Predictive Density Construction and Accuracy Testing with Multiple Possibly Misspecified Models

April 2009

Valentina Corradi and Norman R. Swanson

Warwick University and Rutgers University

This presentation has been prepared for seminars in fall 2009 at Queen's University and University of British Columbia.

Outline

- ♦♦ 1. Summary of Results
- ♦♦ 2. Introduction and Overview
- ♦♦ 3. Setup & Simulation Mthds
- ♦♦ 4. Specification Tests
- ♦♦ 5. Predictive Density Tests
- ♦♦ 6. SGMM and NPSQMLE
- ♦♦ 7. Monte Carlo & Empirical
- ♦♦ 8. Concluding Remarks

1 Summary of Results

- *Financial assets are typically modelled as diffusion processes: hedging strategies, pricing of bonds, derivative assets, etc.*

- *Simulation based framework for constructing conditional distributions for multi factor diffusion models.*

- *Functional form of conditional distribution unknown.*

- *KS type specification tests with patametric rate.*

- *Empirical process version of block bootstrap.*

- *Model selection prediction type tests, misspecified models allowed for.*

- *Allow for jumps, multiple factors under null.*

- *Recursive bootstrap, recentering, and results on recursive SGMM and NPQML estimators.*

- $V_T = \sup_{u\times v\in U\times V} |V_T(u,v)|$

$$V_T(u,v) = \frac{1}{(T-\tau)^{1/2}}.$$

$$\sum_{t=1}^{T-\tau}\left(\frac{1}{S}\sum_{s=1}^{S} 1\left\{X_{s,t+\tau}^{\widehat{\theta}_{T,N,h}} \leq u\right\} - 1\{X_{t+\tau}\leq u\}\right) 1\left\{X_t \leq v\right\},$$

where U and V are compact sets on the real line.

-

$$D_{k,P,N}(u) = \frac{1}{P^{1/2}}\sum_{t=R}^{T-\tau}$$

$$\left(\left[\frac{1}{N}\sum_{i=1}^{N} 1\left\{X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t)\leq u\right\} - 1\{X_{t+\tau}\leq u\}\right]^2\right.$$

$$\left. - \left[\frac{1}{N}\sum_{i=1}^{N} 1\left\{X_{k,t+\tau,i}^{\widehat{\vartheta}_{k,t,N,h}}(X_t)\leq u\right\} - 1\{X_{t+\tau}\leq u\}\right]^2\right),$$

or

$$D_{k,P,N}^{Max}(u) = \max_{k=2,\ldots,m} D_{k,P,N}(u)$$

2 Introduction and Overview

■ *There has been much focus on testing for correct specification of diffusions, particularly since the key paper by Aït-Sahalia (1996), who compares marginal density implied by joint specification of drift and variance with nonparametric density.*

■ *Compare empirical (marginal or joint) distribution with corresponding distribution implied by the model (Corradi and Swanson (2005))*

■ *No power against alternatives with same marginal (or joint). Need to test for correct specification of transition function.*

■ *If transition density known in closed form, can use the probability integral trasform approach of Diebold, Gunther and Tai (1998), Kolmogorov test of Bai (2003), Cross-spectrum approach of Hong (2001), Hong and*

Li (2005), Comparison of empirical and model based characteristic functions Hong and Chen (2005).

- *Often the transition density of a diffusion is not known in closed form. Compare conditional density of simulated and historical data (Altissimo and Mele, (2005)), compare conditional distributions of simulated and historical data (current papers). Approximation of closed form of transition density via Hermite polynomial (Aït-Sahalia (2002) and Aït-Sahalia, Fan and Peng (2005)).*

- *Models are approximations of reality and are likely to be misspecified.*

- *In contrast to specification testing, in several circumstances such as risk management and value at risk assessment, interest may lie in predictive densities and predictive intervals.*

■■*Summary of Some Related Papers:*

* probability integral transform approach: Diebold, Gunther and Tay (1998)

* cross spectrum approach: Hong (2001), Hong and Li (2005), Hong, Li and Zhao (2004)

* martingalization/Kolmogorov test approach: Bai (2003)

* normality transformation approach: Bontemps and Meddahi (2005,2006) and Duan (2004)

* CDF approach: Corradi and Swanson (2005)

* DSGE models: Corradi and Swanson (2007)

* unknown transition density approach: Altissimo and Mele (2002, 2005), and Thompson (2004)

* closed form transition density approximation approach: Aït-Sahalia (1999, 2002)

■■ *Objectives:*

** methodology that can be used for:*

Part (i) in-sample specification testing based on conditional distributions and confidence intervals; and

Part (ii) out-of-sample model selection based on predictive densities and confidence intervals

** applicable to multi-factor and multi-dimensional diffusion processes*

** simulation and bootstrap critical value construction that is computationally simple*

** functional form of conditional distributions assumed unkown; drawing on notion that all models are approximations of reality and likely to be misspecified*

■ Part (i): *Simulation based conditional Kolmogorov type test of correct specification for cases where form of conditional density is unkown, given parameter estimation error.*

* *power against larger class of alternatives than Aït-Sahalia (1996) and Corradi and Swanson (2005)*

* *first order asymptotically valid bootstrap critical values*

■ Part (ii): *Simulation based test for comparison of multiple possibly misspecified diffusion processes via examination of predictive performance.*

* *accuracy defined in terms of distributional generalization of mean square forecast error*

* *empirical evidence supports the presence of jumps in asset prices - need to allow for jump components in returns*

* *while there is some debate on whether short-term rates evolve as one-factor or stochastic volatility models, there is widespread consensus that stock returns are better modeled as stochastic volatility processes; hence need to allow for unobservable factors*

* *prediction approach requires using recursively estimated parameters and simulating* $P - \tau$ *processes of length* τ, *using observed values* τ *periods ago as initial values*

* *for stochastic volatility models, initial values of volatility not observed - hence start the process over large number of random volatilities and average over them*

* with *empirical predictive as well as historical data distributions, consider pairwise (DM type tests), and multiple model (White reality check) type tests*

* *limiting distribution of statistics reflect contribution of recursively estimated parameters*

** appropriate recentering for bootstrap estimators and statistics - establish first order validity of the bootstrap statistics*

** as by-product, establish consistency and asymptotic normality of non-parametric simulated MLE (Fermanian and Salanie 2004, Kristensen and Shin 2006), and of (exactly identified) simulated GMM in a recursive setting under misspecification; and establish first order validity of their bootstrap counterparts*

■■■ *Supplementary Discussion on Specification Testing*

■ *There are many tests for the null of correct specification of a given model (problem of sequential KS testing).*

■■ *Probability Integral Transform* - Rosenblatt (1952) and Diebold, Gunther and Tay (1998).

■ $Z^{t-1} = (y_{t-1}, ..., y_{t-v}, X_t, ..., X_{t-w})$, v, w finite, X_t vector valued

■ $F(y_t | Z^{t-1}, \theta_0) = \int_{-\infty}^{y_t} f_t(y|Z^{t-1}, \theta_0) dy,$ is an iid uniform RV on $[0, 1]$

■■ DGT: difference between the empirical distribution of $F_t\ (y_t\ |Z^{t-1}\ ,\widehat{\theta}_T\ )$ and the 45° – degree line

■

$$H_0\ :\ \Pr\ (y_t \leq y|\Im_{t-1}\ ,\theta_0\ ) = F_t\ (y|\Im_{t-1}\ ,\theta_0\ ),$$
$$H_A\ :\ \text{the negation of}\ H_0,$$

■ Compare $F_t\ (y|\Im_{t-1}\ ,\theta_0\ )$ with CDF of uniform RV on $[0,1]$; differentiability; nonstationarity; Z^{t-1} contains all useful info in $\Im_{t-1}$;

■ use $\widehat{U}_t = F(y_t\ |Z^{t-1}\ ,\widehat{\theta}_T\ )$

$$\widehat{V}_T(r) = \frac{1}{T^{1/2}}\sum\left(1\{\widehat{U}_t \leq r\} - r\right)$$

$$V_{1T} = \sup_{r\in[0,1]}\ |\widehat{V}_{1T}\ (r)|$$

- Many related tests have power against violations of uniformity, but not independence (see e.g. Bai (2003), Diebold, Hahn and Tay (1999), Hong (2001) and Hong and Li (2003) for Kolmogorov-Smirnov versions of this which compare distributions against uniformity).

- A crucial feature of most specification (and predictive accuracy) tests is the following:

Assume correct specification under the null hypothesis; so that actually look at the difference between the empirical distribution of $F_t(y_t|\Im_{t-1}, \widehat{\theta}_T)$ and the $45^{\circ}-$degree line (i.e. the CDF of a uniform on [0,1]) as a measure of "goodness of fit". But how to define $\Im_{t-1}$ (the information set containing all "relevant" past information).

Kolmogorov-Smirnov and Related Statistics

- Empirical Distribution Function - a natural estimator for F which is unbiased, consistent, and asymptotically normal

- $F_T\,(y) = T^{-1} \sum_{t=1}^{T} 1\{y_t \leq y\}$

- Cramer - von Mises Discrepancy Measure [Cramer (1938, 1946), von Mises (1947)]:

- $T \int \, (F_T - F)^2 \, dF$

- Kolmogorov - Smirnov Discrepancy Measure [Kolmogorov (1933), Smirnov (1939)]:

- $T^{1/2} || F_T - F ||_\infty = sup_x T^{1/2} |F_T(y) - F(y)|$

- Glivenko-Cantelli uniform convergence [Glivenko (1933), Cantelli (1933)]:

- $||F_T - F||_\infty \to^{a.s.} 0$

- Donsker uniform or functional CLT (*iid* data) [Donsker (1952)]:

- $T^{1/2}(F_T - F)$ converges to a Gaussian process, and in particular to a Brownian bridge limit process

- Modified Kolmogorov-Smirnov Statistic

- $T^{1/2}||F_T - F_{\widehat{\theta}}||_\infty$

- Convergence in distribution to the supremum of a Gaussian process - limit distribution may depend on the model F, the estimator $\widehat{\theta}$, and even the parameter value θ

■■ *The Kulback-Leibler Information Criterion* [White (1982), Vuong (1989), Giacomini (2002), Kitamura (2002)]:

■ Choose model which minimizes KLIC; choose model 1 over 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 has been recently employed for the evaluation of DSGE models (see e.g. Schorfheide (2000), Fernandez-Villaverde and Rubio-Ramirez (2004), and Chang, Gomes and Schorfheide (2002)).

■ Our measures of distributional accuracy are intended as complements to the KLIC, although evaluation of conditional confidence intervals may be difficult to address using the KLIC, and we shall also use a probability mass weighting function.

Information Sets and Critical Values

- Limit distribution of KS tests affected by dynamic misspecification. Critical values derived under correct specification given $\Im_{t-1}$ are not in general valid in the case of correct specification given a subset of $\Im_{t-1}$. Many authors use Z^{t-1}, and assume that $Z^{t-1} \equiv \Im_{t-1}$.

- Assume interested in testing whether $y_t | y_{t-1}$ is $N(\alpha_1^\dagger y_{t-1}, \sigma_1)$

- Suppose $\Im_{t-1}$ includes y_{t-1} and y_{t-2} : true cond model is

$$y_t | \Im_{t-1} = y_t | y_{t-1}, y_{t-2} = N(\alpha_1 y_{t-1} + \alpha_2 y_{t-2}, \sigma_2)$$

- Then, $\alpha_1^\dagger$ differs from α_1 and correct specification holds wrt information in y_{t-1} ; but there is dynamic misspecification with respect to y_{t-1} , y_{t-2}.

- Even without taking account of PEE, CVs obtained assuming correct dynamic specification are invalid.

- Stated differently, tests that are designed to have power against both uniformity and independence violations (i.e. tests that assume correct dynamic specification under H_0) will reject; an inference which is incorrect, at least in the sense that the "normality" assumption is *not* false (uniformity still holds, but independence does not -> rejection of model)

■■ *Comparison of Many Models* [White (2000), Corradi and Swanson (2003)]:

■ Need mean square error and other measures of distributional discrepancy

■ Issues of sequential test bias, allowance for misspecification, alternative methods to construct CVs are all relevant

■ Many (possibly) misspecified conditional distributions,

$$F_1(u|Z^t, \theta_1^\dagger), \ \ldots, \ F_m(u|Z^t, \theta_m^\dagger),$$

and true conditional distribution,

$$F_0(u|Z^t, \theta_0) = \mathrm{Pr}(Y_{t+1} \leq u|Z^t)$$

■ One accuracy measure; average over $u \in U$, or use interval based on u_{low}, u_{up} of the following moment:

■ $$E\left(\left(F_j(u|Z^{t+1}, \theta_i^\dagger) - F_0(u|Z^{t+1}, \theta_0)\right)^2\right)$$

3 Setup & Simulation Methods

- Consider m jump diffusions for X_t, $k = 1, ..., m$:

$$\begin{aligned} X_t &= \int_0^t b_k(X(s), \theta_k^\dagger) ds - \lambda_k t \int_Y y\phi_k(y) dy \\ &\quad + \int_0^t \sigma_k(X(s), \theta_k^\dagger) dW(s) + \sum_{j=1}^{J_t} y_{k,j} \end{aligned}$$

$$\begin{aligned} dX(t) &= \left(b_k(X(t), \theta_k^\dagger) - \lambda_k \mu_{y,k}\right) dt \\ &\quad + \sigma_k(X(t), \theta_k^\dagger) dW(t) + \int_Y y p(dy, dt), \end{aligned}$$

$p(dy, dt)$ a random Poisson measure giving point mass at y if a jump occurs in the interval dt.

- J_t a Poisson process with state-independent finite intensity parameter λ_k; $y_{k,j}$ the state independent jump size - iid with marginal density ϕ_k.

- J_t, $y_{k,j}$ independent of the driving Brownian motion, W_t. Also, $\mu_{y,k} = \int_Y y\phi_k(y)dy$ denotes the mean jump size under model k. The case of no jumps is that of $J_t = 0$ for all t, and $\lambda_k = 0$.

- Let $\vartheta_k = (\theta_k, \lambda_k, \mu_{y,k})$

- *Correct specification test (no jumps):*

$$dX(t) = b_0(X(t), \theta_0)dt + \sigma_0(X(t), \theta_0)dW(t)$$

$b_k(\cdot, \cdot) = b_0(\cdot, \cdot)$ and $\sigma_k(\cdot, \cdot) = \sigma_0(\cdot, \cdot)$

- Our target: $F_0^\tau(u|X_t, \vartheta_0) = \Pr(X_{t+\tau} \leq u|X_t, \vartheta_0)$

- (i.e. want to estimate the probability that $X_{t+\tau} \leq u$ given that today we observe X_t)

- For prediction type tests, accuracy measure:

$$E\left(\left(F_1^\tau(u|X_t,\vartheta_1^\dagger) - F_0^\tau(u|X_t,\vartheta_0)\right)^2\right)$$

$$\leq E\left(\left(F_k^\tau(u|X_t,\vartheta_k^\dagger) - F_0^\tau(u|X_t,\vartheta_0)\right)^2\right)$$

- Replace $F_0^\tau(u|X_t,\vartheta_0)$ with $1\{X_{t+\tau} \leq u\}$ (unbiased, consistent, asymptotically normal - Donsker).

- Replace $F_k^\tau(u|X_t,\vartheta_k^\dagger)$ with $1\left\{X_{k,t+\tau}^{\vartheta_k^\dagger}(X_t) \leq u\right\}$, where $X_{k,t+\tau}^{\vartheta_k^\dagger}$ is the process simulated under model k and initiated τ periods before at X_t. Also, as $\vartheta_k^\dagger$ is unknown should be replaced with an estimator (consistent for F_k^τ in general, and for F_0^τ under correct specification).

■ Using a Milstein scheme, simulate paths of length N :

$$
\begin{aligned}
X_{qh}^{\vartheta_k} - X_{(q-1)h}^{\vartheta_k} &= b_k(X_{(q-1)h}^{\theta_k}, \theta_k)h + \sigma_k(X_{(q-1)h}^{\theta_k}, \theta_k)\epsilon_{qh} \\
&\quad -\frac{1}{2}\sigma_k(X_{(q-1)h}^{\theta_k}, \theta_k)'\sigma_k(X_{(q-1)h}^{\theta_k}, \theta_k)h \\
&\quad +\frac{1}{2}\sigma_k(X_{(q-1)h}^{\theta_k}, \theta_k)'\sigma_k(X_{(q-1)h}^{\theta_k}, \theta_k)\epsilon_{qh}^2 \\
&\quad -\lambda_k\mu_{y,k}h + \sum_{j=1}^{\mathcal{J}_k} y_k 1\left\{(q-1)h \leq \mathcal{U}_j \leq qh\right\}
\end{aligned}
$$

$\epsilon_{qh} \overset{iid}{\sim} N(0,h)$, $q = 1, \ldots, Q$, with $Qh = N$. Use $X_1, \ldots X_t$, $t = R, \ldots, R+P-1$ and $X_{k,j,h}^{\vartheta_k}$ $j = 1, \ldots, N$ to obtain an estimator $\widehat{\vartheta}_{k,t,N,h}$.

- We can now use $\widehat{\vartheta}_{k,t,N,h}$ to simulate paths under model k. To begin, generate $X^{\widehat{\vartheta}_{k,t,N,h}}_{k,t+\tau,i}(X_t)$, $i = 1, ..., N$, for $t = R, ..., P+R-\tau$ using at each replication i, the same set of randomly drawn errors and the same draws for number of jumps, jump times and jumps size, across t. Thus, only starting values used to initialize simulations change. Overall we simulate $(P-\tau)\times N$ paths of length τ.(The effect of the starting value approaches zero at an exponential rate, as $\tau \to \infty$.)

- Construct the empirical distribution of the simulated data, $\frac{1}{N}\sum_{i=1}^{N} 1\left\{X^{\widehat{\vartheta}_{k,t,N,h}}_{k,t+\tau,i}(X_t) \leq u\right\}$.

- Under mild regularity conditions get consistency

$$\frac{1}{N}\sum_{i=1}^{N} 1\left\{X^{\widehat{\vartheta}_{k,t,N,h}}_{k,t+\tau,i}(X_t) \leq u\right\} \xrightarrow{pr} F^{\tau}_{k}(u|X_t, \vartheta^{\dagger}_{k})$$

for $t = R, ..., T-\tau$

■ **Assumptions**: $X(t), t \in \Re^{+}$, is a strictly stationary, geometric ergodic $\beta-$mixing diffusion, moments on jump component, smoothness of drift and variance. For any fixed h and $\vartheta_k \in \Theta_k$, $X_{qh}^{\vartheta_k}$ is geometrically ergodic and strictly stationary, smoothness and domination condition for $X_{qh}^{\vartheta_k}$ uniformly on Θ_k. Also, as $P, R, N \to \infty$ and $h \to 0$, up to an $o_P(1)$ term

$$\frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \left(\widehat{\vartheta}_{k,t,N,h} - \vartheta_k^{\dagger}\right) = A_k^{\dagger} \frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \psi_{k,t,N,h}\left(\vartheta_k^{\dagger}\right)$$

$\frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \psi_{k,t,N,h}\left(\vartheta_k^{\dagger}\right) \xrightarrow{d} N\left(0, V_k^{\dagger}\right).$

4 Specification Testing

- $H_0 : F(u|X_t, \theta^\dagger) = F_0(u|X_t, \theta_0)$, for all u, a.s.

  H_A : Pr $\left(F(u|X_t, \theta^\dagger) - F_0(u|X_t, \theta_0) \neq 0\right) > 0$, for some u, with non-zero Lebesgue measure.

- Null coincides with correct specification of the conditional distribution, and is implied by correct specification of drift and variance terms used in simulating the paths.

- *Test statistic:*

$$V_T = \sup_{u \times v \in U \times V} |V_T(u,v)|$$

$$V_T(u,v) = \frac{1}{(T-\tau)^{1/2}} \cdot$$

$$\sum_{t=1}^{T-\tau} \left(\frac{1}{S} \sum_{s=1}^{S} 1 \left\{ X_{s,t+\tau}^{\widehat{\theta}_{T,N,h}} \leq u \right\} - 1\{X_{t+\tau} \leq u\} \right) 1 \left\{ X_t \leq v \right\},$$

where U and V are compact sets on the real line.

Theorem: *Let Assumptions A and B hold. Assume that $T, N, S \rightarrow \infty$. Then, if $h \rightarrow 0$, $T/N \rightarrow 0$, $T/S \rightarrow 0$, $T^2/S \rightarrow \infty$, $Nh \rightarrow 0$, and $h^2 T \rightarrow 0$, the following result holds under H_0 :*

$$V_T \xrightarrow{d} \sup_{u \times v \in U \times V} |Z(u,v)|,$$

where $Z(u,v)$ is a Gaussian process with covariance kernel $K(u,u',v,v')$.

■ *Critical Values:*

Step 1*:* At each replication, draw b blocks (with replacement) of length l, where $bl = T$. Thus, each block is equal to $X_{i+1}, ..., X_{i+l}$, for some $i = 0, ..., T-l+1$, with probability $1/(T-l+1)$. More formally, let I_k, $k = 1, ..., b$ be iid discrete uniform random variables on $[0, 1, ..., T-l+1]$. Then, the resampled series, X_t^* is such that $X_1^*, X_2^*, ..., X_l^*, X_{l+1}^*, ..., X_T^* = X_{I_1+1}, X_{I_1+2}, ..., X_{I_1+l}, X_{I_2}, ..., X_{I_b+l}$, and so a resampled series consists of b blocks that are discrete iid uniform random variables, conditional on the sample. Use these data to construct $\widehat{\theta}^*_{T,N,h}$. (note, e.g. as $N/T \to \infty$, GMM and SGMM are asymptotically equivalent).

Step 2*:* Using the same set of random errors used in the construction of the actual statistic, construct $X_{s,t+\tau,*}^{\widehat{\theta}^*_{T,N,h}}$, $s = 1, ..., S$, and $t = 1, ..., T-\tau$. Note that we do not resample the simulated series (as $S/T \to \infty$, simulation error is asymptotically negligible). Instead, simply simulate the series using bootstrap estimators and using bootstrapped values as starting values.

Step 3: Construct the following bootstrap statistic, which is the bootstrap counterpart of V_T:

$$V_T^* = \sup_{u \times v \in U \times V} \left| V_T^*(u,v) \right|,$$

where

$$V_T^*(u,v) = \frac{1}{(T-\tau)^{1/2}} \cdot$$

$$\sum_{t=1}^{T-\tau} \left(\frac{1}{S} \sum_{s=1}^{S} 1 \left\{ X_{s,t+\tau,*}^{\widehat{\theta}_{T,N,h}^*} \leq u \right\} - 1\{X_{t+\tau}^* \leq u\} \right) 1 \left\{ X_t^* \leq v \right\}$$

$$- \frac{1}{(T-\tau)^{1/2}} \cdot$$

$$\sum_{t=1}^{T-\tau} \left(\frac{1}{S} \sum_{s=1}^{S} 1 \left\{ X_{s,t+\tau}^{\widehat{\theta}_{T,N,h}} \leq u \right\} - 1\{X_{t+\tau} \leq u\} \right) 1 \left\{ X_t \leq v \right\}$$

Step 4: Repeat *Steps 1-3* B times, and generate the empirical distribution of the B bootstrap statistics.

Theorem: *Let Assumptions A and B hold. Assume that* $T, N, S \to \infty$. *Then, if* $h \to 0$, $T/N \to 0$, $T/S \to 0$, $T^2/S \to \infty$, $Nh \to 0$, $h^2 T \to 0$, $l \to \infty$, *and* $l^2/T \to 0$, *the following result holds*:

$$P\left[\omega : \sup_{x \in \Re} \left| P^*\left(V_T^*(\omega) \le x\right) - P\left(\left(V_T - E(V_T)\right) \le x\right)\right| > \varepsilon\right]$$

$$\to 0,$$

where P^* *denotes the probability law of the resampled series, conditional on the sample.*

Extension: Stochastic Volatility Models

Use a generalized Milstein scheme rather than simple approximation scheme.

Step 1: Simulate a path of length N using the scheme and estimate θ by SGMM. Also, retrieve $V_{kh}^{\widehat{\theta}_{T,N,h}}$, for $k = 1/h, ..., N/h$, and hence obtain $V_{j,h}^{\widehat{\theta}_{T,N,h}}$, $j = 1, ..., N$ (i.e. we sample the simulated volatility at the same frequency as the data).

Step 2: Simulate $S \times N$ paths of length τ, setting the initial value for the observable state variable to be X_t. As we do not observe data on volatility, use the values simulated in the previous step as the initial value for the volatility process (i.e. as initial values for unobservable state variable, use $V_{j,h}^{\widehat{\theta}_{T,N,h}}$, $j = 1, ..., N$). Also, keep the simulated randomness (i.e. $\epsilon_{1,kh}, \epsilon_{2,kh}$, $\int_{kh}^{(k+1)h} \left(\int_{kh}^{s} dW_{1,\tau} \right) dW_{2,s}$) constant across j (i.e. constant across the different starting values for the unobservable and observable state variable). Define $X_{j,s,t+\tau}^{\widehat{\theta}_{T,N,h}}$ to be the simulated $\tau-$step ahead value for the return series at replication s, and using initial values X_t and $V_{j,h}^{\widehat{\theta}_{T,N,h}}$.

Step 3: As an estimator of $F(u|X_t, \theta^{\dagger})$, construct $\frac{1}{NS} \sum_{j=1}^{N} \sum_{s=1}^{S} 1 \left\{ X_{j,s,t+\tau}^{\widehat{\theta}_{T,N,h}} \leq u \right\}$. Note that, by averaging over the initial value of the volatility process, we have integrated out its effect.

Step 4: Construct the statistic of interest:

$$SV_T = \sup_{u \times v \in U \times V} \left| SV_T(u,v) \right|,$$

where

$$SV_T(u,v) = \frac{1}{(T-\tau)^{1/2}} \cdot$$

$$\sum_{t=1}^{T-\tau} \left(\frac{1}{NS} \sum_{j=1}^{N} \sum_{s=1}^{S} 1 \left\{ X_{j,s,t+\tau}^{\widehat{\theta}_{T,N,h}} \leq u \right\} \right.$$

$$\left. -1\{X_{t+\tau} \leq u\} \right) 1 \left\{ X_t \leq v \right\}$$

- *All above results generalize to this setting.*

■ *Critical Values:*

Resample as above (no need to resample $V^{\theta}_{h,j}$). Then form bootstrap statistic:

$$SV^*_T = \sup_{u \times v \in U \times V} |SV^*_T(u,v)|,$$

$$SV^*_T(u,v) = \frac{1}{(T-\tau)^{1/2}} \cdot \sum_{t=1}^{T-\tau}$$

$$[(\frac{1}{NS}\sum_{j=1}^{N}\sum_{s=1}^{S} 1\left\{X^{\widehat{\theta}^*_{i,T,N,h}}_{j,s,t+\tau,*} \leq u\right\}$$
$$-1\{X^*_{t+\tau} \leq u\}\Big) 1\{X^*_t \leq v\}]$$

$$-[(\frac{1}{NS}\sum_{j=1}^{N}\sum_{s=1}^{S} 1\left\{X^{\widehat{\theta}_{i,T,N,h}}_{j,s,t+\tau} \leq u\right\}$$
$$-1\{X_{t+\tau} \leq u\}) 1\{X_t \leq v\}],$$

where $X^{\widehat{\theta}^*_{i,T,N,h}}_{j,s,t+\tau,*}$ is the simulated value at simulation s, constructed using $\widehat{\theta}^*_{i,T,N,h}$ and using as initial value X^*_t and $V^{\widehat{\theta}^*_{i,T,N,h}}_{j,h}$.

5 Predictive Density Tests

- For notational simplicity, set $u_1 = -\infty$, $u_2 = u$.

- $$H_0 : E_X\left(F_1^{\tau}(u|X_t, \vartheta_1^{\dagger})(u) - F_0^{\tau}(u|X_t)\right)^2$$

$$- E_X\left(F_k^{\tau}(u|X_t, \vartheta_k^{\dagger}) - F_0^{\tau}(u|X_t)\right)^2 = 0$$

or

$$H_0 : \max_{k=2,\ldots,m}\left(E_X\left(F_1^{\tau}(u|X_t, \vartheta_1^{\dagger}) - F_0^{\tau}(u|X_t)\right)^2\right.$$

$$\left.-E_X\left(F_k^{\tau}(u|X_t, \vartheta_k^{\dagger}) - F_0^{\tau}(u|X_t)\right)^2\right) \leq 0$$

versus H_A : negation of H_0.

■ *Test statistics:*

$$D_{k,P,N}(u) = \frac{1}{P^{1/2}} \sum_{t=R}^{T-\tau}$$

$$\left(\left[\frac{1}{N} \sum_{i=1}^{N} 1 \left\{ X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t) \leq u \right\} - 1\{X_{t+\tau} \leq u\} \right]^2 \right.$$

$$\left. - \left[\frac{1}{N} \sum_{i=1}^{N} 1 \left\{ X_{k,t+\tau,i}^{\widehat{\vartheta}_{k,t,N,h}}(X_t) \leq u \right\} - 1\{X_{t+\tau} \leq u\} \right]^2 \right),$$

or

$$D_{k,P,N}^{Max}(u) = \max_{k=2,\ldots,m} D_{k,P,N}(u)$$

Theorem: *Let Assumptions A1-A4 hold. Also assume that models* 1 *and* k *are nonnested. If as* $P, R, N \to \infty$, $h \to 0$, $P/N \to 0$, $h^2 P \to 0$, *and* $P/R \to \pi$, $0 < \pi < \infty$, *then under* H_0, $D_{k,P,N}(u) \xrightarrow{d} N(0, W_k(u))$, *and under* H_A, $\left| D_{k,P,N}(u) \right|$ *diverges at rate* $P^{1/2}$.

Theorem: *Let Assumptions A1-A4 hold. Also assume that models* 1 *and* k *are nonnested. If as* $P, R, N \to$

∞, $h \rightarrow 0$, $P/N \rightarrow 0$, $h^2P \rightarrow 0$, *and* $P/R \rightarrow \pi$, $0 < \pi < \infty$, *then:*

$$\max_{k=2,..,m}\left(D_{k,P,N}(u) - (\mu_1(u) - \mu_k(u))\right)$$

$$\xrightarrow{d} \max_{k=2,...,m} Z_k(u_1, u_2),$$

$$\mu_j(u) = E\left(\left(F_{X_{1,t+\tau}^{\vartheta_1^\dagger}(X_t)}(u) - F_0(u|X_t)\right)^2\right),$$

where $(Z_1(u), ..., Z_m(u))$ *is a* $m-$*dimensional Gaussian vector with covariance matrix with* kk *element given by* $W_k(u)$.

■ The covariance matrix $W_k(u)$ reflects the contribution of recursive parameter estimation error, captured by $\frac{1}{P^{1/2}}\sum_{t=R}^{T}\left(\widehat{\vartheta}_{k,t,N,h} - \vartheta_k^\dagger\right)$. Hence, the limiting distribution is not nuisance parameter free and we need to rely on bootstrap critical values.

■ Need bootstrap procedure able to mimic the limiting distribution of $\frac{1}{P^{1/2}}\sum_{t=R}^{T}\left(\widehat{\vartheta}_{k,t,N,h} - \vartheta_k^\dagger\right)$.

Critical Values:

In the recursive case, observations at the beginning of the sample are used more frequently than observations at the end of the sample. This introduces a location bias to the usual block bootstrap. Also, the bias term varies across samples and can be either positive or negative.

Corradi and Swanson (2007): $\{$boostrap $m-$estimators for recursive schemes$\}$ address the issue of bootstrapping SGMM estimators in a recursive setting.

Resample b blocks of length l from the full sample, with $lb = T$. For any given τ, we need to jointly resample $X_t, X_{t+1}, ..., X_{t+\tau}$. More precisely, let $Z^{t,\tau} = (X_t, X_{t+1}, ..., X_{t+\tau})$, $t = 1, ..., T - \tau$, we resample b overlapping blocks of length l from $Z^{t,\tau}$. This yields $Z^{t,*} = (X_t^*, X_{t+1}^*, ..., X_{t+\tau}^*)$, $t = 1, ..., T - \tau$. Use these data to construct $\widehat{\theta}^*_{k,t,N,h}$.

■ Assume that $\frac{1}{P^{1/2}} \sum_{t=R}^{T} \left(\widehat{\vartheta}^{*}_{k,t,N,h} - \widehat{\vartheta}_{k,t,N,h} \right)$ has the same limiting distribution as $\frac{1}{P^{1/2}} \sum_{t=R}^{T} \left(\widehat{\vartheta}_{k,t,N,h} - \vartheta^{\dagger}_{k} \right)$, conditional on sample.

■ Note that we used recentering around sample mean calculated over full sample. This ensures that

$E^{*} \left(G^{*}_{k,t,N,h}(\widehat{\theta}_{k,t,N,h}) \right) = O(l/T)$ (e.g. bootstrap moment conditions used in SGMM have zero mean, up to negligble terms).

■ As $N/R, N/P \to \infty$, do not need to resample simulated series. Simulation error asymptotically negligible.

■ Let $X^{\widehat{\theta}_{k,t,N,h}}_{k,t+\tau,i}(X_t) = X^{N}_{k,t+\tau}$, $X^{\widehat{\theta}^{*}_{k,t,N,h}}_{k,t+\tau,i}(X^{*}_{t}) = X^{*,N}_{k,t+\tau}$,

$$X_{k,t+\tau,i}^{\widehat{\theta}_{k,t,N,h}}(X_j) = X_{k,j+\tau}^{N}.$$

$$D_{k,P,N}^{*}(u) = \frac{1}{P^{1/2}} \sum_{t=R}^{T-\tau}$$
$$\{([\frac{1}{N}\sum_{i=1}^{N} 1\left\{X_{1,t+\tau}^{*,N} \leq u\right\} - 1\{X_{t+\tau}^{*} \leq u\}\right]^2$$
$$-[\frac{1}{T}\sum_{j=1}^{T}[\frac{1}{N}\sum_{i=1}^{N} 1\left\{X_{1,j+\tau}^{N} \leq u\right\} - 1\{X_{j+\tau} \leq u\}]^2])$$

$$-([\frac{1}{N}\sum_{i=1}^{N} 1\left\{X_{k,t+\tau}^{*,N} \leq u\right\} - 1\{X_{t+\tau}^{*} \leq u\}\right]^2$$
$$-[\frac{1}{T}\sum_{j=1}^{T}[\frac{1}{N}\sum_{i=1}^{N} 1\left\{X_{k,j+\tau}^{N} \leq u\right\} - 1\{X_{j+\tau} \leq u\}]^2])\}$$

- Note that each bootstrap term is recentered around the (full) sample mean.

- This is necessary as the bootstrap statistic is constructed using the last P resampled observations, which in turn have been resampled from the full sample.

■ If $P/R \rightarrow 0$, then we do not need to mimic parameter estimation error, and so could simply use $\widehat{\theta}_{k,t,N,h}$ instead of $\widehat{\theta}^*_{k,t,N,h}$, but we still need to recenter any bootstrap term around the (full) sample mean.

Theorem: *Let Assumptions A1-A5 hold. Also assume that models* 1 *and* k *are nonnested. If as* $P, R, N \rightarrow \infty$, $h \rightarrow 0$, $P/N \rightarrow 0$, $h^2 P \rightarrow 0$, $l \rightarrow \infty$, $l/T^{1/4} \rightarrow 0$, *and* $P/R \rightarrow \pi$, $0 < \pi < \infty$, *then:*

$$P\left(\omega : \sup_{v \in \Re^{\varrho}} \left| P_T^*\left(\max_{k=2,\ldots,m} D^*_{(\cdot)}(u) \leq v\right) \right.\right.$$

$$\left.\left. - P\left(\max_{k=2,\ldots,m}\left(D_{(\cdot)}(u) - [\mu_1(u) - \mu_k(u)]\right) \leq v\right) \right| > \varepsilon\right)$$

$$\rightarrow 0$$

■ Boot CV provides test with correct asymptotic size for least favorable case under H_0; $\mu_1(u) - \mu_k(u) = 0$, for all k. CVs upper bounds whenever $\mu_1(u) - \mu_k(u) < 0$, for some k.

■ Extension: Stochastic Volatility Models:

■ Estimation when the volatility process is not observable.

■ Constructing and evaluating predictive densities when the volatility process is not observable.

■ In the one-dimensional case, the diffusion process $X(t)$ can be expressed as a function of the driving Brownian motion $W(t)$. In the multidimensional case, $X(t)$ is function of $(W_v(t), \int_0^t W_v(s) dW_w(s)$, $v, w = 1, ...p$ (Pardoux and Talay (1985)), unless the covariance matrix is commutative. Typical stochastic volatility with leverage are not commutative. In this case one has to approximate stochastic integral when simulating the return path, see Kloeden and Platen (1999).

■ **SGMM** does not require conditioning on observables, thus SGMM for SV models carries through as in one factor case, matching moments of observables. Problem with exact identification harder here.

NPSQMLE. For each $\theta_k \in \Theta_k$, simulate $N \times S \times T$ draws for volatility, $V_s^{\theta_k}$, $s = 1, ..., S, ..., S \times T$, using the volatility equation under model k. Then, we generate $S \times T$ paths of length one for the return process, using observables and $V_s^{\theta_k}$ as initial values. Basically, for any initial value X_t we simulate $N \times S$ paths of length one, using $V_s^{\theta_k}$, $tS \leq s < (t+1)S$. Define,

$$
\begin{aligned}
&\widetilde{f}_{k,N,S,h}\left(X_t|X_{t-1},\theta_k\right) \\
&= \frac{1}{S}\sum_{s=S(t-1)}^{St-1}\frac{1}{N\xi_N}\sum_{j'=1}^{N}K\left(\frac{X_{t,j'}^{\theta_k}(X_{t-1},V_{s-1}^{\theta_k})-X_t}{\xi_N}\right).
\end{aligned}
$$

Finally, define for $t \geq R$,

$$
\begin{aligned}
\widetilde{\theta}_{k,t,N,S,h} &= \arg\min_{\theta_k\in\Theta_k}\frac{1}{t}\sum_{l=2}^{t}\log\widetilde{f}_{k,N,S,h}\left(X_l|X_{l-1},\theta_k\right) \\
&\quad \times\tau_N\left(\widetilde{f}_{k,N,S,h}\left(X_l|X_{l-1},\theta_k\right)\right)
\end{aligned}
$$

■ Simulate $(P-\tau)\times S\times N$ paths of lenght τ setting the initial values for the observable state variable equal to the initial value X_t, $t=R+1,...,R+P-\tau$ and for each X_t, using the S different starting values for volatility (i.e. $V_j^{\widehat{\theta}_{k,t,N,h}}$, $j=1,...,S$). For any initial value X_t and $V_j^{\widehat{\theta}_{k,t,N,h}}$, $t=R+1,...,R+P-\tau$ and $j=1,...,S$ we generate N independent paths of length τ.

■ We keep the simulated randomness constant across the different starting values for the unobservable and observable state variables.

Call $X_{k,t+\tau,i,j}^{\widehat{\theta}_{k,t,N,h}}(X_t,V_j^{\widehat{\theta}_{k,t,N,h}})$. the $\tau-$step ahead, simulated (under model k), value for the return series, at replication i, $i=1,...,N$ using initial values X_t and $V_j^{\widehat{\theta}_{k,t,N,h}}$.

■ As an estimator of $F_{X_{k,t+\tau}^{\theta_k^\dagger}(X_t,V_j^{\widehat{\theta}_{k,t,N,h}})}(u)$, construct:

$\frac{1}{NS}\sum_{j=1}^{S}\sum_{i=1}^{N} 1\left\{X_{k,t+\tau,i,j}^{\widehat{\theta}_{k,t,N,h}}(X_t, V_{k,t,j}^{\widehat{\theta}_{k,t,N,h}}) \leq u\right\}$

■ By averaging across different starting values for volatility, while simulated randomness is kept constant across different starting values for observable and unobservable, we integrate out the effect of initial volatility.

$$DV_{k,P,S,N}(u_1,u_2) = \frac{1}{P^{1/2}}\times$$

$$\sum_{t=R}^{T-\tau}\left(\left(\frac{1}{NS}\sum_{j=1}^{S}\sum_{i=1}^{N} 1\left\{X_{1,t+\tau,i,j}^{\widehat{\theta}_{1,t,N,h}}(X_t, V_{1,j}^{\widehat{\theta}_{1,t,N,h}}) \leq u\right\}\right.\right.$$
$$\left.-1\{X_{t+\tau} \leq u\}\right)^2$$

$$-\left(\frac{1}{NS}\sum_{j=1}^{S}\sum_{i=1}^{N} 1\left\{X_{k,t+\tau,i,j}^{\widehat{\theta}_{k,t,N,h}}(X_t, V_{k,j}^{\widehat{\theta}_{k,t,N,h}}) \leq u\right\}\right.$$
$$\left.\left.-1\{X_{t+\tau} \leq u\}\right)^2\right)$$

Statement in Theorems above apply to $DV_{k,P,S,N}(u)$ and to its bootstrap counterpart.

6 SGMM and NPSQMLE

■ *Bootstrap SGMM:*

$$\widehat{\vartheta}^{*}_{k,t,N,h} = \arg\min_{\theta_k \in \Theta_k} G^{*}_{k,t,N,h}(\vartheta_k)' \widehat{\Omega}^{*}_{k,t} G^{*}_{k,t,N,h}(\vartheta_k)$$

where

$$\begin{aligned} & G^{*}_{k,t,N,h}(\vartheta_k) \\ = \; & \frac{1}{t}\sum_{j=1}^{t}\left(\left(g_k(X^{*}_{j}) - \frac{1}{T}\sum_{j'=1}^{T} g_k(X_{j'})\right)\right. \\ & \left.- \left(\frac{1}{N}\sum_{j=1}^{N} g_k(X^{\vartheta_k}_{j,h}) - \frac{1}{N}\sum_{j=1}^{N} g_k(X^{\widehat{\vartheta}_{k,t,N,h}}_{j,h})\right)\right) \end{aligned}$$

■ This recentering ensures that $E^{*}\left(G^{*}_{k,t,N,h}(\widehat{\vartheta}_{k,t,N,h})\right) = O(l/T)$ (i.e. bootstrap moments conditions have zero mean, up to negligble term).

As $N/R, N/P \rightarrow \infty$, do not need to resample simulated series. Simulation error asymptotically negligible.

■ Under mild regularity conditions,

$\frac{1}{P^{1/2}}\sum_{t=R}^{T}\left(\widehat{\vartheta}^{*}_{k,t,N,h}-\widehat{\vartheta}_{k,t,N,h}\right)$ has the same limiting distribution as

$\frac{1}{P^{1/2}}\sum_{t=R}^{T}\left(\widehat{\vartheta}_{k,t,N,h}-\vartheta^{\dagger}_{k}\right)$, conditional on sample.

■ *Bootstrap NPSQMLE:*

For each simulation replication, generate $N\times(T-1)$ paths of length one, using as starting values $X^{*}_{1},...,X^{*}_{T-1}$; and so obtaining $X^{\vartheta_k}_{k,t,j}(X^{*}_{t-1})$, for $t=2,...,T-1$, $j=1,...,N$. Let

$$\begin{aligned} & \widetilde{f}^{*}_{k,N,h}\left(X^{*}_{t}|X^{*}_{t-1},\vartheta_k\right) \\ = \ & \frac{1}{N\xi_N}\sum_{j=1}^{N}K\left(\frac{X^{\vartheta_k}_{t,j,h}(X^{*}_{t-1})-X^{*}_{t}}{\xi_N}\right), \end{aligned}$$

Define for $t = R, ..., R + P - 1$

$$
\begin{aligned}
\widetilde{\vartheta}^*_{k,t,N,h} &= \arg \max_{\vartheta_k \in \Theta_k} \frac{1}{t} \sum_{l=2}^{t} \Big(\log \widetilde{f}_{k,N,h} \left(X^*_l | X^*_{l-1}, \vartheta_k \right) \\
&\quad \times \tau_N \left(\widetilde{f}_{k,N,h} \left(X^*_l | X^*_{l-1}, \vartheta_k \right) \right) \\
&\quad - \vartheta'_k \left(\frac{1}{T} \sum_{l'=2}^{T} \frac{\nabla_{\vartheta_k} \widetilde{f}_{k,N,h} \left(X_{l'} | X_{l'-1}, \widetilde{\vartheta}_{k,t,N,h} \right)}{\widetilde{f}_{k,N,h} \left(X_{s'} | X_{t-s'}, \widetilde{\vartheta}_{k,t,N,h} \right)} \right. \\
&\quad \times \tau_N \left(\widetilde{f}_{k,N,h} \left(X_{l'} | X_{l'-1}, \widetilde{\vartheta}_{k,t,N,h} \right) \right) \\
&\quad + \tau'_N \left(\widetilde{f}_{k,N,h} \left(X_{l'} | X_{l'-1}, \widetilde{\vartheta}_{k,t,N,h} \right) \right) \\
&\quad \times \nabla_{\vartheta_k} \widetilde{f}_{k,N,h} \left(X_{l'} | X_{l'-1}, \widetilde{\vartheta}_{k,t,N,h} \right) \\
&\quad \left. \log \widetilde{f}_{k,N,h} \left(X_{l'} | X_{l'-1}, \widetilde{\vartheta}_{k,t,N,h} \right) \right) \Big)
\end{aligned}
$$

where $\tau'_N(\cdot)$ denotes the derivative with respect to its argument. Note that each term in the simulated likelihood is recentered around the (full) sample mean of the score, evaluated at $\widetilde{\vartheta}_{k,t,N,h}$. This ensure that the bootstrap score, conditional on the sample, has mean zero.

- The recentering term requires the knowledge of

$\nabla_{\theta_k} \widetilde{f}_{k,N,h}\left(X_{l'}|X_{l'-1}, \widetilde{\vartheta}_{k,t,N,h}\right)$, which is not known in closed form. Nevertheless, it can be computed numerically, simply taking the numerical derivative of the simulated likelihood.

- Under analogous conditions as used above,

$\frac{1}{P^{1/2}} \sum_{t=R}^{T} \left(\widetilde{\vartheta}^{*}_{k,t,N,h} - \widehat{\vartheta}_{k,t,N,h}\right)$ has the same limiting distribution as

$\frac{1}{P^{1/2}} \sum_{t=R}^{T} \left(\widetilde{\vartheta}_{k,t,N,h} - \vartheta^{\ddagger}_{k}\right)$, conditional on sample.

7 Monte Carlo and Empirical Results

■ *Monte Carlo using the following specification test statistic*

$$V_T(v) = \frac{1}{(T-\tau)^{1/2}} \cdot \sum_{t=1}^{T-\tau} \left(\frac{1}{S} \sum_{s=1}^{S} 1 \left\{ u_{low} \leq X_{s,t+\tau}^{\widehat{\theta}_{T,N,h}} \leq u^{up} \right\} \right.$$

$$\left. - 1\{u_{low} \leq X_{t+\tau} \leq u^{up}\} \right) 1 \{X_t \leq v\}$$

■ *Also, construct* $D^{Max}_{k,P,N}(u_1,u_2)$ *and* $DV^{Max}_{k,P,S,N}(u_1,u_2)$ *in empirical excercises. In our tables, we also report the so-called "predictive density" mean square forecast errors* $(PDMSFE)$ *terms in these statistics, which are constructed using the following formulae:*

$$\frac{1}{P}\sum_{t=R}^{T-\tau}\Big(\frac{1}{NS}\sum_{j=1}^{S}\sum_{i=1}^{N}1\left\{u_1 \leq X^{\widehat{\theta}_{1,t,N,S,h}}_{1,t+\tau,i,j}(X_t, V^{\widehat{\theta}_{1,t,N,S,h}}_{1,j}) \leq u_2\right\}$$

$$-1\{u_1 \leq X_{t+\tau} \leq u_2\}\Big)^2$$

and

$$\frac{1}{P}\sum_{t=R}^{T-\tau}\Big(\frac{1}{N}\sum_{i=1}^{N}1\left\{u_1 \leq X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_t) \leq u_2\right\}$$

$$-1\{u_1 \leq X_{t+\tau} \leq u_2\}\Big)^2,$$

CIR: $dX(t) = \kappa_1 (\alpha_1 - X(t))\, dt + \gamma_1 X^{1/2}(t) dW_1(t)$, where $\kappa_1 > 0$, $\gamma_1 > 0$ and $2\kappa_1\alpha_1 \geq \gamma_1^2$,

SV: $dX(t) = \kappa_2 (\alpha_2 - X(t))\, dt + V^{1/2}(t) dW_1(t)$, and $dV(t) = \kappa_3 (\alpha_3 - V(t))\, dt + \gamma_2 V^{1/2}(t) dW_2(t)$, where $W_1(t)$ and $W_2(t)$ are independent Brownian motions, and where $\kappa_2 > 0$, $\kappa_3 > 0$, $\gamma_2 > 0$, and $2\kappa_3\alpha_3 \geq \gamma_2^2$.

SVJ: $dX(t) = \kappa_4 (\alpha_4 - X(t))\, dt + V^{1/2}(t) dW_1(t) + J_u dq_u - J_d dq_d$, and $dV(t) = \kappa_5 (\alpha_5 - V(t))\, dt + \gamma_3 V^{1/2}(t) dW_2(t)$, where $W_r(t)$ and $W_v(t)$ are independent Brownian motions, and where $\kappa_4 > 0$, $\kappa_5 > 0$, $\gamma_3 > 0$, and $2\kappa_5\alpha_5 \geq \gamma_3^2$. Further q_u and q_d are Poisson processes with jump intensity λ_u and λ_d, and are independent of the Brownian motions $W_1(t)$ and $W_2(t)$. Jump sizes are iid and are controlled by jump magnitudes $\zeta_u, \zeta_d > 0$, which are drawn from exponential distributions, with densities: $f(J_u) = \frac{1}{\zeta_u} \exp\left(-\frac{J_u}{\zeta_u}\right)$ and $f(J_d) = \frac{1}{\zeta_d} \exp\left(-\frac{J_d}{\zeta_d}\right)$. Here, λ_u is the probability of a jump up, $\Pr(dq_u(t) = 1) = \lambda_u$, and jump up size is controlled by J_u; while λ_d and J_d control jump down intensity and size.

- Specification test experiments calibrated using one-month Eurodollar deposit rate for the periods January 6, 1971 - April 8, 2005 (1,789 weekly observations) and January 3, 1990 - April 8, 2005 (798 observations). Model selection empirics carried out using two samples of weekly data, one from January 6, 1989 - December 31, 1998 (526 observations) and one from January 8, 1999 - April 30, 2008 (491 observations), chosen arbitrarily. The variable that we model is the effective (or market) federal funds rate (i.e. the inter-bank interest rate), measured at the close.

8 Concluding Remarks

- *Simulation based procedures convenient and easy to apply.*

- *Block bootstrap generalizes nicely to such cases as those considered here.*

- *Results useful for in-sample specification tests as well as in out-of-sample contexts with recursively estimated parameters.*

- *Identification remains an issue.*

Table 1: Predictive Density Model Selection Test Results

Sample period January 6, 1989 - December 31, 1998

(CIR model is the benchmark, bootstrap block length=5)

τ	u_1, u_2	$D^{Max}_{k,P,S,N}(u_1,u_2)$	$PDMSFE_{CIR}$	$PDMSFE_{SV}$	$PDMSFE_{SVJ}$	10% CV
1	$\overline{X} \pm 0.5\sigma_X$	2.82927*	5.66205	3.62009	2.83278	1.65848
	$\overline{X} \pm \sigma_X$	1.31996	1.58636	0.3691	0.2664	1.64695
2	$\overline{X} \pm 0.5\sigma_X$	1.57134*	4.13194	2.62781	2.56061	0.85015
	$\overline{X} \pm \sigma_X$	0.53925	0.85434	0.34105	0.31509	0.8354
3	$\overline{X} \pm 0.5\sigma_X$	0.80223*	4.26257	3.87959	3.46034	0.20535
	$\overline{X} \pm \sigma_X$	1.19189*	1.82012	0.93572	0.62823	0.40461
4	$\overline{X} \pm 0.5\sigma_X$	1.23058*	4.32896	3.82788	3.09838	0.28591
	$\overline{X} \pm \sigma_X$	0.48079*	1.02194	0.76792	0.54115	0.28204
5	$\overline{X} \pm 0.5\sigma_X$	-0.00077	3.71976	3.72053	3.97788	0.2032
	$\overline{X} \pm \sigma_X$	0.18502	1.09725	1.01962	0.91223	0.2164
6	$\overline{X} \pm 0.5\sigma_X$	1.52213*	4.949	3.83724	3.42687	0.08187
	$\overline{X} \pm \sigma_X$	0.58406*	1.63659	1.05253	1.18955	0.12362
12	$\overline{X} \pm 0.5\sigma_X$	0.56293*	4.58393	4.37846	4.021	0.03085
	$\overline{X} \pm \sigma_X$	0.41295*	1.30048	1.5585	0.88753	0.01912

(*) Notes: Numerical entries in the table are test statistics, predicitve density type $PDMSFE$s (see Section 7 for further discussion), and associated bootstrap critical values, constructed using intervals given in the second column of the table, and for predictive horizons, τ $=1,2,3,4,5,6,12$. Starred entries denote rejection of the null hypothesis that the CIR model yields predictive densities at least as accurate as the competitor SV and SVJ models. Weekly data are used in all estimations, and the sample period across which predictive densities are constructed is $T/2$, where T is the sample size. Predictive densities are constructed using simulations of length $S = 10T$. Empirical bootstrap distributions are constructed using 100 bootstrap replications, and critical values are reported for the 95^{th}, 90^{th}, 85^{th}, and 80^{th} percentiles of the bootstrap distribution. $\overline{X}$ and σ_X are the mean and variance of an initial sample of data used in the first in-sample estimation, prior to the construction of the first predictive density (i.e. using $T/2$ observations). Finally, the predictive density type "mean square forecast errors" ($MSFE$s) reported in the fourth through sixth columns of the table are defined above, and reported entries are multiplied by $P^{1/2}$, where $P = T/2$ is the *ex ante* prediction period.

Table 2: Predictive Density Model Selection Test Results

Sample period January 6, 1989 - December 31, 1998

(CIR model is the benchmark, bootstrap block length=10)

τ	u_1, u_2	$D^{Max}_{k,P,S,N}(u_1, u_2)$	$PDMSFE_{CIR}$	$PDMSFE_{SV}$	$PDMSFE_{SVJ}$	10% CV
1	$\overline{X} \pm 0.5\sigma_X$	2.82927*	5.66205	3.62009	2.83278	1.87189
	$\overline{X} \pm \sigma_X$	1.31996	1.58636	0.3691	0.2664	1.94914
2	$\overline{X} \pm 0.5\sigma_X$	1.57134*	4.13194	2.62781	2.56061	1.12574
	$\overline{X} \pm \sigma_X$	0.53925	0.85434	0.34105	0.31509	1.12383
3	$\overline{X} \pm 0.5\sigma_X$	0.80223*	4.26257	3.87959	3.46034	0.26336
	$\overline{X} \pm \sigma_X$	1.19189*	1.82012	0.93572	0.62823	0.61716
4	$\overline{X} \pm 0.5\sigma_X$	1.23058*	4.32896	3.82788	3.09838	0.31387
	$\overline{X} \pm \sigma_X$	0.48079*	1.02194	0.76792	0.54115	0.45501
5	$\overline{X} \pm 0.5\sigma_X$	-0.00077	3.71976	3.72053	3.97788	0.18285
	$\overline{X} \pm \sigma_X$	0.18502	1.09725	1.01962	0.91223	0.29925
6	$\overline{X} \pm 0.5\sigma_X$	1.52213*	4.949	3.83724	3.42687	0.10103
	$\overline{X} \pm \sigma_X$	0.58406*	1.63659	1.05253	1.18955	0.14107
12	$\overline{X} \pm 0.5\sigma_X$	0.56293*	4.58393	4.37846	4.021	0.04347
	$\overline{X} \pm \sigma_X$	0.41295*	1.30048	1.5585	0.88753	0.03183

(*) Notes: see Table 1

Table 3: Predictive Density Model Selection Test Results

Sample period January 8, 1999 - April 30, 2008

(CIR model is the benchmark, bootstrap block length=5)

τ	u_1, u_2	$D^{Max}_{k,P,S,N}(u_1, u_2)$	$PDMSFE_{CIR}$	$PDMSFE_{SV}$	$PDMSFE_{SVJ}$	10% CV
1	$\overline{X} \pm 0.5\sigma_X$	3.36528*	3.93191	0.56663	2.35979	2.31001
	$\overline{X} \pm \sigma_X$	0.39113	0.39172	0.00059	0.13535	1.99902
2	$\overline{X} \pm 0.5\sigma_X$	1.8218*	2.32377	0.50197	2.04596	1.71781
	$\overline{X} \pm \sigma_X$	0.59514	0.60979	0.01464	0.26331	2.09447
3	$\overline{X} \pm 0.5\sigma_X$	1.2709	1.86856	0.59766	2.29788	1.33248
	$\overline{X} \pm \sigma_X$	0.97425	1.04645	0.0722	0.46272	1.77604
4	$\overline{X} \pm 0.5\sigma_X$	1.33461*	1.86611	0.5315	2.50816	1.03895
	$\overline{X} \pm \sigma_X$	0.59446	0.78217	0.18771	0.23341	1.31151
5	$\overline{X} \pm 0.5\sigma_X$	1.55731*	1.92318	0.36586	2.3208	0.72157
	$\overline{X} \pm \sigma_X$	0.62454*	0.92698	0.30244	0.42899	0.91251
6	$\overline{X} \pm 0.5\sigma_X$	1.07981	1.5355	0.45569	2.23224	0.81358
	$\overline{X} \pm \sigma_X$	1.0877*	1.3928	0.39654	0.3051	0.88946
12	$\overline{X} \pm 0.5\sigma_X$	1.06647*	1.72738	0.66091	2.59892	0.7709
	$\overline{X} \pm \sigma_X$	0.74472*	0.9282	0.43853	0.18348	0.73613

(*) Notes: see Table 1

Table 4: Predictive Density Model Selection Test Results

Sample period January 8, 1999 - April 30, 2008

(CIR model is the benchmark, bootstrap block length=10)

τ	u_1, u_2	$D^{Max}_{k,P,S,N}(u_1, u_2)$	$PDMSFE_{CIR}$	$PDMSFE_{SV}$	$PDMSFE_{SVJ}$	10% CV
1	$\overline{X} \pm 0.5\sigma_X$	3.36528*	3.93191	0.56663	2.35979	2.79456
	$\overline{X} \pm \sigma_X$	0.39113	0.39172	0.00059	0.13535	2.30575
2	$\overline{X} \pm 0.5\sigma_X$	1.8218	2.32377	0.50197	2.04596	2.41921
	$\overline{X} \pm \sigma_X$	0.59514	0.60979	0.01464	0.26331	2.67829
3	$\overline{X} \pm 0.5\sigma_X$	1.2709	1.86856	0.59766	2.29788	2.25422
	$\overline{X} \pm \sigma_X$	0.97425	1.04645	0.0722	0.46272	2.8359
4	$\overline{X} \pm 0.5\sigma_X$	1.33461	1.86611	0.5315	2.50816	1.91697
	$\overline{X} \pm \sigma_X$	0.59446	0.78217	0.18771	0.23341	2.56512
5	$\overline{X} \pm 0.5\sigma_X$	1.55731	1.92318	0.36586	2.3208	1.80572
	$\overline{X} \pm \sigma_X$	0.62454	0.92698	0.30244	0.42899	2.30651
6	$\overline{X} \pm 0.5\sigma_X$	1.07981	1.5355	0.45569	2.23224	1.64939
	$\overline{X} \pm \sigma_X$	1.0877	1.3928	0.39654	0.3051	2.08945
12	$\overline{X} \pm 0.5\sigma_X$	1.06647*	1.72738	0.66091	2.59892	1.00359
	$\overline{X} \pm \sigma_X$	0.74472	0.9282	0.43853	0.18348	0.98574

(*) Notes: see Table 1

Figure 1: Predictive Densities for CIR, SV and SVJ Models - 01:1989-12:1998

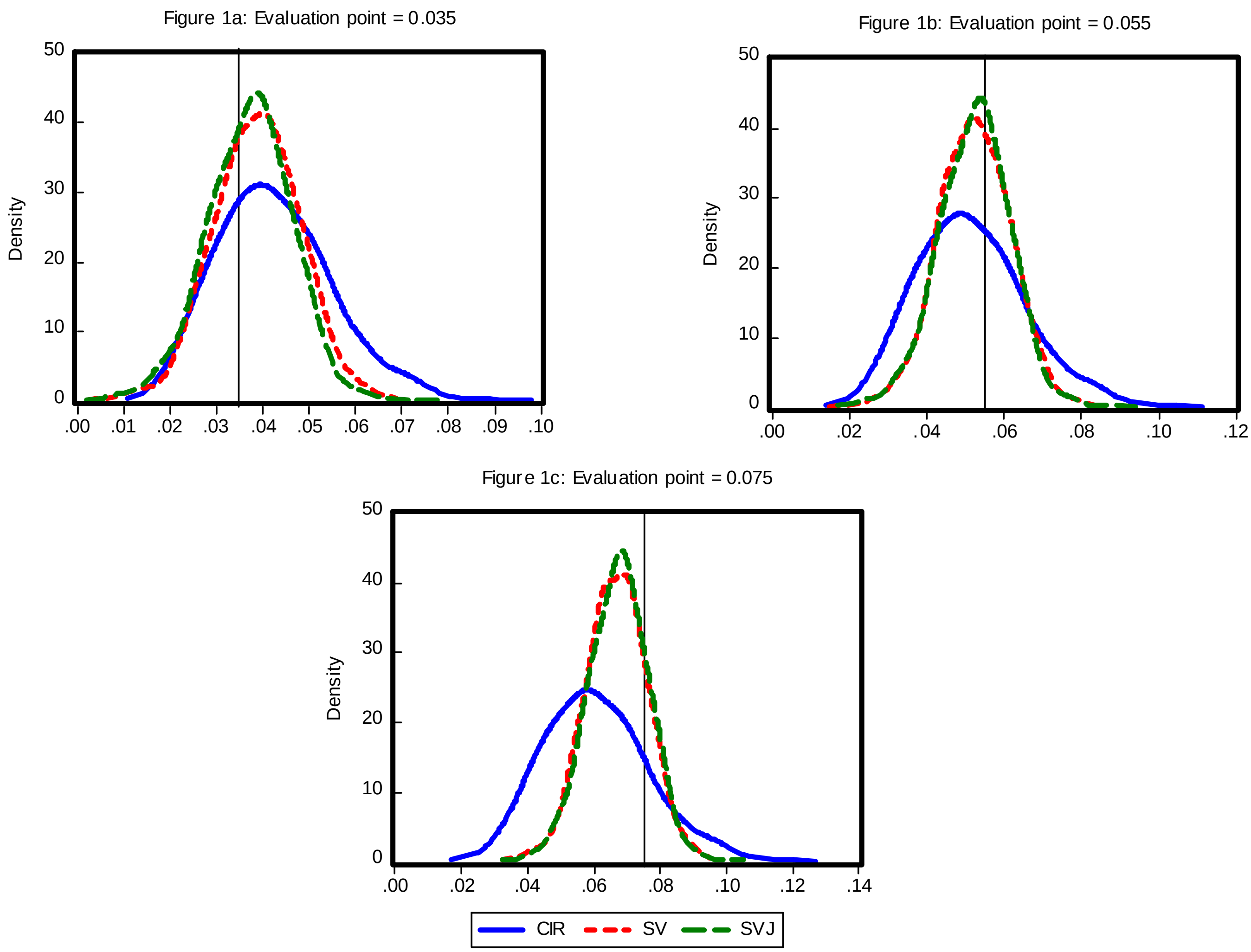

Figure 2: Predictive Densities for CIR, SV and SVJ Models - 01:1999-04:2008

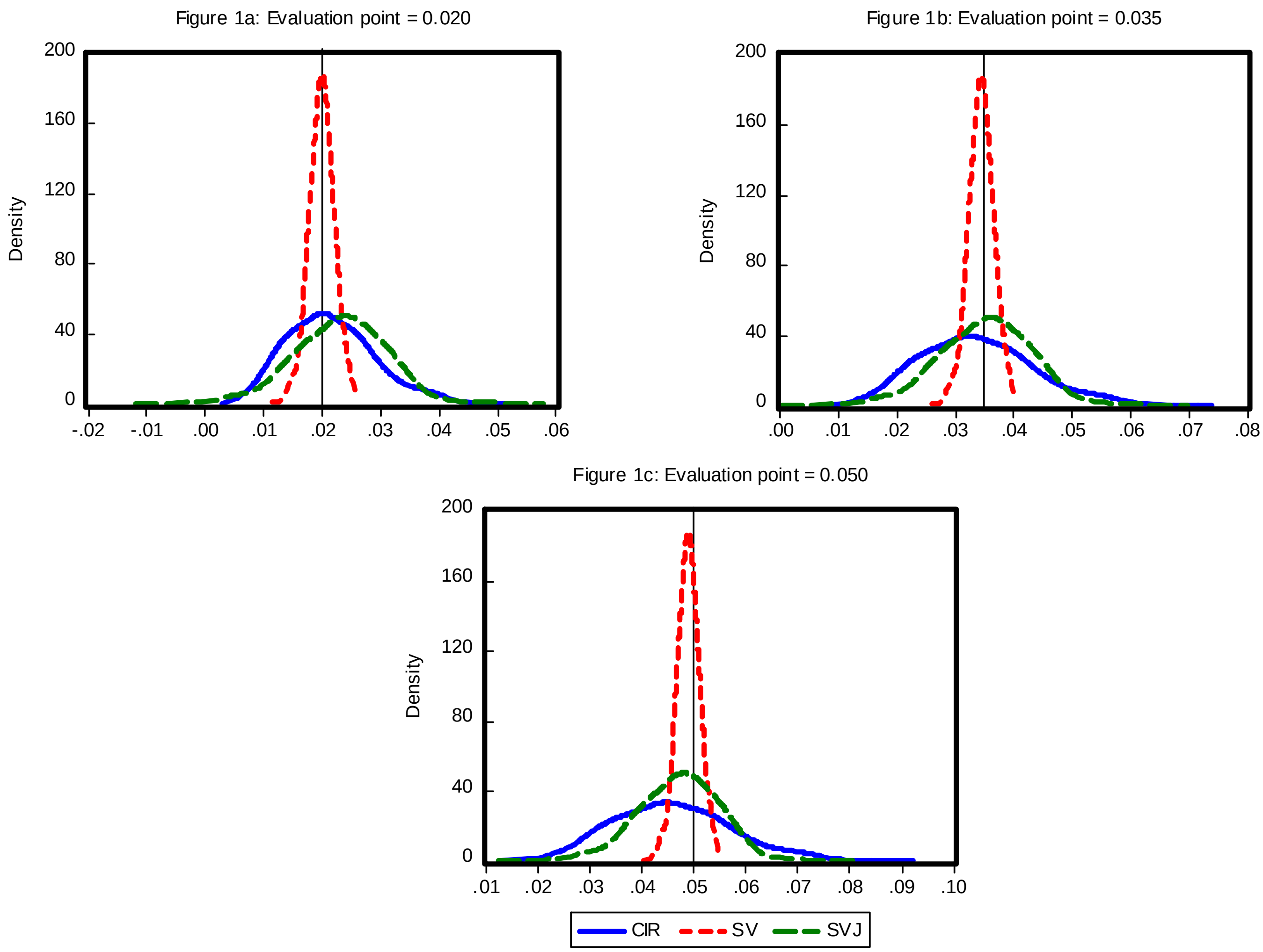

Table 1: Specification Test Rejection Frequencies for the One Factor Model - Empirical Size[(*)]

Data Generated using the CIR $_{(0.15, 0.05, 0.10)}$ Model

τ	$(\underline{u}, \overline{u})$	S,l					
		10T,10	20T,10	10T,20	20T,20	10T,50	20T,50
		Panel A: $T = 400$					
1	$\overline{X} \pm 0.5\sigma_X$	0.1559	0.2383	0.177	0.1503	0.2191	0.2831
	$\overline{X} \pm \sigma_X$	0.1559	0.2211	0.1638	0.1539	0.2105	0.1611
2	$\overline{X} \pm 0.5\sigma_X$	0.1796	0.2534	0.2303	0.2827	0.1492	0.1441
	$\overline{X} \pm \sigma_X$	0.1459	0.2009	0.1854	0.2256	0.1040	0.1468
4	$\overline{X} \pm 0.5\sigma_X$	0.1826	0.1963	0.1849	0.1341	0.1560	0.2107
	$\overline{X} \pm \sigma_X$	0.1933	0.1545	0.1868	0.1658	0.1158	0.1484
12	$\overline{X} \pm 0.5\sigma_X$	0.2142	0.2697	0.2372	0.283	0.3113	0.2443
	$\overline{X} \pm \sigma_X$	0.1374	0.1804	0.2836	0.2439	0.2666	0.1949
		Panel B: $T = 800$					
1	$\overline{X} \pm 0.5\sigma_X$	0.1149	0.1266	0.1267	0.1197	0.1260	0.1177
	$\overline{X} \pm \sigma_X$	0.1313	0.1199	0.1203	0.1291	0.1110	0.1513
2	$\overline{X} \pm 0.5\sigma_X$	0.1332	0.1058	0.1151	0.1291	0.1313	0.1110
	$\overline{X} \pm \sigma_X$	0.1462	0.1068	0.1217	0.1383	0.1388	0.1078
4	$\overline{X} \pm 0.5\sigma_X$	0.1089	0.1327	0.1193	0.1227	0.1274	0.1016
	$\overline{X} \pm \sigma_X$	0.1222	0.1056	0.1231	0.1071	0.1275	0.1210
12	$\overline{X} \pm 0.5\sigma_X$	0.1373	0.1409	0.132	0.1154	0.1028	0.1225
	$\overline{X} \pm \sigma_X$	0.1269	0.1369	0.134	0.1091	0.1032	0.1067

Data Generated using the CIR $_{(0.30, 0.05, 0.10)}$ Model

τ	$(\underline{u}, \overline{u})$	S,l					
		10T,10	20T,10	10T,20	20T,20	10T,50	20T,50
		Panel A: $T = 400$					
1	$\overline{X} \pm 0.5\sigma_X$	0.1804	0.2433	0.1747	0.1523	0.2382	0.2848
	$\overline{X} \pm \sigma_X$	0.1744	0.2204	0.1868	0.1649	0.2325	0.1781
2	$\overline{X} \pm 0.5\sigma_X$	0.1715	0.2772	0.2502	0.2887	0.1509	0.1577
	$\overline{X} \pm \sigma_X$	0.1654	0.2084	0.2077	0.2391	0.1381	0.1714
4	$\overline{X} \pm 0.5\sigma_X$	0.1971	0.2010	0.1725	0.1562	0.1455	0.2260
	$\overline{X} \pm \sigma_X$	0.2085	0.1609	0.1998	0.1770	0.1316	0.1692
12	$\overline{X} \pm 0.5\sigma_X$	0.2199	0.2681	0.2626	0.2988	0.3112	0.2617
	$\overline{X} \pm \sigma_X$	0.1588	0.1869	0.2881	0.2478	0.2824	0.1994
		Panel B: $T = 800$					
1	$\overline{X} \pm 0.5\sigma_X$	0.1205	0.1291	0.1382	0.1203	0.1412	0.1106
	$\overline{X} \pm \sigma_X$	0.1291	0.1199	0.1253	0.1447	0.1133	0.1417
2	$\overline{X} \pm 0.5\sigma_X$	0.1354	0.1118	0.1198	0.1265	0.1390	0.1180
	$\overline{X} \pm \sigma_X$	0.1397	0.1248	0.1139	0.1313	0.1443	0.1168
4	$\overline{X} \pm 0.5\sigma_X$	0.1161	0.1251	0.1298	0.1313	0.1289	0.1259
	$\overline{X} \pm \sigma_X$	0.1363	0.1108	0.1324	0.1230	0.1189	0.1206
12	$\overline{X} \pm 0.5\sigma_X$	0.1355	0.1421	0.1297	0.1246	0.1180	0.1254
	$\overline{X} \pm \sigma_X$	0.1373	0.1474	0.1283	0.1130	0.1148	0.1348

Data Generated using the CIR $_{(0.50, 0.05, 0.10)}$ Model

τ	$(\underline{u}, \overline{u})$	S,l					
		10T,10	20T,10	10T,20	20T,20	10T,50	20T,50
				Panel A: $T = 400$			
1	$\overline{X} \pm 0.5\sigma_X$	0.2011	0.247	0.1886	0.1633	0.244	0.2842
	$\overline{X} \pm \sigma_X$	0.1908	0.2383	0.1944	0.1814	0.2403	0.1756
2	$\overline{X} \pm 0.5\sigma_X$	0.1904	0.2783	0.2492	0.2978	0.1452	0.1666
	$\overline{X} \pm \sigma_X$	0.1781	0.219	0.2175	0.261	0.1547	0.181
4	$\overline{X} \pm 0.5\sigma_X$	0.2141	0.2109	0.1868	0.1783	0.1563	0.2221
	$\overline{X} \pm \sigma_X$	0.2204	0.1593	0.2075	0.2005	0.1523	0.1977
12	$\overline{X} \pm 0.5\sigma_X$	0.2257	0.2766	0.2767	0.3171	0.3345	0.2653
	$\overline{X} \pm \sigma_X$	0.162	0.1917	0.3012	0.2662	0.2962	0.2035
				Panel B: $T = 800$			
1	$\overline{X} \pm 0.5\sigma_X$	0.1261	0.1368	0.1615	0.1324	0.1568	0.1215
	$\overline{X} \pm \sigma_X$	0.118	0.1248	0.1381	0.1469	0.1225	0.1682
2	$\overline{X} \pm 0.5\sigma_X$	0.1437	0.1378	0.1231	0.1483	0.1608	0.1209
	$\overline{X} \pm \sigma_X$	0.1625	0.1429	0.1124	0.1553	0.1585	0.1202
4	$\overline{X} \pm 0.5\sigma_X$	0.1319	0.1248	0.1451	0.1525	0.1388	0.1541
	$\overline{X} \pm \sigma_X$	0.1428	0.1263	0.1517	0.1317	0.1364	0.1167
12	$\overline{X} \pm 0.5\sigma_X$	0.1406	0.1432	0.128	0.1202	0.145	0.1293
	$\overline{X} \pm \sigma_X$	0.1399	0.1688	0.1409	0.1216	0.1305	0.1294

$^{(*)}$ Notes: Entries in the table are empirical rejection frequencies for tests constructed using intervals given in the second column of the table, and for τ =1, 2, 4, 12. (S, l) combinations used in test construction are given in the second row of the table, so that simulation periods considered are $S = (10T, 20T)$ and block lengths considered are $l = (10, 20, 50)$, where T is the sample size, and T =400, 800. Empirical bootstrap distributions are constructed using 100 bootstrap replications, and critical values are set equal to the 90^{th} percentile of the bootstrap distribution. Finally, $\overline{X}$ and σ_X are the mean and variance of an initial sample of data. All results are based on 500 Monte Carlo simulations. See Section 6.1 for further details.

Table 2: Specification Test Rejection Frequencies For the Two Factor Models - Empirical Size[(*)]

Data Generated using the SV Model

τ	$(\underline{u}, \overline{u})$	S,l					
		10T,10	20T,10	10T,20	20T,20	10T,50	20T,50
		Panel A: $T = 400$					
1	$\overline{X} \pm 0.5\sigma_X$	0.2995	0.1698	0.1752	0.1765	0.3532	0.2132
	$\overline{X} \pm \sigma_X$	0.2266	0.2467	0.2585	0.1550	0.1506	0.2040
2	$\overline{X} \pm 0.5\sigma_X$	0.2141	0.2856	0.2328	0.2814	0.1912	0.3257
	$\overline{X} \pm \sigma_X$	0.1396	0.2107	0.2219	0.2264	0.1597	0.2196
4	$\overline{X} \pm 0.5\sigma_X$	0.2074	0.1941	0.2803	0.1313	0.2057	0.1475
	$\overline{X} \pm \sigma_X$	0.1770	0.1117	0.2266	0.1205	0.1550	0.1974
12	$\overline{X} \pm 0.5\sigma_X$	0.1910	0.2917	0.2056	0.2317	0.2435	0.2680
	$\overline{X} \pm \sigma_X$	0.1355	0.1497	0.1748	0.1990	0.2117	0.1168
		Panel B: $T = 800$					
1	$\overline{X} \pm 0.5\sigma_X$	0.1405	0.1584	0.1299	0.1366	0.1491	0.1409
	$\overline{X} \pm \sigma_X$	0.1282	0.1140	0.1271	0.1430	0.1208	0.1192
2	$\overline{X} \pm 0.5\sigma_X$	0.1048	0.1493	0.1169	0.1228	0.1203	0.1107
	$\overline{X} \pm \sigma_X$	0.1167	0.1548	0.1159	0.1275	0.1112	0.1165
4	$\overline{X} \pm 0.5\sigma_X$	0.1035	0.1183	0.1312	0.1416	0.1055	0.1276
	$\overline{X} \pm \sigma_X$	0.1173	0.1269	0.1329	0.1196	0.1123	0.1017
12	$\overline{X} \pm 0.5\sigma_X$	0.1207	0.1324	0.1043	0.1584	0.1033	0.1104
	$\overline{X} \pm \sigma_X$	0.1178	0.1071	0.1258	0.1058	0.1121	0.1277

Data Generated using the SVJ Model

τ	$(\underline{u}, \overline{u})$	S,l					
		10T,10	20T,10	10T,20	20T,20	10T,50	20T,50
		Panel A: $T = 400$					
1	$\overline{X} \pm 0.5\sigma_X$	0.2768	0.2686	0.2340	0.2114	0.3675	0.1569
	$\overline{X} \pm \sigma_X$	0.1939	0.1626	0.2192	0.2439	0.2342	0.1355
2	$\overline{X} \pm 0.5\sigma_X$	0.2112	0.3244	0.2442	0.1924	0.1727	0.2838
	$\overline{X} \pm \sigma_X$	0.1456	0.2375	0.2978	0.1346	0.1572	0.2124
4	$\overline{X} \pm 0.5\sigma_X$	0.2078	0.2898	0.1467	0.2099	0.1929	0.1839
	$\overline{X} \pm \sigma_X$	0.2927	0.1512	0.1189	0.1381	0.2961	0.1453
12	$\overline{X} \pm 0.5\sigma_X$	0.2102	0.1667	0.1881	0.1228	0.2757	0.3071
	$\overline{X} \pm \sigma_X$	0.2029	0.1459	0.3656	0.1784	0.2463	0.2439
		Panel B: $T = 800$					
1	$\overline{X} \pm 0.5\sigma_X$	0.1533	0.1328	0.1433	0.1280	0.1320	0.1099
	$\overline{X} \pm \sigma_X$	0.1068	0.1214	0.1397	0.1151	0.1228	0.1363
2	$\overline{X} \pm 0.5\sigma_X$	0.1263	0.1259	0.1191	0.1389	0.1350	0.1187
	$\overline{X} \pm \sigma_X$	0.1179	0.1122	0.1134	0.1164	0.1115	0.1183
4	$\overline{X} \pm 0.5\sigma_X$	0.1403	0.1541	0.1595	0.1262	0.1597	0.1394
	$\overline{X} \pm \sigma_X$	0.1178	0.1248	0.1185	0.1152	0.1131	0.1130
12	$\overline{X} \pm 0.5\sigma_X$	0.1248	0.1042	0.1249	0.1432	0.1110	0.1515
	$\overline{X} \pm \sigma_X$	0.1187	0.1120	0.1135	0.1188	0.1187	0.1062

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

Table 3: Specification Test Rejection Frequencies - Empirical Power(*)

Data Generated using the CIR Model

τ	$(\underline{u},\overline{u})$	S,l					
		10T,10	20T,10	10T,20	20T,20	10T,50	20T,50
		Panel A: $T = 400$					
1	$\overline{X} \pm 0.5\sigma_X$	0.3784	0.3425	0.4643	0.4662	0.4323	0.4582
	$\overline{X} \pm \sigma_X$	0.2547	0.2786	0.3222	0.3444	0.3443	0.3287
2	$\overline{X} \pm 0.5\sigma_X$	0.3782	0.3766	0.4483	0.4486	0.4568	0.4643
	$\overline{X} \pm \sigma_X$	0.2361	0.2786	0.3444	0.3328	0.3584	0.3587
4	$\overline{X} \pm 0.5\sigma_X$	0.3482	0.3727	0.4429	0.4584	0.4580	0.4267
	$\overline{X} \pm \sigma_X$	0.2345	0.2621	0.3189	0.3484	0.3169	0.3525
12	$\overline{X} \pm 0.5\sigma_X$	0.3747	0.3681	0.4644	0.4381	0.4248	0.4583
	$\overline{X} \pm \sigma_X$	0.2580	0.2485	0.3282	0.3161	0.3480	0.3681
		Panel B: $T = 800$					
1	$\overline{X} \pm 0.5\sigma_X$	0.8544	0.9261	0.8225	0.8442	0.8023	0.9285
	$\overline{X} \pm \sigma_X$	0.7125	0.7342	0.8727	0.8164	0.9404	0.8166
2	$\overline{X} \pm 0.5\sigma_X$	0.8067	0.8164	0.8864	0.9348	0.8185	0.9028
	$\overline{X} \pm \sigma_X$	0.8382	0.7667	0.9200	0.9360	0.9465	0.9188
4	$\overline{X} \pm 0.5\sigma_X$	0.8567	0.8884	0.8181	0.8969	0.9168	0.9062
	$\overline{X} \pm \sigma_X$	0.8329	0.7768	0.9367	0.8840	0.8383	0.9183
12	$\overline{X} \pm 0.5\sigma_X$	0.8744	0.9328	0.9442	0.9024	0.8924	0.8481
	$\overline{X} \pm \sigma_X$	0.7244	0.7223	0.9383	0.8163	0.8167	0.8143

Data Generated using the SV Model

τ	$(\underline{u}, \overline{u})$	S,l					
		10T,10	20T,10	10T,20	20T,20	10T,50	20T,50
		Panel A: $T = 400$					
1	$\overline{X} \pm 0.5\sigma_X$	0.5611	0.5613	0.5534	0.5491	0.5299	0.5421
	$\overline{X} \pm \sigma_X$	0.4589	0.4355	0.4248	0.5007	0.4461	0.4546
2	$\overline{X} \pm 0.5\sigma_X$	0.4555	0.5396	0.5204	0.5200	0.5374	0.4363
	$\overline{X} \pm \sigma_X$	0.4528	0.3486	0.4202	0.4361	0.4018	0.4062
4	$\overline{X} \pm 0.5\sigma_X$	0.5423	0.5666	0.4707	0.4851	0.5085	0.5612
	$\overline{X} \pm \sigma_X$	0.4217	0.4226	0.4445	0.4301	0.4279	0.4461
12	$\overline{X} \pm 0.5\sigma_X$	0.4561	0.4485	0.4364	0.4820	0.4307	0.4534
	$\overline{X} \pm \sigma_X$	0.3781	0.3431	0.3937	0.3537	0.3348	0.3719
		Panel B: $T = 800$					
1	$\overline{X} \pm 0.5\sigma_X$	0.6483	0.6885	0.6272	0.6962	0.6708	0.6362
	$\overline{X} \pm \sigma_X$	0.5309	0.5728	0.5139	0.5457	0.5859	0.6005
2	$\overline{X} \pm 0.5\sigma_X$	0.6324	0.6021	0.5930	0.6038	0.6048	0.6194
	$\overline{X} \pm \sigma_X$	0.5057	0.5169	0.5258	0.5160	0.5466	0.5580
4	$\overline{X} \pm 0.5\sigma_X$	0.5931	0.6143	0.6098	0.5909	0.6539	0.6115
	$\overline{X} \pm \sigma_X$	0.5039	0.5286	0.5508	0.5651	0.5783	0.5569
12	$\overline{X} \pm 0.5\sigma_X$	0.6598	0.6104	0.6242	0.6319	0.636	0.6936
	$\overline{X} \pm \sigma_X$	0.5614	0.5385	0.5725	0.5995	0.5891	0.6061

Data Generated using the SVJ Model

τ	$(\underline{u}, \overline{u})$	S,l					
		10T,10	20T,10	10T,20	20T,20	10T,50	20T,50
		Panel A: $T = 400$					
1	$\overline{X} \pm 0.5\sigma_X$	0.4581	0.5514	0.4988	0.4942	0.4867	0.4901
	$\overline{X} \pm \sigma_X$	0.4261	0.4358	0.4080	0.4035	0.3553	0.3857
2	$\overline{X} \pm 0.5\sigma_X$	0.4724	0.4724	0.5173	0.5236	0.5477	0.5030
	$\overline{X} \pm \sigma_X$	0.3650	0.3461	0.4072	0.4182	0.4020	0.4189
4	$\overline{X} \pm 0.5\sigma_X$	0.5364	0.5659	0.5599	0.5442	0.4206	0.5604
	$\overline{X} \pm \sigma_X$	0.3912	0.3470	0.3945	0.4259	0.3202	0.4005
12	$\overline{X} \pm 0.5\sigma_X$	0.4327	0.4406	0.5387	0.4920	0.4657	0.5130
	$\overline{X} \pm \sigma_X$	0.4192	0.3343	0.4005	0.3420	0.4492	0.4466
		Panel B: $T = 800$					
1	$\overline{X} \pm 0.5\sigma_X$	0.6053	0.7094	0.7019	0.6936	0.5708	0.5944
	$\overline{X} \pm \sigma_X$	0.5566	0.6406	0.5721	0.6083	0.5351	0.5765
2	$\overline{X} \pm 0.5\sigma_X$	0.6029	0.6983	0.6716	0.6260	0.5841	0.6327
	$\overline{X} \pm \sigma_X$	0.5511	0.5362	0.5762	0.5471	0.4741	0.5956
4	$\overline{X} \pm 0.5\sigma_X$	0.6176	0.6980	0.6109	0.6919	0.6546	0.6748
	$\overline{X} \pm \sigma_X$	0.4757	0.5663	0.5254	0.5675	0.4627	0.5117
12	$\overline{X} \pm 0.5\sigma_X$	0.6964	0.6304	0.5953	0.6155	0.7165	0.7053
	$\overline{X} \pm \sigma_X$	0.5657	0.5960	0.4882	0.5848	0.6158	0.6036

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

Table 4: Empirical Illustration - Specification Testing Using 1 and 2-Factor Models(*)

Specification Test Results - CIR Model

	$(\underline{u},\overline{u})$	$S=10T$		$S=20T$		$S=30T$	
		V_T	10% CV	V_T	10% CV	V_T	10% CV
				Panel A: $l=25$			
1	$\overline{X}\pm 0.5\sigma_X$	0.5274***	0.3545	0.5046***	0.3980	0.4923**	0.4768
	$\overline{X}\pm \sigma_X$	0.4289***	0.3178	0.4524***	0.3568	0.4655***	0.3635
2	$\overline{X}\pm 0.5\sigma_X$	0.6824***	0.4911	0.6973***	0.5509	0.6075**	0.5134
	$\overline{X}\pm \sigma_X$	0.4897*	0.5182	0.4601	0.5040	0.4985**	0.3560
4	$\overline{X}\pm 0.5\sigma_X$	0.8662**	0.8491	0.8813**	0.7962	0.8726**	0.6247
	$\overline{X}\pm \sigma_X$	0.8539*	0.9389	0.8153*	0.9330	0.8595**	0.8581
12	$\overline{X}\pm 0.5\sigma_X$	1.1631*	1.3009	1.2236*	1.2932	1.2432**	1.1562
	$\overline{X}\pm \sigma_X$	1.0429	2.0222	1.0731	2.0401	1.0387	2.0335
				Panel B: $l=50$			
1	$\overline{X}\pm 0.5\sigma_X$	0.5274***	0.3523	0.5046***	0.4440	0.4923*	0.5749
	$\overline{X}\pm \sigma_X$	0.4289**	0.3325	0.4524***	0.3584	0.4655***	0.297
2	$\overline{X}\pm 0.5\sigma_X$	0.6824***	0.4915	0.6973***	0.5141	0.6075**	0.4683
	$\overline{X}\pm \sigma_X$	0.4897*	0.5594	0.4601**	0.4574	0.4985**	0.3960
4	$\overline{X}\pm 0.5\sigma_X$	0.8662*	0.9498	0.8813**	0.7367	0.8726**	0.5917
	$\overline{X}\pm \sigma_X$	0.8539*	0.9371	0.8153*	0.9404	0.8595**	0.8055
12	$\overline{X}\pm 0.5\sigma_X$	1.1631*	1.3256	1.2236*	1.2570	1.2432*	1.2776
	$\overline{X}\pm \sigma_X$	1.0429	2.0165	1.0731	2.0157	1.0387	2.0071

Specification Test Results - *sv* Model

τ	$(\underline{u}, \overline{u})$	$S = 10T$		$S = 20T$		$S = 30T$	
		V_T	10% CV	V_T	10% CV	V_T	10% CV
				Panel A: $l = 25$			
1	$\overline{X} \pm 0.5\sigma_X$	0.9841***	0.9031	0.9453***	0.7986	0.9112**	0.8328
	$\overline{X} \pm \sigma_X$	0.6870	0.7254	0.7276	0.7674	0.7775**	0.7608
2	$\overline{X} \pm 0.5\sigma_X$	0.4113	1.4900	1.0265	1.4124	0.9641	1.4808
	$\overline{X} \pm \sigma_X$	0.3682	1.2243	0.8390	1.4938	0.8295	1.5048
4	$\overline{X} \pm 0.5\sigma_X$	1.2840	2.6109	1.0835	2.5397	1.6839	2.5685
	$\overline{X} \pm \sigma_X$	1.0472	2.2745	1.0110	2.2695	1.1328	2.3104
12	$\overline{X} \pm 0.5\sigma_X$	1.7687	5.2832	1.7135	5.3526	2.6901	5.345
	$\overline{X} \pm \sigma_X$	1.7017	5.6522	1.4404	5.6279	1.7675	5.6733
				Panel B: $l = 50$			
1	$\overline{X} \pm 0.5\sigma_X$	0.9841***	0.8988	0.9453**	0.8093	0.9112**	0.7988
	$\overline{X} \pm \sigma_X$	0.6870**	0.6597	0.7276	0.7903	0.7775*	0.8269
2	$\overline{X} \pm 0.5\sigma_X$	0.4113	1.4466	1.0265	1.3969	0.9641	1.4365
	$\overline{X} \pm \sigma_X$	0.3682	1.4673	0.8390	1.4975	0.8295	1.3444
4	$\overline{X} \pm 0.5\sigma_X$	1.2840	2.5657	1.0835	2.6108	1.6839	2.4884
	$\overline{X} \pm \sigma_X$	1.0472	2.2299	1.0110	2.3095	1.1328	2.3672
12	$\overline{X} \pm 0.5\sigma_X$	1.7687	5.2820	1.7135	5.3347	2.6901	5.4429
	$\overline{X} \pm \sigma_X$	1.7017	5.6487	1.4404	5.6249	1.7675	5.6787

Specification Test Results - SVJ Model

τ	$(\underline{u}, \overline{u})$	$S = 10T$		$S = 20T$		$S = 30T$	
		V_T	10% CV	V_T	10% CV	V_T	10% CV
				Panel A: $l = 25$			
1	$\overline{X} \pm 0.5\sigma_X$	1.1319	2.1957	1.1787	2.1342	1.1655	2.1594
	$\overline{X} \pm \sigma_X$	1.2272*	1.3031	1.0220	1.1669	0.9906	1.2893
2	$\overline{X} \pm 0.5\sigma_X$	0.9615*	1.1334	1.0150*	1.0677	1.0528*	1.0903
	$\overline{X} \pm \sigma_X$	1.2571	1.4096	1.1491	1.4264	1.1562	1.5162
4	$\overline{X} \pm 0.5\sigma_X$	1.5012*	1.6856	1.3255	1.6501	1.3545*	1.5410
	$\overline{X} \pm \sigma_X$	0.9901*	1.0507	1.0180*	1.0400	0.6941	0.9318
12	$\overline{X} \pm 0.5\sigma_X$	2.4237*	3.0640	2.3428*	2.9880	2.3622*	3.0997
	$\overline{X} \pm \sigma_X$	1.4522	2.1684	1.4766	2.1625	1.4668	2.1360
				Panel B: $l = 50$			
1	$\overline{X} \pm 0.5\sigma_X$	1.1319	2.1109	1.1787	2.0323	1.1655	2.0733
	$\overline{X} \pm \sigma_X$	1.2272*	1.2574	1.0220*	1.2657	0.9906	1.2276
2	$\overline{X} \pm 0.5\sigma_X$	0.9615*	1.1571	1.0150*	1.0640	1.0528**	1.0296
	$\overline{X} \pm \sigma_X$	1.2571	1.3863	1.1491	1.4236	1.1562	1.5230
4	$\overline{X} \pm 0.5\sigma_X$	1.5012*	1.6471	1.3255	1.5650	1.3545*	1.5775
	$\overline{X} \pm \sigma_X$	0.9901	1.0296	1.0180**	0.9835	0.6941	0.8466
12	$\overline{X} \pm 0.5\sigma_X$	2.4237*	3.0895	2.3428*	3.0086	2.3622*	2.9668
	$\overline{X} \pm \sigma_X$	1.4522	2.2239	1.4766	2.1317	1.4668	2.2222

(*) Notes: See notes to Table 1. Tabulated entries are test statistics and 5%, 10% and 20% level critical values. Test intervals are given in the second column of the table, for $\tau = 1, 2, 4, 12$. All tests are carried out using historical one-month Eurodollar deposit rate data for the period January 1971 - September 2005, measured at a weekly frequency. Single, double, and triple starred entries denote rejection at the 20%,10%, and 5% levels, respectively. Additionally, $\overline{X}$ and σ_X are the mean and standard deviation of the historical data. See Section 6.2 for complete details.

Seminar Notes

Predictive Density Evaluation

Norman R. Swanson

Rutgers University

2009

1 Overview

- *There is a clear need, when forming macroeconomic policies and when managing financial risk in the insurance and banking industries, to examine predictive confidence intervals or entire predictive conditional distributions.*

- *In this chapter focus is on specification testing of and model selection amongst predictive densities, with brief discussion of estimation*

■■■ *Specification Testing*

■■ *Probability Integral Transform* - Rosenblatt (1952) and Diebold, Gunther and Tay (1998)

■ $Z^{t-1} = (y_{t-1}, ..., y_{t-v}, X_t, ..., X_{t-w})$, v, w finite, X_t vector valued

■ $F(y_t | Z^{t-1}, \theta_0) = \int_{-\infty}^{y_t} f_t(y|Z^{t-1}, \theta_0) dy$, is an iid uniform RV on $[0, 1]$

■ Tests with power against departures from uniformity and independence (see e.g. Bai (2003) for KS version of this comparing against uniformity)

Kolmogorov-Smirnov and Related Statistics

- Empirical Distribution Function - a natural estimator for F which is unbiased, consistent, and asymptotically normal

- $F_T\,(y) = T^{-1}\sum_{t=1}^{T} 1\{y_t \leq y\}$

- Cramer - von Mises Discrepancy Measure [Cramer (1938, 1946), von Mises (1947)]:

- $T\int\,(F_T - F)^2\,dF$

- Kolmorov - Smirnov Discrepancy Measure [Kolmogorov (1933), Smirnov (1939)]:

- $T^{1/2}||F_T - F||_{\infty} = sup_x T^{1/2}|F_T(y) - F(y)|$

- Glivenko-Cantelli uniform convergence [Glivenko (1933), Cantelli (1933)]:

- $||F_T - F||_\infty \to^{a.s.} 0$

- Donsker uniform or functional CLT (*iid* data) [Donsker (1952)]:

- $T^{1/2}(F_T - F)$ converges to a Gaussian process, and in particular to a Brownian bridge limit process

- Modified Kolmogorov-Smirnov Statistic

- $T^{1/2}||F_T - F_{\widehat{\theta}}||_\infty$

- Convergence in distribution to the supremum of a Gaussian process - limit distribution may depend on the model F, the estimator $\widehat{\theta}$, and even the parameter value θ

■ Hence the use of the bootstrap for CVs: time series data versus *iid*; data dependence

■ In economics and finance, data dependence suggests that conditional models may be more useful than unconditional models, and hence the use of conditional distributions as predictive densities

■■ *The Kulback-Leibler Information Criterion* [White (1982), Vuong (1989), Giacomini (2002), Kitamura (2002)]:

■ Choose model which minimizes KLIC; choose model 1 over 2 if

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

Model Selection

Information Sets and Critical Values

- Limit distribution of KS tests affected by dynamic misspecification. Critical values derived under correct specification given $\Im_{t-1}$ are not in general valid in the case of correct specification given a subset of $\Im_{t-1}$. Many authors use Z^{t-1}, and assume that $Z^{t-1} \equiv \Im_{t-1}$.

- Assume interested in testing whether $y_t \, | y_{t-1}$ is $N(\alpha_1^\dagger \, y_{t-1} \, , \sigma_1 \,)$

- Suppose $\Im_{t-1}$ includes y_{t-1} and y_{t-2} : true cond model is

$$y_t \, | \Im_{t-1} = y_t \, | y_{t-1} \, , y_{t-2} = \; N(\alpha_1 \, y_{t-1} + \alpha_2 \, y_{t-2} \, , \sigma_2 \,)$$

- Then, $\alpha_1^\dagger$ differs from α_1 and correct specification holds wrt information in y_{t-1} ; but there is dynamic misspecification with respect to y_{t-1} , y_{t-2}.

- Even without taking account of PEE, CVs obtained assuming correct dynamic specification are invalid.

- Stated differently, tests that are designed to have power against both uniformity and independence violations (i.e. tests that assume correct dynamic specification under H_0) will reject; an inference which is incorrect, at least in the sense that the "normality" assumption is *not* false (uniformity still holds, but independence does not -> rejection of model)

■■ *Selecting from Amongst Many Models* [White (2000), Corradi and Swanson (2003)]:

■ Need mean square error and other measures of distributional discrepancy

■ Issues of sequential test bias, allowance for mis-specification, alternative methods to construct CVs are all relevant

- Many (possibly) misspecified conditional distributions,

$$F_1(u|Z^t, \theta_1^\dagger), \ ..., \ F_m(u|Z^t, \theta_m^\dagger),$$

and true conditional distribution,

$$F_0(u|Z^t, \theta_0) = \Pr(Y_{t+1} \leq u|Z^t)$$

- One accuracy measure; average over $u \in U$, or use interval based on u_{low}, u_{up}

- $E\left(\left(F_i(u|Z^{t+1}, \theta_i^\dagger) - F_0(u|Z^{t+1}, \theta_0)\right)^2\right)$

2 Specification Testing

■■■ *Probability Integral Transform KS Type Tests*

■■ DGT: PIT: difference between the empirical distribution of $F_t(y_t | Z^{t-1}, \widehat{\theta}_T)$ and the $45^\circ -$ degree line

■

$$
\begin{aligned}
H_0 &: \Pr(y_t \leq y | \Im_{t-1}, \theta_0) = F_t(y | \Im_{t-1}, \theta_0), \\
H_A &: \text{the negation of } H_0,
\end{aligned}
$$

■ Compare $F_t(y|\Im_{t-1}, \theta_0)$ with CDF of uniform RV on $[0, 1]$; differentiability; nonstationarity; Z^{t-1} contains all useful info in $\Im_{t-1}$;

■ $\widehat{U}_t = F(y_t | Z^{t-1}, \widehat{\theta}_T)$

■■ Bai (2003): $\widehat{V}_T(r) = \frac{1}{T^{1/2}} \sum \left(1\{\widehat{U}_t \leq r\} - r \right)$

■ use martingalization of Khmaladze to account for PEE

■ Theorem 2.1 (from Corollary 1 in Bai (2003)): Let BAI1-BAI4 hold, then under H_0,

$$\sup_{r \in [0,1]} \widehat{W}_T(r) \xrightarrow{d} \sup_{r \in [0,1]} W(r),$$

where $W(r)$ is a standard Brownian motion. Therefore, the limiting distribution is nuisance parameter free and critical values can be tabulated.

■ Suppose $\Pr(y_t \leq y | \Im_{t-1}, \theta_0) \neq \Pr(y_t \leq y | Z^{t-1}, \theta^{\dagger})$. Then CVs not valid. However, if $F(y_t | Z^{t-1}, \theta^{\dagger})$ correctly specified for $\Pr(y_t \leq y | Z^{t-1}, \theta^{\dagger})$, uniformity still holds $=>$ no guarantee statistic diverges. Thus, test does not have unit asymptotic power against violations of independence.

Hong and Li (2003):

$$\widehat{\phi} = (n-j)^{-1} \sum_{\tau=j+1}^{n} K_h(u_1, \widehat{U}_\tau) K_h(u_2, \widehat{U}_{\tau-j}),$$

where

$$K_h(x,y) = \begin{cases} h^{-1}\left(\frac{x-y}{h}\right) / \int_{-(x/h)}^{1} k(u)du \text{ if } x \in [0,h) \\ h^{-1}\left(\frac{x-y}{h}\right) \quad \text{if } x \in [h, 1-h) \\ h^{-1}\left(\frac{x-y}{h}\right) / \int_{-1}^{(1-x)/h} k(u)du, \text{ if } x \in [0,h) \end{cases}$$

h is bandwidth parameter. As an example, one might use,

$$k(u) = \frac{15}{16}(1-u^2)^2 \, 1\{|u| \leq 1\}.$$

Also, define

$$\widehat{M}(j) = \int_0^1 \int_0^1 \left(\widehat{\phi}(u_1, u_2) - 1 \right)^2 du_1 \, du_2$$

and

$$\widehat{Q}\ (j) = \Big(\ (n-j)\,\widehat{M}\ (j) - A_h^0\Big)\ /V_0^{1/2},$$

with

$$A_h^0 = ((h^{-1}-2)\int_{-1}^{1}\ k^2(u)du + 2\int_0^1\int_{-1}^{b} k_b(u)dudb)^2-1,$$

$$k_b\ (\cdot) = k(\cdot)/\int_{-1}^{b}\ k(v)dv,$$

and

$$V_0 = 2\left(\int_{-1}^{1}\left(\int_{-1}^{1}\ k(u+v)k(v)dv\right)^2\ du\right)^2$$

■ Theorem 2.2 (from Theorem 1 in Hong and Li (2003): Let HL1-HL4 hold. If $h = cT^{-\delta}$, $\delta \in (0, 1/5)$, then under H_0, for any $j > 0$, $j = o(T^{1-\delta(5-2/v)})$, $\widehat{Q}\ (j) \xrightarrow{d}\ N(0,1)$

■■ Corradi and Swanson (2003a): parametric rate, no bandwidth parameter, easy to construct, but uses bootstrap

■ Bai's martingale transformation argument does not apply to the case in which the score is not a martingale difference process (i.e. dyn miss not allowed for with his test).

■ $V_{1T} = \sup_{r\in[0,1]} |V_{1T}(r)|$

■ $V_{1T}(r) = \frac{1}{T^{1/2}} \sum_{t=1}^{T} \left(1\{\widehat{U}_t \leq r\} - r\right)$

■ Theorem 2.3 (from Theorem 1 in Corradi and Swanson (2003a)): Let CS1, CS2(i)–(ii) and CS3 hold. Then: (i) Under H_0, $V_{1T} \Rightarrow \sup_{r\in[0,1]} |V_1(r)|$, where V is a zero mean Gaussian process with covariance kernel $K_1(r, r')$.

■ a related test not based on PIT, but directly on KS:

- Compare empirical joint distribution of $y_t,\ Z^{t-1}$ with product of dist of $y_t\ |Z^t$ and empirical CDF of Z^{t-1} .

- Joint:

$$\widehat{H}_T\ (u,v) = \frac{1}{T}\sum_{t=1}^{T}\ 1\{y_t \leq u\}1\{Z^{t-1} < v\}$$

and semi-empirical/semi-parametric analog of

$F(u,v,\theta_0\ ),$

$$\widehat{F}_T\ (u,v,\widehat{\theta}_T\ ) = \frac{1}{T}\sum_{t=1}^{T}\ F(u|Z^{t-1}\ ,\widehat{\theta}_T\ )1\{Z^{t-1} < v\}$$

- $$V_{2T} = \sup_{u\times v\in U\times V}\ |V_{2T}\ (u,v)|,$$

where U and V are compact subsets of $\Re$ and $\Re^d$, respectively, and

■

$$V_{2T}(u,v) =$$

$$\frac{1}{T^{1/2}}\sum_{t=1}^{T}\Big((1\{y_t\le u\} - F(u|Z^{t-1},\widehat{\theta}_T))1\{Z^{t-1}\le v\}\Big)$$

■ Theorem 2.4 (from Theorem 2 in Corradi and Swanson (2003a)): Let CS1, CS2(iii)–(iv) and CS3 hold. Then:

(i) Under H_0, $V_{2T} \Rightarrow \sup_{u\times v\in U\times V} |Z(u,v)|$, where Z is a zero mean Gaussian process with covariance kernel $K_2(u,v,u',v')$;

(ii) Under H_A, there exists an $\varepsilon > 0$ such that $\lim_{T\to\infty} \Pr(\frac{1}{T^{1/2}} V_{2T} > \varepsilon) = 1$.

■ No iid bootstrap; parametric bootstrap only under correct dynamic specification: crucial to properly mimic long run variance of statistic

- $\widehat{U}_t = F(y_t \,|Z^{t-1}, \widehat{\theta}_T), \quad \widehat{U}^*_t = F(y^*_t \,|Z^{t-1}, \widehat{\theta}^*_T)$

- $V^*_{1T}(r) = \frac{1}{T^{1/2}} \sum_{t=1}^{T} \left(1\{\widehat{U}^*_t \leq r\} - 1\{\widehat{U}_t \leq r\} \right)$

-

$$V^*_{2T}(u, v) =$$

$$\frac{1}{T^{1/2}} \sum_{t=1}^{T} ((1\{y^*_t \leq u\} - F(u|Z^{*,t-1}, \widehat{\theta}^*_T))1\{Z^{*,t-1} \leq v\}$$

$$-(1\{y_t \leq u\} - F(u|Z^{t-1}, \widehat{\theta}_T))1\{Z^{t-1} \leq v\})$$

■ In summary, $V^*_{1T}(\omega)$ (resp. $V^*_{2T}(\omega)$) has a well defined limiting distribution, conditional on the sample, that coincides with that of V_{1T} (resp. V_{2T}), under H_0. Thus, for any bootstrap replication, compute the bootstrapped statistic, V^*_{1T} (resp. V^*_{2T}). Perform B bootstrap replications (B large) and compute the percentiles of the empirical distribution of the B bootstrapped statistics. Reject H_0 if V_{1T} (V_{2T}) is greater than the $(1-\alpha)th$ -percentile. Otherwise, do not reject H_0. This approach ensures that the test has asymptotic size equal to α. Under the alternative, V_{1T} (V_{2T}) diverges to infinity, while the corresponding bootstrap statistic has a well defined limiting distribution. This ensures unit asymptotic power.

■■ In sample model evaluation - PEE contribution summarized by limiting distribution of $T^{1/2}\left(\widehat{\theta}_T - \theta^\dagger\right)$

■ Recursive and rolling estimation schemes - PEE contribution summarized by the limiting distribution of $\frac{1}{P^{1/2}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_t - \theta^\dagger\right)$

- Bai; Hong and Li; limit distributions of appropriate statistics same. Corradi and Swanson, covariance kernel structure changes.

- V_{1T} , V_{2T} tests: split and full sample bootstrap statistics available - for example:

-

$$\widehat{\theta}_t^* = \arg \min_{\theta_i \in \Theta_i} \frac{1}{t} \sum_{j=1}^{t} q(y_j^*, Z^{*,j-1}, \theta), \quad R \leq t \leq T-1$$

Further, define: $\Psi^*_{R,P} =$

$$\frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \left(\widehat{\theta}_t^* - \widehat{\theta}_t \right) + \left(-\frac{1}{T} \sum_{t=1}^{T} \nabla^2_{\theta_i} q(y_t, Z^{t-1}, \widehat{\theta}_{T-1}) \right)^{-1}$$
$$\times \frac{1}{P^{1/2}} \sum_{j=1}^{P-1} a_{R,j} (\nabla_\theta q(y_{R+j}, Z^{R+j-1}, \widehat{\theta}_{T-1})$$
$$- \frac{1}{P} \sum_{j=1}^{P} \nabla_\theta q(y_{R+j}, Z^{R+j-1}, \widehat{\theta}_{T-1})),$$

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

If as $P, R \rightarrow \infty$, $P/R \rightarrow \pi = 0$, then both terms on the RHS above approach zero and so $\Psi^*_{R,P} \overset{pr}{\rightarrow} 0$. This is not surprising, as in this case parameter estimation error vanishes and so $\Psi^*_{R,P}$ properly mimics something going to zero in probability. Therefore, if $\pi = 0$, there is no need at all to use a correction term when constructing the bootstrap statistic.

- Theorem 3.6 (from Theorem 1 in Corradi and Swanson (2003b): Let CS1 and CS3 hold. Also, assume that as $P, R \rightarrow \infty$, $P/R \rightarrow \pi$, $0 < \pi < \infty$, and that as $l_1, l_2 \rightarrow \infty$, $\frac{l_2}{P^{1/4}} \rightarrow 0$ and $\frac{l_1}{R^{1/4}} \rightarrow 0$. Then,

$$P(\omega : \sup_{v \in \Re^{\varrho(i)}} |P^*_{R,P}\left(\Psi^*_{R,P} \leq v\right)$$

$$-P(\frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \left(\widehat{\theta}_{i,t} - \theta_i^\dagger\right) \leq v)| > \varepsilon) \rightarrow 0$$

- add PEE adjustment terms to boot statistics

3 Model Selection

■ Choose a model which provides the best (loss function specific) out of sample predictions, from amongst a set of potentially misspecified models, and not just from amongst models that may only be dynamically misspecified, as is the case with some of the tests discussed above.

■■■ *Point forecast comparison*

■ Nested versus non-nested models? Differentiability?

■

$$H_0 : E(g(u_{0,t}) - g(u_{1t})) = 0$$

$$H_A : E(g(u_{0,t}) - g(u_{1t})) \neq 0$$

■■ *Non-nested Models:*

■ $DM_P = \frac{1}{P^{1/2}} \frac{1}{\widehat{\sigma}_P} \sum_{t=R}^{T-1} \left(g(\widehat{u}_{0,t+1}) - g(\widehat{u}_{1,t+1}) \right)$

■ Proposition 4.1 (from Theorem 4.1 in West (1996)): Let W1-W2 hold. Also, assume that g is continuously differentiable, then, if as $P \to \infty$, $l_p \to \infty$ and $l_P / P^{1/4} \to 0$, then as P, $R \to \infty$, under H_0, $DM_P \xrightarrow{d} N(0,1)$ and under H_A,

$$\Pr\left(P^{-1/2} |DM_P| > \varepsilon \right) \to 1, \text{ for any } \varepsilon > 0.$$

■ Note that when the two models are nested, so that $u_{0,t} = u_{1,t}$ under H_0, both the numerator of the DM_P statistic and $\widehat{\sigma}_P$ approach zero in probability at the same rate, if $P/R \to 0$, so that the DM_P statistic no longer has a normal limiting distribution under the null.

Nested Models: (Granger causality)

$$y_t = \sum_{j=1}^{q} \beta_j \, y_{t-j} + \epsilon_t$$

and unrestricted model is

$$y_t = \sum_{j=1}^{q} \beta_j \, y_{t-j} + \sum_{j=1}^{k} \alpha_j \, x_{t-j} + u_t$$

$$ENC - T = (P-1)^{1/2} \frac{\varsigma}{(P^{-1} \sum_{t=R}^{T-1} (c_{t+1} - \varsigma))^{1/2}},$$

where $c_{t+1} = \widehat{\epsilon}_{t+1} \left(\widehat{\epsilon}_{t+1} - \widehat{u}_{t+1} \right)$,

$\varsigma = P^{-1} \sum_{t=R}^{T-1} c_{t+1}$, and where $\widehat{\epsilon}_{t+1}$ and $\widehat{u}_{t+1}$ are residuals from the LS estimation.

One sided test; Assumes larger model dynamically correctly specified (i.e. mds).

- Convergence to functional of standard Brownian motion or N(0,1) if PEE vanishes.

- To allow for dynamic misspecification and/or conditional heteroskedasticity, use two sided Bierens type test:

- CCS: $m_P = P^{-1/2}\sum_{t=R}^{T-1} \widehat{\epsilon}_{t+1} X_t$, appropriately scaled

- Version of this test also available for testing against generic alternatives

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

 and

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

Multiple Models: Data Snooping Methods

- Data mining and sequential test bias?

- Reality Check:

-

$$S_P = \max_{k=2,\ldots,m} S_P(1,k),$$

where

$$S_P(1,k) = \frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \Big(g(\widehat{u}_{1,t+1}) - g(\widehat{u}_{k,t+1}) \Big),\ \ k = 2, \ldots, m$$

The hypotheses are formulated as

$$H_0 : \max_{k=2,\ldots,m} E(g(u_{1,t+1}) - g(g_{k,t+1})) \leq 0$$

$$H_A : \max_{k=2,\ldots,m} E(g(u_{1,t+1}) - g(u_{k,t+1})) > 0,$$

- Convergence to the max of a m-1 dimensional Gaussian process with complicated covariance kernel.

■ Critical values obtained using stationary bootstrap, block bootstrap, Hansen modified CVs, subsampling methods, false discovery rate approach.

■■■ *Density (distribution) comparison*

■■ *The KLIC*

■ A well known measure of distributional accuracy is the Kullback-Leibler Information Criterion (KLIC), according to which we choose the model which minimizes the KLIC (see e.g. White (1982), Vuong (1989), Giacomini (2002), and Kitamura (2002)). In particular, choose model 1 over model 2, if

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

■ For the iid case, Vuong (1989) suggests a likelihood ratio test for choosing the conditional density model that is closer to the "true" conditional density in terms of the KLIC. 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.

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

- Furthermore, the KLIC approach has recently been employed for the evaluation of dynamic stochastic general equilibrium models (see e.g. Schorfheide (2000), Fernandez-Villaverde and Rubio-Ramirez (2004), and Chan, Gomes and Schorfheide (2002)). For example, Fernandez-Villaverde and Rubio-Ramirez (2004) show that the KLIC-best model is also the model with the highest posterior probability.

Evaluation of Predictive Density Based on OOS Distributional MSE

$$H_0' : \max_{k=2,\dots,m} E\left((\Upsilon_1 - \Upsilon_0)^2 - (\Upsilon_i - \Upsilon_0)^2 \right) \leq 0$$

versus

$$H_A' : \max_{k=2,\dots,m} E(\Upsilon_1 - \Upsilon_0)^2 - (\Upsilon_i - \Upsilon_0)^2 > 0$$

where $\Upsilon_i = \left(F_i \left(u_{up} | Z^t , \theta_i^\dagger \right) - F_i \left(u_{low} | Z^t , \theta_i^\dagger \right)\right)$, $i = 1, \dots, m$

and $\Upsilon_0 = \left(F_0 \left(u_{up} | Z^t , \theta_0 \right) - F_0 \left(u_{low} | Z^t , \theta_0 \right)\right)$

- $$Z_{P,j} = \max_{k=2,\ldots,m} \int_U Z_{P,u,j}(1,k)\phi(u)du,\ \ j = 1,2$$

and

$$Z_{P,u,2}(1,k) = \frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \Big(\Big(1\{y_{t+1} \leq u\} - F_1(u|Z^t, \widehat{\theta}_{1,t}) \Big)^2$$

$$- \Big(1\{y_{t+1} \leq u\} - F_k(u|Z^t, \widehat{\theta}_{k,t}) \Big)^2 \Big),$$

- Limit distribution is max of functional of zero mean Gaussian process. Can use bootstrap for CVs.

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

■ When all competing models provide an approximation to the true conditional distribution that is as (mean square) accurate as that provided by the benchmark (i.e. when $\int_U \left(\mu_1^2(u) - \mu_k^2(u)\right) \phi(u)du = 0, \forall k$), then the limiting distribution is a zero mean Gaussian process with a covariance kernel which is not nuisance parameter free. Additionally, when all competitor models are worse than the benchmark, the statistic diverges to minus infinity at rate $P^{1/2}$. Finally, when only some competitor models are worse than the benchmark, the limiting distribution provides a conservative test, as Z_P will always be smaller than

$$\max_k \int_U \left(Z_{P,u}(1,k) - P^{1/2}\left(\mu_1^2(u) - \mu_k^2(u)\right)\right) \phi(u)du$$

asymptotically. Of course, when H_A holds, the statistic diverges to plus infinity at rate $P^{1/2}$.

4 Concluding Remarks

- *Large number of available tests and procedures.*

- *Many approaches to specification and selection are specialized to nested/nonested models; limit results hinge on PEE assumptions, differentiability, and assumptions concerning misspecification (either dynamic or general).*

- *Predictive density is one of the relatively uncharted areas, and recent theoretical advances in bootstrap theory and other limit theory results are allowing for much new work.*

- *End* □