	0.725	0.825	0.840	0.875	0.880	0.865	0.845	0.805	0.830
			$l=4$	$l=10$	$l=15$	$l=20$	$l=25$	$l=30$	$l=40$	$l=50$	$l=60$	$l=100$
1200	*rec*	$\hat{\rho}$	0.035	0.520	0.670	0.755	0.800	0.810	0.810	0.905	0.865	0.840
		$\hat{\rho}_1$	0.215	0.780	0.880	0.895	0.895	0.895	0.905	0.940	0.915	0.845
		$\hat{\rho}_2$	0.220	0.775	0.830	0.830	0.915	0.930	0.890	0.875	0.910	0.860
	no adj	$\hat{\rho}$	0.035	0.515	0.670	0.720	0.770	0.795	0.770	0.885	0.870	0.810
		$\hat{\rho}_1$	0.215	0.775	0.855	0.870	0.880	0.875	0.880	0.925	0.895	0.815
		$\hat{\rho}_2$	0.225	0.740	0.780	0.790	0.875	0.910	0.850	0.845	0.880	0.820
			$l=4$	$l=10$	$l=20$	$l=30$	$l=40$	$l=50$	$l=60$	$l=80$	$l=100$	$l=120$
2400	*rec*	$\hat{\rho}$	0.000	0.310	0.670	0.780	0.840	0.890	0.855	0.910	0.885	0.905
		$\hat{\rho}_1$	0.025	0.645	0.860	0.870	0.885	0.950	0.890	0.925	0.935	0.920
		$\hat{\rho}_2$	0.020	0.480	0.825	0.855	0.930	0.880	0.890	0.890	0.870	0.900
	no adj	$\hat{\rho}$	0.000	0.290	0.635	0.770	0.800	0.865	0.840	0.855	0.865	0.890
		$\hat{\rho}_1$	0.030	0.665	0.815	0.850	0.850	0.920	0.875	0.880	0.920	0.890
		$\hat{\rho}_2$	0.035	0.465	0.805	0.860	0.905	0.885	0.855	0.850	0.835	0.875

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