Bootstrap Out of Sample Predictive Evaluation Tests*

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

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

June 2004 - incomplete

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, predictive evaluation tests have a limiting distribution which 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 an appropriate recentering of the bootstrap score. The usefulness of our approach is illustrated via two applications: one is an extension of White (2000) reality check to the case of non vanishing parameter estimation error. The oother is an out of sample version of integrated conditional moment tests of Bierens (1982, 1990) and Bierens and Ploberger (1997). The main findings from a small Monte Carlo experiment indicate that the ....

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 for earlier version of this paper. Corradi gratefully acknowledges ESRC grant RES-000-23-0006.

1 Introduction

In the predictive evaluation literature, multiple misspecified models are compared in terms of their out of sample predictive ability, for given loss function. 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. Furthermore, it is common practice to split the sample T into $T = R+P$ observations, where only 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 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 constructing tests which have power against generic alternatives, the statistic has a limiting distribution can be shown to be a functional of a Gaussian process with a covariance kernel that reflects both (dynamic) misspecification as well as the contribution of (recursive) parameter estimation error. Such a limiting distribution is 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 procedure that allows us to form statistics which properly mimic the contribution of $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_t - \theta^{\dagger}\right)$. The first 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. On the other hand, in standard block bootstrap, any block from the original sample has the same probability of being selected. Thus, the bootstrap estimator constructed as a direct analog of $\widehat{\theta}_t$, is characterized by a location bias that can be either positive or negative, depending on the sample that we observe. In order to circumvent this problem, we suggest a recentering of the bootstrap score which ensures that the new bootstrap estimator, which is no longer the analog of $\widehat{\theta}_t$, is asymptotically unbiased. Now, the idea of recentering is not new at all in the bootstrap literature for the case of full sample estimation. In fact, recentering is necessary even for first order validity in the case of overidentified generalized method of moments (GMM) estimators, see e.g. Hall and Horowitz (1996), Andrews (2002, 2004), Goncalves and White (2004) and Inoue and Shintani (2004). This is due to the fact that, in the overidentified case, the boostrap moment conditions are not equal to zero, even if the population moment conditions are. However, in the context of $m-$estimators using the full sample, recentering is needed only for high order asymptotics, but not for first order validity, in the sense that the bias term is of smaller order than $T^{-1/2}$, see e.g. Andrews (2002). In the case of recursive $m-$estimators the bias term is instead if order $T^{-1/2}$ and so it does contribute to the limiting distribution.

The recursive PEE bootstrap can be used to provide valid critical values in a variety of interesting testing contexts, and two such leading applications are developed. The first is a generalization of the reality check test of White (2000) that allows for non vanishing parameter estimation error. The second 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.

To be more specific, the first application concerns the reality check of White (2000), which extends the Diebold and Mariano (1995) and West (1996) test for the relative predictive accuracy of two models by allowing for the joint comparison of multiple misspecified models against a given benchmark. White obtains valid asymptotic critical values for his test via use of the Politis and Romano (1994) stationary bootstrap for the case in which parameter estimation error is asymptotically negligible. Using the bootstrap for recursive estimator, we generalize the reality check to the case in which parameter estimation error does not vanish asymptotically.

The objective of the second application is to test the predictive accuracy of a given (non)linear model against generic (non)linear alternatives. The ICM type test used for this purpose differs from those developed by Bierens (1982,1990) and Bierens and Ploberger (1997) because parameters are estimated recursively, out-of-sample prediction models are analyzed, and the null hypothesis is that the reference model is the best "loss function specific" predictor, for a given information set. This application builds on previous work by Corradi and Swanson (2002) who use a conditional p-value method for constructing critical values in this context, extending earlier work by Hansen (1996) and Inoue (2001). However, the conditional p-value approach they use suffers from the fact that under the alternative, the simulated statistics diverge (at rate as high as $\sqrt{\widetilde{l}}$), conditional on the sample, where $\widetilde{l}$ plays a role analogous to the block length in the block bootstrap. As this feature may lead to reduced power in finite samples, we establish in the second application that the bootstrap for recursive $m-$estimators can yield a $\sqrt{P}$-consistent test.

MONTE CARLO DISCUSS

The rest of the paper is organized as follows. Section 2 explains the block bootstrap for recursive $m-$estimators and establishes its first order validity. Sections 3 and 4 outline the two applications of the recursive block boostrap: White's reality check and out of sample integrated conditional moment tests. 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. Analogously, $O_{a.s.^*}(1)$ and $o_{a.s.^*}(1)$ denote a term almost surely bounded and a term zero almost surely according the the probability law P^*, conditionally on the sample.

2 Bootstrap for Recursive $m-$Estimators

In this section, we establish the first order validity of a block bootstrap estimator that allows us to properly capture the effect of parameter estimation error in contexts of *recursive* m-estimators 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 application outlined in Section 3 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.

We use the overlapping block resampling scheme of Künsch (1989), as follows:[1] At each replication, draw b blocks (with replacement) of length l from the sample $W_t = (y_t, Z^{t-1})$, where $bl = T - s$. Thus, the first block is equal to $W_{i+1}, ..., W_{i+l}$, for some $i = s-1, ..., T-l+1$, with probability $1/(T-s-l+1)$, the second block is equal to $W_{i+1}, ..., W_{i+l}$, again for some $i = s-1, ..., T-l+1$, with

[1] The main difference between the block bootstrap and the stationary bootstrap of Politis and Romano (PR: 1994a) is that the former uses a deterministic block length, which may be either overlapping as in Künsch (1989) or non-overlapping as in Carlstein (1986), while the latter resamples using blocks of random length. One important feature of the PR bootstrap is that the resampled series, conditional on the sample, is stationary, while a series resampled from the (overlapping or non overlapping) block bootstrap is nonstationary, even if the original sample is strictly stationary. However, Lahiri (1999) shows that all block boostrap methods, regardless of whether the block length is deterministic or random, have a first order bias of the same magnitude, but the bootstrap with deterministic block length has a smaller first order variance. In addition, the overlapping block boostrap is more efficient than the non overlapping block bootstrap.

probability $1/(T-s-l+1)$, and so on for all blocks. More formally, let I_k, $k=1,...,b$ be *iid* discrete uniform random variables on $[s=s-1,s,...,T-l+1]$. Then, the resampled series, $W_t^*=(y_t^*,Z^{*,t-1})$, is such that $W_1^*,W_2^*,...,W_l^*,W_{l+1}^*,...,W_T^*=W_{I_1+1},W_{I_1+2},...,W_{I_1+l},W_{I_2},...,W_{I_b+l}$, and so a resampled series consists of b blocks that are discrete *iid* uniform random variables, conditional on the sample.

Let $W_t=(y_t,Z^{t-1})$, and draw b overlapping blocks of length l, where $bl=T-s$. The resampled observations, $W_s^*,W_{s+1}^*,...,W_{s+l-1}^*,...,W_T^*$, are equal to $W_{I_1},W_{I_1+1},...,W_{I_1+l-1},...,W_{I_b+l-1}$, where I_i, $i=1,...,b$ are independent discrete uniform random draws on the interval $s,...,T-l+1$.

Suppose we define the bootstrap estimator $\widehat{\theta}_{i,t}^*$ as the direct analog of $\widehat{\theta}_{i,t}$, that is

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

By the first order conditions, $\frac{1}{t}\sum_{j=s}^{t}\nabla_\theta q_i(y_j^*,Z^{*,j-1},\widehat{\theta}_{i,t}^*)=0$, and via a mean value expansion of $\frac{1}{t}\sum_{j=s}^{t}\nabla_\theta q_i(y_j^*,Z^{*,j-1},\widehat{\theta}_{i,t}^*)$ around $\widehat{\theta}_{i,t}$, after a few simple manipulations we have that

$$\begin{aligned}
&\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_{i,t}^*-\widehat{\theta}_{i,t}\right)\\
=\ &\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\left(\frac{1}{t}\sum_{j=s}^{t}\nabla_\theta^2 q_i(y_j^*,Z^{*,j-1},\overline{\theta}_{i,t}^*)\right)^{-1}\frac{1}{t}\sum_{j=s}^{t}\nabla_\theta q_i(y_j^*,Z^{*,j-1},\widehat{\theta}_{i,t})\right)\\
=\ &B_i^\dagger\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\frac{1}{t}\sum_{j=s}^{t}\nabla_\theta q_i(y_j^*,Z^{*,j-1},\widehat{\theta}_{i,t})\right)+o_{P^*}(1)\ \ \Pr-P\\
=\ &B_i^\dagger\frac{a_{R,0}}{\sqrt{P}}\sum_{t=s}^{R}\nabla_\theta q_i(y_j^*,Z^{*,j-1},\widehat{\theta}_{i,t})+B_i^\dagger\frac{1}{\sqrt{P}}\sum_{j=1}^{P-1}a_{R,j}\nabla_\theta q_i(y_{R+j}^*,Z^{*,R+j-1},\widehat{\theta}_{i,t})\\
&+o_{P^*}(1)\ \ \Pr-P,
\end{aligned} \tag{4}$$

where $\overline{\theta}_{i,t}^*\in\left(\widehat{\theta}_{i,t}^*,\widehat{\theta}_{i,t}\right)$, $B_i^\dagger=E\left(\nabla_\theta^2 q_i(y_j,Z^{j-1},\theta_i^\dagger)\right)$, $a_{R,j}=\frac{1}{R+j}+\frac{1}{R+j+1}+...+\frac{1}{R+P-1}$, $j=0,1,...,P-1$, and where the last equality on the right hand side of (4) comes straightforwardly by the same argument as in Lemma A5 in West (1996). Analogously,

$$\begin{aligned}
&\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_{i,t}-\theta_i^\dagger\right)\\
=\ &B_i^\dagger\frac{a_{R,0}}{\sqrt{P}}\sum_{t=s}^{R}\nabla_\theta q_i(y_j,Z^{j-1},\theta_i^\dagger)+B_i^\dagger\frac{1}{\sqrt{P}}\sum_{j=1}^{P-1}a_{R,j}\nabla_\theta q_i(y_{R+j},Z^{R+j-1},\theta_i^\dagger)+o_P(1)
\end{aligned} \tag{5}$$

Now, given (2), $E\left(\nabla_\theta q_i(y_j, Z^{j-1}, \theta_i^\dagger)\right) = 0$ for all j, and in fact $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_{i,t} - \theta_i^\dagger\right)$ has a zero mean normal limiting distribution (see Theorem 4.1 in West (1996)). On the other hand, as any block of observations has the same chance of being drawn,

$$E^*\left(\nabla_\theta q_i(y_j^*, Z^{*,j-1}, \widehat{\theta}_{i,t})\right) = \frac{1}{T-s}\sum_{k=1}^{T-1}\nabla_\theta q_i(y_k, Z^{k-1}, \widehat{\theta}_{i,t}) + O\left(\frac{l}{T}\right) \ \Pr -P, \tag{6}$$

where the $O\left(\frac{l}{T}\right)$ comes from the fact that the first and last l observations have lower chance of being drawn (see e.g. Fitzenberger (1997)).[2] Now, $\frac{1}{T-s}\sum_{k=1}^{T-1}\nabla_\theta q_i(y_k, Z^{k-1}, \widehat{\theta}_{i,t}) \neq 0$, and is instead of order $O_P\left(T^{-1/2}\right)$ and so, $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\frac{1}{T-s}\sum_{k=1}^{T-1}\nabla_\theta q_i(y_k, Z^{k-1}, \widehat{\theta}_{i,t}) = O_P(1)$ and does not vanish. This clearly contrasts with the full sample case, in which $\frac{1}{T-s}\sum_{k=1}^{T-1}\nabla_\theta q_i(y_k, Z^{k-1}, \widehat{\theta}_{i,T}) = 0$ because of the first order conditions. Thus, $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_{i,t}^* - \widehat{\theta}_{i,t}\right)$ cannot have a zero mean normal limiting distribution, but is instead characterized by a location bias that can be either positive or negative depending on the sample.

Given (6), we want to have the bootstrap score centered around $\frac{1}{T-s}\sum_{k=1}^{T-1}\nabla_\theta q_i(y_k, Z^{k-1}, \widehat{\theta}_{i,t})$. Thus, define a new bootstrap estimator $\widetilde{\theta}_{i,t}^*$ as:

$$\widetilde{\theta}_{i,t}^* = \arg\min_{\theta_i\in\Theta_i}\frac{1}{t}\sum_{j=s}^{t}\left(q_i(y_j^*, Z^{*,j-1}, \theta_i) - \theta_i'\left(\frac{1}{T}\sum_{k=s}^{T-1}\nabla_{\theta_i} q_i(y_k, Z^{k-1}, \widehat{\theta}_{i,t})\right)\right), \tag{7}$$

$R \leq t \leq T-1$, $i = 1, ..., n$.[3]

By the first order conditions, $\frac{1}{t}\sum_{j=s}^{t}\left(\nabla_\theta q_i(y_j^*, Z^{*,j-1}, \widetilde{\theta}_{i,t}^*) - \left(\frac{1}{T}\sum_{k=s}^{T-1}\nabla_{\theta_i} q_i(y_k, Z^{k-1}, \widehat{\theta}_{i,t})\right)\right) = 0$, and via a mean value expansion of $\frac{1}{t}\sum_{j=s}^{t}\nabla_\theta q_i(y_j^*, Z^{*,j-1}, \widetilde{\theta}_{i,t}^*)$ around $\widehat{\theta}_{i,t}$, after a few manipulations,

$$\begin{aligned}
&\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widetilde{\theta}_{i,t}^* - \widehat{\theta}_{i,t}\right) \\
= \ & B^\dagger\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(\frac{1}{t}\sum_{j=s}^{t}\left(\nabla_\theta q(y_j^*, Z^{*,j-1}, \widehat{\theta}_{i,t}) - \left(\frac{1}{T}\sum_{k=s}^{T-1}\nabla_\theta q(y_k, Z^{k-1}, \widehat{\theta}_{i,t})\right)\right)\right) \\
&+ o_{P^*}(1) \ \Pr -P
\end{aligned}$$

[2] In fact, the first and last observation in the sample, can appear only at the beginning and end of the block, for example.

[3] More precisely, we should have defined

$$\widetilde{\theta}_{i,t}^* = \arg\min_{\theta_i\in\Theta_i}\frac{1}{t-s}\sum_{j=s}^{t}\left(q_i(y_j^*, Z^{*,j-1}, \theta_i) - \theta_i'\left(\frac{1}{T-s}\sum_{k=s}^{T-1}\nabla_{\theta_i} q_i(y_k, Z^{k-1}, \widehat{\theta}_{i,t})\right)\right)$$

However, for notational simplicity we approximate $\frac{1}{t-s}$ and $\frac{1}{T-s}$ with $\frac{1}{t}$ and $\frac{1}{T}$.

Given (6), it is immediate to see that the bias asociated with $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widetilde{\theta}_{i,t}^{*}-\widehat{\theta}_{i,t}\right)$ is of order $O\left(lT^{-1/2}\right)$, conditional on the sample, and so it is negligible for first order asymptotics, as $l = o(T^{1/2})$.

Theorem 1 below requires the following assumptions.

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

Assumption A2: (i) $\theta_i^{\dagger}$ is uniquely identified (i.e. $E(q_i(y_t, Z^{t-1}, \theta_i)) > E(q_i(y_t, Z^{t-1}, \theta_i^{\dagger}))$ 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 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.[4]

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. In fact, if P grew at a slower rate than R, i.e. $P/R \to 0$, then $\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(\widehat{\theta}_{i,t}-\theta_i^{\dagger}\right) = o_P(1)$ and so there were no need to capture the contribution of parameter estimation error.

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

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

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

Broadly speaking, Theorem 1 states that $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widetilde{\theta}_{i,t}^{*}-\theta_i^{\dagger}\right)$ 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.

[4]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 $\left|\nabla_{\theta_i} q_i(y_t, Z^{t-1}, \theta_i)\right|_j \leq D_t$, and $E(|D_t|^{2r}) < \infty$. For more details on domination conditions, see Gallant and White (1988, pp. 33).

3 White's Reality Check for Data Snooping

In this section, we extend the White (2000) reality check to the case in which the effect of parameter estimation error does not vanish asymptotically. In particular, we show that the block bootstrap for recursive m-estimators properly mimics the contribution of parameter estimation error to the covariance kernel of the limiting distribution of the original reality check test. Although we focus our attention in this paper on the recursive PEE bootstrap, which is based on resampling blocks of deterministic length, we conjecture that the same approach can be used to extend the stationary bootstrap employed by White (2000) to the case of nonvanishing parameter estimation error.

Let the generic forecast error be $u_{i,t+1} = y_{t+1} - \kappa_i(Z^t, \theta_i^\dagger)$, and let $\widehat{u}_{i,t+1} = y_{t+1} - \kappa_i(Z^t, \widehat{\theta}_{i,t})$, where $\kappa_i(Z^t, \widehat{\theta}_{i,t})$ is the estimated conditional mean function under model i. We assume that the set of regressors may vary across different models, so that Z^t is meant to denote the collection of all potential regressors. Following White (2000), define the statistic

$$S_P = \max_{k=2,\dots,n} S_P(1,k),$$

where

$$S_P(1,k) = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(g(\widehat{u}_{1,t+1}) - g(\widehat{u}_{k,t+1}) \right), \ k = 2, ..., n,$$

with g a given loss function. Recall that in this test, parameter estimation error need not be accounted for in the covariance kernel of the limiting distribution unless $g \neq q_i$ for some i. This follows upon examination of the results of both West (1996) and White (2000). In particular, in West (1996), the parameter estimation error components that enter into the covariance kernel of the limiting distribution of his test statistic are zero whenever the same loss function is used for both predictive evaluation and in-sample estimation. The same argument holds for the reality check test. This means that as long as $g = q_i \ \forall i$, the White test can be applied regardless of the rate of growth of P and R. When we write the covariance kernel of the limiting distribution of the statistic (see below), however, we include terms capturing the contribution of parameter estimation error, thus implicitly assuming that $g \neq q_i$ for some i. In practice, one reason why we allow for cases where $g \neq q_i$ is that least squares is sometimes better behaved in finite samples and/or easier to implement than more generic $m-$estimators, so that practitioners sometimes use least squares for estimation and more complicated (possibly asymmetric) loss functions for predictive evaluation.[5]

[5]Consider linex loss, where $g(u) = e^{au} - au - 1$, so that for $a > 0$ ($a < 0$) positive (negative) errors are more

Of course, there are also applications for which parameter estimation error does not vanish, even if the same loss function is used for parameter estimation and predictive evaluation. One such application is discussed in the next section.

For a given loss function, the reality check tests the null hypothesis that a benchmark model (defined as model 1) performs equal to or better than all competitor models (i.e. models 2,...,n). The alternative is that at least one competitor performs better than the benchmark.[6] Formally, the hypotheses are:

$$H_0 : \max_{k=2,\dots,n} E\left(g(u_{1,t+1}) - g(u_{k,t+1})\right) \leq 0$$

and

$$H_A : \max_{k=2,\dots,n} E\left(g(u_{1,t+1}) - g(u_{k,t+1})\right) > 0.$$

In order to derive the limiting distribution of S_P we require the following additional assumption.

Assumption A4: (i) κ_i is twice continuously differentiable on the interior of Θ_i and the elements of $\nabla_{\theta_i}\kappa_i(Z^t, \theta_i)$ and $\nabla^2_{\theta_i}\kappa_i(Z^t, \theta_i)$ are $p-$dominated on Θ_i, for $i = 2, ..., n$, with $p > 2(2 + \psi)$, where ψ is the same positive constant as that defined in Assumption A1; (ii) g is positive valued, twice continuously differentiable on Θ_i, and g, g' and g'' are $p-$dominated on Θ_i with p defined as in (i); and (iii) let $c_{kk} =$ $\lim_{T\to\infty} Var\left(\frac{1}{\sqrt{T}}\sum_{t=s}^{T}\left(g(u_{1,t+1}) - g(u_{k,t+1})\right)\right)$, $k = 2, ..., n$, define analogous covariance terms, $c_{j,k}$, $j, k = 2, ..., n$, and assume that $[c_{j,k}]$ is positive semi-definite.

Assumptions A4(i)-(ii) are standard smoothness and domination conditions imposed on the conditional mean functions of the models. Assumption A4(iii) is standard in the literature that uses DM type tests (see e.g. West (1996) and White (2000)), and states that at least one of the competing models has to be nonnested with (and not nesting) the benchmark.

Proposition 2: Let assumptions A1-A4 hold, then

$$\max_{k=2,\dots,n}\left(S_P(1,k) - \sqrt{P}E\left(g(u_{1,t+1}) - g(u_{k,t+1})\right)\right) \xrightarrow{d} \max_{k=2,\dots,n} S(1,k),$$

(less) costly than negative (positive) errors. Here, the errors are exponentiated, so that in this particular case, laws of large numbers and central limit theorems may require a large number of observations before providing satisfactory approximations. This feature of linex loss is illustrated in the Monte Carlo findings of Corradi and Swanson (2002). (Linex loss is studied in Zellner (1986), Christoffersen and Diebold (1996, 1997) and Granger (1999), for example.)

[6] In the current context, we are interested in choosing the model which is more accurate for given loss function. An alternative approach is to combine different forecasting model in some optimal way. For very recent contributions along these lines, see Elliott and Timmermann (2004a,b).

where $S = (S(1,2), ..., S(1,n))$ is a zero mean Gaussian process with covariance kernel given by V, with V a $n \times n$ matrix with i, i element

$v_{i,i} = S_{g_i g_i} + 2\Pi \mu_1' B_1^\dagger C_{11} B_1^\dagger \mu_1 + 2\Pi \mu_i' B_i^\dagger C_{ii} B_i^\dagger \mu_i - 4\Pi \mu_1' B_1^\dagger C_{1i} B_i^\dagger \mu_i + 2\Pi S_{g_i q_1} B_1^\dagger \mu_1 - 2\Pi S_{g_i q_i} B_i^\dagger \mu_i$,

where $S_{g_i g_i} = \sum_{\tau=-\infty}^{\infty} E\left((g(u_{1,1}) - g(u_{i,1}))\left(g(u_{1,1+\tau}) - g(u_{i,1+\tau})\right)\right)$,

$C_{ii} = \sum_{\tau=-\infty}^{\infty} E\left(\left(\nabla_{\theta_i} q_i(y_{1+s}, Z^s, \theta_i^\dagger)\right)\left(\nabla_{\theta i} q_i(y_{1+s+\tau}, Z^{s+\tau}, \theta_i^\dagger)\right)'\right)$,

$S_{g_i q_i} = \sum_{\tau=-\infty}^{\infty} E\left((g(u_{1,1}) - g(u_{i,1}))\left(\nabla_{\theta i} q_i(y_{1+s+\tau}, Z^{s+\tau}, \theta_i^\dagger)\right)'\right)$,

$B_i^\dagger = \left(E\left(-\nabla_{\theta i}^2 q_i(y_t, Z^{t-1}, \theta_i^\dagger)\right)\right)^{-1}$, $\mu_i = E\left(\nabla_{\theta_i} g(u_{i,t+1})\right)$, and $\Pi = 1 - \pi^{-1}\ln(1+\pi)$.

Just as in White (2000), note that under the null, the least favorable case arises when $E\left(g(u_{1,t+1}) - g(u_{k,t+1})\right) = 0$, $\forall\ k$. In this case, the distribution of S_P coincides with that of $\max_{k=2,...,n}\left(S_P(1,k) - \sqrt{P}E\left(g(u_{1,t+1}) - g(u_{k,t+1})\right)\right)$, so that S_P has the above limiting distribution, which is a functional of a Gaussian process with a covariance kernel that reflects uncertainty due to parameter estimation error and dynamic mispecification. 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 S_P will always be smaller than $\max_{k=2,...,n}\left(S_P(1,k) - \sqrt{P}E\left(g(u_{1,t+1}) - g(u_{k,t+1})\right)\right)$, asymptotically. Of course, when H_A holds, the statistic diverges to plus infinity at rate $\sqrt{P}$.

In a recent paper, Hansen (2004) 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 $E\left(g(u_{1,t+1}) - g(u_{k,t+1})\right) = 0$, $\forall\ k$), and suggests a mean correction for S_P in order to address this feature of the test. Our version of the reality check has the same features and can also be modified using the method proposed by Hansen.[7]

Recall that the maximum of a Gaussian process is not Gaussian in general, so that standard

[7] 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 $\max_{k=2,...,m} E\left(g(u_{1,t+1}) - g(u_{k,t+1})\right) = 0$, 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.

critical values cannot be used to conduct inference on S_P. As pointed out by White (2000), one possibility in this case is to first estimate the covariance structure and then draw 1 realization from an $(n-1)$-dimensional normal with covariance equal to the estimated covariance structure. From this realization, pick the maximum value over $k = 2, \ldots, n$. Repeat this a large number of times, form an empirical distribution using the maximum values over $k = 2, \ldots, n$, and obtain critical values in the usual way. A drawback to this approach is that we need to rely on an estimator of the covariance structure based on the available sample of observations, which in many cases may be small relative to the number of models being compared. Furthermore, whenever the forecasting errors are not martingale difference sequences (as in our context), heteroskedasticity and autocorrelation consistent covariance matrices should be estimated, and thus a lag truncation parameter must be chosen. Another approach which avoids these problems involves using the stationary bootstrap of Politis and Romano (1994). This is the approach used by White (2000), and in general, bootstrap procedures have been shown to perform well in a variety of finite sample contexts (see e.g. Diebold and Chen (1996)). Our approach is to apply the recursive PEE bootstrap outlined above.

Hereafter, let $u^*_{i,t+1} = y^*_{t+1} - \kappa_i(Z^{*,t}, \theta_i^\dagger)$ and $\widetilde{u}^*_{i,t+1} = y^*_{t+1} - \kappa_i(Z^{*,t}, \widetilde{\theta}^*_{i,t})$, where

$$\widetilde{\theta}^*_{i,t} = \arg\min_{\theta_i \in \Theta_i} \frac{1}{t} \sum_{j=s}^{t} \left(q_i(y^*_j, Z^{*,j-1}, \theta_i) - \theta_i' \left(\frac{1}{T} \sum_{h=s}^{T-1} \nabla_{\theta_i} q_i(y_h, Z^{h-1}, \widehat{\theta}_{i,t}) \right) \right), \tag{8}$$

with $R \le t \le T-1$, $i = 1, ..., n$. Define

$$S^*_P = \max_{k=2,\ldots,n} S^*_P(1,k),$$

where

$$S^*_P(1,k) = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\left(g(\widetilde{u}^*_{1,t+1}) - g(\widetilde{u}^*_{k,t+1}) \right) - \frac{1}{T} \sum_{i=1}^{T-1} \left(g(\widehat{u}_{1,i+1}) - g(\widehat{u}_{k,i+1}) \right) \right).$$

NOTE THAT $g(\widehat{u}_{1,i+1})$ and $g(\widehat{u}_{k,i+1})$ ARE INSIDE THE SUMMATION OVER t, AS THEY DEPEND ON t, VIA $\widehat{\theta}_{1,t}$ and $\widehat{\theta}_{k,t}$.

Proposition 3: Let assumptions A1-A4 hold. Also, assume that as $T, P, R \to \infty$, $l \to \infty$, and that $\frac{l}{T^{1/4}} \to 0$. Then, as T, P and $R \to \infty$,

$$P\left(\omega : \sup_{v \in \Re} \left| P^*_T \left(\max_{k=2,\ldots,n} S^*_P(1,k) \le v \right) - P\left(\max_{k=2,\ldots n} S^\mu_P(1,k) \le v \right) \right| > \varepsilon \right) \to 0,$$

and

$$S_P^{\mu}(1,k) = S_P(1,k) - \sqrt{P}E\left(g(u_{1,t+1}) - g(u_{k,t+1})\right),$$

The above result suggests proceeding in the following manner. For any bootstrap replication, compute the bootstrap statistic, S_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, S_P has the same limiting distribution as the corresponding bootstrapped statistic when $E\left(g(u_{1,t+1}) - g(u_{k,t+1})\right) = 0 \; \forall \; k$, ensuring asymptotic size equal to α. On the other hand, when one or more competitor models are strictly dominated by the benchmark, the rule provides a test with asymptotic size between 0 and α. Under the alternative, S_P diverges to (plus) infinity, while the corresponding bootstrap statistic has a well defined limiting distribution, ensuring unit asymptotic power.

4 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). There are at least two reasons for using an ICM type test rather than a Diebold and Mariano (DM: 1995), West (1996) or reality check type test. The first reason is that ICM type tests have power against generic alternatives, while the tests cited above have power only against a fixed number of alternatives. The second reason is, whenever the same loss function is used for both estimation and testing, then Diebold Mariano type tests do no longer have a gaussian limiting distribution in the nested case.[8]

[8] McCracken (2004) shows that a particular version of the DM has a nonstandard limiting distribution which is a functionals over Brownian motions. His results relies on a quadratic loss function (and OLS estimator) and on martingale difference score. Other tests for comparing a fixed number of nested models are Clark and McCracken (2001 and 2003) for one-step and multi-step prediction, and Chao, Corradi and Swanson (2001). The latter allows for dynamic misspecification under the null. Recently, Giacomini and White (2003) provides a test for conditional predictive ability valid for both nested and nonnested models. The key ingredient of their test is the fact that parameters are estimated using a fixed rolling window.

As shown in the appendix, 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. 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 statistic diverges (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.[9]

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).[10] 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:

$$y_t = \theta^{\dagger}_{1,1} + \theta^{\dagger}_{1,2} y_{t-1} + u_{1,t}, \tag{9}$$

where $\theta_1^{\dagger} = (\theta^{\dagger}_{1,1}, \theta^{\dagger}_{1,2})' = \arg\min_{\theta_1 \in \Theta_1} E(q_1(y_t - \theta_{1,1} - \theta_{1,2} y_{t-1}))$, $\theta_1 = (\theta_{1,1}, \theta_{1,2})'$, 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:

$$y_t = \theta^{\dagger}_{2,1}(\gamma) + \theta^{\dagger}_{2,2}(\gamma) y_{t-1} + \theta^{\dagger}_{2,3}(\gamma) w(Z^{t-1}, \gamma) + u_{2,t}(\gamma), \tag{10}$$

[9] About CS (2002), extensive Monte Carlo findings as well as an empirical illustration to U.S. macroeconomic series have been provided in Corradi and Swanson (2004).

[10] 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.

where $\theta_2^\dagger(\gamma) = (\theta_{2,1}^\dagger(\gamma), \theta_{2,2}^\dagger(\gamma), \theta_{2,3}^\dagger(\gamma))' = \arg\min_{\theta_2 \in \Theta_2} E(q_1(y_t - \theta_{2,1} - \theta_{2,2}y_{t-1} - \theta_{2,3}w(Z^{t-1}, \gamma)))$, $\theta_2(\gamma) = (\theta_{2,1}(\gamma), \theta_{2,2}(\gamma), \theta_{2,3}(\gamma))'$, $\theta_2 \in \Theta_2$, 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 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:

$$H_0 : E(g(u_{1,t+1}) - g(u_{2,t+1}(\gamma))) = 0 \text{ versus } H_A : E(g(u_{1,t+1}) - g(u_{2,t+1}(\gamma))) > 0. \tag{11}$$

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:

$$H_0 : E(g'(u_{1,t+1})w(Z^t, \gamma)) = 0 \text{ versus } H_A : E(g'(u_{1,t+1})w(Z^t, \gamma)) \neq 0, \tag{12}$$

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:

$$M_P = \int_\Gamma m_P(\gamma)^2 \phi(\gamma) d\gamma, \tag{13}$$

and

$$m_P(\gamma) = \frac{1}{P^{1/2}} \sum_{t=R}^{T-1} g'(\widehat{u}_{1,t+1}) w(Z^t, \gamma), \tag{14}$$

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 A5: (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 4: Let assumptions A1-A3 and A5 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 (14) and Z is a Gaussian process with covariance kernel given by:

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

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 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[11]

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

$$\begin{aligned}\widetilde{\theta}^*_{1,t} &= (\widetilde{\theta}^*_{1,1,t}, \widetilde{\theta}^*_{1,2,t})' = \arg\min_{\theta_1 \in \Theta_1} \frac{1}{t} \sum_{j=2}^{t} \Big(q_1(y^*_j - \theta_{1,1} - \theta_{1,2} y^*_{j-1}) \\ &\quad - \theta'_1 \frac{1}{T} \sum_{i=2}^{T-1} \nabla_\theta q_1 (y_i - \widehat{\theta}_{1,1,t} - \widehat{\theta}_{1,2,t} y_{i-1}) \Big) \end{aligned} \quad (15)$$

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

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

where, recalling that $g = q_1$,

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

NOTE TO SELF! $g'(\widehat{u}_{1,i+1})w(Z^i,\gamma)$ IS INSIDE THE SUMMATION OVER t, AS THEY DEPEND ON t, VIA $\widehat{\theta}_{1,t}$.

Proposition 5: Let assumptions A1-A3 and A5 hold. Also, assume that as $T, P, R \to \infty$, $l \to \infty$, and that $\frac{l}{T^{1/4}} \to 0$. Then, as T, P and $R \to \infty$,

$$P\left(\omega : \sup_{v \in \Re} \left| P^*_T \left(\int_\Gamma m^*_P(\gamma)^2 \phi(\gamma) d\gamma \leq v \right) - P\left(\int_\Gamma m^\mu_P(\gamma)^2 \phi(\gamma) d\gamma \leq v \right) \right| > \varepsilon \right) \to 0,$$

where $m^\mu_P(\gamma) = m_P(\gamma) - \sqrt{P} E\left(g'(u_{1,t+1}) w(Z^t, \gamma) \right)$.

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.

5 Monte Carlo

to be completed

6 Appendix

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.

Proof of Theorem 1: Given (7), by the first order conditions,

$$\frac{1}{t}\sum_{j=s}^{t}\left(\nabla_\theta q(y_j^*, Z^{*,j-1}, \widetilde{\theta}_t^*) - \left(\frac{1}{T}\sum_{k=s}^{T-1}\nabla_\theta q(y_k, Z^{k-1}, \widehat{\theta}_t)\right)\right) = 0$$

and so, via aTaylor expansion, around $\widehat{\theta}_t$,

$$\begin{aligned}\left(\widetilde{\theta}_t^* - \widehat{\theta}_t\right) &= \left(\frac{1}{t}\sum_{j=s}^{t}\nabla_\theta^2 q(y_j^*, Z^{*,j-1}, \overline{\theta}_t^*)\right)^{-1} \\ &\times \left(\frac{1}{t}\sum_{j=s}^{t}\left(\nabla_\theta q(y_j^*, Z^{*,j-1}, \widehat{\theta}_t) - \left(\frac{1}{T}\sum_{k=s}^{T-1}\nabla_\theta q(y_k, Z^{k-1}, \widehat{\theta}_t)\right)\right)\right),\end{aligned}$$

where $\overline{\theta}_t^* \in \left(\widetilde{\theta}_t^*, \widehat{\theta}_t\right)$. Hereafter, let $B^\dagger = \left(E\left(\nabla_\theta q(y_j, Z^{j-1}, \theta^\dagger)\right)\right)^{-1}$. Recalling that, regardless the value of t, we resample from the entire sample,

$$\frac{1}{t}\sum_{j=s}^{t}E^*\left(\nabla_\theta^2 q(y_j^*, Z^{*,j-1}, \theta)\right) = \frac{1}{T}\sum_{k=s}^{T-1}\nabla_\theta^2 q(y_k, Z^{k-1}, \theta) + O_{P^*}\left(\frac{l}{T}\right), \ \ \Pr -P, \tag{16}$$

where the $O_{P^*}\left(\frac{l}{T}\right)$ is due to the end effect, i.e. to the contribution of the first and last l observations, as shown in Lemma A1 in Fitzenberger (1997). Thus,

$$\begin{aligned}&\sup_{t\geq R}\sup_{\theta\in\Theta}\left|\left(\frac{1}{t}\sum_{j=s}^{t}\nabla_\theta^2 q(y_j^*, Z^{*,j-1}, \theta)\right)^{-1} - B^\dagger\right| \\ \leq\ &\sup_{t\geq R}\sup_{\theta\in\Theta}\left|\left(\frac{1}{t}\sum_{j=s}^{t}\nabla_\theta^2 q(y_j^*, Z^{*,j-1}, \theta)\right)^{-1} - \left(\frac{1}{t}\sum_{j=s}^{t}E^*\left(\nabla_\theta^2 q(y_j^*, Z^{*,j-1}, \theta)\right)\right)^{-1}\right| \\ &+\sup_{t\geq R}\sup_{\theta\in\Theta}\left|\left(\frac{1}{t}\sum_{j=s}^{t}E^*\left(\nabla_\theta^2 q(y_j^*, Z^{*,j-1}, \theta)\right)\right)^{-1} - B^\dagger\right|\end{aligned} \tag{17}$$

Given (16), and assumptions A1-A2, the second term on the RHS of (17) is $o_P(1)$. Recalling, that the resampled series consists of b independent and identically distributed blocks, and $b/T^{1/2} \to \infty$, the first term on on the RHS of (17) is $o_{P^*}(1)$ $\Pr -P$ because of the uniform law of large number for *iid* random variables. Thus,

$$
\begin{aligned}
& \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\widetilde{\theta}_t^* - \widehat{\theta}_t \right) \\
= \; & B^{\dagger} \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\frac{1}{t} \sum_{j=s}^{t} \left(\nabla_{\theta} q(y_j^*, Z^{*,j-1}, \widehat{\theta}_t) - \left(\frac{1}{T} \sum_{k=s}^{T-1} \nabla_{\theta} q(y_k, Z^{k-1}, \widehat{\theta}_t) \right) \right) \right) \\
& + o_{P^*}(1) \ \Pr -P \qquad (18)
\end{aligned}
$$

and, by taking a first order expansion of the RHS in (18) around $\theta^{\dagger}$,

$$
\begin{aligned}
& \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\widetilde{\theta}_t^* - \widehat{\theta}_t \right) \\
= \; & B^{\dagger} \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\frac{1}{t} \sum_{j=s}^{t} \left(\nabla_{\theta} q(y_j^*, Z^{*,j-1}, \theta^{\dagger}) - \left(\frac{1}{T} \sum_{k=s}^{T-1} \nabla_{\theta} q(y_k, Z^{k-1}, \theta^{\dagger}) \right) \right) \right) \\
& + B^{\dagger} \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\left(\frac{1}{t} \sum_{j=s}^{t} \left(\nabla_{\theta}^2 q(y_j^*, Z^{*,j-1}, \overline{\theta}_t) - \left(\frac{1}{T} \sum_{k=s}^{T-1} \nabla_{\theta}^2 q(y_k, Z^{k-1}, \overline{\theta}_t) \right) \right) \right) \right. \\
& \left. \times \left(\widehat{\theta}_t - \theta^{\dagger} \right) \right) + o_{P^*}(1) \ \Pr -P \qquad (19)
\end{aligned}
$$

We need to show that the second term on the RHS of (19) is $o_{P^*}(1)$ $\Pr -P$. Note that it is majorized by

$$
B^{\dagger} \sup_{t \geq R} t^{\vartheta} \left| \widehat{\theta}_t - \theta^{\dagger} \right| \sup_{t \geq R} \sup_{\theta \in \Theta} \frac{\sqrt{P}}{t^{1+\vartheta}} \left| \sum_{j=s}^{t} \left(\nabla_{\theta}^2 q(y_j^*, Z^{*,j-1}, \theta) - \left(\frac{1}{T} \sum_{k=s}^{T-1} \nabla_{\theta}^2 q(y_k, Z^{k-1}, \theta) \right) \right) \right|,
$$

with $1/3 < \vartheta < 1/2$. Recalling that $bl = T$ and $l = o(T^{1/4})$, it follows that $b/T^{3/4} \to \infty$, thus by the same argument used in Lemma 1(i) in Altissimo and Corradi (2002), given (16),

$$
\sup_{t \geq R} \sup_{\theta \in \Theta} \left| \frac{1}{t} \sum_{j=s}^{t} \left(\nabla_{\theta}^2 q(y_j^*, Z^{*,j-1}, \theta) - \left(\frac{1}{T} \sum_{k=s}^{T-1} \nabla_{\theta}^2 q(y_k, Z^{k-1}, \theta) \right) \right) \right| = O_{a.s.^*} \left(\sqrt{\frac{\log \log b}{b}} \right),
$$

Thus,

$$
\sup_{t \geq R} \sup_{\theta \in \Theta} \frac{\sqrt{P}}{t^{1+\vartheta}} \left| \sum_{j=s}^{t} \left(\nabla_{\theta}^2 q(y_j^*, Z^{*,j-1}, \theta) - \left(\frac{1}{T} \sum_{j=s}^{T} \nabla_{\theta}^2 q(y_j, Z^{j-1}, \theta) \right) \right) \right| = o_{P^*}(1)
$$

for $\vartheta > 1/3$. Finally, for all $\vartheta < 1/2$, $\sup_{t \geq R} t^{\vartheta} \left| \widehat{\theta}_t - \theta^{\dagger} \right| = o_P(1)$ by Lemma A3 in West (1996). Recalling that

$$
\frac{1}{t} \sum_{j=s}^{t} E^* \left(\nabla_{\theta} q(y_j^*, Z^{*,j-1}, \theta^{\dagger}) \right) = \frac{1}{T} \sum_{k=s}^{T-1} \nabla_{\theta} q(y_k, Z^{k-1}, \theta^{\dagger}) + O_P \left(\frac{l}{T} \right),
$$

the right hand side of (19) writes as:

$$
\begin{aligned}
& \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\widetilde{\theta}_t^* - \widehat{\theta}_t \right) \\
= \quad & B^\dagger \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\frac{1}{t} \sum_{j=s}^{t} \left(\nabla_\theta q(y_j^*, Z^{*,j-1}, \theta^\dagger) - E^* \left(\nabla_\theta q(y_j^*, Z^{*,j-1}, \theta^\dagger) \right) \right) \right) + o_{P^*}(1) \ \operatorname{Pr}-P \\
= \quad & B^\dagger \frac{a_{R,0}}{\sqrt{P}} \sum_{j=1}^{R} \left(\nabla_\theta q(y_j^*, Z^{*,j-1}, \theta^\dagger) - E^* \left(\nabla_\theta q(y_j^*, Z^{*,j-1}, \theta^\dagger) \right) \right) \\
& + B^\dagger \frac{1}{\sqrt{P}} \sum_{j=1}^{P-1} a_{R,j} \left(\nabla_\theta q(y_{R+j}^*, Z^{*,R+j-1}, \theta^\dagger) - E^* \left(\nabla_\theta q(y_{R+j}^*, Z^{*,R+j-1}, \theta^\dagger) \right) \right) \\
& + o_{P^*}(1) \ \operatorname{Pr}-P, \qquad (20)
\end{aligned}
$$

where $a_{R,j} = a_{R,i} = (R+i)^{-1} + ... + (R+P-1)^{-1}$, for $0 \leq i < P-1$. The first term of the second equality on the RHS of (20) follows directly from Lemma A5 in West (1996).

Now, $\frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\widetilde{\theta}_t^* - \widehat{\theta}_t \right)$ satisfies a central limit theorem for triangular independent arrays (see e.g. White and Wooldridge (1988)), as thus, conditionally on the sample, it converges in distribution to a zero mean normal random variable.

By Theorem 4.1 in West (1996),

$$\frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left(\widehat{\theta}_t - \theta^\dagger \right) \xrightarrow{d} N\left(0, 2\Pi B^\dagger C_{00} B^\dagger\right),$$

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 $\Pi = 1 - \pi^{-1} \ln(1+\pi)$. Therefore, the statement in the Theorem will follow once we have shown that

$$Var^* \left(\frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left(\widetilde{\theta}_t^* - \widehat{\theta}_t \right) \right) = 2\Pi B^\dagger C_{00} B^\dagger, \operatorname{Pr}-P \qquad (21)$$

For notational simplicity, let $\nabla_\theta q(y_j^*, Z^{*,j-1}, \theta^\dagger) = h_j^*$, and $\nabla_\theta q(y_j, Z^{j-1}, \theta^\dagger) = h_j$, and also let $\overline{h}_T = \frac{1}{T} \sum_{t=s}^{T} h_t$. Thus, given (20),

$$
\begin{aligned}
Var^* \left(\frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\widetilde{\theta}_t^* - \widehat{\theta}_t \right) \right) &= \frac{R}{P} Var^* \left(a_{R,0} \frac{1}{\sqrt{R}} \sum_{j=1}^{R} h_j^* \right) \\
+ \frac{1}{P} Var^* \left(\sum_{j=1}^{P-1} a_{R,j} h_{R+j}^* \right) &+ \frac{1}{P} Cov^* \left(a_{R,0} \sum_{j=1}^{R} h_j^*, \sum_{j=1}^{P-1} a_{R,j} h_{R+j}^* \right)
\end{aligned}
$$

As blocks are independent conditionally on the sample, the covariance term is equal to zero. Without loss of generality, set $R = b_1 l$ and $P = b_2 l$, where $b_1 + b_2 = b$. Thus up to a term of order $O(l/R^{1/2})$,

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

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

So,

$$\begin{aligned} &\frac{R}{P}Var^*\left(a_{R,0}\frac{1}{\sqrt{R}}\sum_{j=1}^{R}h_j^*\right) \\ = \ &\frac{Ra_{R,0}^2}{P}\sum_{j=-l}^{l}\gamma_j + \frac{Ra_{R,0}^2}{P}\left(\frac{1}{R}\sum_{t=l}^{R-l}\sum_{j=-l}^{l}((h_t-\overline{h}_T)(h_{t+j}-\overline{h}_T)'-\gamma_j)\right) + O\left(\frac{l^2}{R}\right), \end{aligned} \tag{22}$$

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

$$Var^*\left(\frac{1}{\sqrt{P}}\sum_{j=1}^{P-1}a_{R,j}h_{R+j}^*\right) = Var^*\left(\frac{1}{\sqrt{P}}\sum_{k=b_1+1}^{b_2}\sum_{i=1}^{l}a_{R,((k-1)l+i)}h_{I_k+i}\right)$$

$$\begin{aligned} &= \frac{1}{P}E^*\left(\sum_{k=b_1+1}^{b_2}\sum_{i=1}^{l}\sum_{j=1}^{l}a_{R,((k-b_1-1)l+i)}a_{R,((k-b_1-1)l+j)}(h_{I_k+i}-\overline{h}_T)(h_{I_k+j}-\overline{h}_T)'\right) \\ &= \frac{1}{P}\sum_{k=b_1+1}^{b_2}\sum_{i=1}^{l}\sum_{j=1}^{l}a_{R,((k-b_1-1)l+i)}a_{R,((k-b_1-1)l+j)}E^*\left((h_{I_k+i}-\overline{h}_T)(h_{I_k+j}-\overline{h}_T)'\right) \end{aligned}$$

$$\begin{aligned} &= \frac{1}{P}\sum_{k=b_1+1}^{b_2}\sum_{i=1}^{l}\sum_{j=1}^{l}a_{R,((k-b_1-1)l+i)}a_{R,((k-b_1-1)l+j)}\left(\frac{1}{T}\sum_{t=l}^{T-l}(h_{t+i}-\overline{h}_P)(h_{+t+j}-\overline{h}_P)'\right) + O(l/P^{1/2}) \ \Pr -P \\ &= \frac{1}{P}\sum_{k=b_1+1}^{b_2}\sum_{i=1}^{l}\sum_{j=1}^{l}a_{R,((k-b_1-1)l+i)}a_{R,((k-b_1-1)l+j)}\gamma_{i-j} \\ &\quad + \frac{1}{P}\sum_{k=b_1+1}^{b_2}\sum_{i=1}^{l}\sum_{j=1}^{l}a_{R,((k-b_1-1)l+i)}a_{R,((k-b_1-1)l+j)}\left(\frac{1}{T}\sum_{t=l}^{T-l}\left((h_{t+i}-\overline{h}_T)(h_{t+j}-\overline{h}_T)'-\gamma_{i-j}\right)\right) \\ &\quad + O(l/P^{1/2}) \ \Pr -P \end{aligned} \tag{23}$$

We need to show that the last term on the last equality in (23) is $o(1)$ Pr $-P$. First note that it is majorized by

$$\left|\frac{b_2}{P}\sum_{i=1}^{l}\sum_{j=1}^{l}\left(\frac{1}{T}\sum_{t=l}^{T-l}\left((h_{t+i}-\overline{h}_T)(h_{t+j}-\overline{h}_T)'-\gamma_{i-j}\right)\right)\right|$$

$$=\left|\frac{1}{T}\sum_{t=l}^{T-l}\sum_{j=-l}^{l}\left((h_t-\overline{h}_T)(h_{t+j}-\overline{h}_T)'-\gamma_j\right)\right|+O(l/P^{1/2})\ \text{Pr}-P \tag{24}$$

The first term on the RHS of (24) goes to zero in probability, by the same argument as in Lemma 2 in Corradi (1999)[12]. As for the first term on the RHS of the last equality in (23),

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

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

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

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

As the first term on the RHS of (22) converges to $\pi^{-1}\ln^2(1+\pi)C_{00}$ (see West (1996), p.1082), the desired outcome then follows.

Proof of Proposition 2: Let $\overline{u}_{i,t}=y_t-\kappa(Z^{t-1},\overline{\theta}_{i,t})$, with $\overline{\theta}_{i,t}\in\left(\widehat{\theta}_{i,t},\theta^{\dagger}\right)$. Via a mean value

[12] 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.

expansion, and given A1,A2

$$
\begin{aligned}
S_P(1,k) &= \frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(g(\widehat{u}_{1,t+1})-g(\widehat{u}_{k,t+1})\right) \\
&= \frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(g(u_{1,t+1})-g(u_{k,t+1})\right) \\
&\quad +\frac{1}{P}\sum_{t=R}^{T-1} g'(\overline{u}_{1,t+1})\nabla_{\theta_1}\kappa_1(Z^t,\overline{\theta}_{1,t})P^{1/2}\left(\widehat{\theta}_{1,t}-\theta_1^{\dagger}\right) \\
&\quad -\frac{1}{P}\sum_{t=R}^{T-1} g'(\overline{u}_{k,t+1})\nabla_{\theta_k}\kappa_k(Z^t,\overline{\theta}_{k,t})P^{1/2}\left(\widehat{\theta}_{k,t}-\theta_k^{\dagger}\right) \\
&= \frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(g(u_{1,t+1})-g(u_{k,t+1})\right) \\
&\quad +\mu_1\frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_{1,t}-\theta_1^{\dagger}\right)-\mu_k\frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_{k,t}-\theta_k^{\dagger}\right)+o_P(1),
\end{aligned}
$$

where $\mu_1 = E\left(g'(u_{1,t+1})\nabla_{\theta_1}\kappa_1(Z^t,\theta_1^{\dagger})\right)$, and μ_k is defined analogously. Now, when all competitors have the same predictive accuracy as the benchmark model, by the same argument as in Theorem 4.1 in West (1996),

$$(S_P^{\mu}(1,2),...,S_P^{\mu}(1,n)) \xrightarrow{d} N(0,V),$$

where $S_P^{\mu}(1,k) = S_P(1,k)-\sqrt{P}E\left(g(u_{1,t+1})-g(u_{k,t+1})\right)$, V is a $n\times n$ matrix with i,j element $v_{i,j} = S_{g_ig_j}+2\Pi C_{ii}+2\Pi C_{jj}-2\Pi C_{ij}$, where $C_{ij} = \sum_{j=-\infty}^{\infty} E\left(\left(\nabla_\theta q_i(y_{1+s},Z^s,\theta_i^{\dagger})\right)\left(\nabla_\theta q_j(y_{1+s+j},Z^{s+j},\theta_j^{\dagger})\right)'\right)$, $\Pi = 1-\pi^{-1}\ln(1+\pi)$. The distribution of S_P then follows straightforwardly from the continuous mapping theorem.

Proof of Proposition 3: Let $\widehat{u}^*_{i,t+1} = y^*_{t+1}-\kappa_i(Z^{*,t},\widehat{\theta}_{i,t})$ and $\overline{u}^*_{i,t+1} = y^*_{t+1}-\kappa_i(Z^{*,t},\overline{\theta}^*_{i,t})$, with

$\overline{\theta}^*_{i,t} \in \left(\widetilde{\theta}^*_{i,t}, \widehat{\theta}_{i,t}\right)$, $u^*_{i,t+1} = y^*_{t+1} - \kappa_i(Z^{*,t}, \theta^\dagger_i)$, $u_{i,t+1} = y_{t+1} - \kappa_i(Z^t, \theta^\dagger_i)$. We have:

$$\begin{aligned}
S^*_P(1,k) &= \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\left(g(\widetilde{u}^*_{1,t+1}) - g(\widehat{u}_{1,t+1}) \right) - \left(g(\widetilde{u}^*_{k,t+1}) - g(\widehat{u}_{k,t+1}) \right) \right) \\
&= \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\left(g(\widehat{u}^*_{1,t+1}) - g(\widehat{u}_{1,t+1}) \right) - \left(g(\widehat{u}^*_{k,t+1}) - g(\widehat{u}_{k,t+1}) \right) \right) \\
&\quad + \frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \left(\nabla_{\theta_1} g(\overline{u}^*_{1,t+1}) \left(\widehat{\theta}^*_{1,t} - \widehat{\theta}_{1,t} \right) - \nabla_{\theta_1} g(\overline{u}^*_{k,t+1}) \left(\widehat{\theta}^*_{k,t} - \widehat{\theta}_{k,t} \right) \right) \\
&= \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\left(g(u^*_{1,t+1}) - g(u_{1,t+1}) \right) - \left(g(u^*_{k,t+1}) - g(u_{k,t+1}) \right) \right) \\
&\quad + \frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \left(\nabla_{\theta_1} g(\overline{u}^*_{1,t+1}) \left(\widehat{\theta}^*_{1,t} - \widehat{\theta}_{1,t} \right) - \nabla_{\theta_1} g(\overline{u}^*_{k,t+1}) \left(\widehat{\theta}^*_{k,t} - \widehat{\theta}_{k,t} \right) \right) + o_P(1),
\end{aligned} \quad (25)$$

where the $o_P(1)$ term on the last line of (25), comes from a Taylor expansion around $\theta^\dagger_i$ and from the fact that

$$\frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\nabla_{\theta_i} g(\overline{\overline{u}}^*_{i,t+1}) - \nabla_{\theta_i} g(\overline{\overline{u}}_{i,t+1}) \right) \left(\widehat{\theta}_{i,t} - \theta^\dagger_i \right) = o_{P^*}(1) \ \Pr -P,$$

for $\overline{\overline{u}}^*_{i,t+1} = y^*_{t+1} - \kappa_i(Z^{*,t}, \overline{\theta}_{i,t})$ and $\overline{\overline{u}}_{i,t+1} = y_{t+1} - \kappa_i(Z^t, \overline{\theta}_{i,t})$, with $\overline{\theta}_{i,t} \in \left(\widehat{\theta}_{i,t} \theta^\dagger_i\right)$, by the same argument as in the proof of Theorem 1. Now, the first term on the RHS of the last equality in (25) has the same limiting distribution of $\frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\left(g(u_{1,t+1}) - g(u_{k,t+1}) \right) - E\left(g(u_{1,t+1}) - g(u_{k,t+1}) \right) \right)$ conditionally on the sample and for all sample but a subset of measure zero, by Theorem 3.5 in Künsch. Also, $+\frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \nabla_{\theta_i} g(\overline{u}^*_{i,t+1}) \left(\widehat{\theta}^*_{i,t} - \widehat{\theta}_{i,t} \right)$ has the same limiting distribution of $E\left(\nabla_{\theta_i} g(u_{i,t+1}) \right) \frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \left(\widehat{\theta}_{i,t} - \theta^\dagger_i \right)$, conditionally on the sample and for all sample but a subset of measure zero.

Proof of Proposition 4: From Theorem 1 in Corradi and Swanson (2002).

Proof of Proposition 5: Recall that $g = q_1$, also let $\widetilde{u}^*_{1,t+1} = y^*_{t+1} - \begin{pmatrix} 1 & y^*_t \end{pmatrix} \widetilde{\theta}^*_{1,t}$, $\widehat{u}^*_{1,t+1} = y^*_{t+1} - \begin{pmatrix} 1 & y^*_t \end{pmatrix} \widehat{\theta}_{1,t}$, $\overline{u}^*_{1,t+1} = y^*_{t+1} - \begin{pmatrix} 1 & y^*_t \end{pmatrix} \overline{\theta}^*_{1,t}$ where $\overline{\theta}^*_{1,t} \in \left(\widetilde{\theta}^*_{1,t}, \theta^\dagger_1\right)$, and $u^*_{1,t+1} = y^*_{t+1} -$

$\begin{pmatrix} 1 & y_t^* \end{pmatrix} \theta_{1,t}^\dagger$. Then,

$$\begin{aligned}
& \frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \Big(g'(\widetilde{u}_{1,t+1}^*) w(Z^{*,t},\gamma) - g'(\widehat{u}_{1,t+1}) w(Z^t,\gamma) \Big) \\
= & \frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \Big(g'(\widehat{u}_{1,t+1}^*) w(Z^{*,t},\gamma) - g'(\widehat{u}_{1,t+1}) w(Z^t,\gamma) \Big) \\
& + \frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \Big(\nabla_\theta g'(\overline{u}_{1,t+1}^*) w(Z^{*,t},\gamma) \Big) \Big(\widetilde{\theta}_{1,t}^* - \widehat{\theta}_{1,t} \Big) \\
= & \frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \Big(g'(u_{1,t+1}^*) w(Z^{*,t},\gamma) - g'(\widehat{u}_{1,t+1}) w(Z^t,\gamma) \Big) \\
& + \frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \Big(\nabla_\theta g'(\overline{u}_{1,t+1}^*) w(Z^{*,t},\gamma) \Big) \Big(\widetilde{\theta}_{1,t}^* - \widehat{\theta}_{1,t} \Big) + o_{P^*}(1) \ \ \mathrm{Pr} - P
\end{aligned}$$

Now, pointwise in γ, the first term on the RHS of the last equality in (**??**), 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 $\frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \Big(g'(u_{1,t+1}^*) w(Z^{*,t},\gamma) - g'(u_{1,t+1}) w(Z^t,\gamma) \Big)$ has the same limiting distribution of the same functional of $\frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \left(g'(u_{1,t+1}) w(Z^t,\gamma) - E\left(g'(u_{1,t+1}) w(Z^t,\gamma) \right) \right)$.

Finally, $\frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \Big(\nabla_\theta g'(\overline{\overline{u}}_{1,t+1}^*) w(Z^{*,t},\gamma) \Big) \Big(\widetilde{\theta}_{1,t}^* - \widehat{\theta}_{1,t} \Big)$ properly captures the contribution of recursive parameter estimation error to the covariance kernel.

7 References

Altissimo, F. and V. Corradi, (2002), Bounds for Inference with Nuisance Parameters Present only under the Alternative", *Econometrics Journal*, 5, 494-518.

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

Andrews, D.W.K., (2004), The Block-Block Bootstrap: Improved Asymptotic Refinements, *Econometrica,* 72, 673-700.

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.

Chao, J.C., V. Corradi, and N.R. Swanson (2001), An Out of Sample Test for Granger Causality", *Macroeconomic Dynamics,* v.5, 598-620.

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.*

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

Clark, T.E., and M.W., McCracken (2003), Evaluating Long-Horizon Forecasts, Working Paper, University of Missouri-Columbia

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

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

Corradi, V. and N.R. Swanson, (2004), Some Recent Developments in Predictive Accuracy Testing with Nested Models and (Generic) Nonlinear Alternatives", *International Journal of Forecasting*, 20, 185-199.

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

Diebold, F.X. and C. Chen, (1996), Testing Structural Stability with Endogenous Breakpoint: A Size Comparison of Analytic and Bootstrap Procedures, *Journal of Econometrics*, 70, 221-241.

Elliott, G. and A. Timmerman, (2004a), Optimal Forecast Combinations Under General Loss Functions and Forecast Error Distributions, *Journal of Econometrics,* forthcoming.

Elliott, G. and A. Timmerman, (2004b), Optimal Forecast Combination Under Regime Switching, *International Economic Review,* fothcoming.

Fitzenberger, B. (1997), The Moving Block Bootstrap and Robust Inference for Linear Least Square and Quantile Regressions, *Journal of Econometrics*, 82, 235-287.

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

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

Goncalves, S., and H. White, (2004), Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models, *Journal of Econometrics,* 119, 199-219.

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, P.R., (2004), An Unbiased and Powerful Test for Superior Predictive Ability, Manuscript, Brown University.

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

Inoue, A., and M. Shintani, (2004), Bootstrapping GMM Estimators for Time Series, *Journal of Econometrics,* forthcoming.

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., (2004), 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.