

Corradi, Valentina; Swanson, Norman R.

**Working Paper**

## A Test for Comparing Multiple Misspecified Conditional Distributions

Working Paper, No. 2003-14

**Provided in Cooperation with:**

Department of Economics, Rutgers University

*Suggested Citation:* Corradi, Valentina; Swanson, Norman R. (2003) : A Test for Comparing Multiple Misspecified Conditional Distributions, Working Paper, No. 2003-14, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:

<https://hdl.handle.net/10419/23171>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# A Test for Comparing Multiple Misspecified Conditional Distributions\*

Valentina Corradi  
University of Exeter

Norman R. Swanson  
Rutgers University

March 2003

## Abstract

This paper introduces a conditional Kolmogorov test, in the spirit of Andrews (1997), that allows for comparison of multiple misspecified conditional distribution models, for the case of dependent observations. A conditional confidence interval version of the test is also discussed. Model accuracy is measured using a distributional analog of mean square error, in which the squared (approximation) error associated with a given model, say model  $i$ , 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  $U$  is a possibly unbounded set on the real line,  $Z^t$  is the conditioning information set,  $F_i$  is the distribution function of a particular candidate model, and  $F_0$  is the true (unknown) distribution function. When comparing more than two models, a “benchmark” model is specified, and the test is constructed along the lines of the “reality check” of White (2000). Valid asymptotic critical values are obtained via a version of the block bootstrap which properly captures the effect of parameter estimation error. The results of a small Monte Carlo experiment indicate that the conditional confidence interval version of the test has reasonable finite sample properties even for samples with as few as 60 observations.

*JEL classification:* C22, C52.

*Keywords:* block bootstrap, conditional Kolmogorov test, data snooping, misspecified conditional distribution

---

\*Valentina Corradi, Department of Economics, University of Exeter, Exeter, EX4 4PU, U.K., v.corradi@exeter.ac.uk; and Norman R. Swanson, Department of Economics, Rutgers University, New Brunswick, NJ 08901-1248, U.S.A., nswanson@econ.rutgers.edu. We would like to thank Don Andrews, Russell Davidson, Clive Granger, Christelle Viaroux and seminar participants at the 2002 UK Econometrics Group meeting in Bristol, the 2002 European Econometric Society meetings, the 2002 University of Pennsylvania NSF-NBER time series conference, the 2002 EC<sup>2</sup> Conference in Bologna, Cornell University, the State University of New York at Stony Brook and the University of California at Davis for many helpful comments and suggestions on a previous version of this paper.

# 1 Introduction

There are several instances in which a “good” model for the conditional mean and/or variance is not adequate for the task at hand. For example, financial risk management involves tracking the entire distribution of a portfolio; or measuring certain distributional aspects, such as value at risk (see e.g. Duffie and Pan (1997)). In these cases, the choice of the best loss function specific model for the conditional mean may not be of too much help. Important contributions that go beyond the examination of models of conditional mean include assessing the correctness of conditional interval prediction (Christoffersen (1998)) and assessing volatility predictability by comparing unconditional and conditional interval forecasts (Christoffersen and Diebold (2000)).<sup>1</sup> Needless to say, correct specification of the conditional distribution implies correct specification of all conditional aspects of the model. Perhaps in part for this reason, there has been growing interest in recent years in providing tests for the correct specification of conditional distributions. One contribution in this direction is the conditional Kolmogorov (CK) test of Andrews (1997), which is based on the comparison of the empirical joint distribution of  $y_t$  and  $X_t$  with the product of a given distribution of  $y_t|X_t$  and the empirical CDF of  $X_t$ . Other contributions in this direction include, for example, Zheng (2000), who suggests a nonparametric test based on a first-order, linear, expansion of the Kullback Leibler Information Criterion (KLIC), and Li and Tkacz (2002), who propose a test based on the comparison of a nonparametric kernel estimate of the conditional density with the density implied under the null hypothesis.<sup>2</sup> Following a different route based on use of the probability integral transform, Diebold, Gunther and Tay (1998) suggest a simple and effective means by which predictive densities can be evaluated (see also Bai (2001), Diebold, Hahn and Tay (1999), and Hong (2001)). All of the papers cited above consider a null hypothesis of correct dynamic specification of the conditional distribution.<sup>3</sup> However, we would like to argue that as all models are arguably approximations of the truth, it is likely that they are all misspecified. For this reason, it is our objective in this paper to provide a conditional Kolmogorov test, in the

---

<sup>1</sup>Prediction confidence intervals are also discussed in Granger, White and Kamstra (1989), Chatfield (1993), Diebold, Tay and Wallis (1998), Clements and Taylor (2001), and the references cited therein.

<sup>2</sup>Whang (2000,2001) proposes a CK type test for the correct specification of the conditional mean.

<sup>3</sup>One exception is the approach taken by Corradi and Swanson (2003), who consider testing the null of correct specification of the conditional distribution for a given information set, thus allowing for dynamic misspecification under both hypotheses.

spirit of Andrews (1997), that allows for the joint comparison of multiple misspecified conditional distribution models, for the case of dependent observations. In particular, assume that the object of interest is the conditional distribution of a scalar,  $Y_t$ , given a (possibly vector valued) conditioning set,  $Z^t$ , where  $Z^t$  contains lags of  $Y_t$  and/or other variables. Now, given a group of (possibly) misspecified conditional distributions,  $F_1(u|Z^t, \theta_1^\dagger), \dots, F_m(u|Z^t, \theta_m^\dagger)$ , assume that the objective is to compare these models in terms of their closeness to the true conditional distribution,  $F_0(u|Z^t, \theta_0) = \Pr(Y_t \leq u|Z^t)$ . If  $m > 2$ , we follow White (2000). Namely, we choose a particular conditional distribution model as the “benchmark” and test the null hypothesis that no competing model can provide a more accurate approximation of the “true” conditional distribution, against the alternative that at least one competitor outperforms the benchmark model. Needless to say, pairwise comparison of alternative models, in which no benchmark need be specified, follows as a special case. In our context, accuracy is measured using a distributional analog of mean square error. More precisely, the squared (approximation) error associated with model  $i$ ,  $i = 1, \dots, m$ , 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. It should be pointed out that one well known measure of distributional accuracy is the Kullback-Leibler Information Criterion (KLIC), in the sense that the “most accurate” model can shown to be that which minimizes the KLIC (see Section 2 for a more precise discussion). For the iid case, Vuong (1989) suggests a likelihood ratio test for choosing the conditional density model closer to the “true” conditional density in terms of the KLIC. Additionally, Giacomini (2002) suggests a weighted version of the Vuong likelihood ratio test for the case of dependent observations, while Kitamura (2002) employs a KLIC based approach to select among misspecified conditional models that satisfy given moment conditions.<sup>4</sup> Furthermore, the KLIC approach has been recently employed for the evaluation of dynamic stochastic general equilibrium models (see e.g. Schorfheide (2000), Fernandez-Villaverde and Rubio-Ramirez (2001), and Chan, Gomes and Schorfheide (2002)). For example, Fernandez-Villaverde and Rubio-Ramirez (2001) show that the KLIC-best model is also the model with the highest posterior probability. In general, there is no reason why our measure of accuracy is more “natural” than the KLIC, or vice-versa. However, in the next section we outline how certain problems (such as comparing conditional confidence intervals) that are difficult to address using the KLIC can be handled quite

---

<sup>4</sup>Of note is that White (1982) shows that quasi maximum likelihood estimators (QMLEs) minimize the KLIC, under mild conditions.

easily using our measure of distributional accuracy.

The rest of the paper is organized as follows. Section 2 states the hypothesis of interest and describes the test statistic which will be examined in the sequel. In Section 3.1, it is shown that the limiting distribution of the statistic (properly recentered) is a functional of a zero mean Gaussian process with a covariance kernel that reflects both the contribution of parameter estimation error and the effect of misspecification. As a result, the limiting distribution is not nuisance parameter free and therefore critical values cannot be tabulated. Valid asymptotic critical values are instead constructed using an appropriate version of the block bootstrap. In particular, in Section 3.2 we show the first order validity of an empirical process version of the block bootstrap which properly captures the contribution of parameter estimation error to the covariance kernel associated with the limiting distribution of the test statistic. Finally, the results from a small Monte Carlo experiment are collected in Section 4, and concluding remarks are given in Section 5. Proofs of results stated in the text are outlined in the Appendix.

## 2 Set-Up and Test Statistics

Our objective is to form parametric conditional distributions for a scalar random variable,  $Y_t$ , given  $Z^t$ , where  $Z^t = (Y_{t-1}, \dots, Y_{t-s_1}, X_t, \dots, X_{t-s_2+1})$  with  $s_1, s_2$  finite, and to select among these. Define the group of conditional distribution models from which we want to make our selection as  $F_1(u|Z^t, \theta_1^\dagger), \dots, F_m(u|Z^t, \theta_m^\dagger)$ , and define the true conditional distribution as

$$F_0(u|Z^t, \theta_0) = \Pr(Y_t \leq u|Z^t).$$

Hereafter, assume that  $\theta_i^\dagger \in \Theta_i$ , where  $\Theta_i$  is a compact set in a finite dimensional Euclidean space, and let  $\theta_i^\dagger$  be the probability limit of a quasi maximum likelihood estimator (QMLE) of the parameters of the conditional distribution under model  $i$ . If model  $i$  is correctly specified, then  $\theta_i^\dagger = \theta_0$ . As mentioned in the introduction, accuracy is measured in terms of a distributional analog of mean square error. In particular, we say that model 1 is more accurate than model 2, if

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

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

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

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

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

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. The null and the alternative hypotheses are:

$$H_0 : \max_{k=2, \dots, m} \int_U E \left( \left( F_1(u|Z^t, \theta_1^\dagger) - F_0(u|Z^t, \theta_0) \right)^2 - \left( F_k(u|Z^t, \theta_k^\dagger) - F_0(u|Z^t, \theta_0) \right)^2 \right) \phi(u) du \leq 0 \quad (1)$$

versus

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

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  $U$ . If interest focuses on conditional confidence intervals, so that the objective is to “approximate”  $\Pr(\underline{u} \leq Y_t \leq \bar{u}|Z^t)$ , then the null and alternative hypotheses can

be stated as:

$$H'_0 : \max_{k=2,\dots,m} E \left( \left( \left( F_1(\bar{u}|Z^t, \theta_1^\dagger) - F_1(\underline{u}|Z^t, \theta_1^\dagger) \right) - \left( F_0(\bar{u}|Z^t, \theta_0) - F_0(\underline{u}|Z^t, \theta_0) \right) \right)^2 \right. \\ \left. - \left( \left( F_k(\bar{u}|Z^t, \theta_k^\dagger) - F_k(\underline{u}|Z^t, \theta_k^\dagger) \right) - \left( F_0(\bar{u}|Z^t, \theta_0) - F_0(\underline{u}|Z^t, \theta_0) \right) \right)^2 \right) \leq 0.$$

versus

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

Alternatively, if interest focuses on testing the null of equal accuracy of two 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 order to test  $H_0$  versus  $H_A$ , we propose the following statistic<sup>5</sup>:

$$Z_T = \max_{k=2,\dots,m} \int_U Z_{T,u}(1, k) \phi(u) du, \quad (3)$$

where

$$Z_{T,u}(1, k) = \frac{1}{\sqrt{T}} \sum_{t=s}^T \left( \left( 1\{Y_t \leq u\} - F_1(u|Z^t, \hat{\theta}_{1,T}) \right)^2 - \left( 1\{Y_t \leq u\} - F_k(u|Z^t, \hat{\theta}_{k,T}) \right)^2 \right), \quad (4)$$

with  $s = \max\{s_1, s_2\}$ ,

$$\hat{\theta}_{i,T} = \arg \max_{\theta_i \in \Theta_i} \frac{1}{T} \sum_{t=s}^T \ln f_i(Y_t|Z^t, \theta_i), \quad i = 1, \dots, m, \quad (5)$$

and

$$\theta_i^\dagger = \arg \max_{\theta_i \in \Theta_i} E(\ln f_i(Y_t|Z^t, \theta_i)), \quad i = 1, \dots, m,$$

---

<sup>5</sup>In the Monte Carlo experiment reported on below,  $Z_T$  is constructed as follows:

$$Z_T = \max_{k=2,\dots,m} \frac{1}{N_u} \sum_{i=1}^{N_u} \left( \frac{1}{\sqrt{T}} \sum_{t=s}^T \left( \left( 1\{Y_t \leq u_i\} - F_1(u_i|Z^t, \hat{\theta}_{1,T}) \right)^2 - \left( 1\{Y_t \leq u_i\} - F_k(u_i|Z^t, \hat{\theta}_{k,T}) \right)^2 \right) \right),$$

where  $N_u$  is taken sufficiently large.

where  $f_i(Y_t|Z^t, \theta_i)$  is the conditional density under model  $i$ . As  $f_i(\cdot|\cdot)$  does not in general coincide with the true conditional density,  $\hat{\theta}_{i,T}$  are QMLE, and  $\theta_i^\dagger \neq \theta_0$ . It is worth noting that the statistic in equation (3) is very easy to compute, as integration is only over  $U$ , regardless the dimensionality of  $Z^t$ .

The intuition behind equation (4) is very simple. First, note that for any given  $u$ ,  $E(1\{Y_t \leq u\}|Z^t) = \Pr(Y_t \leq u|Z^t) = F_0(u|Z^t, \theta_0)$ . Thus,  $1\{Y_t \leq u\} - F_i(u|Z^t, \theta_i^\dagger)$  can be interpreted as an “error” term associated with computation of the conditional expectation under  $F_i$ . Now, write the statistic in equation (4) as:

$$\begin{aligned} & \frac{1}{\sqrt{T}} \sum_{t=s}^T \left( \left( 1\{Y_t \leq u\} - F_1(u|Z^t, \hat{\theta}_{1,T}) \right)^2 - \mu_1^2(u) \right) - \left( \left( 1\{Y_t \leq u\} - F_k(u|Z^t, \hat{\theta}_{k,T}) \right)^2 - \mu_k^2(u) \right) \\ & + \frac{T-s}{\sqrt{T}} (\mu_1^2(u) - \mu_k^2(u)), \end{aligned} \quad (6)$$

where  $\mu_j^2(u) = E \left( \left( 1\{Y_t \leq u\} - F_j(u|Z^t, \theta_j^\dagger) \right)^2 \right)$ ,  $j = 1, \dots, m$ . In the appendix, it is shown that the first term in equation (6) weakly converges as a process on  $U$ . Also, for  $j = 1, \dots, m$ :

$$\begin{aligned} \mu_j^2(u) &= E \left( \left( 1\{Y_t \leq u\} - F_j(u|Z^t, \theta_j^\dagger) \right)^2 \right) \\ &= E \left( \left( (1\{Y_t \leq u\} - F_0(u|Z^t, \theta_0)) - (F_j(u|Z^t, \theta_j^\dagger) - F_0(u|Z^t, \theta_0)) \right)^2 \right) \\ &= E \left( (1\{Y_t \leq u\} - F_0(u|Z^t, \theta_0))^2 \right) + E \left( \left( F_j(u|Z^t, \theta_j^\dagger) - F_0(u|Z^t, \theta_0) \right)^2 \right), \end{aligned}$$

given that the expectation of the cross product is zero (which follows because  $1\{Y_t \leq u\} - F_0(u|Z^t, \theta_0)$  is uncorrelated with any measurable function of  $Z^t$ ). Therefore,

$$\mu_1^2(u) - \mu_k^2(u) = E \left( \left( F_1(u|Z^t, \theta_1^\dagger) - F_0(u|Z^t, \theta_0) \right)^2 \right) - E \left( \left( F_k(u|Z^t, \theta_k^\dagger) - F_0(u|Z^t, \theta_0) \right)^2 \right). \quad (7)$$

Before outlining the asymptotic properties of the statistic in (3) a few comments are worth making. First, following the reality check approach of White (2000), the problem of testing multiple hypotheses has been reduced to a single comparison test by applying the (single valued) max function to the multiple hypotheses. This approach has the advantage that it avoids sequential testing bias and also captures the correlation across the various models. On the other hand, if we reject the null, we can conclude that there is at least one model that outperforms the benchmark, but we do not have available to us a complete picture concerning which model(s) contribute to the rejection of the null. Of course, some information can be obtained by looking at the distributional



analog of mean square error associated with the various models, and forming a crude ranking of the models, although the usual cautions associated with using a MSE type measure to rank models should be taken. Alternatively, our approach can be complemented by a multiple comparison approach, such as the false discovery rate (FDR) approach of Benjamini and Hochberg (1995), which allows one to select among alternative groups of models, in the sense that one can assess which group(s) contribute to the rejection of the null. The FDR approach has the objective of controlling the expected number of false rejections and in practice one computes p-values associated with the  $m$  hypotheses and orders these p-values in increasing fashion, say  $P_1 \leq \dots \leq P_i \leq \dots \leq P_m$ . Then, all hypotheses characterized by  $P_i \leq (1 - (i - 1)/m)\alpha$  are rejected, where  $\alpha$  is a given significance level. Such an approach, though less conservative than Hochberg's (1988) approach, is still conservative as it provides bounds on p-values. Overall, we think that a sound practical strategy could be to first implement our reality check type tests. These tests can then be complemented by using a multiple comparison approach, yielding a better overall understanding concerning which model(s) contribute to the rejection of the null, if it is indeed rejected. If the null is not rejected, then we simply choose the benchmark model. Nevertheless, even in this case, it may not hurt to see whether some of the individual hypotheses in our joint null hypothesis are rejected via a multiple test comparison approach.

### 3 Asymptotic Results

The results stated below require the following assumption.

**Assumption A:** (i)  $(Y_t, X_t)$ , is a strictly stationary and absolutely regular  $\beta$ -mixing process with size  $-4$ , for  $i = 1, \dots, m$ ; (ii)  $F_i(u|Z^t, \theta_i)$  is continuously differentiable on the interior of  $\Theta_i$ , where  $\Theta_i$  is a compact set in  $\mathbb{R}^{p_i}$ , and  $\nabla_{\theta_i} F_i(u|Z^t, \theta_i^\dagger)$  is  $2r$ -dominated on  $\Theta_i$ , uniformly in  $u$ ,  $r > 2$ ;<sup>6</sup> (iii)  $\theta_i^\dagger$  is uniquely identified (i.e.  $E(\ln f_i(Y_t|Z^t, \theta_i^\dagger)) > E(\ln f_i(Y_t|Z^t, \theta_i))$ , for any  $\theta_i \neq \theta_i^\dagger$ ); (iv) the density,  $f_i$ , associated with  $F_i$ , is twice continuously differentiable on the interior of  $\Theta_i$ ,  $\nabla_{\theta_i} \ln f_i(Y_t|Z^t, \theta_i)$  and  $\nabla_{\theta_i}^2 \ln f_i(Y_t|Z^t, \theta_i)$  are  $2r$ -dominated on  $\Theta_i$ , with  $r > 2$ ; (v)  $E(-\nabla_{\theta_i}^2 \ln f_i(Y_t|Z^t, \theta_i))$  is negative definite, uniformly on  $\Theta_i$ ,  $\lim_{T \rightarrow \infty} \text{Var}(\nabla_{\theta_i} \ln f_i(Y_t|Z^t, \theta_i))$  is positive definite; and (vi) let  $v_{kk} =$

---

<sup>6</sup>We say that  $\nabla_{\theta_i} F(u|Z^t, \theta_i)$  is  $2r$ -dominated on  $\Theta_i$  uniformly in  $u$ , if its  $k^{th}$ -element,  $k = 1, \dots, p_i$ , is such that  $|\nabla_{\theta_i} F_i(u|Z^t, \theta_i)|_k \leq D_t(u)$ , and  $\sup_{u \in R} E(|D_t(u)|^{2r}) < \infty$ . For more details on domination conditions, see Gallant and White (1988, pp. 33).

$\lim_{T \rightarrow \infty} Var \left( \frac{1}{\sqrt{T}} \sum_{t=s}^T \left( \left( 1\{Y_t \leq u\} - F_1(u|Z^t, \theta_1^\dagger) \right)^2 - \mu_1^2(u) \right) - \left( \left( 1\{Y_t \leq u\} - F_k(u|Z^t, \theta_k^\dagger) \right)^2 - \mu_k^2(u) \right) \right) \right)$ ,  
 $k = 2, \dots, m$ , define analogous covariance terms,  $v_{jk}$ ,  $j, k = 2, \dots, m$ , and assume that  $COV = [v_{jk}]$  is positive semi-definite.

Recalling that  $Z^t = (Y_{t-1}, \dots, Y_{t-s_1}, X_t, \dots, X_{t-s_2+1})$ , A1(i) ensures that  $Z^t$  is strictly stationary mixing with size  $-4$ . Note that A(vi) requires that at least one of the competing models to be neither nested in nor nesting the benchmark model. However assumption A(vi) can be relaxed, in which case the limiting distribution of the test statistic takes exactly the same form as given in Theorem 1 below, except that covariance kernel contains only terms which reflect parameter estimation error.<sup>7</sup> This is the case considered in the Monte Carlo section.

### 3.1 Limiting Distributions

**Theorem 1:** Let Assumption A hold. Then:

$$\max_{k=2, \dots, m} \int_U \left( Z_{T,u}(1, k) - \sqrt{T} (\mu_1^2(u) - \mu_k^2(u)) \right) \phi(u) du \xrightarrow{d} \max_{k=2, \dots, m} \int_U Z_{1,k}(u) \phi(u) du,$$

where  $Z_{1,k}(u, v)$  is a zero mean Gaussian process with covariance  $C_k(u, u')$  equal to<sup>8</sup>:

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

---

<sup>7</sup>In particular, using the notation of Theorem 1 below,  $C_k(u, u') =$

$$\begin{aligned} & 4m_{\theta_1^\dagger}(u)' A(\theta_1^\dagger) E \left( \sum_{j=-\infty}^{\infty} \nabla_{\theta_1} \ln f_1(Y_s|Z^s, \theta_1^\dagger) \nabla_{\theta_1} \ln f_1(Y_{s+j}|Z^{s+j}, \theta_1^\dagger)' \right) A(\theta_1^\dagger) m_{\theta_1^\dagger}(u') \\ & + 4m_{\theta_k^\dagger}(u)' A(\theta_k^\dagger) E \left( \sum_{j=-\infty}^{\infty} \nabla_{\theta_k} \ln f_k(Y_s|Z^s, \theta_k^\dagger) \nabla_{\theta_k} \ln f_k(Y_{s+j}|Z^{s+j}, \theta_k^\dagger)' \right) A(\theta_k^\dagger) m_{\theta_k^\dagger}(u') \\ & - 8m_{\theta_1^\dagger}(u)' A(\theta_1^\dagger) E \left( \sum_{j=-\infty}^{\infty} \nabla_{\theta_1} \ln f_1(Y_s|Z^s, \theta_1^\dagger) \nabla_{\theta_k} \ln f_k(Y_{s+j}|Z^{s+j}, \theta_k^\dagger)' \right) A(\theta_k^\dagger) m_{\theta_k^\dagger}(u'). \end{aligned}$$

<sup>8</sup>Note that the recentered statistics is actually

$$\max_{k=2, \dots, m} \int_U \left( Z_{T,u}(1, k) - \frac{T-s}{\sqrt{T}} (\mu_1^2(u) - \mu_k^2(u)) \right) \phi_U(u) du,$$

however, because of notational simplicity, and given that the two are asymptotically equivalent, we "approximate"  $\frac{T-s}{\sqrt{T}}$  with  $\sqrt{T}$ , both in the text and in the appendix.

$$\begin{aligned}
& +E \sum_{j=-\infty}^{\infty} \left( \left( 1\{Y_s \leq u\} - F_k(u|Z^s, \theta_k^\dagger) \right)^2 - \mu_k^2(u) \right) \left( \left( 1\{Y_{s+j} \leq u'\} - F_k(u'|Z^{s+j}, \theta_k^\dagger) \right)^2 - \mu_k^2(u') \right) \\
& -2E \sum_{j=-\infty}^{\infty} \left( \left( 1\{Y_s \leq u\} - F_1(u|Z^s, \theta_1^\dagger) \right)^2 - \mu_1^2(u) \right) \left( \left( 1\{Y_{s+j} \leq u'\} - F_k(u'|Z^{s+j}, \theta_k^\dagger) \right)^2 - \mu_k^2(u') \right) \\
& +4m_{\theta_1^\dagger}(u)' A(\theta_1^\dagger) E \left( \sum_{j=-\infty}^{\infty} \nabla_{\theta_1} \ln f_1(Y_s|Z^s, \theta_1^\dagger) \nabla_{\theta_1} \ln f_1(Y_{s+j}|Z^{s+j}, \theta_1^\dagger)' \right) A(\theta_1^\dagger) m_{\theta_1^\dagger}(u') \\
& +4m_{\theta_k^\dagger}(u)' A(\theta_k^\dagger) E \left( \sum_{j=-\infty}^{\infty} \nabla_{\theta_k} \ln f_k(Y_s|Z^s, \theta_k^\dagger) \nabla_{\theta_k} \ln f_k(Y_{s+j}|Z^{s+j}, \theta_k^\dagger)' \right) A(\theta_k^\dagger) m_{\theta_k^\dagger}(u') \\
& -8m_{\theta_1^\dagger}(u,)' A(\theta_1^\dagger) E \left( \sum_{j=-\infty}^{\infty} \nabla_{\theta_1} \ln f_1(Y_s|Z^s, \theta_1^\dagger) \nabla_{\theta_k} \ln f_k(Y_{s+j}|Z^{s+j}, \theta_k^\dagger)' \right) A(\theta_k^\dagger) m_{\theta_k^\dagger}(u') \\
& -4m_{\theta_1^\dagger}(u)' A(\theta_1^\dagger) E \left( \sum_{j=-\infty}^{\infty} \nabla_{\theta_1} \ln f_1(Y_s|Z^s, \theta_1^\dagger) \left( \left( 1\{Y_{s+j} \leq u\} - F_1(u|Z^{s+j}, \theta_1^\dagger) \right)^2 - \mu_1^2(u) \right) \right) \\
& +4m_{\theta_1^\dagger}(u)' A(\theta_1^\dagger) E \left( \sum_{j=-\infty}^{\infty} \nabla_{\theta_1} \ln f_1(Y_s|Z^s, \theta_1^\dagger) \left( \left( 1\{Y_{s+j} \leq u\} - F_k(u|Z^{s+j}, \theta_k^\dagger) \right)^2 - \mu_k^2(u) \right) \right) \\
& -4m_{\theta_k^\dagger}(u)' A(\theta_k^\dagger) E \left( \sum_{j=-\infty}^{\infty} \nabla_{\theta_k} \ln f_k(Y_s|Z^s, \theta_k^\dagger) \left( \left( 1\{Y_{s+j} \leq u\} - F_k(u|Z^{s+j}, \theta_k^\dagger) \right)^2 - \mu_k^2(u) \right) \right) \\
& +4m_{\theta_k^\dagger}(u)' A(\theta_k^\dagger) E \left( \sum_{j=-\infty}^{\infty} \nabla_{\theta_k} \ln f_k(Y_s|Z^s, \theta_k^\dagger) \left( \left( 1\{Y_{s+j} \leq u\} - F_1(u|Z^{s+j}, \theta_1^\dagger) \right)^2 - \mu_1^2(u) \right) \right) \quad (8)
\end{aligned}$$

with  $m_{\theta_i^\dagger}(u)' = E \left( \nabla_{\theta_i} F_i(u|Z^t, \theta_i^\dagger)' \left( 1\{Y_t \leq u\} - F_i(u|Z^t, \theta_i^\dagger) \right) \right)$  and  $A(\theta_i^\dagger) = \left( E \left( -\ln \nabla_{\theta_i}^2 f_i(y_t|Z^t, \theta_i^\dagger) \right) \right)^{-1}$ .

From Theorem 1, 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  $\int_U (\mu_1^2(u) - \mu_k^2(u)) \phi(u) du = 0, \forall k$ ), then the limiting distribution is a functional of a zero mean Gaussian process with a covariance kernel that reflects both the contribution of parameter estimation error and the fact that  $\left( \left( 1\{y_t \leq u'\} - F_i(u'|Z^t, \theta_i^\dagger) \right)^2 - \mu_i^2(u') \right)$  is not a martingale difference sequence in the presence of misspecification. Therefore, the limiting

distribution is not nuisance parameter free and critical values cannot be tabulated. Additionally, when all competitor models are worse than the benchmark, the statistic diverges to minus infinity at rate  $\sqrt{T}$ . Finally, when only some competitor models are worse than the benchmark, the limiting distribution provides a conservative test, as  $Z_T$  will always be smaller than  $\max_{k=2,\dots,m} \int_U \left( Z_{T,u}(1, k) - \sqrt{T} (\mu_1^2(u) - \mu_k^2(u)) \right) \phi(u) du$ , asymptotically, and therefore the critical values of  $\max_{k=2,\dots,m} \int_U Z_{1,k}(u) \phi_U(u) du$  provide upper bounds for the critical values of  $\max_{k=2,\dots,m} \int_U Z_{1,k}(u) \phi_U(u) du$ . Of course, when  $H_A$  holds, the statistic diverges to plus infinity at rate  $\sqrt{T}$ .

### 3.2 Bootstrap Critical Values

In this subsection we outline how to obtain valid critical values for the asymptotic distribution of  $\max_{k=2,\dots,m} \int_U \left( Z_{T,u}(1, k) - \sqrt{T} (\mu_1^2(u) - \mu_k^2(u)) \right) \phi_U(u) du$ , via use of an empirical process version of the block bootstrap that properly captures the contribution of parameter estimation error to the covariance kernel.

In order to show the first order validity of the bootstrap, we shall obtain the limiting distribution of the bootstrapped statistic and show that it coincides with the limiting distribution in Theorem 1. As all candidate models are potentially misspecified under both hypotheses, the parametric bootstrap is not generally applicable in our context. In fact, if observations are resampled from one of the candidate models, then we cannot ensure that the resampled statistic has the limiting distribution described in Theorem 1. Our approach is thus to establish the first order validity of the block bootstrap in the presence of parameter estimation error, using the results of Goncalves and White (2002a and 2002b) for block bootstrap QMLE estimators.<sup>9</sup> Let  $W_t = (Y_t, Z^t)$ . Draw  $b$  overlapping blocks of length  $l$  from  $W_s, \dots, W_T$ , where  $s = \max\{s_1, s_2\}$ , so that  $bl = T - s$ . Thus,  $W_s^*, \dots, W_{s+l}^*, \dots, W_{T-l+1}^*, \dots, W_T^*$  is equal to  $W_{I_1+1}, \dots, W_{I_1+l}, \dots, W_{I_b+1}, \dots, W_{I_b+l}$ , where  $I_i$ ,  $i = 1, \dots, b$  are identically and independently distributed discrete uniform random variates on  $s - 1, s, \dots, T - l$ . It follows that, conditional on the sample, the pseudo time series  $W_t^*$ ,  $t = s, \dots, T$ , consists of  $b$  independent and identically distributed blocks of length  $l$ . We begin by considering the bootstrap analog of  $Z_T$ . Define the block bootstrap QMLE as,

$$\hat{\theta}_{i,T}^* = \arg \max_{\theta_i \in \Theta_i} \frac{1}{T} \sum_{t=s}^T \ln f_i(Y_t^* | Z^{*t}, \theta_i), \quad i = 1, \dots, m,$$

---

<sup>9</sup>Goncalves and White (2002a,b) indeed consider the more general case of heterogeneous and near epoch dependent observations.

and define the bootstrap statistic as:

$$Z_T^* = \max_{k=2,\dots,m} \int_U Z_{T,u}^*(1, k) \phi(u) du,$$

where

$$\begin{aligned} Z_{T,u}^*(1, k) &= \frac{1}{\sqrt{T}} \sum_{t=s}^T \left( \left( 1\{Y_t^* \leq u\} - F_1(u|Z^{*t}, \hat{\theta}_{1,T}^*) \right)^2 - \left( 1\{Y_t \leq u\} - F_1(u|Z^t, \hat{\theta}_{1,T}) \right)^2 \right) \\ &\quad - \left( \left( 1\{Y_t^* \leq u\} - F_k(u|Z^{*t}, \hat{\theta}_{k,T}^*) \right)^2 - \left( 1\{Y_t \leq u\} - F_k(u|Z^t, \hat{\theta}_{k,T}) \right)^2 \right). \end{aligned}$$

**Theorem 2:** Let Assumption A hold. If  $l \rightarrow \infty$  and  $l/T^{1/2} \rightarrow 0$ , as  $T \rightarrow \infty$ , then,

$$\begin{aligned} &P \left( \omega : \sup_{v \in \mathbb{R}} \left| P^* \left( \max_{k=2,\dots,m} \int_U Z_{T,u}^*(1, k) \phi(u) du \leq v \right) \right. \right. \\ &\quad \left. \left. - P \left( \max_{k=2,\dots,m} \int_U \left( Z_{T,u}(1, k) - \sqrt{T} (\mu_1^2(u) - \mu_k^2(u)) \right) \phi(u) du \leq v \right) \right| > \varepsilon \right) \rightarrow 0, \end{aligned}$$

where  $P^*$  denotes the probability law of the resampled series, conditional on the sample, and  $\mu_1^2(u) - \mu_k^2(u)$  is defined as in equation (7).

The above result suggests proceeding in the following manner. For any bootstrap replication, compute the bootstrap statistic,  $Z_T^*$ . Perform  $B$  bootstrap replications ( $B$  large) and compute the quantiles of the empirical distribution of the  $B$  bootstrap statistics. Reject  $H_0$  if  $Z_T$  is greater than the  $(1 - \alpha)th$ -quantile. Otherwise, do not reject. Now, for all samples except a set with probability measure approaching zero,  $Z_T$  has the same limiting distribution as the corresponding bootstrapped statistic, when  $\int_U (\mu_1^2(u) - \mu_k^2(u)) \phi(u) du = 0, \forall k$ , which is the least favorable case under the null hypothesis. Thus, the above approach ensures that the test has asymptotic size equal to  $\alpha$ . On the other hand, when one or more, but not all, competing models are strictly dominated by the benchmark, the approach above ensures that the test has an asymptotic size between 0 and  $\alpha$ . When all models are dominated by the benchmark, the statistic vanishes to minus infinity, so that the rule above implies zero asymptotic size. Finally, under the alternative,  $Z_T$  diverges to (plus) infinity, while the corresponding bootstrap statistic has a well defined limiting distribution. This ensures unit asymptotic power. From the above discussion, we see that the bootstrap distribution provides correct asymptotic critical values only for the least favorable case under the null hypothesis; that is, when all competitor models are as good as the benchmark model. When  $\max_{k=2,\dots,m} \int_U (\mu_1^2(u) - \mu_k^2(u)) \phi(u) du = 0$ , but  $\int_U (\mu_1^2(u) - \mu_k^2(u)) \phi(u) du < 0$  for

some  $k$ , then the bootstrap critical values lead to conservative inference. An alternative to our bootstrap critical values in this case is the construction of critical values based on subsampling (see e.g. Politis, Romano and Wolf (1999), Ch.3). Heuristically, construct  $T - 2b_T$  statistics using subsamples of length  $b_T$ , where  $b_T/T \rightarrow 0$ . The empirical distribution of these statistics computed over the various subsamples properly mimics the distribution of the statistic. Thus, subsampling provides valid critical values even for the case where  $\max_{k=2,\dots,m} \int_U (\mu_1^2(u) - \mu_k^2(u)) \phi(u) du = 0$ , but  $\int_U (\mu_1^2(u) - \mu_k^2(u)) \phi(u) du < 0$  for some  $k$ . This is the approach used by Linton, Maasoumi and Whang (2003), for example, in the context of testing for stochastic dominance. Needless to say, one problem with subsampling is that unless the sample is very large, the empirical distribution of the subsampled statistics may yield a poor approximation of the limiting distribution of the statistic. An alternative approach for addressing the conservative nature of our bootstrap critical values is suggested in Hansen (2001). Hansen's idea is to recenter the bootstrap statistics using the sample mean, whenever the latter is larger than (minus) a bound of order  $\sqrt{2T \log \log T}$ . Otherwise, do not recenter the bootstrap statistics. In the current context, his approach leads to correctly sized inference when  $\max_{k=2,\dots,m} \int_U (\mu_1^2(u) - \mu_k^2(u)) \phi(u) du = 0$ , but  $\int_U (\mu_1^2(u) - \mu_k^2(u)) \phi(u) du < 0$  for some  $k$ . Additionally, his approach has the feature that if all models are characterized by a sample mean below the bound, the null is "accepted" and no bootstrap statistic is constructed.

## 4 Monte Carlo Findings

In this section we carry out a small series of Monte Carlo experiments using data generated as follows:

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_T \\ X \\ W \\ Q \end{pmatrix} \sim t_v \left( \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0_T \\ 0_T \\ 0_T \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2 \rho & \dots & \sigma^2 \rho^{T-1} \\ \sigma^2 \rho & \sigma^2 & \dots & \sigma^2 \rho^{T-2} \\ \dots & \dots & \sigma^2 & \dots \\ \sigma^2 \rho^{T-1} & \sigma^2 \rho^{T-2} & \dots & \sigma^2 \end{pmatrix}, \begin{pmatrix} I_T \cdot 0 & I_T \cdot 0 & I_T \cdot 0 \\ I_T \cdot \sigma_X^2 & I_T \cdot 0 & I_T \cdot 0 \\ I_T \cdot 0 & I_T \cdot \sigma_W^2 & I_T \cdot 0 \\ I_T \cdot 0 & I_T \cdot 0 & I_T \cdot \sigma_Q^2 \end{pmatrix} \right),$$

where  $t_v$  denotes a Student's t-distribution with  $v$  degrees of freedom,  $T$  is the sample size,  $y_t$ ,  $t = 1, \dots, T$  is the scalar random variable of interest (called  $Y_t$  above),  $X = (x_1, \dots, x_T)'$ ,  $W = (w_1, \dots, w_T)'$ , and  $Q = (q_1, \dots, q_T)'$  are  $T \times 1$  vector random variables to be used in the conditioning

sets of the competing models of  $y_t$ ,  $0_T$  is a  $T \times 1$  vector of zeros,  $I_T$  is a  $T \times T$  identity matrix, and  $\sigma^2$ ,  $\sigma_X^2$ ,  $\sigma_W^2$ ,  $\sigma_Q^2$ , and  $\rho$  are scalars. It follows that:

$$y_t|y_{t-1} \sim t_v \left( \alpha y_{t-1}, \frac{v}{v-1} (\sigma^2 - \sigma^2 \alpha^2) \left( 1 + \frac{y_{t-1}^2}{\sigma^2} \right) \right),$$

where  $\alpha = \text{cov}(y_t, y_{t-1}) / \text{var}(y_t, y_{t-1}) = \rho$ , for any  $t > 1$ . In our experiments, we impose misspecification upon all models by assuming normality (i.e. assume that  $F_i$ ,  $i = 1, \dots, m$ , is the normal CDF). Our objective is to ascertain whether a given benchmark model is “better”, in the sense of having lower squared approximation error, than two given alternative models. Thus,  $m = 3$ . Level and power experiments are defined by adjusting the conditioning information sets used to estimate (via QMLE) the parameters of each conditional model, and subsequently to form  $F_i(u|Z^t, \hat{\theta}_{i,T})$ ,  $F_i(u|Z^{*t}, \hat{\theta}_{i,T}^*)$ ,  $Z_T$ , and  $Z_T^*$ . In all experiments, values of  $\alpha = \{0.4, 0.6, 0.8, 0.9\}$  are used, samples of  $T = 60$  and  $120$  are tried,  $v = 3$ ,  $\sigma^2 = 1$ , and  $\sigma_X^2 = \sigma_W^2 = \sigma_Q^2 = \{0.1, 1.0, 10.0\}$ . Throughout, the conditional confidence interval version of the test is constructed, and the upper and lower bounds of the interval are fixed at  $\mu_Y + \gamma\sigma_Y$  and  $\mu_Y - \gamma\sigma_Y$ , respectively, where  $\mu_Y$  and  $\sigma_Y$  are the mean and variance of  $y_t$ , and where  $\gamma = \{\frac{1}{16}, \frac{1}{8}, \frac{1}{2}\}$ . Additionally, 5% and 10% nominal level bootstrap critical values are constructed using 100 bootstrap replications, block lengths of  $l = \{2, 3, 5, 6\}$  are tried, and all reported rejection frequencies are based on 5000 Monte Carlo simulations.<sup>10</sup> Given  $Z^t = (y_{t-1}, x_t, w_t, q_t)$ , the experiments reported on are organized as follows:

*Empirical Level Experiments:* In these experiments, we define the conditioning variable sets as follows: For the benchmark model ( $F_1$ ), use  $\tilde{Z}^t = (y_{t-1}, x_t)$ , where  $\tilde{Z}^t$  is a proper subset of  $Z^t$ . For the two alternative models ( $F_2$  and  $F_3$ ) we set  $\tilde{Z}^t = (y_{t-1}, w_t)$  and  $\tilde{Z}^t = (y_{t-1}, q_t)$ , respectively. In this case, the estimated coefficients associated with  $x_t$ ,  $w_t$ , and  $q_t$  have probability limits equal to zero, as none of these variables enters into the true conditional mean function. In addition, all models are misspecified, as conditional normality is assumed throughout. Therefore, the benchmark and the two competitors are equally misspecified. Finally, the limiting distribution of the test statistic in this case is driven by parameter estimation error, as assumption A(vi) does not hold.

*Empirical Power Experiments:* In these experiments, we set the conditioning variable sets as follows: For the benchmark model ( $F_1$ ),  $\tilde{Z}^t = (w_t)$ . For the two alternative models ( $F_2$  and  $F_3$ ) we set  $\tilde{Z}^t = (y_{t-1})$  and  $\tilde{Z}^t = (q_t)$ , respectively. In this manner, it is ensured that the first of the two

---

<sup>10</sup> Additional results for cases where  $x = \{\frac{1}{4}, 1\}$ ,  $l = \{10, 12\}$ , and critical values are constructed using 250 bootstrap replications are available upon request, and yield qualitatively similar results to those reported in Tables 1-6..

alternative models has smaller squared approximation error than the benchmark model.

Our findings are summarized in Tables 1-3 (empirical level experiments) and Tables 4-6 (empirical power experiments). In these tables, the first column reports the value of  $\rho = \alpha$  used in a particular experiment, while all remaining entries are rejection frequencies of the null hypothesis that the benchmark model is not outperformed by any of the alternative models. A number of conclusions emerge upon inspection of the tables. Turning first to the empirical level results given in Tables 1-3, note, for example, that empirical level varies from values grossly above nominal levels (when the block length is 2), to values slightly below nominal levels (when the block length is 6). Additionally, when  $l = 2$ , it is often the case that moving from 60 to 120 observations results in a slight increase in size, although the larger sample results in smaller rejection frequencies when  $l$  is larger, in many cases. In particular, notice that when  $\alpha = 0.4$  (low persistence) a block length of two usually suffices to capture the dependence structure of the series, while for  $\alpha = 0.9$  (high persistence) a larger block length is necessary. Finally, in many cases, the empirical rejection frequencies are not too distant from nominal levels, a result which is somewhat surprising given the small sample sizes used in our experiments. Empirical power results reported on in Tables 4-6 yield similar conclusions. In particular, rejection frequencies are sometimes rather low (when the level of dependence in the data, as given by  $\alpha$ , is smallest), but rejection frequencies are above 0.5 in many cases, and, broadly speaking, they increase with sample size and with  $\alpha$ . In summary, even when very small samples are used, the conditional confidence interval version of the test appears to perform adequately well, and our evidence points to reasonably steep tradeoff between sample size and test performance.

## 5 Concluding Remarks

We have provided a conditional Kolmogorov test, in the spirit of Andrews (1997), that allows for the joint comparison of multiple misspecified conditional distribution models, for the case of dependent observations, and for the case where accuracy is measured using a distributional analog of mean square error. We also outlined the construction of valid asymptotic critical values based on an empirical process version of the block bootstrap. Findings from a small Monte Carlo study of a conditional confidence interval version of the test indicate that the statistic has reasonable finite sample properties even for samples with as few as 60 observations.



## 6 Appendix

**Proof of Theorem 1:**

$$\begin{aligned}
& \text{Recall that } \mu_i^2(u) = E \left( \left( 1\{Y_t \leq u\} - F_i(u|Z^t, \theta_i^\dagger) \right)^2 \right) \\
& = E \left( \left( 1\{Y_t \leq u\} - F_0(u|Z^t, \theta_0) \right)^2 \right) + E \left( \left( F_0(u|Z^t, \theta_0) - F_i(u|Z^t, \theta_i^\dagger) \right)^2 \right). \text{ Thus,} \\
& Z_{T,u}(1, k) = \frac{1}{\sqrt{T}} \sum_{t=s}^T \left( \left( 1\{Y_t \leq u\} - F_1(u|Z^t, \hat{\theta}_{1,T}) \right)^2 - \left( 1\{Y_t \leq u\} - F_k(u|Z^t, \hat{\theta}_{k,T}) \right)^2 \right) \\
& = \frac{1}{\sqrt{T}} \sum_{t=s}^T \left( \left( 1\{Y_t \leq u\} - F_1(u|Z^t, \hat{\theta}_{1,T}) \right)^2 - \mu_1^2(u) \right) \\
& \quad - \frac{1}{\sqrt{T}} \sum_{t=s}^T \left( \left( 1\{Y_t \leq u\} - F_k(u|Z^t, \hat{\theta}_{k,T}) \right)^2 - \mu_k^2(u) \right) + \sqrt{T-s}(\mu_1^2(u) - \mu_k^2(u)) \\
& = \frac{1}{\sqrt{T}} \sum_{t=s}^T \left( \left( 1\{Y_t \leq u\} - F_1(u|Z^t, \theta_1^\dagger) \right)^2 - \mu_1^2(u) \right) - \frac{1}{\sqrt{T}} \sum_{t=s}^T \left( \left( 1\{Y_t \leq u\} - F_k(u|Z^t, \theta_k^\dagger) \right)^2 - \mu_k^2(u) \right) \\
& \quad - \frac{2}{T} \sum_{t=s}^T \nabla_{\theta_1} F_1(u|Z^t, \bar{\theta}_{1,T})' \left( 1\{Y_t \leq u\} - F_1(u|Z^t, \theta_1^\dagger) \right) \sqrt{T} \left( \hat{\theta}_{1,T} - \theta_1^\dagger \right) \\
& \quad + \frac{2}{T} \sum_{t=s}^T \nabla_{\theta_k} F_k(u|Z^t, \bar{\theta}_{k,T})' \left( 1\{Y_t \leq u\} - F_k(u|Z^t, \theta_k^\dagger) \right) \sqrt{T} \left( \hat{\theta}_{k,T} - \theta_k^\dagger \right) \\
& \quad + \sqrt{T}(\mu_1^2(u) - \mu_k^2(u)) + o_P(1) \\
& = I_T(u) + \sqrt{T}(\mu_1^2(u) - \mu_k^2(u)) + o_P(1), \tag{9}
\end{aligned}$$

where  $\bar{\theta}_{i,T} \in (\hat{\theta}_{i,T}, \theta_i^\dagger)$ , and where the  $o_P(1)$  term holds uniformly in  $u \in U$ . In fact, given Assumption A, for  $i = 1, \dots, m$ ,

$$\begin{aligned}
& \text{Vec} \left( \frac{1}{\sqrt{T}} \sum_{t=s}^T \nabla_{\theta_i} F_i(u|Z^t, \bar{\theta}_{i,T})' \left( \hat{\theta}_{i,T} - \theta_i^\dagger \right) \left( \hat{\theta}_{i,T} - \theta_i^\dagger \right)' \nabla_{\theta_i} F_i(u|Z^t, \bar{\theta}_{i,T}) \right) \\
& = \left[ \frac{1}{T} \sum_{t=s}^T \nabla_{\theta_i} F_i(u|Z^t, \bar{\theta}_{i,T})' \otimes \nabla_{\theta_i} F_i(u|Z^t, \bar{\theta}_{i,T}) \right] \sqrt{T} \text{vec} \left( \left( \hat{\theta}_{i,T} - \theta_i^\dagger \right) \left( \hat{\theta}_{i,T} - \theta_i^\dagger \right)' \right) \\
& = o_P(1), \tag{10}
\end{aligned}$$

uniformly in  $u \in U$ , as the term in square brackets satisfies the uniform strong law of large numbers, and so is  $O_P(1)$  (see e.g. Theorem A.2.2 in White(1994)) for any given  $u$ , and A(ii) ensures that the term is also  $O_P(1)$ , uniformly in  $u$ , and as  $\left( \hat{\theta}_{i,T} - \theta_i^\dagger \right) = O_p(T^{-1/2})$  (since it satisfies the central limit theorem (see e.g. Theorem 6.2 in White (1994))). Note that, given Assumption A(iii)-(v), for

$i = 1, \dots, m,$

$$\sqrt{T} \left( \widehat{\theta}_{i,T} - \theta_i^\dagger \right) = A(\theta_i^\dagger) \frac{1}{\sqrt{T}} \sum_{t=s}^T \nabla_{\theta_i} \ln f_i(Y_t | Z^t, \theta_i^\dagger) + o_P(1),$$

where  $A(\theta_i^\dagger) = \left( E \left( -\nabla_{\theta_i}^2 f_i(y_t | Z^t, \theta_i^\dagger) \right) \right)^{-1}$ . Thus,  $I_T(u)$  converges in distribution pointwise in  $u$ , and, by the Cramer Wold device, it follows that for given  $u$  and  $u'$ ,  $(I_T(u), I_T(u'))'$  converges in distribution to a bivariate normal with covariance  $\begin{pmatrix} C_k(u, u) & C_k(u, u') \\ C_k(u', u) & C_k(u', u') \end{pmatrix}$ , as defined in (8).

Now, in order to show that  $I_T(u)$  weakly converges as a process on  $U$ , we need to show that it is stochastic equicontinuous on  $U$ . First note that,

$$\begin{aligned} I_T(u) &= \frac{1}{\sqrt{T}} \sum_{t=s}^T \left( F_1^2(u | Z^t, \theta_1^\dagger) - E \left( F_1^2(u | Z^t, \theta_1^\dagger) \right) \right) \\ &\quad - \frac{2}{\sqrt{T}} \sum_{t=s}^T \left( F_1(u | Z^t, \theta_1^\dagger) 1\{Y_t \leq u\} - E \left( F_1(u | Z^t, \theta_1^\dagger) 1\{Y_t \leq u\} \right) \right) \\ &\quad - \frac{1}{\sqrt{T}} \sum_{t=s}^T \left( F_k^2(u | Z^t, \theta_k^\dagger) - E \left( F_k^2(u | Z^t, \theta_k^\dagger) \right) \right) \\ &\quad + \frac{2}{\sqrt{T}} \sum_{t=s}^T \left( F_k(u | Z^t, \theta_k^\dagger) 1\{Y_t \leq u\} - E \left( F_k(u | Z^t, \theta_k^\dagger) 1\{Y_t \leq u\} \right) \right) \\ &\quad - 2m_{\theta_1^\dagger}(u)' \sqrt{T} \left( \widehat{\theta}_{1,T} - \theta_1^\dagger \right) + 2m_{\theta_k^\dagger}(u)' \sqrt{T} \left( \widehat{\theta}_{k,T} - \theta_k^\dagger \right) + o_P(1), \end{aligned} \quad (11)$$

where  $m_{\theta_i^\dagger}(u)' = E \left( \nabla_{\theta_i} F_i(u | Z^t, \theta_i^\dagger)' \left( 1\{Y_t \leq u\} - F_i(u | X_t, \theta_i^\dagger) \right) \right)$ , and the  $o_P(1)$  term holds uniformly in  $u \in U$ . Let  $I_{i,T}(u)$  be the term in the  $i$ -th line of (11). Then,

$$\begin{aligned} I_T(u) &= \frac{1}{\sqrt{T}} \sum_{t=s}^T I_{1,t}(u) - \frac{2}{\sqrt{T}} \sum_{t=s}^T I_{2,t}(u) - \frac{1}{\sqrt{T}} \sum_{t=s}^T I_{3,t}(u) + \frac{2}{\sqrt{T}} \sum_{t=s}^T I_{4,t}(u) \\ &\quad - 2m_{\theta_1^\dagger}(u)' \sqrt{T} \left( \widehat{\theta}_{1,T} - \theta_1^\dagger \right) + 2m_{\theta_k^\dagger}(u)' \sqrt{T} \left( \widehat{\theta}_{k,T} - \theta_k^\dagger \right) + o_P(1), \end{aligned} \quad (12)$$

and noting that  $m_{\theta_i^\dagger}(u)'$  is equicontinuous on  $U$ , it suffices to show that the first four terms on the right hand side of (12) are stochastic equicontinuous on  $U$ . Thus, it suffices to show that,

$$\lim_{T \rightarrow \infty} \sup \sum_{k=1}^4 P \left( \sup_{\substack{u, u_j \in U \\ \rho_k(u, u_j) < \delta}} \left| \frac{1}{\sqrt{T}} \sum_{t=s}^T I_{k,t}(u) - \frac{1}{\sqrt{T}} \sum_{t=s}^T I_{k,t}(u_j) \right| > \varepsilon \right) = 0, \text{ as } \delta \rightarrow 0,$$

where  $\rho_k(u, u_j) = \left( E((I_{k,t}(u) - I_{k,t}(u_j))^4) \right)^{1/4}$ . Now, define the bracketing number,  $N_{k,4}(\epsilon, U)$ , to be the smallest number,  $n \in \mathbb{N}$ , for which there exists  $(u_1, \dots, u_n) \in U$ , such that for any  $u \in U$  there exists  $u_j$ ,  $j = 1, \dots, n$  ensuring that  $\left( E((I_{k,t}(u) - I_{k,t}(u_j))^4) \right)^{1/4} \leq \epsilon$ . Once we have shown that for

$k = 1, \dots, 4$ ,  $\int_0^1 \sqrt{\log N_{k,4}(\epsilon, U)} d\epsilon < \infty$ , then stochastic equicontinuity of  $I_T(u)$  on  $U$  follows from Theorem 1 (Application 1) in Doukhan, Massart and Rio (DMR: 1995). In fact, given the size of the mixing coefficients in A(i),  $\sum_{j=1}^\infty j^{-1} \beta_j$  is a convergent series, and thus condition (2.10) in DMR can be replaced with the condition that  $\int_0^1 \sqrt{\log N_{k,4}(\epsilon, U)} d\epsilon < \infty$ , for  $k = 1, \dots, 4$ . Now,

$$\begin{aligned} (E((I_{1,t}(u_j) - I_{1,t}(u))^4))^{1/4} &\leq \left( E \left( \left( F_1^2(u|Z^t, \theta_1^\dagger) - F_1^2(u_j|Z^t, \theta_1^\dagger) \right)^4 \right) \right)^{1/4} \\ &\quad + \left| \int_V \left( F_1^2(u|z, \theta_1^\dagger) - F_1^2(u_j|z, \theta_1^\dagger) \right) f_0(z) dz \right|, \end{aligned} \quad (13)$$

where  $f_0(\cdot)$  denotes the “true” marginal density of the conditioning variable(s), and  $V$  is the support of  $Z^t$ . With regard to the first term on the RHS of (13), note that there exists a  $\tilde{z} \in V$  such that,

$$\left( E \left( \left( F_1^2(u|Z^t, \theta_1^\dagger) - F_1^2(u_j|Z^t, \theta_1^\dagger) \right)^4 \right) \right)^{1/4} = C_1 \left( \int_V \left( F_1^2(u|z, \theta_1^\dagger) - F_1^2(u_j|\tilde{z}, \theta_1^\dagger) \right)^4 f_0(z) dz \right)^{1/4},$$

with  $C_1 = \frac{\left( \int_V \left( F_1^2(u|z, \theta_1^\dagger) - F_1^2(u_j|z, \theta_1^\dagger) \right)^4 f_0(z) dz \right)^{1/4}}{\left( \int_V \left( F_1^2(u|z, \theta_1^\dagger) - F_1^2(u_j|\tilde{z}, \theta_1^\dagger) \right)^4 f_0(z) dz \right)^{1/4}} < \infty$ . We can choose  $(u_1, \dots, u_n) \in U$ , such that  $F_1^2(u_j|\tilde{z}, \theta_1^\dagger) = j\delta$ ,  $j = 1, \dots, 1/\delta$ , so that the LHS of the last inequality above is majorized by  $C_1\delta$ . With regard to the second term on the RHS of (13),

$$\left| \int_V \left( F_1^2(u|z, \theta_1^\dagger) - F_1^2(u_j|z, \theta_1^\dagger) \right) f_0(z) dz \right| \leq C_2 \int_V \left| F_1^2(u|z, \theta_1^\dagger) - F_1^2(u_j|\tilde{z}, \theta_1^\dagger) \right| f_0(z) dz, \quad (14)$$

where  $C_2 = \frac{\int_V \left| F_1^2(u|z, \theta_1^\dagger) - F_1^2(u_j|z, \theta_1^\dagger) \right| f_0(z) dz}{\int_V \left| F_1^2(u|z, \theta_1^\dagger) - F_1^2(u_j|\tilde{z}, \theta_1^\dagger) \right| f_0(z) dz} < \infty$ . For the same  $(u_1, \dots, u_n)$  as above, set  $F_1^2(u_j|\tilde{z}, \theta_1^\dagger) = j\delta$ ,  $j = 1, \dots, 1/\delta$ , so that for any  $u$  we can find an  $u_j$  ensuring that the last term in (14) is majorized by  $C_2\delta$ . Now, set  $\delta(\epsilon) = \epsilon/(C_1 + C_2)$ , so that  $N_{1,4}(\epsilon, U) = 1/\delta(\epsilon) = (C_1 + C_2)\epsilon^{-1}$ , for  $0 < \epsilon < 1/2$ , and  $\int_0^1 \sqrt{\log((C_1 + C_2)/\epsilon)} d\epsilon \leq \int_0^1 \log((C_1 + C_2)/\epsilon) d\epsilon = \log(C_1 + C_2) + 1 < \infty$ . Finally, note that  $I_{2,T}(u)$ ,  $I_{3,T}(u)$ , and  $I_{4,T}(u)$  can be treated in an analogous manner. It has been already shown in Section 2 that

$$\begin{aligned} &\int_U (\mu_1^2(u) - \mu_k^2(u)) \phi(u) du \\ &= \int_U \left( E \left( \left( F_1(u|Z^t, \theta_1^\dagger) - F_0(u|Z^t, \theta_0) \right)^2 \right) - E \left( \left( F_k(u|Z^t, \theta_k^\dagger) - F_0(u|Z^t, \theta_0) \right)^2 \right) \right) \phi(u) du. \end{aligned}$$

Thus, by the continuous mapping theorem it follows that,

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

where  $Z_{1,k}(u)$  is a zero mean Gaussian process with covariance  $C_k(u, u')$  defined as in (8), for  $k = 2, \dots, m$ .

**Proof of Theorem 2:** In the sequel,  $P^*$ ,  $E^*$ , and  $Var^*$  denote the probability law of the resampled series, conditional on the sample, the expectation, and the variance operators associated with  $P^*$ , respectively. With the notation  $o_{P^*}(1) \Pr - P$ , and  $O_{P^*}(1) \Pr - P$ , we mean a term approaching zero in  $P^*$ -probability and a term bounded in  $P^*$ -probability, conditional on the sample and for all samples except a set of measure zero, respectively. Write  $Z_{T,u}^*(1, k)$  as

$$\begin{aligned} Z_{T,u}^*(1, k) &= \frac{1}{\sqrt{T}} \sum_{t=s}^T \left( \left( \left( 1\{Y_t^* \leq u\} - F_1(u|Z^{*t}, \theta_1^\dagger) \right) - \nabla_{\theta_1} F_1(u|Z^{*t}, \bar{\theta}_{1,T}^*) \left( \hat{\theta}_{1,T}^* - \theta_1^\dagger \right) \right)^2 \right. \\ &\quad - \left( \left( 1\{Y_t \leq u\} - F_1(u|Z^t, \theta_1^\dagger) \right) - \nabla_{\theta_1} F_1(u|Z^t, \bar{\theta}_{1,T}) \left( \hat{\theta}_{1,T} - \theta_1^\dagger \right) \right)^2 \Big) \\ &\quad - \left( \left( \left( 1\{Y_t^* \leq u\} - F_k(u|Z^{*t}, \theta_k^\dagger) \right) - \nabla_{\theta_k} F_k(u|Z^{*t}, \bar{\theta}_{k,T}^*) \left( \hat{\theta}_{k,T}^* - \theta_k^\dagger \right) \right)^2 \right. \\ &\quad \left. \left. - \left( \left( 1\{Y_t \leq u\} - F_k(u|Z^t, \theta_k^\dagger) \right) - \nabla_{\theta_k} F_k(u|Z^t, \bar{\theta}_{k,T}) \left( \hat{\theta}_{k,T} - \theta_k^\dagger \right) \right)^2 \right) \right), \end{aligned} \quad (15)$$

where  $\bar{\theta}_{i,T}^* \in \left( \hat{\theta}_{i,T}^*, \theta_i^\dagger \right)$ ,  $\bar{\theta}_{i,T} \in \left( \hat{\theta}_{i,T}, \theta_i^\dagger \right)$ . Now,

$$\begin{aligned} &Vec \left( \frac{1}{\sqrt{T}} \sum_{t=s}^T \nabla_{\theta_i} F_i(u|Z^{*t}, \bar{\theta}_{i,T}^*)' \left( \hat{\theta}_{i,T}^* - \theta_i^\dagger \right) \left( \hat{\theta}_{i,T}^* - \theta_i^\dagger \right)' \nabla_{\theta_i} F_i(u|Z^{*t}, \bar{\theta}_{i,T}^*) \right) \\ &= \left[ \frac{1}{T} \sum_{t=s}^T \nabla_{\theta_i} F_i(u|Z^{*t}, \bar{\theta}_{i,T}^*)' \otimes \nabla_{\theta_i} F_i(u|Z^{*t}, \bar{\theta}_{i,T}^*) \right] \sqrt{T} vec \left( \hat{\theta}_{i,T}^* - \theta_i^\dagger \right) \left( \hat{\theta}_{i,T}^* - \theta_i^\dagger \right)' \\ &= o_{P^*}(1), \Pr - P, \end{aligned} \quad (16)$$

because  $\sqrt{T} \left( \hat{\theta}_{i,T}^* - \theta_i^\dagger \right) = \sqrt{T} \left( \hat{\theta}_{i,T}^* - \hat{\theta}_{i,T} \right) + \sqrt{T} \left( \hat{\theta}_{i,T} - \theta_i^\dagger \right) = O_{P^*}(1) + O(1) = O_{P^*}(1) \Pr - P$ , by Theorem 2.2 in Goncalves (GW, (2002)), and  $\sqrt{T} \left( \hat{\theta}_{i,T}^* - \hat{\theta}_{i,T} \right) = O_{P^*}(1) \Pr - P$ , as it converges in  $P^*$ -distribution, and because the term in square brackets is  $O_{P^*}(1) \Pr - P$ . Thus,  $Z_{T,u}^*(1, k)$  can be written as,

$$\begin{aligned} &\frac{1}{\sqrt{T}} \sum_{t=s}^T \left( F_1^2(u|Z^{*t}, \theta_1^\dagger) - F_1^2(u|Z^t, \theta_1^\dagger) \right) - \frac{2}{\sqrt{T}} \sum_{t=s}^T \left( F_1(u|Z^{*t}, \theta_1^\dagger) 1\{Y_t^* \leq u\} - F_1(u|Z^t, \theta_1^\dagger) 1\{Y_t \leq u\} \right) \\ &- \frac{2}{T} \sum_{t=s}^T \left( \left( 1\{Y_t^* \leq u\} - F_1(u|Z^{*t}, \theta_1^\dagger) \right) \nabla_{\theta_1} F_1(u|Z^{*t}, \bar{\theta}_{1,T}^*)' \right) \sqrt{T} \left( \hat{\theta}_{1,T}^* - \theta_1^\dagger \right) \end{aligned}$$

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

where the  $o_{P^*}(1)$  term holds uniformly in  $u \in U$ . We begin by showing that for  $i = 1, \dots, m$ , conditional on the sample and for all samples except a set of measure approaching zero,

- (a)  $\frac{1}{\sqrt{T}} \sum_{t=s}^T \left( F_i^2(u|Z^{*t}, \theta_i^\dagger) - F_i^2(u|Z^t, \theta_i^\dagger) \right)$  has the same limiting distribution as  $\frac{1}{\sqrt{T}} \sum_{t=s}^T \left( F_i^2(u|Z^t, \theta_i^\dagger) - E \left( F_i^2(u|Z^t, \theta_i^\dagger) \right) \right)$ , as a process over  $U$ .  
(b)  $\frac{1}{\sqrt{T}} \sum_{t=s}^T \left( F_i(u|Z^{*t}, \theta_i^\dagger) 1\{Y_t^* \leq u\} - F_i(u|Z^t, \theta_i^\dagger) 1\{Y_t \leq u\} \right)$  has the same limiting distribution as  $\frac{1}{\sqrt{T}} \sum_{t=s}^T \left( F_i(u|Z^t, \theta_i^\dagger) 1\{Y_t \leq u\} - E \left( F_i(u|Z^t, \theta_i^\dagger) 1\{Y_t \leq u\} \right) \right)$ , as a process over  $U$ .

We first show that (a) and (b) above hold pointwise in  $u$ .

Thereafter, we show that  $\frac{1}{\sqrt{T}} \sum_{t=s}^T \left( F_i^2(u|Z^{*t}, \theta_i^\dagger) - F_i^2(u|Z^t, \theta_i^\dagger) \right)$  and  $\frac{1}{\sqrt{T}} \sum_{t=s}^T \left( F_i(u|Z^{*t}, \theta_i^\dagger) 1\{Y_t^* \leq u\} - F_i(u|Z^t, \theta_i^\dagger) 1\{Y_t \leq u\} \right)$  are  $P^*$ -stochastic equicontinuous on  $U$ , conditional on the sample and for all sample except a set of measure approaching zero. Without loss of generality, we begin by showing (a), letting  $i = 1$ . Given the block resampling scheme described in Section 3.2, it is easy to see that,

$$E^* \left( \frac{1}{\sqrt{T}} \sum_{t=s}^T F_1^2(u|Z^{*t}, \theta_1^\dagger) \right) = \frac{1}{\sqrt{T}} \sum_{t=s}^T F_1^2(u|Z^t, \theta_1^\dagger) + O \left( \frac{l}{\sqrt{T}} \right), \Pr - P.$$

Now, recalling that each block, conditional on the sample, is identically and independently distributed,

$$\begin{aligned}
& Var^* \left( \frac{1}{\sqrt{T}} \sum_{t=s}^T F_1^2(u|Z^{*t}, \theta_1^\dagger) \right) = E^* \left( \left( \frac{1}{\sqrt{T}} \sum_{t=s}^T \left( F_1^2(u|Z^{*t}, \theta_1^\dagger) - E^* \left( F_1^2(u|Z^{*t}, \theta_1^\dagger) \right) \right) \right)^2 \right) \\
& = E^* \left( \left( \frac{1}{\sqrt{T}} \sum_{t=s}^T \left( F_1^2(u|Z^{*t}, \theta_1^\dagger) - \left( \frac{1}{T} \sum_{t=s}^T F_1^2(u|Z^t, \theta_1^\dagger) \right) \right) \right)^2 \right) + O \left( \frac{l}{\sqrt{T}} \right) \\
& = \frac{1}{bl} E^* \left( \left( \sum_{k=1}^b \sum_{i=1}^l \left( F_1^2(u|Z^{I_k+i}, \theta_1^\dagger) - \left( \frac{1}{T} \sum_{t=s}^T F_1^2(u|Z^t, \theta_1^\dagger) \right) \right) \right)^2 \right) + O \left( \frac{l}{\sqrt{T}} \right)
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{l} E^* \left( \left( \sum_{i=1}^l \left( F_1^2(u|Z^{I_1+i}, \theta_1^\dagger) - \left( \frac{1}{T} \sum_{t=s}^T F_1^2(u|Z^t, \theta_1^\dagger) \right) \right) \right)^2 \right) + O\left(\frac{l}{\sqrt{T}}\right) \\
&= \frac{1}{T} \sum_{t=l}^{T-l} \sum_{i=-l}^l \left( F_1^2(u|Z^t, \theta_1^\dagger) - \left( \frac{1}{T} \sum_{t=s}^T F_1^2(u|Z^t, \theta_1^\dagger) \right) \right) \left( F_1^2(u|Z^{t+i}, \theta_1^\dagger) - \left( \frac{1}{T} \sum_{t=s}^T F_1^2(u|Z^t, \theta_1^\dagger) \right) \right) \\
&\quad + O\left(\frac{l}{\sqrt{T}}\right) \\
&= \lim_{T \rightarrow \infty} Var \left( \frac{1}{\sqrt{T}} \sum_{t=s}^T F_1^2(u|Z^t, \theta_1^\dagger) \right), \text{ Pr} - P, \tag{18}
\end{aligned}$$

where the last equality follows from Theorem 1 in Andrews (1991), given Assumption A, and given the growth rate conditions on  $l$ . Therefore, given Assumption A, by Theorem 3.5 in Künsch (1989), (a) holds pointwise in  $u$ .

Now, let  $\frac{1}{\sqrt{T}} \sum_{t=s}^T \left( F_1^2(u|Z^{*t}, \theta_1^\dagger) - F_1^2(u|Z^t, \theta_1^\dagger) \right) = \frac{1}{\sqrt{T}} \sum_{t=s}^T I_{1,t}^*(u)$ . We show that, conditional on the sample and for samples except a set of measure zero,

$$\lim_{T \rightarrow \infty} \sup P^* \left( \sup_{\substack{u, u_j \in U \\ \rho(u, u_j) < \delta}} \left| \frac{1}{\sqrt{T}} \sum_{t=s}^T I_{1,t}^*(u) - \frac{1}{\sqrt{T}} \sum_{t=s}^T I_{1,t}^*(u_j) \right| > \varepsilon \right) = 0, \text{ as } \delta \rightarrow 0,$$

where  $\rho(u, u_j)$  is a semi-norm, which will be defined below. Now,

$$\begin{aligned}
&\frac{1}{\sqrt{T}} \sum_{t=s}^T (I_{1,t}^*(u) - I_{1,t}^*(u_j)) = \frac{1}{\sqrt{lb}} \sum_{k=1}^b \sum_{i=1}^l \left( \left( F_1^2(u|Z^{I_k+i}, \theta_1^\dagger) - F_1^2(u|Z^{(k-1)l+i}, \theta_1^\dagger) \right) \right. \\
&\quad \left. - \left( F_1^2(u_j|Z^{I_k+i}, \theta_1^\dagger) - F_1^2(u_j|Z^{(k-1)l+i}, \theta_1^\dagger) \right) \right) \\
&= \frac{1}{\sqrt{lb}} \sum_{k=1}^b \left( \widetilde{UN}_k(u) - \widetilde{UN}_k(u_j) \right) + O\left(\frac{l}{\sqrt{T}}\right), \text{ Pr} - P,
\end{aligned}$$

where  $\widetilde{UN}_k(u) = UN_k(u) - E^*(UN_k(u))$ , and  $UN_k(u)$  is an *iid* uniform random variable taking values  $F_1^2(u|Z^{i+1}, \theta_1^\dagger) + \dots + F_1^2(u|Z^{i+l}, \theta_1^\dagger)$ ,  $i = s, s+1, \dots, T-l$ , with probability  $1/(T-s-l+1)$ .

Given that all blocks are independent of one another,

$$\begin{aligned}
&Var^* \left( \frac{1}{\sqrt{lb}} \sum_{k=1}^b \left( \widetilde{UN}_k(u) - \widetilde{UN}_k(u_j) \right) \right) = Var^* \left( \frac{1}{\sqrt{l}} \left( \widetilde{UN}_1(u) - \widetilde{UN}_1(u_j) \right) \right) \\
&= \frac{1}{l} \left( E^* \left( \widetilde{UN}_1(u) - \widetilde{UN}_1(u_j) \right)^2 \right) \\
&= \frac{1}{l} Var^* \left( \widetilde{UN}_1(u) \right) + \frac{1}{l} Var^* \left( \widetilde{UN}_1(u_j) \right) - \frac{2}{l} Cov^* \left( \widetilde{UN}_1(u), \widetilde{UN}_1(u_j) \right) \tag{19}
\end{aligned}$$

As  $UN_1(u)$  is an independent sequence,  $P^*$ -stochastic equicontinuity in  $u \in U$  follows from DMR (1995, Theorem 1), once we have shown that  $\int_0^1 \sqrt{\log(N_2(U, \epsilon))} d\epsilon < \infty$ , where  $N_2(U, \epsilon)$  denotes the smallest number  $n \in N$ , for which there exists  $(u_1, \dots, u_n)$  such that  $\rho_u(u, u_j) = \left( \frac{1}{l} E^* \left( \widetilde{UN}_1(u) - \widetilde{UN}_1(u_j) \right)^2 \right)^{1/2} < \epsilon$ , conditional on the sample and for all samples except a set with probability measure approaching zero. In fact, in the independent case, the  $L^{2,\beta}$  norm in Theorem 1 of DMR is equivalent to the  $L^2$ -norm (DMR, pp. 393). By a similar argument as that used to derive equation (18), the RHS of equation (19) can be written as

$$\begin{aligned} & \sum_{i=-l}^l \gamma_i(u, u) + \sum_{i=-l}^l \gamma_i(u_j, u_j) - 2 \sum_{i=-l}^l \gamma_i(u, u_j) + o_{P^*}(1), \quad \text{Pr} - P \\ & \leq \sum_{i=0}^{\infty} |\gamma_i(u, u) - \gamma_i(u_j, u_j)| + 2 \sum_{i=0}^{\infty} |\gamma_i(u, u_j) - \gamma_i(u_j, u_j)| + o_{P^*}(1), \quad \text{Pr} - P, \end{aligned} \quad (20)$$

where for  $i = -l, \dots, l$ ,

$$\gamma_i(u, u_j) = E \left( \left( F_1^2(u|Z^t, \theta_1^\dagger) - E \left( F_1^2(u|Z^t, \theta_1^\dagger) \right) \right) \left( F_1^2(u_j|Z^{t+i}, \theta_1^\dagger) - E \left( F_1^2(u_j|Z^{t+i}, \theta_1^\dagger) \right) \right) \right).$$

Given that  $\beta$ -mixing processes are also  $\alpha$ -mixing (as  $\beta_i \leq \alpha_i$  for all  $i$ ), by Corollary 14.3 in Davidson (1994) we see that letting (in Davidson's notation)  $p = 2$  and  $r = 4$  yields

$$\begin{aligned} |\gamma_i(u, u_j)| & \leq 2(\sqrt{2} + 1)\beta_i^{1/4} \left\| F_1^2(u|Z^t, \theta_1^\dagger) - E \left( F_1^2(u|Z^t, \theta_1^\dagger) \right) \right\|_2 \left\| F_1^2(u_j|Z^{t+i}, \theta_1^\dagger) - E \left( F_1^2(u_j|Z^{t+i}, \theta_1^\dagger) \right) \right\|_4 \\ & \leq 2(\sqrt{2} + 1)\beta_i^{1/4}. \end{aligned}$$

Now, we can choose  $(u_1, \dots, u_n) \in U$  in such a way that  $\gamma_i(u_j, u_j) = 2(\sqrt{2} + 1)\beta_i^{1/4}j\delta$ ,  $j = -1/\delta, \dots, 1/\delta$ . Thus, for any  $u$  we can find a  $u_j$  such that,

$$\sum_{i=0}^{\infty} |\gamma_i(u, u) - \gamma_i(u_j, u_j)| + 2 \sum_{i=0}^{\infty} |\gamma_i(u, u_j) - \gamma_i(u_j, u_j)| \leq 12(\sqrt{2} + 1) \sum_{i=0}^{\infty} \beta_i^{1/4} \delta = \epsilon,$$

if we set  $\delta = \delta(\epsilon) = \frac{\epsilon}{12(\sqrt{2}+1) \sum_{i=0}^{\infty} \beta_i^{1/4}}$ . Thus,  $N_2(U, \epsilon) = 24(\sqrt{2} + 1) \sum_{i=0}^{\infty} \beta_i^{1/4} \epsilon^{-1}$ , and given the mixing size condition in A(i),  $\int_0^1 \sqrt{\log(24(\sqrt{2} + 1) \sum_{i=0}^{\infty} \beta_i^{1/4} \epsilon^{-1})} d\epsilon < \infty$ . This completes the proof of (a). Part (b) can be shown along the same lines.

We now need to establish that the sum of the third and fourth terms in (17) has the same limiting distribution as  $-2m_{\theta_1^\dagger}(u)' \sqrt{T} \left( \hat{\theta}_{1,T} - \theta_1^\dagger \right)$ , conditional on the sample and for all samples except a subset with measure approaching zero. We can write the sum of the third and fourth

terms in (17) as,

$$-\frac{2}{T} \sum_{t=s}^T \nabla_{\theta_1} F_1(u|Z^{*t}, \bar{\theta}_{1,T}^*)' \left( 1\{Y_t^* \leq u\} - F_1(u|X_t^*, \theta_1^\dagger) \right) \sqrt{T} \left( \hat{\theta}_{1,T}^* - \hat{\theta}_{1,T} \right) + o_{P^*}(1), \quad \text{Pr} - P.$$

This follows given the mixing and domination conditions in Assumption A. Then, given Lemma 4 and 5 in Goncalves and White (GW: 2002b),

$$\left| \frac{1}{T} \sum_{t=s}^T \left( \left( 1\{Y_t^* \leq u\} - F_1(u|Z^{*t}, \theta_1^\dagger) \right) \nabla_{\theta_1} F_1(u|Z^{*t}, \bar{\theta}_{1,T}^*)' \right) \right. \\ \left. - \frac{1}{T} \sum_{t=s}^T \left( \left( 1\{Y_t \leq u\} - F_1(u|Z^t, \theta_1^\dagger) \right) \nabla_{\theta_1} F_1(u|Z^t, \bar{\theta}_{1,T})' \right) \right| = o_P^*(1), \quad \text{Pr} - P.$$

Also, by Theorem 2.2 in GW (2002b), there exists an  $\varepsilon > 0$  such that,

$$\text{Pr} \left( \sup_{x \in \mathbb{R}^{p_1}} \left| P^* \left( \sqrt{T} \left( \hat{\theta}_{1,T}^* - \hat{\theta}_{1,T} \right) \leq x \right) - P \left( \sqrt{T} \left( \hat{\theta}_{1,T} - \theta_1^\dagger \right) \leq x \right) \right| > \varepsilon \right) \rightarrow 0.$$

Thus,  $\sqrt{T} \left( \hat{\theta}_{1,T}^* - \hat{\theta}_{1,T} \right)$  has the same asymptotic normal distribution as  $\sqrt{T} \left( \hat{\theta}_{1,T} - \theta_1^\dagger \right)$ , conditional on the sample and for all samples except a set with measure approaching zero. Finally, again by the same argument used in Lemma A4 and A5 in GW (2002b),

$$\frac{1}{T} \sum_{t=s}^T \left( \left( 1\{Y_t^* \leq u\} - F_1(u|Z^{*t}, \theta_1^\dagger) \right) \nabla_{\theta_1} F_1(u|Z^{*t}, \bar{\theta}_{1,T}^*)' \right) = m_{\theta_1^\dagger}(u)' + o_{P^*}(1), \quad \text{Pr} - P,$$

where  $m_{\theta_1^\dagger}(u)' = E \left( \nabla_{\theta_1} F_1(u|Z^t, \theta_1^\dagger)' \left( 1\{Y_t \leq u\} - F_1(u|X_t, \theta_1^\dagger) \right) \right)$ . Needless to say, the sum of the last two terms in (17) can be treated in the same manner. Thus, the sum of the third, fourth, seventh and eighth terms on the RHS of (17) has the same limiting distribution as the sum of the fifth and sixth terms on the RHS of (11). The desired result then follows via application of the continuous mapping theorem.



## 7 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.
- Bai, J., (2001), Testing Parametric Conditional Distributions of Dynamic Models, *Review of Economics and Statistics*, forthcoming.
- Benjamini, Y., and Y. Hochberg, (1995), Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, *Journal of the Royal Statistical Society Series B*, 57, 289-300.
- Chang, Y.S., J.F. Gomes, and F. Schorfheide, (2002), Learning-by-Doing as a Propagation Mechanism, *American Economic Review*, 92, 1498-1520.
- Chatfield, C., (1993), Calculating Interval Forecasts, *Journal of Business and Economic Statistics*, 11, 121-135.
- Christoffersen, P.F., (1998), Evaluating Interval Forecasts, *International Economic Review*, 39, 841-862.
- Christoffersen, P. and F.X. Diebold, (2000), How Relevant is Volatility Forecasting For Financial Risk Management?, *Review of Economics and Statistics*, 82, 12-22.
- Clements, M.P. and N. Taylor, (2001), Bootstrapping Prediction Intervals for Autoregressive Models, *International Journal of Forecasting*, 17, 247-276.
- Corradi, V. and N.R. Swanson, (2003), Bootstrap Conditional Distribution Tests in the Presence of Dynamic Misspecification, Working Paper, University of Exeter and Rutgers University.
- Davidson, J., (1994), *Stochastic Limit Theory*, Oxford University Press, Oxford.
- 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., A.S. Tay and K.D. Wallis, (1998), Evaluating Density Forecasts of Inflation: The Survey of Professional Forecasters, in *Festschrift in Honor of C.W.J. Granger*, eds. R.F. Engle and H. White, Oxford University Press, Oxford.
- 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.

- Doukhan, P., P. Massart and E. Rio, (1995), Invariance Principles for Absolutely Regular Empirical Processes, *Annales de l'Institut Henri Poincaré: Probabilités et Statistique*, 31, 393-427.
- Duffie, D. and J. Pan, (1997), An Overview of Value at Risk, *Journal of Derivatives*, 4, 7-49.
- Fernandez-Villaverde, J., and J.F. Rubio-Ramirez, (2001), Comparing Dynamic Equilibrium Models to Data, Working Paper, University of Pennsylvania.
- Hansen, P.R., (2001), An Unbiased Test for Superior Predictive Ability, Working Paper, Brown University.
- Hochberg, Y., (1988), A Sharper Bonferroni Procedure for Multiple Significance Tests, *Biometrika*, 75, 800-803.
- Hong, Y., (2001), Evaluation of Out of Sample Probability Density Forecasts with Applications to S&P 500 Stock Prices, Working Paper, Cornell University.
- Giacomini, R. (2002), Comparing Density Forecasts via Weighted Likelihood Ratio Tests: Asymptotic and Bootstrap Methods, 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.
- 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, Working Paper, Université de Montréal.
- Granger, C.W.J., H. White, and M. Kamstra, (1989), Interval Forecasting - An Analysis Based Upon ARCH-Quantile Estimators, *Journal of Econometrics*, 40, 87-96.
- Kitamura, Y., (2002), Econometric Comparisons of Conditional Models, Working Paper, University of Pennsylvania.
- Künsch H.R., (1989), The Jackknife and the Bootstrap for General Stationary Observations, *Annals of Statistics*, 17, 1217-1241.
- Li, F. and G. Tkacz, (2002), A Consistent Test for Conditional Density Functions with Time Dependent Data, Working Paper, Bank of Canada.
- Linton, O., E. Maasoumi and Y.J. Whang, (2003), Consistent Testing for Stochastic Dominance Under General Sampling Schemes, Working Paper, London School of Economics, Southern Methodist University and Ewha University.
- Politis, D.N., J.P. Romano and M. Wolf, (1999), *Subsampling*, Springer and Verlag, New York.

- Schorfheide, F., (2000), Loss Function Based Evaluation of DSGE Models, *Journal of Applied Econometrics*, 15, 645-670.
- Vuong, Q. (1989), Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses, *Econometrica*, 57, 307-333.
- Whang, Y.J., (2000), Consistent Bootstrap Tests of Parametric Regression Functions, *Journal of Econometrics*, 27-46.
- Whang, Y.J., (2001), Consistent Specification Testing for Conditional Moment Restrictions, *Economics Letters*, 71, 299-306.
- White, H., (1982), Maximum Likelihood Estimation of Misspecified Models, *Econometrica*, 50, 1-25.
- White, H., (1994), *Estimation, Inference and Specification Analysis*, Cambridge University Press, Cambridge.
- White, H., (2000), A Reality Check for Data Snooping, *Econometrica*, 68, 1097-1126.
- Zheng, J.X., (2000), A Consistent Test of Conditional Parametric Distribution, *Econometric Theory*, 16, 667-691.

Table 1: Empirical Level Experiments: Interval =  $\mu_Y + \frac{1}{16}\sigma_Y$

$\alpha$	Sample Size = 60 Observations						Sample Size = 120 Observations															
	$l = 2$			$l = 5$			$l = 6$			$l = 2$			$l = 3$			$l = 5$			$l = 6$			
	$t = 3$			$t = 3$			$t = 3$			$t = 3$			$t = 3$			$t = 3$			$t = 3$			
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
	Nominal Level																					
	Panel A: Exogenous Variate Variance = 0.1																					
0.4	0.055	0.117	0.028	0.075	0.007	0.039	0.009	0.031	0.128	0.231	0.045	0.085	0.009	0.037	0.007	0.033						
0.6	0.137	0.220	0.075	0.146	0.025	0.069	0.026	0.068	0.181	0.260	0.090	0.151	0.026	0.077	0.024	0.059						
0.8	0.212	0.279	0.115	0.177	0.044	0.101	0.033	0.078	0.211	0.284	0.115	0.192	0.038	0.090	0.042	0.089						
0.9	0.239	0.303	0.127	0.195	0.077	0.129	0.055	0.115	0.252	0.323	0.140	0.209	0.051	0.096	0.037	0.075						
	Panel B: Exogenous Variate Variance = 1.0																					
0.4	0.071	0.148	0.029	0.084	0.016	0.038	0.004	0.024	0.132	0.223	0.050	0.126	0.012	0.054	0.013	0.042						
0.6	0.145	0.227	0.056	0.141	0.025	0.070	0.017	0.061	0.198	0.278	0.105	0.171	0.040	0.078	0.021	0.055						
0.8	0.193	0.278	0.116	0.193	0.041	0.091	0.027	0.071	0.215	0.264	0.117	0.186	0.034	0.086	0.024	0.070						
0.9	0.201	0.281	0.116	0.198	0.062	0.119	0.048	0.087	0.213	0.272	0.111	0.172	0.046	0.109	0.041	0.081						
	Panel C: Exogenous Variate Variance = 10.0																					
0.4	0.069	0.135	0.041	0.086	0.007	0.024	0.011	0.033	0.127	0.206	0.055	0.112	0.011	0.048	0.007	0.039						
0.6	0.127	0.211	0.062	0.147	0.025	0.075	0.018	0.066	0.198	0.265	0.109	0.181	0.023	0.078	0.014	0.067						
0.8	0.187	0.270	0.125	0.181	0.037	0.083	0.037	0.074	0.191	0.276	0.102	0.166	0.041	0.087	0.022	0.063						
0.9	0.179	0.259	0.117	0.184	0.045	0.105	0.051	0.108	0.195	0.259	0.093	0.142	0.037	0.086	0.031	0.079						

Notes: Empirical rejection frequencies are reported in the 2nd through 17th columns of entries in the table. The first column reports the value of  $\alpha$ , the autoregressive parameter in the DGP. In all experiments,  $v = 3$ ,  $\sigma^2 = 1$ , and  $\sigma_X^2 = \sigma_W^2 = \sigma_Q^2 = \{0.1, 1.0, 10.0\}$  (corresponding to the 3 panels in the table). Throughout, the conditional confidence interval version of the test is constructed, and the upper and lower bounds of the interval are fixed at  $\mu_Y + \gamma\sigma_Y$  and  $\mu_Y - \gamma\sigma_Y$ , respectively, where  $\gamma = \frac{1}{16}$  (Tables 1 and 4),  $\frac{1}{8}$  (Tables 2 and 5), and  $\frac{1}{2}$  (Tables 3 and 6). The 5% and 10% nominal level bootstrap critical values used in the experiments are constructed using 100 bootstrap replications, block lengths of  $l = \{2, 3, 5, 6\}$  are tried, and all reported rejection frequencies are based on 5000 Monte Carlo simulations. See above for further details.

Table 2: Empirical Level Experiments: Interval =  $\mu_Y + \frac{1}{8}\sigma_Y$

$\alpha$	Sample Size = 60 Observations						Sample Size = 120 Observations									
	$l = 2$			$l = 3$			$l = 5$			$l = 6$						
	<i>Nominal Level</i>															
<i>Panel A: Exogenous Variate Variance = 0.1</i>																
0.4	0.091	0.154	0.035	0.085	0.009	0.028	0.008	0.022	0.121	0.207	0.051	0.116	0.017	0.054	0.015	0.045
0.6	0.131	0.203	0.075	0.144	0.025	0.073	0.025	0.067	0.184	0.262	0.102	0.178	0.027	0.077	0.019	0.069
0.8	0.209	0.282	0.108	0.177	0.037	0.085	0.036	0.084	0.224	0.285	0.131	0.199	0.039	0.077	0.034	0.077
0.9	0.224	0.287	0.132	0.204	0.058	0.131	0.051	0.103	0.221	0.285	0.138	0.191	0.062	0.119	0.040	0.083
<i>Panel B: Exogenous Variate Variance = 1.0</i>																
0.4	0.078	0.140	0.025	0.062	0.009	0.036	0.009	0.032	0.122	0.230	0.052	0.113	0.010	0.054	0.005	0.033
0.6	0.159	0.226	0.080	0.145	0.032	0.078	0.015	0.051	0.207	0.290	0.086	0.147	0.025	0.073	0.014	0.068
0.8	0.200	0.285	0.106	0.182	0.049	0.103	0.036	0.084	0.237	0.314	0.102	0.162	0.053	0.109	0.027	0.073
0.9	0.204	0.265	0.103	0.175	0.062	0.132	0.042	0.101	0.227	0.283	0.113	0.178	0.056	0.108	0.029	0.070
<i>Panel C: Exogenous Variate Variance = 10.0</i>																
0.4	0.066	0.124	0.025	0.071	0.010	0.029	0.007	0.036	0.138	0.232	0.045	0.092	0.006	0.039	0.006	0.030
0.6	0.130	0.226	0.051	0.131	0.022	0.078	0.021	0.065	0.222	0.293	0.077	0.156	0.021	0.071	0.021	0.062
0.8	0.187	0.265	0.109	0.181	0.051	0.102	0.034	0.083	0.258	0.326	0.113	0.176	0.031	0.077	0.028	0.069
0.9	0.195	0.265	0.118	0.195	0.052	0.099	0.049	0.115	0.246	0.313	0.128	0.198	0.042	0.094	0.047	0.079

Notes: See notes to Table 1.

Table 3: Empirical Level Experiments: Interval =  $\mu_Y + \frac{1}{2}\sigma_Y$

$\alpha$	Sample Size = 60 Observations						Sample Size = 120 Observations							
	$l = 2$			$l = 5$			$l = 6$			$l = 2$				
	$l = 3$			$l = 5$			$l = 3$			$l = 5$				
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
Nominal Level														
Panel A: Exogenous Variate Variance = 0.1														
0.4	0.024	0.049	0.013	0.029	0.014	0.029	0.033	0.053	0.034	0.073	0.021	0.047	0.008	0.018
0.6	0.044	0.066	0.035	0.059	0.042	0.059	0.050	0.074	0.051	0.083	0.033	0.060	0.029	0.046
0.8	0.082	0.118	0.084	0.105	0.126	0.153	0.143	0.173	0.071	0.094	0.076	0.101	0.088	0.113
0.9	0.132	0.157	0.131	0.159	0.198	0.229	0.252	0.278	0.096	0.110	0.114	0.134	0.147	0.172
Panel B: Exogenous Variate Variance = 1.0														
0.4	0.066	0.128	0.022	0.067	0.017	0.046	0.005	0.030	0.106	0.207	0.037	0.106	0.009	0.038
0.6	0.131	0.212	0.059	0.121	0.027	0.069	0.009	0.042	0.166	0.242	0.076	0.149	0.021	0.062
0.8	0.174	0.249	0.081	0.136	0.035	0.085	0.028	0.058	0.177	0.258	0.079	0.142	0.024	0.069
0.9	0.154	0.235	0.061	0.119	0.030	0.080	0.032	0.071	0.145	0.211	0.052	0.106	0.023	0.057
Panel C: Exogenous Variate Variance = 10.0														
0.4	0.060	0.124	0.023	0.070	0.005	0.034	0.008	0.027	0.136	0.218	0.033	0.085	0.009	0.044
0.6	0.119	0.200	0.066	0.146	0.019	0.061	0.022	0.074	0.182	0.245	0.084	0.153	0.027	0.067
0.8	0.170	0.258	0.085	0.145	0.034	0.085	0.024	0.067	0.183	0.259	0.095	0.155	0.036	0.086
0.9	0.161	0.235	0.073	0.131	0.043	0.094	0.036	0.069	0.153	0.213	0.062	0.117	0.031	0.066

Notes: See notes to Table 1.

Table 4: Empirical Power Experiments: Interval =  $\mu_Y + \frac{1}{16}\sigma_Y$

$\alpha$	Sample Size = 60 Observations						Sample Size = 120 Observations					
	$l = 2$			$l = 5$			$l = 6$			$l = 2$		
	$t = 3$			$t = 3$			$t = 3$			$t = 3$		
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
<i>Nominal Level</i>												
<i>Panel A: Exogenous Variate Variance = 0.1</i>												
0.4	0.126	0.193	0.077	0.139	0.033	0.093	0.025	0.073	0.229	0.303	0.145	0.256
0.6	0.620	0.667	0.558	0.607	0.493	0.529	0.482	0.538	0.679	0.714	0.595	0.647
0.8	0.455	0.513	0.349	0.413	0.209	0.298	0.185	0.278	0.476	0.501	0.380	0.447
0.9	0.515	0.542	0.418	0.487	0.267	0.356	0.245	0.327	0.553	0.571	0.403	0.470
											0.283	0.361
											0.204	0.279
<i>Panel B: Exogenous Variate Variance = 1.0</i>												
0.4	0.158	0.231	0.065	0.161	0.035	0.099	0.031	0.094	0.289	0.373	0.142	0.255
0.6	0.589	0.638	0.501	0.584	0.449	0.502	0.424	0.484	0.649	0.684	0.529	0.593
0.8	0.505	0.558	0.358	0.448	0.205	0.317	0.182	0.290	0.532	0.585	0.396	0.461
0.9	0.600	0.642	0.449	0.532	0.273	0.385	0.232	0.355	0.607	0.644	0.502	0.557
											0.253	0.364
											0.198	0.316
<i>Panel C: Exogenous Variate Variance = 10.0</i>												
0.4	0.160	0.241	0.075	0.144	0.031	0.079	0.023	0.068	0.297	0.391	0.161	0.281
0.6	0.612	0.665	0.512	0.579	0.473	0.534	0.459	0.513	0.661	0.695	0.554	0.623
0.8	0.516	0.561	0.371	0.465	0.173	0.291	0.167	0.275	0.567	0.594	0.420	0.481
0.9	0.568	0.607	0.467	0.536	0.263	0.388	0.214	0.315	0.628	0.659	0.481	0.540
											0.286	0.387
											0.236	0.335

Notes: See notes to Table 1.

Table 5: Empirical Power Experiments: Interval =  $\mu_Y + \frac{1}{8}\sigma_Y$

$\alpha$	Sample Size = 60 Observations						Sample Size = 120 Observations					
	$l = 2$			$l = 5$			$l = 6$			$l = 2$		
	$t = 3$			$t = 3$			$t = 3$			$t = 5$		
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
<i>Nominal Level</i>												
<i>Panel A: Exogenous Variate Variance = 0.1</i>												
0.4	0.146	0.250	0.083	0.154	0.036	0.098	0.017	0.069	0.282	0.380	0.157	0.275
0.6	0.619	0.662	0.529	0.604	0.427	0.489	0.423	0.490	0.664	0.717	0.566	0.630
0.8	0.569	0.619	0.372	0.456	0.229	0.335	0.154	0.283	0.591	0.619	0.413	0.481
0.9	0.635	0.672	0.508	0.583	0.293	0.408	0.209	0.324	0.647	0.683	0.514	0.577
											0.305	0.409
											0.229	0.308
<i>Panel B: Exogenous Variate Variance = 1.0</i>												
0.4	0.124	0.207	0.064	0.148	0.033	0.082	0.029	0.084	0.277	0.366	0.153	0.275
0.6	0.592	0.648	0.523	0.588	0.440	0.497	0.460	0.509	0.637	0.674	0.555	0.616
0.8	0.476	0.520	0.359	0.436	0.191	0.297	0.153	0.263	0.506	0.539	0.380	0.457
0.9	0.569	0.597	0.414	0.488	0.264	0.366	0.230	0.328	0.577	0.605	0.436	0.487
											0.206	0.303
<i>Panel C: Exogenous Variate Variance = 10.0</i>												
0.4	0.131	0.210	0.073	0.157	0.032	0.080	0.023	0.083	0.292	0.370	0.137	0.245
0.6	0.613	0.657	0.527	0.585	0.461	0.518	0.460	0.514	0.670	0.697	0.563	0.626
0.8	0.496	0.540	0.378	0.468	0.207	0.322	0.154	0.263	0.533	0.571	0.390	0.462
0.9	0.566	0.594	0.427	0.494	0.316	0.416	0.239	0.337	0.584	0.625	0.476	0.525
											0.284	0.382
											0.230	0.331

Notes: See notes to Table 1.



Table 6: Empirical Power Experiments: Interval =  $\mu_Y + \frac{1}{2}\sigma_Y$ 

$\alpha$	Sample Size = 60 Observations $l = 3$				Sample Size = 120 Observations $l = 3$				$l = 6$					
	$l = 2$		$l = 5$		$l = 2$		$l = 5$		$l = 2$		$l = 5$		$l = 6$	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
Nominal Level														
Panel A: Exogenous Variate Variance = 0.1														
0.4	0.177	0.263	0.075	0.169	0.037	0.105	0.023	0.074	0.250	0.345	0.142	0.271	0.058	0.132
0.6	0.616	0.666	0.525	0.605	0.453	0.510	0.442	0.501	0.626	0.673	0.516	0.597	0.423	0.468
0.8	0.496	0.557	0.365	0.461	0.217	0.327	0.170	0.279	0.591	0.635	0.441	0.527	0.213	0.264
0.9	0.635	0.676	0.473	0.541	0.263	0.375	0.227	0.330	0.651	0.687	0.514	0.577	0.309	0.338
Panel B: Exogenous Variate Variance = 1.0														
0.4	0.168	0.272	0.081	0.187	0.035	0.093	0.038	0.090	0.317	0.415	0.163	0.285	0.050	0.104
0.6	0.614	0.667	0.498	0.574	0.420	0.499	0.414	0.494	0.671	0.706	0.550	0.616	0.421	0.507
0.8	0.577	0.624	0.395	0.503	0.219	0.325	0.177	0.290	0.614	0.656	0.426	0.505	0.230	0.260
0.9	0.630	0.670	0.479	0.565	0.280	0.374	0.206	0.312	0.663	0.694	0.521	0.599	0.298	0.299
Panel C: Exogenous Variate Variance = 10.0														
0.4	0.171	0.278	0.083	0.171	0.035	0.101	0.028	0.090	0.349	0.437	0.178	0.310	0.053	0.121
0.6	0.639	0.691	0.480	0.558	0.380	0.469	0.402	0.472	0.662	0.707	0.530	0.596	0.401	0.460
0.8	0.571	0.638	0.398	0.503	0.208	0.345	0.169	0.289	0.608	0.648	0.454	0.540	0.217	0.267
0.9	0.639	0.682	0.487	0.591	0.279	0.416	0.238	0.351	0.652	0.699	0.505	0.585	0.290	0.329

---

Notes: See notes to Table 1.