Available online at www.sciencedirect.com
ScienceDirect

Journal of Econometrics 141 (2007) 736–776

JOURNAL OF Econometrics

www.elsevier.com/locate/jeconom

Can the random walk model be beaten in out-of-sample density forecasts? Evidence from intraday foreign exchange rates

Yongmiao Hong[a,b,*], Haitao Li[c], Feng Zhao[d]

[a]Department of Economics, Cornell University, Ithaca, NY 14850, USA
[b]Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen 361005, China
[c]Stephen M. Ross School of Business, University of Michigan, Ann Arbor, MI 48109, USA
[d]Rutgers Business School, Rutgers University, Newark, NJ 07102, USA

Available online 26 February 2007

Abstract

It has been documented that random walk outperforms most economic structural and time series models in out-of-sample forecasts of the conditional mean dynamics of exchange rates. In this paper, we study whether random walk has similar dominance in out-of-sample forecasts of the conditional probability density of exchange rates given that the probability density forecasts are often needed in many applications in economics and finance. We first develop a nonparametric portmanteau test for optimal density forecasts of univariate time series models in an out-of-sample setting and provide simulation evidence on its finite sample performance. Then we conduct a comprehensive empirical analysis on the out-of-sample performances of a wide variety of nonlinear time series models in forecasting the intraday probability densities of two major exchange rates—Euro/Dollar and Yen/Dollar. It is found that some sophisticated time series models that capture time-varying higher order conditional moments, such as Markov regime-switching models, have better density forecasts for exchange rates than random walk or modified random walk with GARCH and Student-*t* innovations. This finding dramatically differs from that on mean forecasts and suggests that sophisticated time series models could be useful in out-of-sample applications involving the probability density.

JEL classification: C4; E4; G0

Keywords: Density forecasts; GARCH; Intraday exchange rate; Jumps; Maximum likelihood estimation; Nonlinear time series; Out-of-sample forecasts; Regime-switching

*Corresponding author. Department of Economics, Cornell University, Ithaca, NY 14850, USA. Tel.: +1 607 255 5130; fax: +1 607 255 2818.
E-mail address: yh20@cornell.edu (Y. Hong).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2006.11.003

1. Introduction

Foreign exchange markets are among the most important financial markets in the world, with trading taking place 24 h a day around the globe and trillions of dollars of different currencies transacted each day. Transactions in foreign exchange markets determine the rates at which currencies are exchanged, which in turn determine the costs of purchasing foreign goods and financial assets. Understanding the evolution of exchange rates is important for many outstanding issues in international economics and finance, such as international trade and capital flows, international portfolio management, currency options pricing, and foreign exchange risk management.

The vast literature on exchange rate dynamics has documented several important stylized facts for nominal exchange rates. First, changes of exchange rates are leptokurtic since their unconditional distributions exhibit a sharper peak and fatter tails than normal distributions (e.g., Boothe and Glassman, 1987; Hsieh, 1988). Second, exchange rate changes exhibit persistent volatility clustering: In periods of turbulence, large changes tend to be followed by large changes; and in periods of tranquility, small changes tend to be followed by small changes (e.g., Diebold, 1988).

A variety of sophisticated nonlinear time series models have been proposed in the literature to capture these stylized behaviors. For example, Bollerslev (1987), Engle and Bollerslev (1986), Baillie and Bollerslev (1989), and Hsieh (1989) show that a GARCH model with i.i.d. Student-t innovations can capture volatility clustering in all major exchange rates, and can explain at least part of the leptokurtosis in exchange rate changes. Engel and Hamilton (1990) show that Hamilton's (1989) Markov regime-switching model can capture the "long swings" in several major dollar exchange rates. It also can capture the leptokurtosis in exchange rate data, because its conditional and unconditional densities are mixtures of normal distributions with different means and/or variances. Jorion (1988) and Bates (1996) show that jumps can capture discontinuities in exchange rate data due to various economic shocks, news announcements, and government interventions in foreign exchange markets.

Although these nonlinear time series models have good in-sample performances in capturing exchange rate data, they fail miserably in forecasting future exchange rate changes. As pointed out in Putnam and Quintana (1994, p. 223), exchange rate movements contain a considerable amount of noise, and the low signal-to-noise ratio in exchange rate changes opens up wide possibilities for spurious in-sample dependence which may not be robust for out-of-sample forecasting. Indeed, the classic paper of Meese and Rogoff (1983) and many subsequent important studies (see, e.g., Diebold and Nason, 1990; Meese and Rose, 1990; Engel, 1994) have shown that during the post-Bretton Woods period most economic structural and time series models of exchange rates underperform a "naive" random walk model in predicting the conditional mean of major exchange rates.[1] While nonlinear time series models improve the modeling of even-ordered moments, the random walk model (which does not attempt to capture any conditional mean dynamics) still dominates in forecasting the conditional mean of exchange rate changes. These empirical

[1]Even though the dominance of the random walk model in forecasting the conditional mean of exchange rates is widely established, there is evidence that certain time series models might have better mean forecasts. For example, Wolff (1985) shows that a state space model for exchange rates performs better than the random walk model used by Meese and Rogoff (1983). See also Wolff (1987).

results cast serious doubts on the relevance of nonlinear time series models for out-of-sample applications, despite their good in-sample performances.

In this paper, we address the important question whether some nonlinear time series models can outperform the random walk model in out-of-sample forecasts of the probability density of exchange rates. As argued by Diebold et al. (1998), Granger (1999), Granger and Pesaran (2000), and Corradi and Swanson (2006a), accurate density forecasts are important for decision making under uncertainty when a forecaster's loss function is asymmetric and the underlying process is non-Gaussian.[2] Density forecasts for exchange rates are useful in many economic and financial applications. For example, density forecasts can be employed to compute probabilities of business turning points which are important inputs for optimal business cycle turning point forecasts (Zellner et al., 1991). For another example, financial risk management is essentially dedicated to providing density forecasts for important economic variables, such as interest rates and exchange rates, and then using certain aspects of the distribution such as value-at-risk (VaR) to quantify the risk exposure of a portfolio (e.g., Duffie and Pan, 1997; Morgan, 1996; Jorion, 2000). Density forecasts also are important for valuing currency options, whose payoffs depend on the entire probability distribution of future exchange rates. Jorion (1988), Bates (1996), and Bollen et al. (2000) show that more realistic exchange rate models generate more accurate currency option prices.

Our paper makes both methodological and empirical contributions to the literature on density forecasts for exchange rates. Methodologically, we develop an out-of-sample omnibus nonparametric evaluation procedure for density forecasts of univariate time series models. The pioneering work of Diebold et al. (1998) shows that if a forecast model coincides with the true data-generating process (DGP), then the probability integral transformed data via the model conditional density, which are often referred to as the "generalized residuals" of the forecast model, should be i.i.d. U[0, 1].[3] While Diebold et al. (1998) separately examine the i.i.d. and U[0, 1] properties of the "generalized residuals" using some intuitive graphical methods, we develop a formal nonparametric portmanteau evaluation test for density forecasts by measuring the distance between the model generalized residuals from i.i.d. U[0, 1]. Our approach extends Hong and Li's (2005) approach for evaluating the in-sample performance of a continuous-time model to an out-of-sample forecasting context. The most appealing feature of our test is its omnibus ability to detect a wide range of suboptimal density forecasts for stationary and nonstationary time series processes. Moreover, we explicitly consider the impact of parameter estimation uncertainty on the evaluation procedure, an issue typically ignored in most existing forecast evaluation methods. We provide a simulation study on the finite sample performances of our tests, and develop a simple and distribution-free method for correcting the finite sample biases of the asymptotic tests.

[2]It is important to point out that there is a long tradition in the Bayesian forecasting literature of explicitly using predictive densities (see, e.g., Harrison and Stevens, 1976; West and Harrison, 1997). The so-called "prequential" Bayesian literature also features density forecasts prominently (see Dawid, 1984). For exchange rates, Putnam and Quintana (1994) and Quintana and Putnam (1996) have considered multivariate state space models for a vector of exchange rates and provide dynamic Bayesian predictive density forecasts for a portfolio of exchange rates. Based on these forecasts, they consider portfolio choices and trading strategies involving multiple currencies.

[3]The term "generalized residual" has been widely used in the econometrics and statistics literatures. For example, see Cox and Snell (1968) and Gourieroux et al. (1987).

Empirically, we provide probably the first comprehensive analysis of the density forecasting performances of a wide variety of nonlinear time series models for intraday high-frequency Euro/Dollar and Yen/Dollar exchange rates. Specifically, we consider density forecasts using random walk, GARCH/EGARCH, and jump-diffusion models with either N(0,1) or Student-t innovations. To examine the contribution of serial dependence in higher order conditional moments, we also examine some nonlinear time series models with non-i.i.d. innovations: a regime-switching model with state-dependent GARCH process and Student-t innovation, and Hansen's (1994) autoregressive conditional density (ARCD) model. These models, whose conditional densities cannot be fully described by the first two conditional moments, have been rarely applied to exchange rate data in the literature, particularly in an out-of-sample setting. While density forecasting has become a standard practice in many areas of economics and finance (e.g., Clements and Smith, 2000; Corradi and Swanson, 2006a; Tay and Wallis, 2000), applications to exchange rate data are still rare. One notable exception is Diebold et al. (1999), who consider forecasts of joint densities of exchange rates using a bivariate RiskMetrics model. Our paper naturally fills the gap in this literature.

Our analysis shows that some sophisticated models provide better out-of-sample density forecasts than the simple random walk model (even augmented with GARCH and Student-t innovations to account for the well-known volatility clustering and heavy tails of Euro/Dollar rates). For Euro/Dollar, it is important to model the heavy tails through a Student-t innovation and the asymmetric time-varying conditional volatility through a regime-switching GARCH model for both in-sample and out-of-sample performances; while modeling the conditional mean and serial dependence in higher order conditional moments (e.g., conditional skewness) is important for in-sample performance, it does not improve out-of-sample density forecasts. Overall, a regime-switching model with zero conditional mean, regime-dependent GARCH, and Student-t innovation provides the best density forecasts for the Euro/Dollar rate, and such a model is optimal in the sense that it cannot be rejected by the data. For the Yen/Dollar rate, it is also important to model heavy tails and volatility clustering, and the best density forecasting model is a RiskMetrics model with a Student-t innovation. However, this best forecast model for Yen/Dollar is still suboptimal, suggesting that there still exists room for further improvement in forecasting the density of Yen/Dollar.

Our empirical results on density forecasts dramatically differ from those on mean forecasts. The exchange rate dynamics is completely characterized by its conditional density, which includes not only the conditional mean but also higher order conditional moments. A model that better forecasts the conditional mean does not necessarily better forecast higher order conditional moments. Our results show that the general perception in the literature that simpler models always do better in out-of-sample applications does not apply to density forecasts. By capturing volatility clustering and time-varying higher order conditional moments, some nonlinear time series models can indeed perform well in forecasting the conditional density of future exchange rates. Our results suggest that certain sophisticated nonlinear time series models are indeed useful in out-of-sample applications that involve the entire probability density.

It should be noted that our empirical comparative study has excluded one important class of dynamic Bayesian models (see Zellner, 1971 for introduction to Bayesian econometrics). The dynamic Bayesian approach can provide predictive density models with time-varying parameters that adapt to the structural changes and regime-shifts. This approach has been proven successful in global macroeconomic forecasting (Zellner et al., 1991) and exchange rate forecasting (Putnam and Quintana, 1994; Quintana and Putnam, 1996). It would be interesting to compare the relative performance between dynamic Bayesian predictive density models and the best density forecast models we find in forecasting exchange rate changes.

The paper is planned as follows. In Section 2, we consider a portmanteau test for evaluating out-of-sample density forecasts of univariate time series models. Section 3 provides the finite sample performance of our nonparametric tests. In Section 4, we introduce a wide variety of time series models for exchange rates and discuss their relative merits. In Section 5, we describe the data, estimation methods, and in-sample and out-of-sample performances of each model. Section 6 concludes. The Appendix provides the mathematical proof for the asymptotic theory.

2. Out-of-sample density forecast evaluation

Density forecasts have become a standard practice in many areas of economics and finance. One of the most important issues in density forecasting is to evaluate the quality of a forecast (Granger, 1999). In a decision-theoretic context, Diebold et al. (1998) and Granger and Pesaran (2000) show that when a density forecast model coincides with the true conditional density of the DGP, it will be preferred by all forecast users regardless of their risk attitudes. However, density forecast evaluation is challenging because we never observe an *ex post* density. Except Diebold et al. (1998), Berkowitz (2001), Hong (2001), and Corradi and Swanson (2006b, c), so far there have been relatively few suitable statistical evaluation procedures for out-of-sample density forecasts.[4] To fill the gap in the literature, we now develop a generally applicable omnibus nonparametric evaluation method for out-of-sample density forecasts.

2.1. Dynamic probability integral transform

In a pioneering work, Diebold et al. (1998) first propose to assess the optimality of density forecasts by examining the dynamic probability integral transform of the data with respect to the density forecast model. Suppose $\{Y_t, t = 0, \pm 1, \ldots\}$ is a possibly nonstationary time series governed by a conditional density $p_0(y|I_{t-1}, t)$. For a given model for Y_t, there is a model-implied conditional density

$$\frac{\partial}{\partial y} \mathrm{P}(Y_t \leqslant y | I_{t-1}, \theta) \equiv p(y|I_{t-1}, t, \theta),$$

where θ is an unknown finite-dimensional parameter vector, $I_{t-1} \equiv \{Y_{t-1}, Y_{t-2}, \ldots, Y_1\}$ is the information set available at time $t-1$. We divide a random sample $\{Y_t\}_{t=1}^{T}$ of size T into two subsets: an estimation sample $\{Y_t\}_{t=1}^{R}$ of size R for estimating model parameters and a forecast sample $\{Y_t\}_{t=R+1}^{T}$ of size $n \equiv T - R$ for density forecast evaluation. We can then define the dynamic probability integral transform of the data with respect to the

[4] See Corradi and Swanson (2006a) for an excellent review of existing methods for evaluating density forecasts. See also Corradi and Swanson (2006b) on evaluation of density as well as confidence interval forecasts.

model forecast density:

$$Z_t(\theta) \equiv \int_{-\infty}^{Y_t} p(y|I_{t-1}, t, \theta)\,\mathrm{d}y, \quad t = R+1, \ldots, T. \tag{2.1}$$

Following Cox and Snell (1968), we refer to $\{Z_t(\theta)\}$ as the generalized residuals of the density model $p(y|I_{t-1}, t, \theta)$. The generalized residuals defined this way also have been used in duration analysis in labor economics (e.g., Lancaster (1990)). Diebold et al. (1998) show that if the model $p(y|I_{t-1}, t, \theta)$ is correctly specified in the sense that there exists some θ_0 such that $p(y|I_{t-1}, t, \theta_0)$ coincides with the true conditional density, then the transformed sequence $\{Z_t(\theta_0)\}$ should be i.i.d. U[0, 1].

To illustrate the idea, consider the popular J.P. Morgan's (1996) RiskMetrics model

$$\begin{cases} Y_t = \sqrt{h_t}\varepsilon_t, \\ h_t = (1-\theta)h_{t-1} + \theta Y_{t-1}^2 = (1-\theta)\sum_{j=1}^{\infty}\theta^j h_{t-j}, \\ \varepsilon_t \sim \text{i.i.d. N}(0,1), \end{cases}$$

where θ measures the dependence of h_t on past volatilities. Here, the conditional density of Y_t given I_{t-1} is

$$p(y|I_{t-1}, t, \theta) = \frac{1}{\sqrt{2\pi h_t}} \exp\left(-\frac{y^2}{2h_t}\right), \quad y \in (-\infty, \infty).$$

If $p(y|I_{t-1}, t, \theta_0) = p_0(y|I_{t-1}, t)$ almost surely for some θ_0, then

$$\begin{aligned} Z_t(\theta_0) &= \int_{-\infty}^{Y_t} p_0(y|I_{t-1}, t)\,\mathrm{d}y \\ &= \Phi(\varepsilon_t) \sim \text{i.i.d. U}[0,1], \end{aligned}$$

where $\Phi(\cdot)$ is the N(0,1) CDF.

The i.i.d. U[0, 1] property provides a convenient approach to evaluating the density forecast model $p(y|I_{t-1}, t, \theta)$. Intuitively, the U[0, 1] property indicates proper specification of the unconditional distribution of Y_t, and the i.i.d. property characterizes correct specification of its dynamic structure. If $\{Z_t(\theta)\}$ is not i.i.d. U[0, 1] for all $\theta \in \Theta$, then $p(y|I_{t-1}, t, \theta)$ is not optimal and there exists room to further improve $p(y|I_{t-1}, t, \theta)$. Thus the quality of density forecasts can be evaluated by testing whether the generalized residuals are i.i.d. U[0, 1].

2.2. Nonparametric evaluation for density forecasts

In this section, we develop a nonparametric procedure for density forecast evaluation by extending Hong and Li's (2005) in-sample nonparametric test for continuous-time models. For notational simplicity, put $Z_t = Z_t(\theta^*)$, where θ^* is the probability limit of some parameter estimator $\hat{\theta}_R$ based on the estimation sample $\{Y_t\}_{t=1}^R$.[5] Following Hong and Li (2005), we measure the distance between a density forecast model and the true conditional density by comparing a kernel estimator $\hat{g}_j(z_1, z_2)$ for the joint density of the pair $\{Z_t, Z_{t-j}\}$

[5]It is possible to extend our asymptotic analysis to allow for rolling and recursive estimations. However, we do not consider these possibilities here for simplicity and space.

with unity, the product of two U[0, 1] densities, where j is a lag order.[6] We further propose a portmanteau test statistic that combines the original Hong and Li's (2005) statistics at different lag orders. Compared to the graphical methods of Diebold et al. (1998) that separately examine the i.i.d. and U[0, 1] properties of the generalized residuals, our single omnibus evaluation criterion takes into account deviations from both i.i.d. and U[0, 1] jointly and provides an overall measure of model performance. We also explicitly consider the impact of parameter estimation uncertainty and the choice of relative sample sizes (R, n) between the estimation and prediction samples on the evaluation procedure. These two issues are typically ignored by most existing evaluation procedures for out-of-sample density forecasts. The most appealing feature of this new test is its omnibus ability to detect a wide range of suboptimal density forecasts for possibly nonstationary time series processes.

Specifically, our kernel estimator of the joint density of the pair $\{Z_t, Z_{t-j}\}$ is, for any $j > 0$,

$$\hat{g}_j(z_1, z_2) \equiv (n-j)^{-1} \sum_{t=R+j+1}^{T} K_h(z_1, \hat{Z}_t) K_h(z_2, \hat{Z}_{t-j}), \tag{2.2}$$

where $\hat{Z}_t = Z_t(\hat{\theta}_R)$, and $K_h(z_1, z_2)$ is a boundary-modified kernel function defined below: For $x \in [0, 1]$, we define

$$K_h(x, y) \equiv \begin{cases} h^{-1} k\left(\dfrac{x-y}{h}\right) / \int_{-(x/h)}^{1} k(u)\,\mathrm{d}u & \text{if } x \in [0, h), \\ h^{-1} k\left(\dfrac{x-y}{h}\right) & \text{if } x \in [h, 1-h], \\ h^{-1} k\left(\dfrac{x-y}{h}\right) / \int_{-1}^{(1-x)/h} k(u)\,\mathrm{d}u & \text{if } x \in (1-h, 1], \end{cases} \tag{2.3}$$

where $k(\cdot)$ is a prespecified symmetric probability density, and $h \equiv h(n)$ is a bandwidth such that $h \to 0, nh \to \infty$ as $n \to \infty$. In practice, the choice of bandwidth h is more important than the choice of the kernel $k(\cdot)$. Like Scott (1992), we choose $h = \hat{S}_Z n^{-1/6}$, where $\hat{S}_Z$ is the sample standard deviation of $\{\hat{Z}_t\}_{t=R+1}^{T}$. This simple bandwidth rule attains the optimal rate for bivariate kernel density estimation.

We use the above modified kernel because a standard kernel density estimator gives biased estimates near the boundaries of the data, due to its asymmetric coverage of the data in the boundary regions. In contrast, the weighting functions in the denominators of $K_h(x, y)$ for $x \in [0, h) \cup (1-h, 1]$ account for the asymmetric coverage and ensure that the kernel density estimator is asymptotically unbiased uniformly over the entire support [0, 1]. The modified-kernel approach allows us to use all the data in estimation. Otherwise, significant amounts of data in the boundary regions might have to be discarded due to the boundary bias problem.[7] For financial time series, one may be particularly interested in the tail distribution of the underlying process, which is exactly contained in (and only in) the boundary regions! Our approach also has advantages over the so-called jackknife

[6] One advantage of this approach is that since there is no serial dependence in $\{Z_t\}$ under correct model specification, nonparametric joint density estimators are expected to perform well in finite samples. There is also no asymptotic bias for nonparametric density estimators under the null hypothesis of correct model specification because the conditional density of Z_t given $\{Z_{t-1}, Z_{t-2}, \ldots\}$ is uniform (i.e., a constant).

[7] For a nearly uniformly distributed transformed sequence $\{Z_t\}$, the data in the boundary region are still about 10% when the sample size is 5000.

kernel used by Chapman and Pearson (2000) to eliminate boundary bias. The jackknife kernel may generate negative density estimates and has relatively large variances for the kernel estimates in the boundary regions, which could result in poor finite sample performance. In contrast, our modified kernel always produces nonnegative density estimates with a smaller variance in the boundary regions than a jackknife kernel.

Extending Hong and Li's (2005) in-sample specification test for a continuous-time model, we obtain the test statistic for out-of-sample density forecasts

$$\hat{Q}(j) \equiv \left[(n-j)h\int_0^1\int_0^1[\hat{g}_j(z_1,z_2)-1]^2\,\mathrm{d}z_1\,\mathrm{d}z_2 - hA_h^0\right]\Big/ V_0^{1/2}, \quad j=1,2,\ldots, \tag{2.4}$$

where the nonstochastic centering and scaling factors

$$A_h^0 \equiv \left[(h^{-1}-2)\int_{-1}^1 k^2(u)\,\mathrm{d}u + 2\int_0^1\int_{-1}^b k_b^2(u)\,\mathrm{d}u\,\mathrm{d}b\right]^2 - 1,$$

$$V_0 \equiv 2\left[\int_{-1}^1\left[\int_{-1}^1 k(u+v)k(v)\,\mathrm{d}v\right]^2\mathrm{d}u\right]^2,$$

and $k_b(\cdot) \equiv k(\cdot)/\int_{-1}^b k(v)\,\mathrm{d}v$. Note that the modification of the kernel $k(\cdot)$ in the boundary regions affects the centering constant A_h^0.

The use of the $\hat{Q}(j)$ statistics with different j's can reveal the information on the lag orders at which we have significant departures from i.i.d. U[0, 1]. However, when comparing two different models, it is desirable to construct a single portmanteau test statistic. Otherwise we would run into difficulty when one model has a smaller $\hat{Q}(j)$ at lag j_1 but the other model has a smaller $\hat{Q}(j)$ at lag $j_2 \neq j_1$. To avoid this, we propose the following portmanteau evaluation test statistic:

$$\hat{W}(p) = \frac{1}{\sqrt{p}}\sum_{j=1}^{p}\hat{Q}(j), \tag{2.5}$$

where p is a lag truncation order. This test can be viewed as a generalization of the popular Box–Pierce–Ljung autocorrelation test from a linear time series in-sample context to a nonlinear time series out-of-sample context. It can check model misspecifications in not only the conditional mean but also the entire conditional distribution of Y_t.

The derivation of the asymptotic distributions of $\hat{Q}(j)$ and $\hat{W}(p)$ is based on the following regularity conditions. Throughout, we use C to denote a generic bounded constant, $|\cdot|$ to denote the usual Euclidean norm, and $\frac{\partial}{\partial\theta}Z_t(\theta_0)$ to denote $\frac{\partial}{\partial\theta}Z_t(\theta)|_{\theta=\theta_0}$.

Assumption A.1. The random sample $\{Y_t\}_{t=1}^T$ is generated from an unknown conditional probability density function $p_0(y|I_{t-1},t) \equiv \frac{\partial}{\partial y}\mathrm{P}(Y_t \leqslant y|I_{t-1})$, where I_{t-1} is an information set (or sigma-field) at time $t-1$.

Assumption A.2. Let Θ be a finite-dimensional parameter space. (i) For each $\theta \in \Theta$, $p(y|I_{t-1},t,\theta)$ is a conditional density model for $\{Y_t\}$, and is a measurable function of (y, I_{t-1}); (ii) with probability one, $p(y|I_{t-1},t,\theta)$ is twice-continuously differentiable with respect to θ in a neighborhood Θ_0 of θ_0, with $\lim_{T-R\to\infty}(T-R)^{-1}\sum_{t=R+1}^T \mathrm{E}\sup_{\theta\in\Theta_0}|\frac{\partial}{\partial\theta}Z_t(\theta_0)|^{2v} \leqslant C$ for some constant $v>1$ and $\lim_{T-R\to\infty}(T-R)^{-1}\sum_{t=R+1}^T \mathrm{E}\sup_{\theta\in\Theta_0}|\frac{\partial^2}{\partial\theta\,\partial\theta'}Z_t(\theta_0)|^2 \leqslant C$, where $Z_t(\theta)$ is defined in (2.1).

Assumption A.3. (i) $G_{t-1}(z) \equiv \mathrm{E}[\frac{\partial}{\partial\theta}Z_t(\theta_0)|Z_t(\theta_0) = z, I_{t-1}]$ is a measurable function of (z, I_{t-1}); (ii) with probability one, $G_{t-1}(z)$ is continuously differentiable with respect to z, and $\lim_{T-R\to\infty}(T-R)^{-1}\sum_{t=R+1}^{T}\mathrm{E}|G'_{t-1}[Z_t(\theta_0)]|^2 \leqslant C$.

Assumption A.4. $\{Y_t, \frac{\partial}{\partial\theta}Z_t(\theta_0)\}'$ is a strong mixing process with strong mixing coefficient $\alpha(j)$ satisfying $\sum_{j=0}^{\infty}\alpha(j)^{(v-1)/v} \leqslant C$, where $v > 1$ is as in Assumption A.2.

Assumption A.5. $\hat{\theta}_R \equiv \hat{\theta}(\{Y_t\}_{t=1}^{R}) \in \Theta$ is a parameter estimator based on the first subsample $\{Y_t\}_{t=1}^{R}$ such that $R^{1/2}(\hat{\theta}_R - \theta^*) = \mathrm{O_P}(1)$, where $\theta^* \equiv p\lim_{R\to\infty}\hat{\theta}_R$ is an interior element in Θ and $\theta^* = \theta_0$ under the hypothesis of optimal density forecasts.

Assumption A.6. The kernel function $k : [-1, 1] \to \mathbb{R}^+$ is a symmetric, bounded, and twice continuously differentiable probability density such that $\int_{-1}^{1} k(u)\,\mathrm{d}u = 1$, $\int_{-1}^{1} uk(u)\,\mathrm{d}u = 0$, and $\int_{-1}^{1} u^2k(u)\,\mathrm{d}u < \infty$.

Assumption A.7. (i) The bandwidth $h = cn^{-\delta}$ for $c \in (0, \infty)$ and $\delta \in (0, \frac{1}{5})$, where $n \equiv T - R$; (ii) $n^{\lambda}/R \to 0$, where $\lambda < \max[1 - \delta, \frac{1}{2}(1 + 5\delta), (5 - \frac{2}{v})\delta]$.

Assumption A.1 is a regularity condition on the DGP of $\{Y_t\}$. We allow the functional form of the conditional density $p_0(y|I_{t-1}, t)$ to be time-varying. Assumptions A.2 and A.3 are regularity conditions on the conditional density model $p(y|I_{t-1}, t, \theta)$. Assumption A.4 characterizes temporal dependence in $\{Y_t, \frac{\partial}{\partial\theta}Z_t(\theta_0)\}$. The strong mixing condition is often used in nonlinear time series analysis, as is the case here. For the definition of the strong mixing condition, see (e.g.) White (1984, p. 45). We note that although $\{Z_t(\theta_0)\}$ is i.i.d. when the density forecast model is optimal, the sequence of its gradients $\{\frac{\partial}{\partial\theta}Z_t(\theta_0)\}$ is generally no longer i.i.d. Assumption A.5 allows for any in-sample $\sqrt{R}$-consistent estimator for θ_0, which need not be asymptotically most efficient. Assumption A.6 is a standard regularity condition on kernel function $k(\cdot)$. Assumption A.7 provides conditions on the bandwidth h and the relative speed between R and n, the sizes of the estimation sample and the prediction sample, respectively. We allow the optimal bandwidth rate (e.g., $h \propto n^{-1/6}$) for bivariate kernel estimation. Moreover, we allow the size of the prediction sample, n, to be larger or smaller than or the same as the size of the estimation sample, R. This offers a wide scope of applicability of our procedure, particularly when the whole sample $\{Y_t\}_{t=1}^{T}$ is relatively small.

Under the above regularity conditions, we have the following asymptotic results for $\hat{Q}(j)$ and $\hat{W}(p)$:

Theorem 1. *Suppose Assumptions* A.1–A.7 *hold. Then for any fixed integer* $j > 0$, *we have* $\hat{Q}(j) \to^{\mathrm{d}} \mathrm{N}(0, 1)$ *when density forecasts are optimal.*

Proof. See the Appendix.

Theorem 2. *Suppose Assumptions* A.1–A.7 *hold. Then* $\hat{W}(p) \to^{\mathrm{d}} \mathrm{N}(0, 1)$ *when density forecasts are optimal.*

Proof. See the Appendix.

Intuitively, $\hat{W}(p) \to^{\mathrm{d}} \mathrm{N}(0, 1)$ because when the density forecast model is optimal, we have that $\hat{Q}(i) \to^{\mathrm{d}} \mathrm{N}(0, 1)$, and $\mathrm{cov}[\hat{Q}(i), \hat{Q}(j)] \to^{\mathrm{p}} 0$ for $i \neq j$ as $R, n \to \infty$. Thus, $\hat{W}(p)$ is a normalized sum of approximately i.i.d. N(0, 1) random variables, and so is asymptotically N(0, 1).

To derive the asymptotic power of our tests when the density forecast model is suboptimal, we impose an additional condition.

Assumption A.8. For each integer $j>0$, the joint density $g_j(z_1, z_2)$ of the transformed random vector $\{Z_t, Z_{t-j}\}$, where $Z_t \equiv Z_t(\theta^*)$ and θ^* is as in Assumption A.5, exists and is continuously differentiable on $[0,1]^2$.

Theorem 3. *Suppose Assumptions* A.1–A.8 *hold. Then* (i) $(nh)^{-1}\hat{Q}(j) \to^p V_0^{-1/2} \int_0^1 \int_0^1 [g_j(z_1,z_2)-1]^2 \, dz_1 \, dz_2$ *for any fixed integer* $j>0$; (ii) *for any sequence of constants* $\{C_n = o(nh)\}$, $P[\hat{W}(p) > C_n] \to 1$ *whenever* Z_t *and* Z_{t-j} *are not independent or* U[0, 1] *at some lag* $j \in \{1, 2, \ldots, p\}$.

Proof. See the Appendix.

Theorem 3 suggests that as long as model misspecification occurs such that $\hat{Q}(j) \to \infty$ at some lag $j \in \{1, 2, \ldots, p\}$, we have $\hat{W}(p) \to \infty$ in probability. Therefore, $\hat{W}(p)$ can be used as an omnibus evaluation procedure for density forecasts.[8]

In fact, our asymptotic theory can be extended to the following general divergence measure which includes the quadratic form as a special case:

$$\hat{D}(j) = \int C[\hat{g}_j(z_1, z_2), 1] \, dz_1 \, dz_2,$$

where $C(f_1,f_2)$ is a divergence measure for two bivariate probability densities $f_1(z_1,z_2)$ and $f_2(z_1,z_2)$ such that $C(f_1,f_2)=0$, $\partial C(f_1,f_2)/\partial f_1 = 0$, and $\partial^2 C(f_1,f_2)/\partial f_1^2 = w(z_1,z_2) \neq 0$. Examples of $C(f_1,f_2)$ include the quadratic form

$$C(f_1,f_2) = [f_1(z_1,z_2) - f_2(z_1,z_2)]^2,$$

the Hellinger distance

$$C(f_1,f_2) = [\sqrt{f_1(z_1,z_2)} - \sqrt{f_2(z_1,z_2)}]^2,$$

and the Kullback–Leibler information criterion

$$C(f_1,f_2) = \ln[f_1(z_1,z_2)/f_2(z_1,z_2)].$$

For this class of divergence measures, we can construct test statistics similar to $\hat{Q}(j)$ and $\hat{W}(j)$ under the same set of regularity conditions.[9]

The model generalized residuals $\{\hat{Z}_t\}$ contain rich information on potential sources of model misspecifications and can be used for diagnostic analysis. For example, the U[0, 1] property of the generalized residual measures how well a density forecast model captures the marginal density of $\{Y_t\}$, while the i.i.d. property of the generalized residuals measures how well a density forecast model captures the dynamics of $\{Y_t\}$. Here, we also extend a

[8] We note that one could also consider a chi-square test, such as $C(p) = \sum_{j=1}^{p} \hat{Q}^2(j)$. This statistic is asymptotically χ_p^2 when the density forecast model is optimal. However, we expect it to be less powerful than $W(p)$, because the latter exploits the one-sided nature of the $\hat{Q}(j)$ statistic under the alternative hypothesis (i.e., $\hat{Q}(j)$ diverges to positive infinity under suboptimal density forecasts).

[9] We emphasize that our tests compare the relative performance between any two models based on their distances to the true DGP. To compare the relative performance between two potentially misspecified models directly, we need to develop a test similar to that of Diebold and Mariano (1995) or Giacomini and White (2003). The derivation of the asymptotic distribution for such a test statistic is not trivial in the present context, because nonparametric estimation is involved. The approach by Corradi and Swanson (2005) is expected to be very useful here. We leave this to future research.

class of rigorous in-sample separate inference procedures considered in Hong and Li (2005) to the out-of-sample setting. Specifically, we consider the following test statistics:

$$\mathrm{M}(m,l) \equiv \left[\sum_{j=1}^{n-1} w^2(j/p)(n-j)\hat{\rho}_{ml}^2(j) - \sum_{j=1}^{n-1} w^2(j/p)\right] \Bigg/ \left[\sum_{j=1}^{n-2} w^4(j/p)\right], \tag{2.6}$$

where $\hat{\rho}_{ml}(j)$ is the sample cross-correlation between $\hat{Z}_t^m$ and $\hat{Z}_{t-|j|}^l$, and $w(\cdot)$ is a weighting function for lag order j.[10] The $\mathrm{M}(m,l)$ test is an out-of-sample extension of Hong's (1996) spectral density tests for the adequacy of linear time series models. Extending the proof of Hong (1996), we can show that for each given pair of positive integers (m,l),

$$\mathrm{M}(m,l) \to^{\mathrm{d}} \mathrm{N}(0,1)$$

under the null hypothesis of optimal density forecasts, provided the lag truncation order $p \equiv p(n) \to \infty, p/n \to 0$. Moreover, parameter estimation uncertainty in $\hat{\theta}_R$ has no impact on the asymptotic distribution of $\mathrm{M}(m,l)$. Although the moments of the generalized residuals $\{Z_t\}$ are not exactly the same as that of the original time series $\{Y_t\}$, they are highly correlated. In particular, the choice of $(m,l) = (1,1), (2,2), (3,3), (4,4)$ is very sensitive to autocorrelations in level, volatility, skewness, and kurtosis of $\{Y_t\}$, respectively (see, e.g., Diebold et al., 1998). Furthermore, the choice of $(m,l) = (1,2)$ and $(2,1)$ is sensitive to ARCH-in-mean and leverage effects of $\{Y_t\}$, respectively. Different choices of orders (m,l) can thus examine various dynamic aspects of the underlying process $\{Y_t\}$. Like $\hat{Q}(j)$ and $\hat{W}(p)$, upper-tailed N(0,1) critical values are suitable for $\mathrm{M}(m,l)$.

3. Finite sample performances

3.1. *A simple and distribution-free correction for finite sample bias*

We now study the finite sample performances of the $\hat{Q}(j)$ and $\hat{W}(p)$ tests for density forecasts, using some nonlinear time series models to be used in our empirical study. Hong and Li (2005) show that the in-sample $\hat{Q}(j)$ test has excellent finite sample performance for both univariate and multivariate continuous-time models: The test has excellent size and power performances for a sample size as small as 250. However, we find that for density forecasts, both tests, especially $\hat{W}(p)$, tend to overreject the null hypothesis when asymptotic critical values are used. For example, the rejection rates of $\hat{W}(p)$ can be about 20% (10%) at the 10% (5%) significance level even for $n = 1000$. The fact that the out-of-sample generalized residuals are computed based on the parameter estimates obtained from the in-sample observations could generate larger variations in $\hat{Q}(j)$ and $\hat{W}(p)$ in finite samples, which could lead to the observed overrejection. Moreover, although $\{\hat{Q}(j)\}$ should be independent from each other asymptotically, we find nontrivial correlations among $\hat{Q}(j)$'s in finite samples. As a result, the asymptotic distribution of $\hat{W}(p)$ tends to underestimate the finite sample variance of the test statistic and thus leads to overrejection.

[10]We assume that $w(\cdot)$ is symmetric around 0 and continuous on the real line except for a finite number of points. An example is the Bartlett kernel $w(z) = (1-|z|)\mathbf{1}(|z| \leqslant 1)$. If $w(\cdot)$ has bounded support, p is a lag truncation order; if $w(\cdot)$ has unbounded support, all $n-1$ lags in the prediction sample are used. Usually $w(\cdot)$ discounts higher order lags. This will give better power than equal weighting when $|\rho_{ml}(j)|$ decays to zero as lag order j increases. This is typically the case for most financial markets, where more recent events tend to have bigger impact than remote past events.

To deal with this issue, we propose a simple and distribution-free method for correcting the finite sample biases of the proposed tests. Based on the fact that the generalized residuals of an optimal density forecast model follow i.i.d. U[0, 1], we can obtain critical values of $\hat{Q}(j)$ and $\hat{W}(p)$ by using simulated i.i.d. U[0, 1] random variables. Specifically, the method can be described as follows:

- Step 1: For each forecast sample size n, generate B (a large number) data sets from the i.i.d. U[0, 1] random sample of size n. That is, for $b = 1, 2, \ldots, B$, we generate $\{Z_t^{(b)}\}_{t=1}^n \sim$ i.i.d. U[0, 1].
- Step 2: For each simulated data set $\{Z_t^{(b)}\}_{t=1}^n, b = 1, 2, \ldots, B$, compute the test statistics $\hat{Q}^{(b)}(j)$ and $\hat{W}^{(b)}(p)$. A set of simulated test statistics $\{\hat{Q}^{(b)}(j), \hat{W}^{(b)}(p)\}_{b=1}^B$ is obtained.
- Step 3: The critical values $Q_\alpha(j)$ and $W_\alpha(p)$ for $\hat{Q}(j)$ and $\hat{W}(p)$ at significance level $\alpha \in (0, 1)$ are defined as, respectively,

$$\frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\{\hat{Q}^{(b)}(j) > Q_\alpha(j)\} = \alpha \quad \text{and} \quad \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\{\hat{W}^{(b)}(p) > W_\alpha(p)\} = \alpha,$$

  where $\mathbf{1}\{\cdot\}$ is an indicator function.
- Step 4: Reject the null hypothesis of optimal density forecasts at significance level α if $\hat{Q}(j) > Q_\alpha(j)$ or $\hat{W}(p) > W_\alpha(p)$.

The critical values obtained this way are exact finite sample critical values for $\hat{Q}(j)$ and $\hat{W}(p)$ if the true parameter value θ_0 were used in computing the generalized residuals. When a $\sqrt{R}$-consistent estimator $\hat{\theta}_R$ rather than θ_0 is used, the obtained critical values are not exact finite sample critical values for $\hat{Q}(j)$ and $\hat{W}(j)$, due to the impact of parameter estimation uncertainty in $\hat{\theta}_R$. Nevertheless, Theorem A.2 of the Appendix shows that parameter estimation uncertainty in $\hat{\theta}_R$ has no impact on the asymptotic distribution of $\hat{Q}(j)$ and so $\hat{W}(p)$ as well; that is, the asymptotic distributions of $\hat{Q}(j)$ and $\hat{W}(p)$ remain unchanged when θ_0 is replaced with $\hat{\theta}_R$. As a result, the critical values obtained from the above simulation procedure are asymptotically valid for $\hat{Q}(j)$ and $\hat{W}(p)$, and so can be used in practice. This method is very easy to implement and is distribution-free. In a simulation study below, we examine the impact of parameter estimation uncertainty in $\hat{\theta}_R$ on this procedure. We find that the simulated critical values provide (i) much better finite sample approximations than the asymptotic critical values and (ii) rather reasonable finite sample performances when $R/n \geqslant 2$ for n as small as 250.[11]

3.2. *Size performances of $\hat{Q}(j)$ and $\hat{W}(p)$*

To examine the size performances of $\hat{Q}(j)$ and $\hat{W}(p)$, we consider the following two models:

- Random-Walk-Normal Model (RW-N):

$$\begin{cases} Y_t = \sigma\varepsilon_t, \\ \varepsilon_t \sim \text{i.i.d. N}(0, 1). \end{cases}$$

[11] One could use bootstrap to obtain more accurate finite sample critical values that take into account parameter estimation uncertainty. However, bootstrap is much more tedious and much more difficult to implement, especially in the current out-of-sample and nonlinear setting we consider. See Corradi and Swanson (2007) for related issues on bootstrap.

- GARCH-Normal Model (GARCH-N):

$$\begin{cases} Y_t = \sqrt{h_t}\varepsilon_t, \\ h_t = \beta_0 + h_{t-1}(\beta_1 + \beta_2\varepsilon_{t-1}^2), \\ \varepsilon_t \sim \text{i.i.d. N}(0,1). \end{cases}$$

The parameters of each model are the same as the parameter estimates obtained from the data on Euro/Dollar exchange rates used in our empirical study below. That is, $\sigma = 2.77$ and $(\beta_0, \beta_2, \beta_3) = (0.76, 0.77, 0.14)$. For each model, we simulate 3000 data sets from the random sample $\{Y_t\}_{t=1}^T$, where $T = R + n$. We consider $n = 250, 500$, and 1000, and $R/n = 1, 2$, and 3 for each n. That is, we choose the forecast sample size n to be 250, 500, and 1000 and the ratio between the estimation and forecast sample sizes to be 1, 2, and 3. For each random sample, we estimate model parameters using the first R observations via the maximum likelihood estimation (MLE) method and compute model generalized residuals using the n observations in the prediction sample.

Panel A of Table 1 reports the size performances of $\hat{Q}(j)$ and $\hat{W}(p)$ under RW-N and GARCH-N, using the simulated critical values. One of the most interesting findings from our simulation studies is that the sample size ratio R/n is crucial for the size performances of both tests, a feature that is unique to the out-of-sample analysis. For example, when $R/n = 1$, both tests overreject the null hypothesis at conventional significance levels even if we increase the out-of-sample size n from 250 to 1000. The rejection rates for $\hat{W}(p)$ at the 10% (5%) level are about 14% (8%) for RW-N, and 16% (10%) for GARCH-N. However, if we increase R/n to 2 and 3, both tests have reasonable size performances under both RW-N and GARCH-N when $n = 1000$. In particular, under RW-N, both tests have rejection rates very close to significance levels even when the sample size n is as small as 250.

3.3. Power performances of $\hat{Q}(j)$ and $\hat{W}(p)$

To investigate the power of the $\hat{Q}(j)$ and $\hat{W}(p)$ tests, we generate data from three alternative DGPs and test the null hypothesis that the data are generated from RW-N using the simulated critical values. The three DGPs are:

- GARCH-Normal Model (GARCH-N):

$$\begin{cases} Y_t = \sqrt{h_t}\varepsilon_t, \\ h_t = \beta_0 + h_{t-1}(\beta_1 + \beta_2\varepsilon_{t-1}^2), \\ \varepsilon_t \sim \text{i.i.d. N}(0,1). \end{cases}$$

- Random-Walk-T (RW-T) Model:

$$\begin{cases} Y_t = \sigma\varepsilon_t, \\ \varepsilon_t \sim \text{i.i.d.} \sqrt{\frac{\nu - 2}{\nu}} t(\nu). \end{cases}$$

Table 1
Finite sample performance of $Q(j)$ and $W(p)$ tests

(A) Size performance of $Q(j)$ and $W(p)$ tests for RW-N and GARCH-N models

		RW-N									GARCH-N								
		$R/n=1$			$R/n=2$			$R/n=3$			$R/n=1$			$R/n=2$			$R/n=3$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
$n=250$	$Q(1)$	0.130	0.073	0.017	0.106	0.056	0.011	0.102	0.048	0.007	0.169	0.104	0.037	0.123	0.066	0.014	0.111	0.058	0.010
	$Q(10)$	0.132	0.074	0.021	0.112	0.057	0.013	0.108	0.058	0.014	0.163	0.103	0.044	0.132	0.075	0.017	0.107	0.060	0.016
	$Q(20)$	0.129	0.069	0.021	0.115	0.061	0.013	0.117	0.059	0.011	0.165	0.100	0.040	0.130	0.072	0.017	0.108	0.055	0.013
	$W(5)$	0.139	0.075	0.020	0.119	0.063	0.012	0.113	0.055	0.010	0.191	0.124	0.045	0.146	0.078	0.018	0.115	0.061	0.013
	$W(10)$	0.144	0.080	0.021	0.121	0.063	0.015	0.113	0.058	0.010	0.196	0.121	0.047	0.149	0.079	0.017	0.114	0.065	0.014
	$W(20)$	0.138	0.079	0.021	0.122	0.062	0.013	0.113	0.059	0.010	0.198	0.120	0.047	0.148	0.080	0.016	0.114	0.067	0.014
$n=500$	$Q(1)$	0.131	0.075	0.021	0.118	0.059	0.014	0.106	0.053	0.009	0.170	0.107	0.035	0.122	0.070	0.015	0.126	0.064	0.015
	$Q(10)$	0.142	0.082	0.019	0.105	0.058	0.012	0.113	0.056	0.011	0.161	0.095	0.027	0.121	0.060	0.012	0.115	0.064	0.012
	$Q(20)$	0.137	0.079	0.018	0.104	0.057	0.012	0.113	0.060	0.009	0.155	0.099	0.031	0.119	0.071	0.017	0.120	0.065	0.016
	$W(5)$	0.143	0.079	0.025	0.120	0.066	0.014	0.112	0.058	0.008	0.183	0.109	0.036	0.137	0.072	0.016	0.136	0.070	0.015
	$W(10)$	0.149	0.082	0.024	0.121	0.062	0.014	0.113	0.058	0.009	0.186	0.107	0.034	0.139	0.078	0.014	0.131	0.072	0.015
	$W(20)$	0.154	0.078	0.024	0.124	0.064	0.014	0.114	0.059	0.010	0.184	0.107	0.035	0.137	0.077	0.014	0.135	0.068	0.014
$n=1000$	$Q(1)$	0.131	0.075	0.021	0.118	0.059	0.014	0.110	0.057	0.013	0.140	0.082	0.031	0.120	0.062	0.017	0.115	0.060	0.014
	$Q(10)$	0.142	0.082	0.019	0.105	0.058	0.012	0.104	0.055	0.011	0.142	0.081	0.022	0.109	0.061	0.015	0.105	0.054	0.015
	$Q(20)$	0.137	0.079	0.018	0.104	0.057	0.012	0.113	0.057	0.013	0.143	0.081	0.028	0.118	0.064	0.015	0.106	0.055	0.012
	$W(5)$	0.143	0.079	0.025	0.120	0.066	0.014	0.112	0.058	0.012	0.161	0.098	0.031	0.126	0.068	0.017	0.115	0.063	0.015
	$W(10)$	0.149	0.082	0.024	0.121	0.062	0.014	0.117	0.056	0.013	0.163	0.101	0.035	0.131	0.069	0.018	0.117	0.062	0.015
	$W(20)$	0.154	0.078	0.024	0.124	0.064	0.014	0.116	0.059	0.014	0.169	0.099	0.033	0.131	0.068	0.017	0.123	0.057	0.015

Table 1 (*continued*)

(B) Power performance of $Q(j)$ and $W(p)$ tests for RW-T, GARCH-N, and RS-T models

		RW-T						GARCH-N						RS-T					
		$R/n=1$		$R/n=2$		$R/n=3$		$R/n=1$		$R/n=2$		$R/n=3$		$R/n=1$		$R/n=2$		$R/n=3$	
		10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
$n=250$	$Q(1)$	0.465	0.368	0.421	0.334	0.399	0.306	0.465	0.368	0.421	0.334	0.399	0.306	0.8010	0.7267	0.8663	0.8063	0.8857	0.8213
	$Q(10)$	0.396	0.313	0.350	0.271	0.323	0.248	0.423	0.341	0.350	0.271	0.323	0.248	0.7957	0.7307	0.8647	0.8037	0.8853	0.8270
	$Q(20)$	0.385	0.304	0.336	0.256	0.318	0.236	0.396	0.313	0.336	0.256	0.318	0.236	0.7857	0.7197	0.8547	0.7930	0.8690	0.8050
	$W(5)$	0.491	0.398	0.449	0.345	0.422	0.330	0.385	0.304	0.449	0.345	0.422	0.330	0.8543	0.7877	0.9167	0.8617	0.9273	0.8787
	$W(10)$	0.473	0.388	0.443	0.333	0.409	0.312	0.491	0.398	0.443	0.333	0.409	0.312	0.8590	0.7940	0.9227	0.8683	0.9320	0.8843
	$W(20)$	0.458	0.379	0.422	0.322	0.394	0.301	0.473	0.388	0.422	0.322	0.394	0.301	0.8590	0.7983	0.9237	0.8710	0.9337	0.8893
$n=500$	$Q(1)$	0.589	0.493	0.544	0.451	0.527	0.416	0.458	0.379	0.544	0.451	0.527	0.416	0.9740	0.9607	0.9907	0.9823	0.9940	0.9890
	$Q(10)$	0.454	0.369	0.410	0.328	0.392	0.297	0.589	0.493	0.410	0.328	0.392	0.297	0.9777	0.9620	0.9903	0.9823	0.9943	0.9900
	$Q(20)$	0.427	0.350	0.376	0.294	0.349	0.261	0.496	0.407	0.376	0.294	0.349	0.261	0.9783	0.9613	0.9883	0.9813	0.9923	0.9873
	$W(5)$	0.615	0.504	0.573	0.473	0.569	0.446	0.454	0.369	0.573	0.473	0.569	0.446	0.9877	0.9783	0.9953	0.9913	0.9983	0.9933
	$W(10)$	0.574	0.471	0.530	0.435	0.533	0.416	0.427	0.350	0.530	0.435	0.533	0.416	0.9897	0.9817	0.9963	0.9927	0.9980	0.9940
	$W(20)$	0.539	0.432	0.496	0.389	0.484	0.368	0.615	0.504	0.496	0.389	0.484	0.368	0.9920	0.9810	0.9963	0.9933	0.9973	0.9943
$n=1000$	$Q(1)$	0.737	0.650	0.745	0.652	0.740	0.641	0.574	0.471	0.745	0.652	0.740	0.641	0.9997	0.9997	1.0000	1.0000	1.0000	1.0000
	$Q(10)$	0.520	0.434	0.485	0.400	0.471	0.379	0.539	0.432	0.485	0.400	0.471	0.379	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$Q(20)$	0.466	0.384	0.426	0.338	0.408	0.317	0.737	0.650	0.426	0.338	0.408	0.317	1.0000	0.9997	1.0000	1.0000	1.0000	1.0000
	$W(5)$	0.776	0.682	0.766	0.664	0.765	0.656	0.609	0.525	0.766	0.664	0.765	0.656	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$W(10)$	0.717	0.619	0.695	0.600	0.697	0.582	0.520	0.434	0.695	0.600	0.697	0.582	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

This table reports the finite sample performance of $Q(j)$ and $W(p)$ tests. For size performance, we simulate 3000 random samples from RW-N and GARCH-N of length T ($T=R+n$, R is the number of in-sample observations and n is the number of out-of-sample observations). For each random sample of each model, we estimate model parameters using the estimation sample and compute the model generalized residuals using the forecast sample based on the estimated model parameters. We reject the null hypothesis if $Q(j)$ or $W(p)$ is greater than the finite sample critical values for n. For power performance, we simulate 3000 random samples from GARCH-N, RW-T, and RS-T and test whether the data are generated from RW-N using finite sample critical values.

- Regime-Switching-T (RS-T):

$$\begin{cases} Y_t = \sigma(s_t)\varepsilon_t, \\ \varepsilon_t \sim m.d.s. \sqrt{\dfrac{v(s_t)-2}{v(s_t)}}\, t(v(s_t)), \end{cases}$$

where $s_t = 1$ and 2, and the transition probability between the two regimes is defined in (4.1) below.

The parameters of each model are the same as the parameter estimates obtained from the data on Euro/Dollar exchange rates used in our empirical analysis below. That is, $(\beta_0, \beta_2, \beta_3) = (0.76, 0.77, 0.14)$, $(\sigma, v) = (2.78, 3.39)$, $(\sigma_1, \sigma_2, v_1, v_2) = (1.81, 3.67, 6.92, 3.88)$, and $(c_1, c_2, d_1, d_2) = (3.12, 2.76, 0, 0)$, where c_1, c_2, d_1, d_2 are given in (4.1). Again, for each model, we simulate 3000 data sets of the random sample $\{Y_t\}_{t=1}^T$, where $T = R + n$. We consider $n = 250, 500$, and 1000, and $R/n = 1, 2$, and 3 for each n. For each data set, we estimate an RW-N model via MLE using the first R observations, and based on the estimated model parameters, we compute the generalized residuals of RW-N using the remaining n observations. We test whether the null hypothesis of RW-N can be rejected by comparing the out-of-sample $\hat{Q}(j)$ and $\hat{W}(p)$ statistics with their corresponding simulated critical values.

Among the above three DGPs, GARCH-N is probably the most difficult to distinguish from RW-N, given that its conditional density is Gaussian. Of course, RW-N ignores volatility clustering in the data and assumes a constant volatility. The other two models should be easier to distinguish from RW-N because their conditional densities exhibit heavy tails due to their Student-t innovations.

Simulation results in Panel B of Table 1 show that both tests have reasonably good powers under GARCH-N. When $n = 250$, the rejection rates of both tests are about 40% (30%) at the 10% (5%) level. When $n = 1000$, the rejection rates increase to about 60% (50%) at the 10% (5%) level. For each sample size n, the rejection rates decline a bit as the ratio R/n increases. Perhaps a longer estimation sample could yield more accurate estimates of average volatility and thus allow RW-N to mimic GARCH-N better because both models have Gaussian conditional densities.

Both tests have excellent power under RW-T and RS-T. Even for $n = 250$, the rejection rates of both tests at the 5% level are close to 90%. The rejection rates increase to 1 when the sample size n increases to 500 and 1000. Unlike under GARCH-N, the rejection rates of both tests become even higher when R/n increases under RW-T and RS-T.

In summary, with the simulated critical values obtained as described in Section 3.1, both the $\hat{Q}(j)$ and $\hat{W}(p)$ tests have reasonable sizes and powers in finite samples, provided the sample size ratio $R/n \geqslant 2$ (the higher the ratio, the better the sizes).

4. Exchange rate models

We now introduce a wide variety of time series models of exchange rates that we will use for density forecasts. These models include geometric random walk, GARCH/EGARCH, jump, regime-switching, and Hansen's (1994) ARCD models. They represent one (or a combination) of three different approaches in capturing the leptokurtic distribution of exchange rates: (i) a Paretian stable or a Student-t distribution which has fatter tails than a normal distribution; (ii) a conditionally normal distribution with time-varying moments (e.g., GARCH models); (iii) a mixture of normal distributions with different means or variances, or a mixture of a normal and jump process.

4.1. Geometric random walk models

The geometric random walk (or lognormal) model has been widely used to capture the dynamics of financial time series. While it has been documented that the random walk model outperforms economic structural and time series models in forecasting the conditional mean of exchange rate changes, whether it also has better density forecasts is unknown. Let $Y_t = 100\ \ln(P_t/P_{t-1})$ be the relative change of an exchange rate from period $t-1$ to period t, where P_t is the nominal exchange rate at time t. Then the geometric random walk model with no drift is given by

$$\begin{cases} Y_t = \sigma\varepsilon_t, \\ \varepsilon_t \sim \text{i.i.d. N}(0,1) \text{ or i.i.d. } \sqrt{\dfrac{v-2}{v}}t(v). \end{cases}$$

While the conventional random walk has an i.i.d. normal innovation, we also consider a Student-t innovation (the degree of freedom v is estimated from data) to check whether modeling leptokurtosis in exchange rate data can improve density forecasts.

4.2. RiskMetrics models

Various conditional variance models have been proposed to capture volatility clustering in exchange rate data. For those models, although the conditional distribution is usually normal, the unconditional distribution has fat tails because of time-varying conditional variance. In practice, one popular way to model serial dependence in conditional variance is J.P. Morgan's (1996) RiskMetrics model in which the conditional variance is a weighted average of past squared changes:

$$\begin{cases} Y_t = \sqrt{h_t}\varepsilon_t, \\ h_t = (1-\theta)\sum_{j=1}^{\infty}\theta^j Y_{t-j}^2, \quad 0<\theta<1, \\ \varepsilon_t \sim \text{i.i.d. N}(0,1) \text{ or i.i.d. } \sqrt{\dfrac{v-2}{v}}t(v), \end{cases}$$

where θ governs the persistence of dependence on past volatility. This exponential smoothing technique is a simple but effective forecasting method in time series analysis.

This model has been widely used in the risk management industry for VaR calculation. In practice, one typically uses normal innovations for $\{\varepsilon_t\}$, with a prespecified value for θ. For example, J.P. Morgan (1996) suggests $\theta = 0.94$ for daily financial series. Here, we also consider a Student-t innovation for $\{\varepsilon_t\}$ and estimate v and θ from data. The RiskMetrics model allows us to examine the incremental contribution of modeling volatility clustering to density forecasts.

4.3. GARCH/EGARCH models

An alternative approach to modeling conditional variance is the ARCH/GARCH models of Engle (1982) and Bollerslev (1986), which have been very successful in modeling persistent volatility clustering in financial time series. Like the RiskMetrics model, ARCH/GARCH models can generate unconditionally leptokurtic distributions, although their conditional distributions may be normal. Many authors, such as Bollerslev (1986), Engle and Bollerslev (1986), Baillie and Bollerslev (1989), and Hsieh (1989), have shown that a GARCH(1,1) model with a Student-t innovation can capture weekly and daily exchange rates well. To understand the incremental contribution of GARCH effect, we consider the following GARCH/EGARCH models:

$$\begin{cases} Y_t = \sqrt{h_t}\varepsilon_t, \\ h_t = \beta_0 + h_{t-1}(\beta_1 + \beta_2\varepsilon_{t-1}^2) \quad \text{for GARCH; or} \\ \ln h_t = \beta_0 + \beta_1 \ln h_{t-1} + \beta_2(\varepsilon_{t-1} + \beta_3|\varepsilon_{t-1}|) \quad \text{for EGARCH,} \\ \varepsilon_t \sim \text{i.i.d. N}(0,1) \text{ or i.i.d. } \sqrt{\dfrac{v-2}{v}}t(v). \end{cases}$$

The EGARCH model, proposed in Nelson (1991), can capture asymmetric behavior in volatility.

We note that the RiskMetrics model is a special case of GARCH(1,1), with $\beta_0 = 0$, $\beta_1 = \theta$, and $\beta_2 = 1 - \theta$. This is essentially an integrated GARCH (1,1) process. Although it has some undesirable probability properties (see, e.g., Nelson, 1991), the RiskMetrics model is very popular in financial risk management due to its simplicity and intuitive appeal.

4.4. Jump models

Various economic shocks, news announcements, and interventions in foreign exchange markets by monetary authorities could have pronounced effects on exchange rate movements and generate jumps in exchange rates. Jorion (1988) and Bates (1996) have shown that Poisson jump models can capture the excess kurtosis of exchange rates and help improve currency option pricing.

Following Jorion (1988), we consider the following Poisson jump model:

$$\begin{cases} Y_t = \begin{cases} \sigma\varepsilon_t + \sum_{i=1}^{N_i} \ln J_i & \text{for jump,} \\ \sqrt{h_t}\varepsilon_t + \sum_{i=1}^{N_i} \ln J_i & \text{for jump-GARCH,} \end{cases} \\ h_t = \beta_0 + \beta_1[Y_{t-1} - \mathrm{E}(Y_{t-1}|Y_{t-2})]^2 + \beta_2 h_{t-1}, \\ \varepsilon_t \sim \text{i.i.d. N}(0,1) \text{ or i.i.d. } \sqrt{\dfrac{v-2}{v}}t(v), \\ \ln J_i \sim \text{i.i.d. N}(0, \delta^2), \\ N_i \sim Poisson(\lambda). \end{cases}$$

While Jorion (1988) only considers a normal innovation for $\{\varepsilon_t\}$, we also consider a Student-t innovation for the "smooth" part. Thus, we can compare the relative contribution of Student-t distribution and jumps in capturing the fat tails of exchange rates.

The conditional density of the above jump model can be written as

$$p(y|I_{t-1}) = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \mathrm{e}^{-\lambda} \frac{1}{\sqrt{2\pi(\sigma^2 + \delta^2 j)}} \exp\left\{ \frac{-y^2}{2(\sigma^2 + \delta^2 j)} \right\}$$

if $\varepsilon_t \sim$i.i.d. N(0, 1), and

$$p(y|I_{t-1}) = \mathrm{e}^{-\lambda} f_v(y) + \sum_{j=1}^{\infty} \frac{\lambda^j}{j!} \mathrm{e}^{-\lambda} \int_{-\infty}^{\infty} \phi_j(y-v) f_v(v)\, \mathrm{d}v$$

if $\varepsilon_t \sim$i.i.d. $\sqrt{\frac{v-2}{v}} t(v)$, where $f_v(v)$ is the normalized $t(v)$ density and $\phi_j(\cdot)$ is the $\mathrm{N}(\mu j, \delta^2 j)$ density. Note that the conditional density of the jump model is a mixture of normal and Student-t distributions, which can easily generate excess kurtosis and heavy tails. We can also obtain similar expressions for the conditional density of a jump-GARCH model.

It is much more difficult to estimate a jump model with a Student-t innovation than one with a normal innovation. This is because to calculate the conditional density of $\{Y_t\}$ for the former we need to compute the convolution between a normal and a Student-t distribution through numerical integration. Furthermore, when we estimate the model via MLE, for a given small change in parameter values as part of the optimization procedure, the numerical integration has to be repeated for every single observation.

4.5. ARCD models

In all the above models the conditional density of Y_t can be completely captured by its first two conditional moments. However, in reality the conditional density of Y_t may depend on higher order moments, such as conditional skewness and kurtosis. Thus, it is important and interesting to examine whether modeling dependence in higher order moments can help improve density forecasts. The most well-known econometric model that explicitly accounts for dependence in higher order moments is Hansen's (1994) ARCD model. This model generalizes Engle's (1982) ARCH model by allowing shape (skewness and kurtosis) parameters of the innovation distribution to depend upon conditioning past information. This is achieved by using a low-dimensional generalized skewed Student-t distribution with time-varying parameters. By modeling serial dependence in higher order moments, Hansen's ARCD model may provide additional benefits in forecasting the probability density of exchange rates. We consider the following ARCD model in which the conditional skewness λ_t and kurtosis v_t follow an autoregressive process:

$$\begin{cases} Y_t = \sqrt{h_t}\varepsilon_t \equiv e_t, \\ h_t = \beta_0 + h_{t-1}(\beta_1 + \beta_2 \varepsilon_{t-1}^2), \\ \varepsilon_t \sim \text{i.i.d.} \sqrt{\dfrac{v-2}{v}} t(v) \text{ or } m.d.s.(0,1) \text{ with } pdf\ g(z|v_t, \lambda_t), \\ \begin{cases} v_t = \beta_0^v + \beta_1^v e_{t-1} + \beta_2^v e_{t-1}^2, \\ \lambda_t = \beta_0^\lambda + \beta_1^\lambda e_{t-1} + \beta_2^\lambda e_{t-1}^2, \end{cases} \end{cases}$$

where

$$g(z|v,\lambda) = \begin{cases} bc\left(1+\frac{1}{v-2}\left(\frac{bz+a}{1-\lambda}\right)^2\right)^{-(v+1)/2}, & z<-a/b, \\ bc\left(1+\frac{1}{v-2}\left(\frac{bz+a}{1-\lambda}\right)^2\right)^{-(v+1)/2}, & z\geqslant -a/b, \end{cases}$$

$2<v<\infty$, $-1<\lambda<1$, $a=4\lambda c(\frac{v-2}{v-1})$, $b^2=1+3\lambda^2-a^2$, and $c=\frac{\Gamma(\frac{v+1}{2})}{\sqrt{\pi(v-2)}\Gamma(\frac{v}{2})}$.

Hansen (1994) applies the ARCD models with a conditional Student-t distribution and a conditional skewed Student-t distribution to the excess returns of one month U.S. Treasury bill and monthly Dollar/Swiss Franc exchange rates, respectively. He finds that the autoregressive parameters for the conditional skewness and degree of freedom are all statistically significant.[12]

4.6. Markov regime-switching models

Another popular nonlinear time series model that has been widely used to model exchange rates and other economic variables is Hamilton's (1989) regime-switching model. Engel and Hamilton (1990) apply the model to capture the "long swings" in major dollar exchange rates in the 1970s and 1980s. In their model, exchange rate changes follow a process governed by an unobservable state variable s_t, which follows a two-state Markov chain. When $s_t=1$, $Y_t \sim \mathrm{N}(\mu_1,\sigma_1)$, and when $s_t=2$, $Y_t \sim \mathrm{N}(\mu_2,\sigma_2)$. Thus, the unconditional distribution of exchange rate changes is a mixture of two normal distributions with different means and/or variances. This model can generate unimodal or bimodal distributions and allows for great flexibility in modeling skewness, kurtosis, and fat tails. It also can capture part of volatility clustering. Engel and Hamilton (1990) show that this model outperforms the random walk model in both in-sample fitting and out-of-sample forecasting. Engel (1994), however, shows that this model does not outperform the random walk model in forecasting the conditional mean of exchange rate changes for 18 currencies in terms of mean squared error.

While Engel and Hamilton (1990) and Engel (1994) consider quarterly data, we are interested in intraday data which might exhibit more significant nonlinear behavior. We extend their models to include regime-dependent Student-t innovations and GARCH effects:

$$\begin{cases} Y_t = \begin{cases} \sigma(s_t)\varepsilon_t & \text{or} \\ \sqrt{h_t}\varepsilon_t, \end{cases} \\ h_t = \beta_0(s_t)+\beta_1(s_t)[Y_{t-1}-\mathrm{E}(Y_{t-1}|Y_{t-2})]^2+\beta_2(s_t)h_{t-1}, \\ \varepsilon_t \sim \text{i.i.d. N}(0,1) \text{ or } m.d.s. \sqrt{\frac{v(s_t)-2}{v(s_t)}}t(v(s_t)). \end{cases}$$

We refer to the regime in which $s_t=1$ ($s_t=2$) as the first (second) regime. Following Ang and Bekaert (1998), we assume that the conditional probability of s_t depends on the one

[12] We thank Bruce Hansen for sharing his GAUSS program for estimating ARCD models, which is used in our empirical analysis.

period-lagged exchange rate change:

$$\mathrm{P}(s_t = l|s_{t-1} = l) = \frac{1}{1 + \exp(-c_l - d_l Y_{t-1})}, \quad l = 1, 2. \tag{4.1}$$

Thus, this model allows dependence in higher order moments since the degree of freedom v depends on the state variable s_t. It also provides a richer characterization of conditional volatility by allowing GARCH parameters to depend on s_t, which could generate asymmetric behavior in volatility.

As pointed out by Hamilton and Susmel (1994), a regime-switching GARCH model is intractable due to the dependence of conditional variance on the entire past history of the data. To avoid such difficulty, we follow Gray (1996) to remove the path dependence nature of the regime-switching GARCH model by averaging over regimes the conditional and unconditional variances at each time point. Thus, the conditional density of the exchange rate change in a regime-switching model is

$$p(Y_t|I_{t-1}) = p(Y_t|s_t = 1, I_{t-1})\mathrm{P}(s_t = 1|I_{t-1}) + \mathrm{P}(Y_t|s_t = 2, I_{t-1})\mathrm{P}(s_t = 2|I_{t-1}),$$

where $\mathrm{P}(s_t = l|I_{t-1})$, the *ex ante* probability that Y_t is generated from regime l, can be obtained using a recursive procedure given in Hamilton (1989) and Gray (1996).

5. Empirical results

5.1. Data and estimation method

We consider two intraday high-frequency exchange rates, Euro/Dollar and Yen/Dollar, from July 1, 2000 to June 30, 2001. Euro and Yen are two of the most important currencies in the world after the U.S. dollar. The launch of the new currency Euro is probably the most important event in the history of international monetary and financial system since the end of the Bretton Woods system in the early 1970s. It has created the world's second-largest single currency area after the United States.[13] In the foreign exchange market, Euro/Dollar will surely be the busiest pair of currencies: It is estimated that 40% of the trading will be between this pair, which is twice as large as the Dollar/DM pair had, and twice as large as the Yen/Dollar pair has. Thus, understanding the evolution of the Euro/Dollar exchange rates will be important to many outstanding issues in international economics and finance. The Japanese economy was in prolonged recession in the last decade, and, as a result, the Yen/Dollar rate might have very different time series properties than that of the Euro/Dollar rate.

The data, obtained from Olsen & Associates, are indicative bid and ask quotes posted by banks. We choose the sample period between July 1, 2000 and June 30, 2001 to wait for the market to stabilize after the introduction of the Euro as a new currency on January 1, 1999 and to avoid the impact of the disaster of September 11, 2001. Similar to Diebold et al. (1999), we sample data over a grid of half-hour intervals, i.e., we obtain quotes nearest to half-hour time stamps. Although currency trading occurs around the clock during weekdays, trading is very thin during weekends. Following Diebold et al. (1999), we

[13]The Euro area comprises 12 countries which account for about 16% of global GDP, and has a total population of 290 million. In comparison, the share of the global output produced in the United States is around 20%, with a slightly smaller population, and the Japanese economy accounts for about 8% of the global GDP, with a total population of about 130 million.

eliminate the observations from Friday 21:30 GMT to Sunday 21:00 GMT. Thus, a trading week is between Sunday 21:30 GMT and Friday 21:00 GMT, and each of the five trading days spans 21:30 GMT on one day to 21:00 GMT the next day, which implies that there are $5 \times 48 = 240$ observations for each full week.

We calculate exchange rate changes in the same way as in Andersen and Bollerslev (1997) and Diebold et al. (1999). We first calculate the average log bid and log ask prices to get a ''log price,'' then calculate changes as the difference between log prices at consecutive time points.[14] The autocorrelations of Y_t^2 and $|Y_t|$ show distinct intraday seasonal patterns (especially those of $|Y_t|$). Diebold et al. (1999) argue that calendar effects in volatility occur because trading is more active at certain times of the day than at others.[15] Following Diebold et al. (1999), we remove the intraday calendar effects in volatility, and the autocorrelations of the deseasoned returns do not exhibit seasonality in both mean and variance.[16] The histograms of the two exchange rate returns after removing intraday seasonality show that both exchange rate returns have fatter tails and a higher peak than the normal distribution. In addition, we find that the return of the Euro/Dollar rate has much higher kurtosis than that of the Yen/Dollar rate.

We choose the first half of the sample, from 07/01/2000 to 12/31/2000 (with a total of 6214 observations) as the estimation sample. We consider two choices of forecast samples. One is from 01/01/2001 to 06/30/2001 (with a total of 6214 observations), and the other is from 01/01/2001 to 03/31/2001 (with a total of 3107 observations). Both choices yield similar conclusions and we only report and discuss results based on the second choice because the simulation study in Section 4 has shown that our tests have better finite sample performances in this case. We first consider in-sample performances of all models; we then evaluate their out-of-sample density forecasts using our portmanteau test. All models are estimated via MLE. The optimization algorithm is the well-known BHHH with STEPBT for step length calculation and is implemented via the constrained optimization code in GAUSS Window Version 3.6. The optimization tolerance level is set such that the gradients of the parameters are less than or equal to 10^{-6}.

5.2. In-sample performances

We now examine the in-sample performances of all models for exchange rate dynamics. We first focus on models with similar structures and then compare the performances of models across different classes. We examine model performances based on estimated parameters and likelihood values, as well as the i.i.d. and uniform properties of the in-sample generalized residuals.[17]

[14] To save space, we omit tables and figures that provide summary information of the two exchange rates. These results are available in Hong et al. (2006).

[15] For instance, trading is much less active during the Japanese lunch hour, and much more active when U.S. markets are open.

[16] Diebold et al. (1999) remove seasonality in volatility as follows. Let $r_t = s_{\tau,i} Z_t$, where Z_t is the unseasoned portion of the process and $s_{\tau,i}$ is the time-of-day dummy at time τ of day i, where i is Monday, Tuesday,..., Friday. To remove volatility calender effects, Diebold et al. (1999) fit $2\log|r_t| = 2\log s_{\tau,i} + 2\log Z_t$, and use the estimated time-of-day dummies, suitably normalized so that $s_{\tau,i}$ summed over the entire sample equals 1, to standardize returns.

[17] To save space, we do not report parameter estimates and likelihood values. These results are available in Hong et al. (2006).

We find that modeling the fat tails of both exchange rates significantly improves the performances of the random walk models: Models with a Student-t innovation have much higher likelihood values than corresponding models with a normal innovation. The estimated degree of freedom v of the Student-t innovation is about 3.3 for Euro and 3.8 for Yen, consistent with the fact that Euro/Dollar has higher kurtosis than Yen/Dollar. There is also a predictable component in the conditional mean of both exchange rates: In all models, the MA(1) coefficients are negative and statistically significant, suggesting that the exchange rates are mean-reverting.

Next, we examine the performances of RiskMetrics, GARCH, and EGARCH models. First, similar to the random walk models, it is also important to model fat tails in these models: RM-T and GARCH/EGARCH-T perform much better than RM-N and GARCH/EGARCH-N, respectively. The estimated degree of freedom of Euro/Dollar is again smaller than that of Yen/Dollar. Second, in addition to fat tails, it is important to model volatility clustering in both exchange rates, and GARCH/EGARCH perform much better than the RiskMetrics models in this respect. Likelihood ratio tests (not reported) easily reject those models that ignore volatility clustering. All GARCH parameter estimates are overwhelmingly significant and the implied GARCH processes are covariance-stationary, i.e., the sum of GARCH parameter estimates is less than one ($\hat{\beta}_1 + \hat{\beta}_2 \approx 0.95$, and 0.91, for Euro/Dollar and Yen/Dollar, respectively, in GARCH-T). Note that the specification of $t(v)$ rather than N(0,1) reduces the persistence in volatility clustering, as can be seen from the sum of $\hat{\beta}_1$ and $\hat{\beta}_2$. EGARCH models perform slightly better than GARCH models, although the asymmetric parameter is not significantly different from zero.[18] Similar to the random walk models, we also find statistically significant and negative components in the conditional mean of the three classes of models.

Now we examine the performances of jump models with and without GARCH, respectively. Consistent with Jorion (1988), we find that Jump-N significantly improves RW-N. We also measure the incremental contribution of jumps beyond Student-t innovation, an issue not considered in Jorion (1988). For models with normal innovation and jumps, replacing normal with Student-t innovation further improves model performances. However, for models with Student-t innovation, introducing jumps provides no further improvement for Euro/Dollar and only marginal improvement for Yen/Dollar. It appears that Student-t innovations have provided an adequate description of the fat tails of both exchange rates that cannot be further improved by including jumps.

Our results indicate that capturing fat tails through Student-t innovation provides the most significant improvement in model performances. GARCH/EGARCH provide better characterization of volatility clustering than the RiskMetrics models. Jumps, although they improve upon models with normal innovation, provide no further contribution beyond Student-t innovation. Our findings so far are consistent with those of Bollerslev (1987), Engle and Bollerslev (1986), Baillie and Bollerslev (1989), and Hsieh (1989), who show that models that best capture daily and weekly exchange rate dynamics are GARCH/EGARCH-T.

Next we study whether we can further improve model performance by allowing dependence in higher order moments through Hansen's (1994) ARCD model and a

[18]This result may not be surprising because unlike stock returns, currency returns do not exhibit a pronounced "leverage" effect.

regime-switching model with regime dependent GARCH process and Student-t innovation. For the more complicated models, we use GARCH/EGARCH-T as benchmarks.

We consider ARCD models in which either the degree of freedom v_t or the conditional skewness λ_t follows an autoregressive process. We do not find significant serial dependence in v_t, so we only consider ARCD models with either unconditional skewness (denoted ARCD-USkew) or autoregressive conditional skewness (denoted ARCH-Skew). The ARCD model with time-varying conditional skewness improves the performances of GARCH/EGARCH-T for both exchange rates. In contrast, the ARCD models with unconditional skewness, while outperforming GARCH-N, behave almost exactly the same as GARCH-T. The additional flexibility provided by the unconditional skewness parameter is almost nonexistent. It seems that there is serial dependence in higher order moments, especially conditional skewness, but the unconditional skewness is close to zero.

Introducing regime switching to GARCH-T further improves model performance. This suggests that regime switching is another important feature of exchange rates besides fat tails and volatility clustering. The improvement, however, does not seem to come from modeling higher order moments, because the degrees of freedom in the two regimes turn out to be not significantly different from each other. Instead, the improvement seems to come from better modeling of the asymmetric behavior of conditional variance through regime-dependent GARCH effect: The second regime has much higher volatility and less persistent dependence in conditional volatility, while the first regime has lower volatility and more persistent dependence in variance. The estimates of the transition probability matrix suggest that the first regime is slightly more persistent than the second one.[19] There is also evidence of a predictable component in the conditional mean.

The above results are consistent with the diagnostic analysis based on the i.i.d. and uniform properties of model generalized residuals. Figs. 1 and 2 contain kernel estimators of the marginal densities of the generalized residuals of all models for Euro/Dollar and Yen/Dollar, respectively. Table 2 reports the in-sample separate inference statistics $M(m, l)$ for Euro/Dollar and Yen/Dollar. It is clear from Figs. 1 and 2 that the generalized residuals of models with Student-t innovations are much closer to $U[0, 1]$ than those of models with normal innovations. Jumps also help capture the heavy tails of both exchange rates. It is interesting that for both currencies, the most sophisticated time series models can capture the marginal densities of the exchange rate changes pretty well. It is also clear from Table 2 that most models with volatility clustering can capture the dependence in conditional variance and kurtosis of both exchange rates reasonably well. The MA(1)/AR(1) component in most models substantially reduces the $M(1, 1)$ statistics for both exchange rates, suggesting that modeling dependence in conditional mean is important for in-sample performances.

To sum up, our in-sample analysis reveals some interesting stylized facts for the two intraday exchange rates:

- Consistent with existing studies, it is extremely important to model the fat tails of both exchange rates and it seems that Student-t distribution does a reasonably good job that cannot be further improved by jumps.

[19] Estimates not reported here show that the transition matrix of the Markov state variable does not significantly depend on the level of previous exchange rate changes.

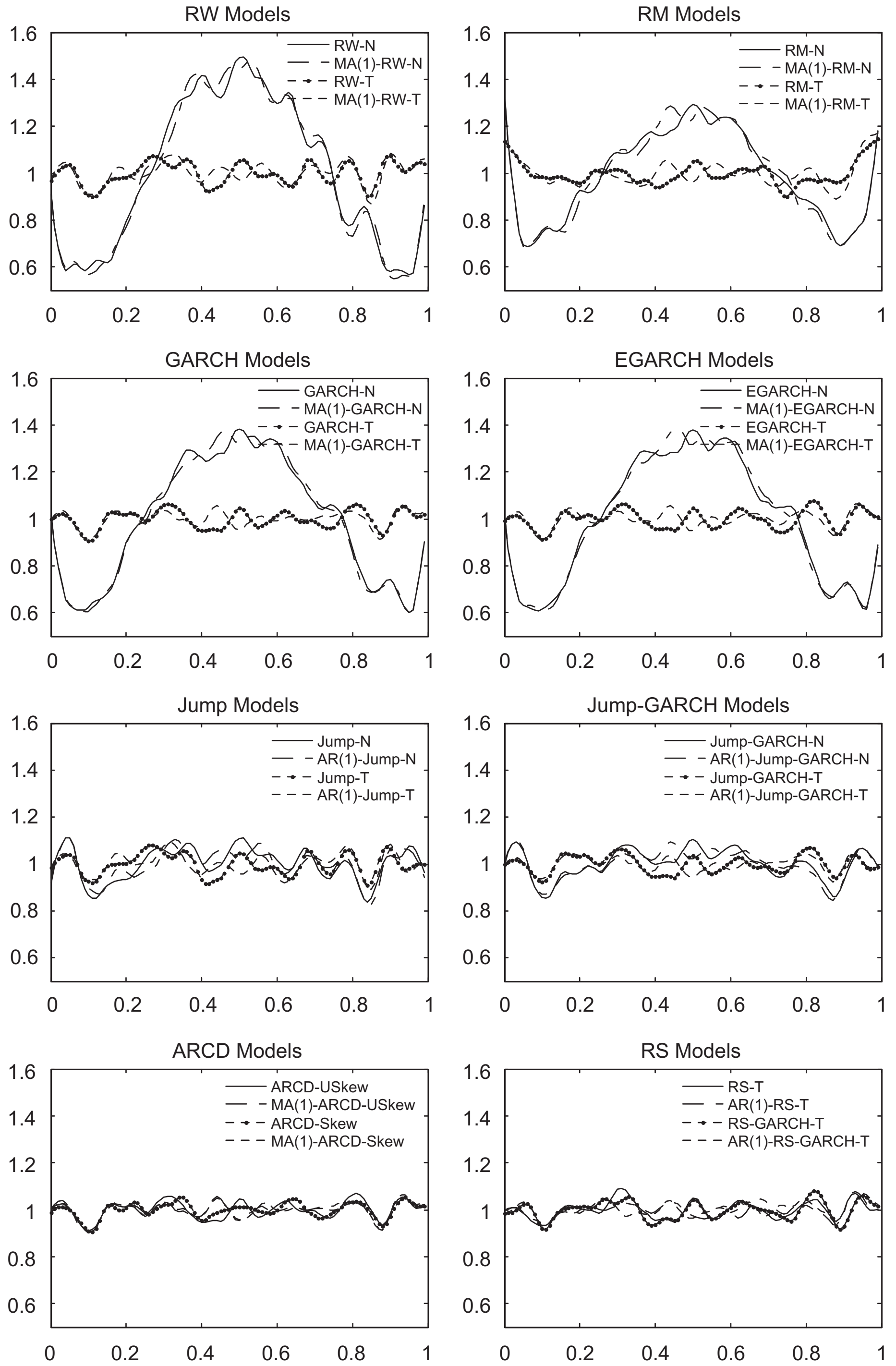

Fig. 1. Nonparametric marginal densities of the in-sample generalized residuals for Euro/Dollar.

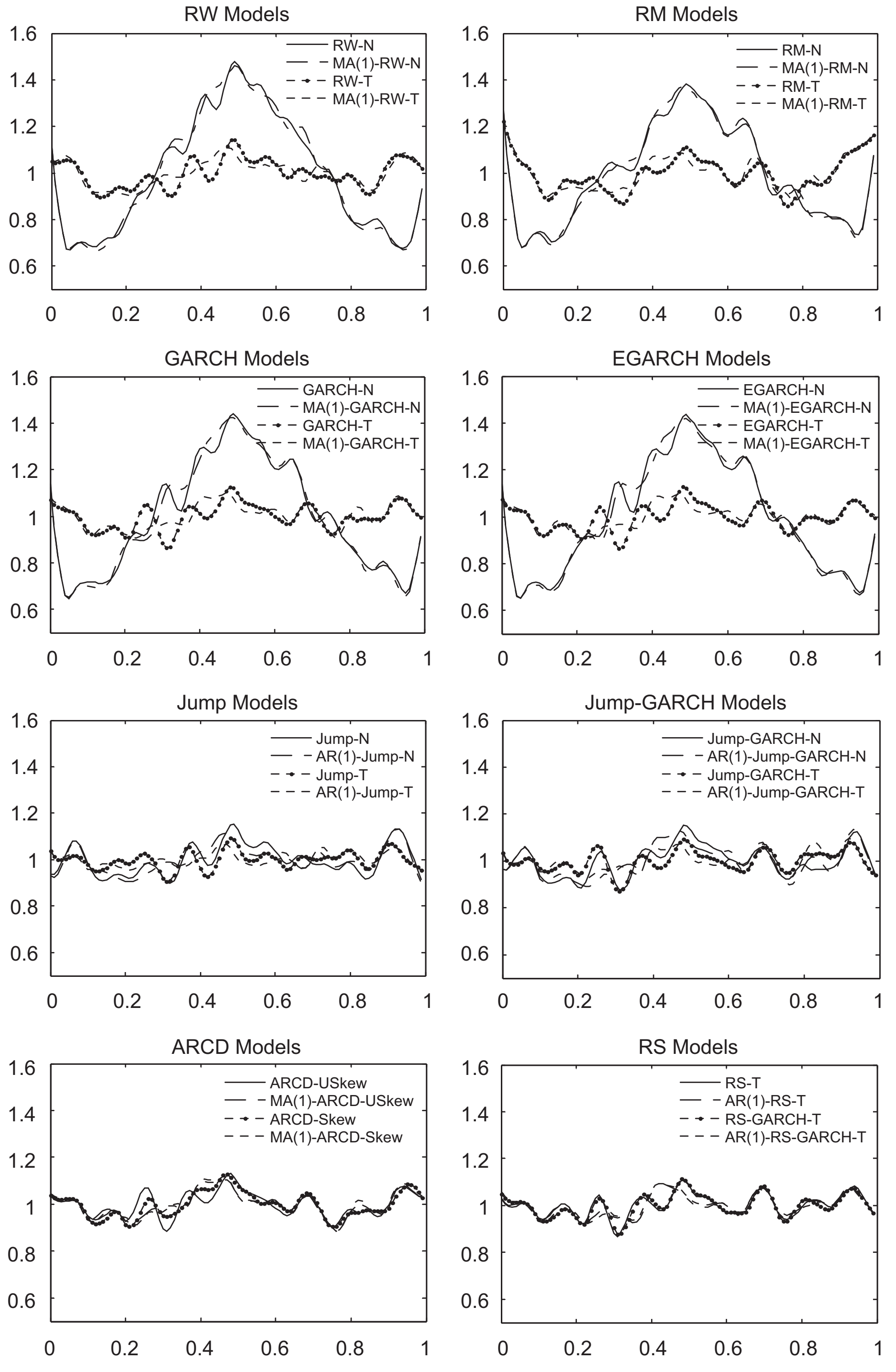

Fig. 2. Nonparametric marginal densities of the in-sample generalized residuals for Yen/Dollar.

Table 2
In-sample separate inference statistics

Model	M(1,1)		M(1,2)		M(2,1)		M(2,2)		M(3,3)		M(4,4)	
	Euro	Yen	Euro	Yen	Euro	Yen	Euro	Yen	Euro	Yen	Euro	Yen
RW-N	11.87	14.47	1.09	0.61	−0.72	−0.61	71.67	51.76	1.54	4.71	75.35	44.67
MA(1)-RW-N	−0.31	2.52	1.32	0.99	−0.60	−0.65	69.76	49.09	1.99	0.64	74.89	42.99
RW-T	14.43	15.81	0.23	−0.54	−1.11	−0.65	60.53	48.42	4.94	8.00	74.35	51.87
MA(1)-RW-T	0.06	2.76	0.60	−0.09	−1.01	−0.59	57.69	44.83	2.62	3.69	71.77	48.77
RM-N	12.41	12.40	0.53	0.57	−0.22	0.09	18.42	23.73	0.95	2.07	12.84	18.91
MA(1)-RM-N	0.29	2.69	0.64	0.97	−0.24	−0.12	17.50	22.73	0.87	0.41	13.34	18.22
RM-T	15.05	13.99	0.34	−0.30	−0.53	−0.11	31.95	22.35	3.92	4.44	31.92	21.44
MA(1)-RM-T	0.13	2.79	0.61	0.26	−0.54	−0.35	29.15	21.26	2.58	4.82	31.05	21.20
GARCH-N	10.13	13.39	0.67	0.53	−0.46	−0.10	0.57	8.87	−0.34	3.12	0.21	4.23
MA(1)-GARCH-N	−0.44	2.30	0.79	0.85	−0.35	−0.23	0.59	7.76	1.41	−0.05	0.27	3.69
GARCH-T	12.96	14.79	0.17	−0.59	−0.82	−0.25	0.58	3.62	2.23	5.74	−0.43	1.71
MA(1)-GARCH-T	−0.09	2.42	0.40	−0.18	−0.76	−0.31	0.43	2.76	2.57	3.00	−0.21	1.41
EGARCH-N	9.72	13.37	0.65	0.27	−0.41	−1.05	2.54	3.04	−0.25	2.71	0.95	1.58
MA(1)-EGARCH-N	−0.40	2.54	0.71	0.55	−0.40	−1.10	2.56	2.14	2.65	−0.22	0.97	1.11
EGARCH-T	12.59	14.72	0.15	−0.66	−0.79	−0.90	1.06	1.45	2.12	5.35	0.12	0.81
MA(1)-EGARCH-T	−0.07	2.47	0.32	−0.28	−0.75	−0.94	0.94	0.54	2.58	3.19	0.37	0.30
Jump-N	14.28	15.83	0.38	−0.58	−1.06	−0.61	60.45	48.31	4.71	8.09	73.96	52.04
AR(1)-Jump-N	0.44	3.35	0.62	−0.14	−0.98	−0.54	57.85	44.84	1.87	3.65	71.64	48.76
Jump-T	14.40	15.82	0.43	−0.50	−1.03	−0.76	60.49	48.30	4.94	8.05	74.36	51.75
AR(1)-Jump-T	0.48	3.27	0.64	−0.13	−0.96	−0.65	58.23	45.07	2.00	3.46	72.34	48.70
Jump-GARCH-N	12.58	14.64	0.34	−0.69	−0.77	−0.11	0.33	2.45	1.87	5.54	−0.27	1.60
AR(1)-Jump-ARCH-N	0.16	2.77	0.42	−0.26	−0.73	−0.16	0.20	1.70	1.50	2.53	−0.11	1.29
Jump-GARCH-T	12.94	14.79	0.27	−0.64	−0.79	−0.32	0.49	2.73	2.23	5.70	−0.48	1.33
AR(1)-Jump-ARCH-T	0.32	2.87	0.40	−0.26	−0.74	−0.29	0.29	1.81	2.01	2.70	−0.34	0.87
ARCD-USkew	12.96	14.80	0.17	−0.60	−0.82	−0.29	0.58	3.47	2.24	5.70	−0.44	1.63
MA(1)-ARCD-USkew	−0.09	2.48	0.39	−0.18	−0.77	−0.36	0.43	2.61	2.55	3.07	−0.22	1.31
ARCD-Skew	5.17	6.52	−0.48	0.03	−0.81	−0.07	0.28	2.45	2.19	5.88	−0.29	1.64
MA(1)-ARCD-Skew	−0.02	2.48	−0.33	0.29	−0.79	−0.15	0.73	2.96	−0.72	0.27	0.04	1.64
RS-T	13.26	14.89	0.19	−0.53	−0.90	−0.38	1.92	2.13	2.78	5.48	1.72	1.66
AR(1)-RS-T	0.48	2.92	0.38	−0.29	−1.00	−0.21	0.31	0.28	0.25	1.88	0.05	0.26
RS-GARCH-T	13.63	14.30	0.23	−0.67	−0.66	−0.35	−0.14	0.03	3.03	4.64	−0.99	0.02
AR(1)-RS-GARCH-T	0.22	2.81	0.48	−0.43	−0.70	−0.18	−0.23	0.02	1.33	1.87	−0.88	−0.12

This table reports the in-sample separate inference statistics $\mathrm{M}(m,l)$ for all models. The statistic $\mathrm{M}(m,l)$ can be used to test whether the cross-correlation between the mth and lth moments of $\{Z_t\}$ is significantly different from zero. The choice of $(m,l)=(1,1)$, (2,2), (3,3), (4,4) is very sensitive to autocorrelations in mean, variance, skewness, and kurtosis of $\{Y_t\}$, respectively. The in-sample data are from July 1, 2000 to December 31, 2000, with a total of 6214 observations. We only show results for lag truncation order $p=20$; the results for $p=10$ and 30 are similar.

- In addition to fat tails, it is also important to model volatility clustering and its asymmetric features. The conditional variance exhibits an asymmetric behavior which is better captured by a regime-switching GARCH model than GARCH/EGARCH and RiskMetrics models.
- Modeling conditional mean and serial dependence in conditional skewness further improves model performances. However, no significant dependence in conditional kurtosis is found.

5.3. Out-of-sample density forecasting performances

Although the more complicated models with conditional mean, conditional heteroskedasticity, fat tails, and regime switching have better in-sample fits, previous studies have shown that they generally underperform the simple random walk model in forecasting the conditional mean of exchange rate changes. In this section, we apply the tests developed in Section 2 to examine whether the features found to be important for in-sample fits remain important for out-of-sample density forecasts, and in particular, whether the random walk model still dominates all other models in density forecasts.

For each model, we first calculate the generalized residuals using the forecast sample based on parameter estimates obtained from the estimation sample. Then we calculate the out-of-sample evaluation statistics $\hat{W}(p)$ ($p = 5, 10$, and 20) and $\hat{Q}(j)$ ($j = 1, 10$, and 20), which are reported in Table 3. The simulated critical values with $n = 3107$ for $\hat{Q}(j)$ are much closer to the asymptotic ones than those of $\hat{W}(p)$. To demonstrate the robustness of our results, we report both $\hat{Q}(j)$ and $\hat{W}(p)$ statistics, but focus our discussions on $\hat{W}(p)$ because both tests yield very similar conclusions.

For Euro/Dollar, modeling fat tails via Student-t innovations is important for both in-sample fits and out-of-sample density forecasts. For example, $\hat{W}(5)$ declines from 53 for RW-N to about 23 for RW-T, and $\hat{W}(5)$ declines from about 50 (40) for GARCH/EGARCH-N (RM-N) to about 5 (1.5) for GARCH/EGARCH-T (RM-T). However, for Yen/Dollar, except for the RiskMetrics models, other models (Random walk, GARCH/EGARCH) with Student-t innovations generally underperform corresponding models with a normal innovation in density forecasts.

In addition to fat tails, modeling volatility clustering provides further improvement for both in-sample and out-of-sample performances. GARCH/EGARCH and RiskMetrics models with a Student-t innovation have much better density forecasts than the random walk model with a Student-t innovation. Interestingly, for both exchange rates, RM-T significantly outperforms GARCH/EGARCH-T in density forecasts, suggesting that the simple exponential smoothing method of RiskMetrics captures volatility clustering better than GARCH/EGARCH for the out-of-sample density forecast purpose.[20]

For Euro/Dollar, similar to in-sample findings, jumps improve the forecasting performances of models with a normal innovation: Jumps reduce $\hat{W}(5)$ from 53 (48) for RW-N (GARCH-N) to single digits. However, the heavy tails of the Euro/Dollar exchange rate have already been well captured by Student-t innovation, and jumps provide no significant improvements in density forecasts. For Yen/Dollar, jumps actually worsen out-of-sample forecasting performances even for RW-N and GARCH-N models. The in-sample evidence shows that jump models have a significant predictable component in mean. However, including an MA or AR component in jump models adversely affects the out-of-sample forecasting performance.

Our analysis so far shows that for both exchange rates, RM-T has the best density forecasts. Next we consider whether we can further improve density forecasts by modeling dependence in higher order moments in the form of ARCD and regime-switching models. Although the ARCD models with time varying conditional skewness have slightly better in-sample fits than GARCH-T, they perform slightly worse in out-of-sample density

[20] However, we need to point out that this result is not very robust for Euro/Dollar. When $R/n = 1$, GARCH/EGARCH-T outperforms RM-T. This is the only material difference between the results for $R/n = 1$ and 2.

Table 3
Density forecast evaluation statistics $W(p)$ and $Q(j)$

Model	$W(5)$		$W(10)$		$W(20)$		$Q(1)$		$Q(10)$		$Q(20)$	
	Euro	Yen	Euro	Yen	Euro	Yen	Euro	Yen	Euro	Yen	Euro	Yen
RW-N	53.1	65.57	72.44	91.57	100.8	126.00	25.27	30.82	21.46	27.02	20.93	27.19
MA(1)-RW-N	55.02	68.30	75.97	96.14	105.6	131.70	25.41	31.73	23.97	28.44	22.19	28.61
RW-T	23.3	193.20	31.49	270.10	42.37	372.80	11.11	88.24	9.857	81.63	7.896	78.09
MA(1)-RW-T	27.96	203.10	38.99	283.90	52.34	389.90	14.41	93.79	13.31	86.05	11.42	79.72
RM-N	39.62	57.04	54.5	79.26	76.29	108.90	17.79	27.33	16.43	24.11	17.78	22.35
MA(1)-RM-N	42.17	59.81	58.18	82.97	81.06	114.00	18.69	28.83	17.4	24.56	17.77	22.97
RM-T	1.568	4.90	1.911	6.78	2.481	8.77	0.9958	3.206	−0.2157	1.581	−0.5106	1.827
MA(1)-RM-T	4.226	5.55	6.128	8.17	8.201	10.03	3.287	4.559	1.022	1.682	1.223	1.602
GARCH-N	48.68	23.76	68.18	33.69	97.21	47.23	21.38	10.82	21.18	9.752	21.57	10.58
MA(1)-GARCH-N	51.11	24.16	71.74	35.01	100.4	49.60	22.14	11.1	22.46	10.11	21.37	10.58
GARCH-T	7.454	45.15	10.13	64.85	14.4	92.14	2.212	20.11	2.617	19.83	2.666	19.87
MA(1)-GARCH-T	11.69	48.44	16.34	69.89	21.75	98.72	5.23	22.52	4.998	21.76	4.384	20.6
EGARCH-N	50.23	22.71	70.36	32.44	100.7	46.58	21.87	10.19	21.82	9.039	22.24	10.75
MA(1)-EGARCH-N	52.08	23.07	73.29	33.48	103.8	48.21	22.61	10.23	22.67	9.399	22.18	10.81
EGARCH-T	5.154	51.04	6.934	73.34	9.83	104.30	1.083	22.98	1.359	22.2	1.127	22.88
MA(1)-EGARCH-T	8.56	54.39	12.03	78.05	15.92	110.60	3.599	25.64	3.466	23.92	2.793	23.59
Jump-N	23.99	183.40	32.5	256.40	43.74	353.90	11.26	83.95	10.34	77.53	8.311	74.11
AR(1)-Jump-N	28.09	187.90	39.28	262.40	52.83	360.30	14.25	86.47	13.57	79.75	11.6	73.66
Jump-T	23.21	177.20	31.4	247.70	42.22	341.70	11.08	81.18	9.86	74.73	7.848	71.81
AR(1)-Jump-T	26.69	186.30	37.39	260.20	50.31	357.20	13.56	85.64	12.74	78.89	10.99	73.42
Jump-GARCH-N	8.575	57.18	11.39	82.33	16.24	117.30	2.629	25.39	3.111	25.4	3.233	25.38
AR(1)-Jump-ARCH-N	10.81	61.16	15.27	88.37	20.8	125.30	3.974	27.76	4.767	27.74	4.417	26.53
Jump-GARCH-T	7.238	47.27	9.851	67.85	14.01	96.43	2.125	21.13	2.593	20.7	2.523	20.85
AR(1)-Jump-ARCH-T	11.12	50.33	15.59	72.59	20.84	102.60	4.743	22.95	4.74	22.46	4.119	21.64
ARCD-USkew	7.598	46.10	10.35	66.22	14.69	94.25	2.275	20.7	2.671	20.28	2.715	20.38
MA(1)-ARCD-USkew	11.8	49.51	16.51	71.46	21.96	100.90	5.282	23.15	5.066	22.22	4.432	21.13
ARCD-Skew	8.357	46.20	11.13	66.76	14.94	94.76	2.796	20.81	2.756	19.99	2.763	19.45
MA(1)-ARCD-Skew	11.72	47.47	16.01	68.52	21.61	96.77	5.359	21.77	4.583	20.99	4.404	19.84
RS-T	4.254	72.18	6.126	102.60	8.457	143.80	1.179	32.93	1.691	30.9	0.4301	31.42
AR(1)-RS-T	8.355	78.30	11.91	112.20	15.95	157.20	3.214	35.63	3.864	34.61	2.189	33.79
RS-GARCH-T	−0.4855	61.35	−0.971	87.87	−1.146	123.80	−0.7883	27.12	−0.1968	26.9	−0.9067	27.5
AR(1)-RS-GARCH-T	0.6355	63.10	0.727	90.83	0.6106	128.40	−0.2752	28.11	0.4112	28.02	−0.4265	27.88

This table reports the evaluation statistics $W(p)$ for the out-of-sample density forecasting performance of the models estimated in Table 3 using the estimation sample (from July 1, 2000 to December 31, 2001, with a total of 6214 observations). The probability integral transforms are obtained using the forecast sample (from January 1, 2001 to March 31, 2001, with a total of 3107 observations). The finite sample critical values of $W(p)$ are obtained via simulation. The finite sample critical values at the 5% level for $W(p)$ for $R = 6214$ and $n = 3107$ are 3.460, 4.674, and 6.414 for $p = 5$, 10, and 20, respectively. The finite sample critical values at the 5% level for $Q(j)$ for $R = 6214$ and $n = 3107$ are 2.160, 2.099, and 2.123 for $j = 1$, 10, and 20, respectively.

forecasts. Therefore, modeling dependence in conditional skewness does not necessarily improve out-of-sample forecasting performances. On the other hand, regime-switching models improve upon GARCH/EGARCH-T for both in-sample and out-of-sample performances. The RS-GARCH-T models with and without drift have the best out-of-sample performances with $\hat{W}(5)$ equal to −0.49 and 0.65, respectively, which are not significant at conventional levels. This suggests that it is important to capture asymmetric conditional volatility for out-of-sample density forecasts. However, for Yen/Dollar, regime switching does not improve forecasts when volatility clustering has already been captured by GARCH or RiskMetrics. Interestingly, it seems that we have found a couple of models that adequately capture the conditional density of Euro/Dollar. However, we have not been able to identify such a model for Yen/Dollar.

Diagnostic analysis based on out-of-sample generalized residuals reveals interesting sources of model misspecifications. While the in-sample generalized residuals of most models with Student-t innovations are close to U[0, 1] for both exchange rates, the out-of-sample residuals are much more nonuniform for most models, and are especially so for Yen/Dollar. As shown in Fig. 3, for Euro/Dollar, most models with Student-t innovations have high peaks at both ends of the distribution. The regime-switching models best capture the uniform property, although their out-of-sample residuals are still not as uniform as the in-sample ones. As shown in Fig. 4, almost all models have pronounced peaks at both ends of the distribution, suggesting that all models cannot capture the extreme movements in the Yen/Dollar exchange rates in the forecast sample.

The out-of-sample separate inference statistics $M(m, l)$ in Table 4 show that models with GARCH or regime switching can capture the dependence in the conditional variance and kurtosis of the generalized residuals pretty well. In contrast, the RiskMetrics models do not perform nearly as well in this regard. Most models with an MA(1) or AR(1) term tend to have higher $M(1, 1)$ statistics. This suggests that while modeling the conditional mean through an MA(1) or AR(1) component is important for in-sample fits, it has adverse effect on out-of-sample density forecasts for both exchange rates. This is consistent with the existing results on mean forecasts.

Our analysis shows that some nonlinear time series models have both good in-sample and out-of-sample performances and they outperform the simple random walk model in density forecasts. In particular, we obtain the following findings:

- Modeling conditional mean and dependence in higher order moments such as conditional skewness, while important for in-sample performances, does not improve density forecasts for both exchange rates.
- For the Euro/Dollar rate, modeling the heavy tails through a Student-t innovation and the asymmetric time-varying conditional volatility through a regime-switching GARCH model improve both in-sample and out-of-sample performances. As a result, a regime-switching model with a zero conditional mean, a regime-dependent GARCH(1,1) volatility, and a Student-t innovation has the best density forecasts.
- For the Yen/Dollar rate, it is also important to model heavy tails and volatility clustering for out-of-sample performances. The best density forecasting model is a RiskMetrics model with a Student-t innovation.

6. Conclusion

It is notoriously difficult to forecast the conditional mean of future changes of exchange rates. Numerous studies have shown that the simple random walk model outperforms most structural and time series models in this regard. In this paper we have asked whether some time series model(s) can outperform the random walk model in forecasting the probability density of exchange rates. The importance of density forecasts can never be over emphasized, because in many important economic and financial applications we usually need to know the entire probability density of exchange rates. Our paper contributes to the literature by (i) developing a nonparametric portmanteau test for out-of-sample density forecast evaluation; and (ii) providing probably the first comprehensive empirical study of

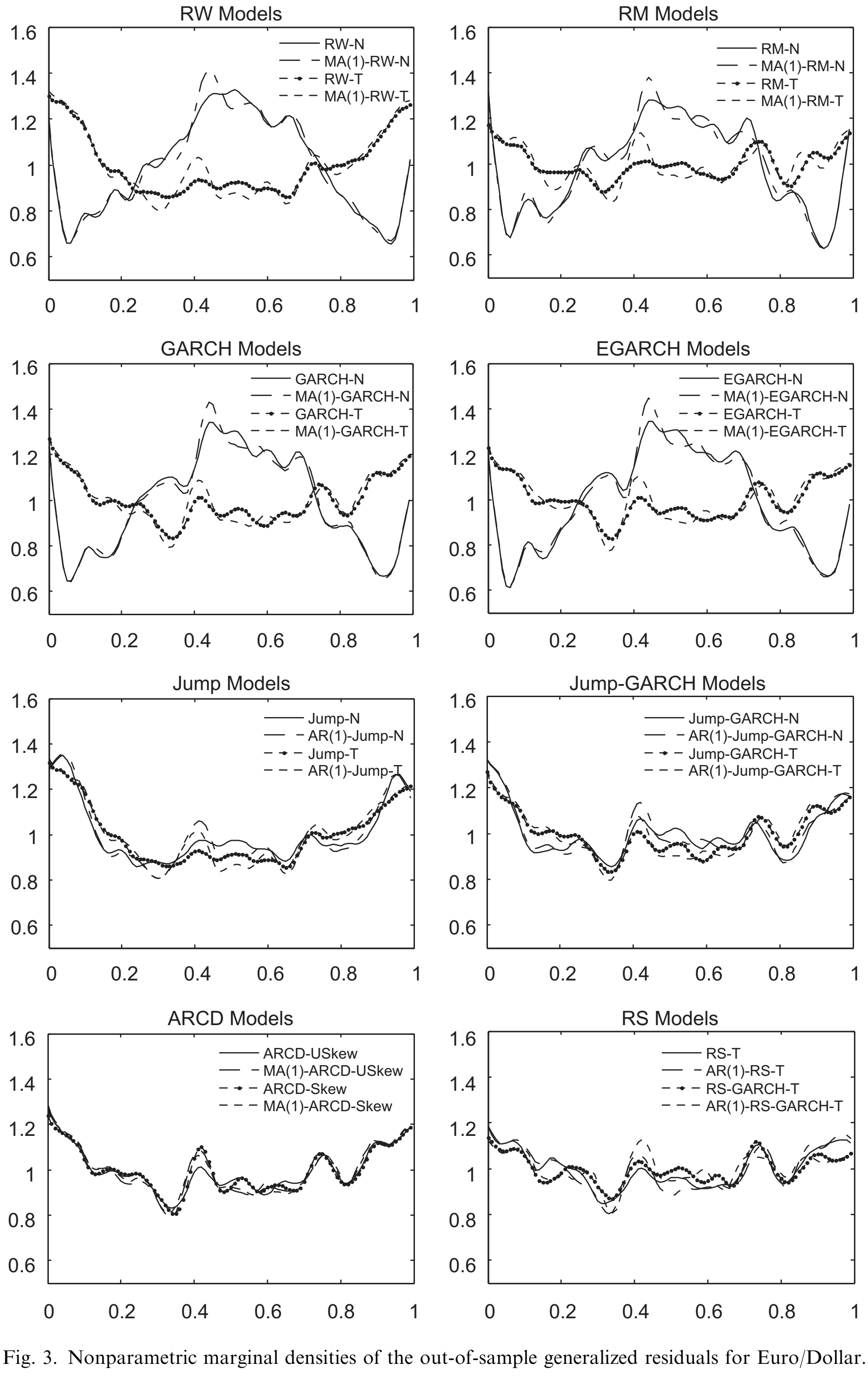

Fig. 3. Nonparametric marginal densities of the out-of-sample generalized residuals for Euro/Dollar.

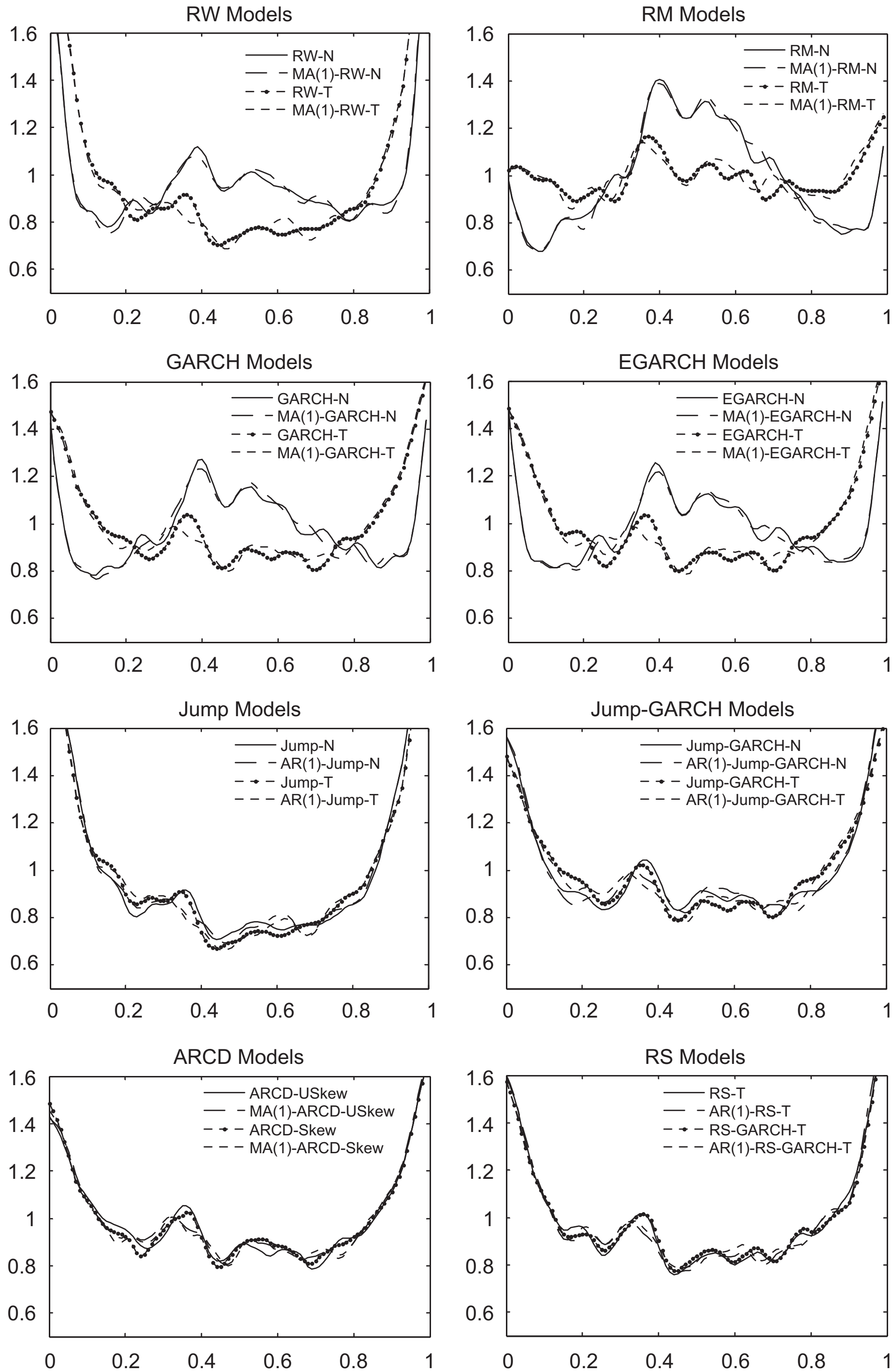

Fig. 4. Nonparametric marginal densities of the out-of-sample generalized residuals for Yen/Dollar.

Table 4
Out-of-sample separate inference statistics

Model	M(1,1)		M(1,2)		M(2,1)		M(2,2)		M(3,3)		M(4,4)	
	Euro	Yen	Euro	Yen	Euro	Yen	Euro	Yen	Euro	Yen	Euro	Yen
RW-N	1.22	2.37	−0.82	−0.52	−1.06	−0.08	19.80	20.86	0.97	1.33	16.82	19.37
MA(1)-RW-N	1.25	0.74	−0.69	−0.46	−1.15	−0.14	19.62	21.45	2.18	1.53	16.92	20.06
RW-T	1.26	2.56	−0.41	−0.71	−1.06	−0.15	18.38	20.47	1.35	1.73	18.95	20.62
MA(1)-RW-T	3.09	2.67	−0.19	−0.68	−1.26	−0.23	18.53	21.03	4.54	4.04	19.23	22.15
RM-N	0.06	2.26	−0.59	−0.85	−0.68	0.09	4.87	8.16	0.08	−0.11	3.84	3.62
MA(1)-RM-N	2.04	0.49	−0.42	−0.70	−0.89	−0.03	5.41	8.19	1.77	0.59	4.18	3.87
RM-T	0.53	2.54	−0.69	−0.93	−0.87	−0.07	7.66	9.39	0.37	0.68	6.68	5.61
MA(1)-RM-T	3.98	2.62	−0.43	−0.71	−1.10	−0.13	8.45	9.79	4.59	3.74	7.63	6.66
GARCH-N	0.21	2.25	−0.90	−1.18	−0.59	−0.01	1.52	1.72	0.08	0.82	2.58	−0.09
MA(1)-GARCH-N	1.51	0.83	−0.78	−1.15	−0.78	−0.09	1.38	1.88	1.55	1.85	2.42	0.23
GARCH-T	0.43	2.38	−0.56	−1.23	−0.73	−0.22	0.19	−0.17	0.42	1.38	1.06	−0.76
MA(1)-GARCH-T	3.20	2.95	−0.29	−1.26	−1.02	−0.20	0.35	0.03	3.77	4.81	0.91	−0.45
EGARCH-N	0.12	2.14	−0.87	−1.12	−0.68	−0.35	2.60	−0.29	−0.08	0.89	3.15	−1.02
MA(1)-EGARCH-N	2.14	0.71	−0.73	−1.12	−0.93	−0.36	2.39	−0.28	1.91	1.82	2.95	−0.82
EGARCH-T	0.35	2.33	−0.50	−1.11	−0.76	−0.43	0.73	−0.40	0.27	1.31	1.76	−0.78
MA(1)-EGARCH-T	2.98	3.01	−0.25	−1.17	−1.07	−0.42	0.71	−0.48	3.56	4.87	1.62	−0.63
Jump-N	1.28	2.55	−0.43	−0.74	−1.06	−0.16	18.25	20.64	1.37	1.69	18.71	20.99
AR(1)-Jump-N	2.46	2.64	−0.24	−0.72	−1.22	−0.23	18.04	21.29	3.73	4.24	18.74	22.62
Jump-T	1.26	2.53	−0.43	−0.72	−1.08	−0.19	18.50	20.67	1.34	1.70	19.11	20.91
AR(1)-Jump-T	2.49	2.48	−0.24	−0.69	−1.24	−0.23	18.35	21.21	3.81	4.05	19.15	22.42
Jump-GARCH-N	0.44	2.37	−0.58	−1.19	−0.73	−0.15	0.09	−0.33	0.49	1.40	0.87	−0.88
AR(1)-Jump-ARCH-N	2.17	2.67	−0.37	−1.23	−0.95	−0.18	0.11	−0.26	2.75	4.37	0.70	−0.76
Jump-GARCH-T	0.42	2.37	−0.56	−1.22	−0.72	−0.26	0.17	−0.32	0.39	1.36	1.06	−0.75
AR(1)-Jump-ARCH-T	2.76	2.77	−0.33	−1.26	−0.98	−0.22	0.25	−0.24	3.33	4.62	0.90	−0.66
ARCD-USkew	0.43	2.40	−0.57	−1.24	−0.72	−0.24	0.19	−0.22	0.42	1.37	1.06	−0.75
MA(1)-ARCD-USkew	3.20	2.97	−0.30	−1.27	−1.02	−0.23	0.35	−0.05	3.78	4.87	0.91	−0.45
ARCD-Skew	−0.48	0.61	−0.47	−1.28	−0.73	−0.13	0.33	−0.29	0.39	1.72	1.58	−0.52
MA(1)-ARCD-Skew	2.29	2.75	−0.27	−1.16	−0.91	−0.13	0.55	−0.03	1.23	1.81	1.16	−0.66
RS-T	0.44	2.53	−0.46	−0.89	−0.92	−0.32	0.47	1.80	0.27	1.48	1.16	2.30
AR(1)-RS-T	1.69	1.59	−0.17	−0.93	−1.02	−0.26	−0.06	0.97	1.42	2.39	0.89	1.33
RS-GARCH-T	0.48	2.51	−0.64	−0.94	−0.99	−0.23	−0.23	0.22	0.15	1.69	0.17	−0.37
AR(1)-RS-GARCH-T	2.24	1.73	−0.43	−1.04	−1.19	−0.22	−0.32	−0.32	2.37	2.54	−0.04	−0.92

This table reports the in-sample separate inference statistics $M(m,l)$ for all models. The statistic $M(m,l)$ can be used to test whether the cross-correlation between the mth and lth moments of $\{Z_t\}$ is significantly different from zero. The choice of $(m,l) = (1,1)$, (2,2), (3,3), (4,4) is very sensitive to autocorrelations in mean, variance, skewness, and kurtosis of $\{Y_t\}$, respectively. The out-of-sample data are from January 1, 2001 to March 31, 2001, with a total of 3107 observations. We only show results for lag truncation order $p = 20$; the results for $p = 10$ and 30 are similar.

the density forecasting performances of a wide variety of time series models for two major exchange rates.

Our empirical analysis of high-frequency intraday Euro/Dollar and Yen/Dollar exchange rates shows that some nonlinear time series models do provide better density forecasts than the simple random walk model. For the Euro/Dollar rate, a regime-switching model with a zero conditional mean, a regime-dependent GARCH, and a Student-t innovation provides the best density forecasts for the Euro/Dollar rate. For the Yen/Dollar rate, while it is also important to model heavy tails and volatility clustering, the best density forecasting model is a RiskMetrics model with a Student-t innovation. Our results strongly suggest that the sophisticated nonlinear time series models that have been developed in the literature are useful for out-of-sample applications involving the entire density.

An interesting future study is to compare the relative exchange rate forecast ability between the best density forecast models documented in this paper and an important class of dynamic Bayesian models along the lines of Zellner et al. (1991). Dynamic Bayesian models provide density forecasts which can adapt to the structural changes and regime-shifts via time-varying parameters and which naturally take into account parameter uncertainty. They have been proven successful in global macroeconomic forecasting (Zellner et al., 1991) and exchange rate forecasting (Putnam and Quintana, 1994; Quintana and Putnam, 1996). Moreover, our out-of-sample density forecast evaluation procedure is a statistical criterion. It also will be interesting to see whether best density forecast models selected by this statistical criterion have best economic performance (e.g., in terms of risk-return criteria) as well. These open issues are left for future research.

Acknowledgments

We would like to thank Arnold Zellner (Coeditor), the associate editor, two anonymous referees, Yangru Wu, and participants of the FMA 2006 Meeting for their insightful comments, suggestions, and references. Hong thanks the support from the Cheung Kong Scholarship by the Chinese Ministry of Education and Xiamen University, and the National Science Foundation of U.S.A. via Grant SES–0111769. Zhao thanks the Whitcomb Center for Research in Financial Services for research support. We are responsible for any remaining errors.

Appendix A. Mathematical appendix

Proof of Theorem 1. Throughout, we put $w \equiv (z_1, z_2) \in \mathbb{I}^2$, where $\mathbb{I} \equiv [0, 1]$. Let $\tilde{g}_j(w)$ be defined in the same way as $\hat{g}_j(w)$ in (2.2) but with $\{Z_\tau\}$ replacing $\{\hat{Z}_\tau\}$, and let $\tilde{Q}(j)$ be defined in the same way as $\hat{Q}(j)$ in (2.4) with $\tilde{g}_j(z)$ replacing $\hat{g}_j(z)$. We shall prove the following theorems.

Theorem A.1. $\hat{Q}(j) - \tilde{Q}(j) \to^p 0$.

Theorem A.2. $\tilde{Q}(j) \to^d \mathrm{N}(0, 1)$.

Proof of Theorem A.1. Put $\hat{M}(j) \equiv \int_{\mathbb{I}^2} [\hat{g}_j(w) - 1]^2 \, \mathrm{d}w$, and let $\tilde{M}(j)$ be defined as $\hat{M}(j)$ with $\{\tilde{g}_j(w)\}$ replacing $\{\hat{g}_j(w)\}$. We write

$$\hat{M}(j) - \tilde{M}(j) = \int_{\mathbb{I}^2} [\hat{g}_j(w) - \tilde{g}_j(w)]^2 \, \mathrm{d}w + 2 \int_{\mathbb{I}^2} [\tilde{g}(w) - 1][\hat{g}(w) - \tilde{g}(w)] \, \mathrm{d}w \equiv \hat{\Delta}_1(j) + 2\hat{\Delta}_2(j). \tag{A.1}$$

We shall show Proposition A.1 and A.2 below. Throughout, put $n_j \equiv n - j = T - R - j$.

Proposition A.1. $n_j h \hat{\Delta}_1(j) \to^p 0$.

Proposition A.2. $n_j h \hat{\Delta}_2(j) \to^p 0$.

To show these propositions, we first state a lemma from Hong and Li (2005, Lemma A.1).

Lemma A.1. *Let* $K_h(\cdot,\cdot)$ *be defined in* (2.3). *Then for* $m = 0, 1, 2$ *and* $\lambda \geqslant 1$, $\int_0^1 |\frac{\partial^m}{\partial^m z_2} K_h(z_1, z_2)|^\lambda \, \mathrm{d}z_1 \leqslant Ch^{1-\lambda(m+1)}$ *for all* $z_2 \in [0,1]$ *and* $\int_0^1 |\frac{\partial^m}{\partial^m z_2} K_h(z_1, z_2)|^\lambda \, \mathrm{d}z_2 \leqslant Ch^{1-\lambda(m+1)}$ *for all* $z_1 \in [0,1]$.

Proof of Proposition A.1. Put $\kappa_h(w, w') \equiv K_h(z_1, z_1')K_h(z_2, z_2') - 1$ and $W_{j\tau}(\theta) \equiv [Z_\tau(\theta), Z_{\tau-j}(\theta)]'$. By a second order Taylor series expansion, we have

$$\begin{aligned}
&\hat{g}_j(w) - \tilde{g}_j(w) \\
&\quad = (\hat{\theta}_R - \theta_0)' n_j^{-1} \sum_{t=R+j+1}^{T} \frac{\partial \kappa_h[w, W_{jt}(\theta_0)]}{\partial \theta} \\
&\qquad + \frac{1}{2}(\hat{\theta}_R - \theta_0)' n_j^{-1} \sum_{t=R+j+1}^{T} \frac{\partial^2 \kappa_h[w, W_{jt}(\bar{\theta}_R)]}{\partial \theta \, \partial \theta'} (\hat{\theta}_R - \theta_0),
\end{aligned} \tag{A.2}$$

where $\bar{\theta}_R$ lies between the segment of $\hat{\theta}_R$ and θ_0. It follows that

$$\begin{aligned}
&\hat{\Delta}_1(j) \\
&\quad \leqslant 2|\hat{\theta}_R - \theta_0|^2 \int_{\mathbb{I}^2} \left| n_j^{-1} \sum_{t=R+j+1}^{T} \frac{\partial \kappa_h[w, W_{j\tau}(\theta_0)]}{\partial \theta} \right|^2 \mathrm{d}w \\
&\qquad + |\hat{\theta}_R - \theta_0|^4 \int_{\mathbb{I}^2} \left| n_j^{-1} \sum_{t=R+j+1}^{T} \frac{\partial^2 \kappa_h[w, W_{j\tau}(\bar{\theta})]}{\partial \theta \, \partial \theta'} \right|^2 \mathrm{d}w \\
&\quad \equiv 2|\hat{\theta}_R - \theta_0|^2 \hat{\Delta}_{11}(j) + |\hat{\theta}_R - \theta_0|^4 \hat{\Delta}_{12}.
\end{aligned} \tag{A.3}$$

Put $\frac{\partial}{\partial \theta} \hat{\kappa}_h(w) \equiv n_j^{-1} \sum_{t=R+j+1}^{T} \frac{\partial}{\partial \theta} \kappa_h[w, W_{j\tau}(\theta_0)]$. Then

$$\hat{\Delta}_{11}(j) \leqslant 2 \int_{\mathbb{I}^2} \left| \mathrm{E} \frac{\partial \hat{\kappa}_h(w)}{\partial \theta} \right|^2 \mathrm{d}w + 2 \int_{\mathbb{I}^2} \left| \frac{\partial \hat{\kappa}_h(w)}{\partial \theta} - \mathrm{E} \frac{\partial \hat{\kappa}_h(w)}{\partial \theta} \right|^2 \mathrm{d}w \equiv 2\hat{D}_1(j) + 2\hat{D}_2(j). \tag{A.4}$$

We now compute the order of magnitude for $\hat{D}_1(j)$. Using the identity that

$$\frac{\partial \kappa_h[w, W_{jt}(\theta)]}{\partial \theta} = \frac{\partial K_h[z_1, Z_t(\theta)]}{\partial \theta} K_h[z_2, Z_{t-j}(\theta)] + K_h[z_1, Z_t(\theta)] \frac{\partial K_h[z_2, Z_{t-j}(\theta)]}{\partial \theta}, \tag{A.5}$$

iterated expectations, and $\mathrm{E}\{K_h[z_1, Z_t(\theta_0)] | I_{t-1}\} = \mathrm{E}K_h[z_1, Z_t(\theta_0)] = 1$ under $\mathbb{H}_0$, we have

$$\begin{aligned}
\mathrm{E}\left\{ \frac{\partial \kappa_h[w, W_{jt}(\theta_0)]}{\partial \theta} \right\} &= \mathrm{E}\left\{ \mathrm{E}\left[\frac{\partial K_h[z_1, Z_t(\theta_0)]}{\partial \theta} | I_{t-1} \right] K_h[z_2, Z_{t-j}(\theta_0)] \right\} \\
&\quad + \mathrm{E}\left\{ \frac{\partial K_h[z_2, Z_{t-j}(\theta_0)]}{\partial \theta} \right\}.
\end{aligned} \tag{A.6}$$

Recall $G_{t-1}(z) \equiv \mathrm{E}\{[\frac{\partial}{\partial \theta} Z_t(\theta)]_{\theta=\theta_0} | Z_t(\theta_0) = z, I_{t-1}\}$ in Assumption A.3. Because

$$\frac{\partial K_h[z_1, Z_t(\theta)]}{\partial \theta} = \frac{\partial K_h[z_1, Z_t(\theta)]}{\partial Z_t(\theta)} \frac{\partial Z_t(\theta)}{\partial \theta} \tag{A.7}$$

and $\frac{\partial}{\partial Z_t(\theta_0)} K_h[z_1, Z_t(\theta_0)]$ is a function of $Z_t(\theta_0)$, which is independent of I_{t-1} under $\mathbb{H}_0$, we have

$$\begin{aligned}\mathrm{E}\left\{\frac{\partial K_h[z_1, Z_t(\theta_0)]}{\partial\theta}|I_{t-1}\right\} &= \int_0^1 \frac{\partial K_h(z_1, z)]}{\partial z} G_{t-1}(z)\,\mathrm{d}z \\ &= [G_{t-1}(z)K_h(z_1, z)]_0^1 - \int_0^1 K_h(z_1, z)G'_{t-1}(z)\,\mathrm{d}z \\ &= -G'_{t-1}(z_1) + \mathrm{o}(1), \end{aligned} \tag{A.8}$$

where the first equality follows by iterated expectations and the i.i.d. U[0,1] property of $\{Z_t(\theta_0)\}$, and the last equality by change of variable $z = z_1 + hu$ and Assumption A.3. For the last equality, we have used the fact that $G_{t-1}(0) = G_{t-1}(1) = 0$ for all t. It follows from (A.6) and (A.8) that

$$\mathrm{E}\left\{\frac{\partial \kappa_h[w, W_{jt}(\theta_0)]}{\partial\theta}\right\} = -\{\mathrm{E}[G_{t-1}(z_1)K_h(z_2, Z_{t-j}(\theta_0)] + \mathrm{E}[G_{t-j-1}(z_2)]\}[1 + \mathrm{o}(1)]. \tag{A.9}$$

Hence, for the first term in (A.4), by change of variable and Assumption A.3, we have

$$\hat{D}_1(j) = \int_{\mathbb{I}^2} \left| n_j^{-1} \sum_{t=R+j+1}^{T} \mathrm{E}\frac{\partial \kappa_h[w, W_{jt}(\theta_0)]}{\partial\theta}\right|^2 \mathrm{d}w = \mathrm{O}(1). \tag{A.10}$$

Next, we consider the second term $\hat{D}_2(j)$ in (A.4). From (A.5) and (A.7), we observe that $\frac{\partial}{\partial\theta}\kappa_h[w, W_{jt}(\theta_0)]$ is a measurable function of at most $\{Z_t(\theta_0), \frac{\partial}{\partial\theta}Z_t(\theta_0), Z_{t-j}(\theta_0), \frac{\partial}{\partial\theta}Z_{t-j}(\theta_0)\}$. Given Assumption A.4 and the fact that $Z_t(\theta_0)$ is independent of I_{t-1} under $\mathbb{H}_0$, $\{\frac{\partial}{\partial\theta}\kappa_h[w, W_{jt}(\theta_0)]\}$ is an α-mixing process with α-mixing coefficient $\alpha_j(l) \leqslant 1$ if $l \leqslant j+1$ and $\alpha_j(l) = \alpha(l-j-1)$ if $l > j+1$ (cf. White, 1984, Proposition 6.1.8, p. 153). By the Cauchy–Schwarz inequality and a standard α-mixing inequality (Hall and Heyde, 1980, Corollary A.2, p. 278), we have

$$\begin{aligned}&\int_{\mathbb{I}^2} \mathrm{E}\left|\frac{\partial\hat{\kappa}_h(w)}{\partial\theta} - \mathrm{E}\frac{\partial\hat{\kappa}_h(w)}{\partial\theta}\right|^2 \mathrm{d}w \\ &\leqslant 2n_j^{-1}\sum_{l=0}^{n-1}\sum_{t=R+l+1}^{T}\int_{\mathbb{I}^2}\mathrm{cov}\left\{\frac{\partial\kappa_h[w, W_{jt}(\theta_0)]}{\partial\theta}, \frac{\partial\kappa_h[w, W_{j(t-l)}(\theta_0)]}{\partial\theta}\right\}\mathrm{d}w \\ &\leqslant Cn_j^{-1}\left[\sum_{l=0}^{\infty}\alpha_j(l)^{v/(v-1)}\right] n_j^{-1}\sum_{t=R+l+1}^{T}\int_{\mathbb{I}^2}\left\{\mathrm{E}\left|\frac{\partial\kappa_h[w, W_{jt}(\theta_0)]}{\partial\theta}\right|^{2v}\right\}^{1/v}\mathrm{d}w \\ &= \mathrm{O}(n_j^{-1}jh^{-6+2/v}),\end{aligned} \tag{A.11}$$

where we made use of the facts that $\sum_{l=0}^{\infty}\alpha_j(l)^{v/(v-1)} \leqslant C(j+1)$ by Assumption A.4, and $n_j^{-1}\sum_{t=R+j+1}^{T}\int_{\mathbb{I}^2}\{\mathrm{E}|(\partial/\partial\theta)\kappa_h[w, W_{jt}(\theta_0)]|^{2v}\}^{1/v}\,\mathrm{d}w \leqslant 2C(v)h^{-6+2/v}[n_j^{-1}\sum_{t=R+j+1}^{T}\mathrm{E}|(\partial/\partial\theta) Z_\tau(\theta_0)|^{2v}]^{1/v}$ by Jensen's inequality, the C_r-inequality, (A.5), (A.7), Lemma A.1 and Assumption A.2. It follows from (A.11) and Markov's inequality that $\hat{D}_2(j) = \mathrm{O_P}(n_j^{-1}jh^{-6+2/v})$. This, (A.4) and (A.10) imply

$$\hat{\varDelta}_{11}(j) = \mathrm{O_P}(1 + n_j^{-1}jh^{-6+2/v}). \tag{A.12}$$

Next, we consider the second term $\hat{\Delta}_{12}$ in (A.3). Noting that

$$\frac{\partial^2 \kappa_h[w, W_{jt}(\theta)]}{\partial\theta\,\partial\theta'} = \frac{\partial^2 K_h[z_1, Z_t(\theta)]}{\partial\theta\,\partial\theta'} K_h[z_2, Z_{t-j}(\theta)] + K_h[z_1, Z_t(\theta)] \frac{\partial^2 K_h[z_2, Z_{t-j}(\theta)]}{\partial\theta\,\partial\theta'} + 2\frac{\partial K_h[z_1, Z_t(\theta)]}{\partial\theta}\frac{\partial K_h[z_2, Z_{t-j}(\theta)]}{\partial\theta'}, \tag{A.13}$$

we write

$$\begin{aligned}
&\hat{\Delta}_{12}(j)\\
&\quad\leqslant 8\int_{\mathbb{I}^2}\left|n_j^{-1}\sum_{t=R+j+1}^{T}\frac{\partial^2 K_h[z_1, Z_t(\theta)]}{\partial\theta\,\partial\theta'}K_h[z_2, Z_{t-j}(\theta)]\right|^2 \mathrm{d}w\\
&\qquad + 8\int_{\mathbb{I}^2}\left|n_j^{-1}\sum_{t=R+j+1}^{T}K_h[z_1, Z_t(\theta)]\frac{\partial^2 K_h[z_2, Z_{t-j}(\theta)]}{\partial\theta\,\partial\theta'}\right|^2 \mathrm{d}w\\
&\qquad + 8\int_{\mathbb{I}^2}\left|n_j^{-1}\sum_{t=R+j+1}^{T}\frac{\partial K_h[z_1, Z_t(\theta)]}{\partial\theta}\frac{\partial K_h[z_2, Z_{t-j}(\theta)]}{\partial\theta'}\right|^2 \mathrm{d}w\\
&\quad\equiv 8\hat{D}_3(j) + 8\hat{D}_4(j) + 8\hat{D}_5(j).
\end{aligned}\tag{A.14}$$

For the first term in (A.14), by the Cauchy–Schwarz inequality, the identity that

$$\frac{\partial^2 K_h[z_1, Z_t(\theta)]}{\partial\theta\,\partial\theta'} = \frac{\partial^2 K_h[z_1, Z_t(\theta)]}{\partial^2 Z_t(\theta)}\frac{\partial Z_t(\theta)}{\partial\theta}\frac{\partial Z_t(\theta)}{\partial\theta'} + \frac{\partial K_h[z_1, Z_t(\theta)]}{\partial Z_t(\theta)}\frac{\partial^2 Z_t(\theta)}{\partial\theta\,\partial\theta'}, \tag{A.15}$$

Lemma A.1, and Assumption A.2, we have $\hat{D}_3(j) = \mathrm{O_P}(h^{-6})$ and $\hat{D}_4(j) = \mathrm{O_P}(h^{-6})$. For $\hat{D}_5(j)$, we have $\hat{D}_5(j) \leqslant \{n_j^{-1}\sum_{t=R+1}^{T}\int_0^1 |\frac{\partial}{\partial\theta}K_h[z_1, Z_t(\theta)]|^2 \mathrm{d}z_1\}^2 = \mathrm{O}(h^{-6})$ by the Cauchy–Schwarz inequality, (A.7), Lemma A.1, and Assumption A.2. It follows from (A.14) that $\hat{\Delta}_{12}(j) = \mathrm{O_P}(h^{-6})$. This, (A.3), (A.12), and Assumptions A.5 and A.7 imply $\hat{\Delta}_1(j) = \mathrm{O_P}(R^{-1} + R^{-1}n_j^{-1}jh^{-6+2/v} + R^{-2}h^{-6}) = \mathrm{o_P}(n_j^{-1}h^{-1})$ for any fixed $j>0$. The proof of Proposition A.1 is finished. □

Proof of Proposition A.2. Using (A.2), we have

$$\begin{aligned}
\hat{\Delta}_2(j) &= (\hat{\theta}_R - \theta_0)'\int_{\mathbb{I}^2}[\tilde{g}_j(w) - 1]n_j^{-1}\sum_{\tau=R+j+1}^{T}\frac{\partial\kappa_h[w, W_{jt}(\theta_0)]}{\partial\theta}\,\mathrm{d}w\\
&\quad + \frac{1}{2}(\hat{\theta}_R - \theta_0)'\int_{\mathbb{I}^2}[\tilde{g}_j(w) - 1]n_j^{-1}\sum_{t=R+j+1}^{T}\frac{\partial^2\kappa_h[w, W_{jt}(\bar{\theta}_R)]}{\partial\theta\,\partial\theta'}\,\mathrm{d}w(\hat{\theta}_R - \theta_0)\\
&\equiv (\hat{\theta}_R - \theta_0)'\hat{\Delta}_{21}(j) + \frac{1}{2}(\hat{\theta}_R - \theta_0)'\hat{\Delta}_{22}(j)(\hat{\theta}_R - \theta_0).
\end{aligned}\tag{A.16}$$

We first consider $\hat{\Delta}_{21}(j)$. Recall the definition of $\frac{\partial}{\partial\theta}\hat{\kappa}(w)$ as used in (A.4). We have

$$\begin{aligned}
\hat{\Delta}_{21}(j) &= \int_{\mathbb{I}^2}\mathrm{E}\frac{\partial\hat{\kappa}_h(w)}{\partial\theta}[\tilde{g}_j(w) - 1]\,\mathrm{d}w + \int_{\mathbb{I}^2}\left[\frac{\partial\hat{\kappa}_h(w)}{\partial\theta} - \mathrm{E}\frac{\partial\hat{\kappa}_h(w)}{\partial\theta}\right][\tilde{g}_j(w) - 1]\,\mathrm{d}w\\
&\equiv \hat{D}_6(j) + \hat{D}_7(j).
\end{aligned}\tag{A.17}$$

We write $\hat{D}_6(j) = n_j^{-1}\sum_{t=R+j+1}^{T} D_{6ht}(j)$, where $D_{6ht}(j) \equiv \int_{\mathbb{I}^2} \kappa_h(w, W_{jt})\mathrm{E}[\frac{\partial}{\partial\theta}\kappa_h(w, W_{jt})]\,\mathrm{d}w$ is a j-dependent process with zero mean given $\mathrm{E}\kappa_h[w, W_{jt}(\theta_0)] = 0$ under $\mathbb{H}_0$. Because $\mathrm{E}\{\kappa_h[w, W_{jt}(\theta_0)]\kappa_h[w, W_{js}(\theta_0)]\} = 0$ unless $t = s, s \pm j$, we have $\mathrm{E}|\hat{D}_6(j)|^2 \leqslant 3n_j^{-1}\sum_{t=R+j+1}^{T} \mathrm{E}|D_{6ht}(j)|^2 = \mathrm{O}(n_j^{-1})$ by (A.9), change of variable, and Assumption A.3. Thus, we have $\hat{D}_6(j) = \mathrm{O_P}(n_j^{-1/2})$.

For the second term in (A.17), by the Cauchy–Schwarz inequality, (A.11), Markov's inequality, and $\sup_{z\in\mathbb{I}^2}|\tilde{g}_j(w) - 1| = \mathrm{O_P}(n_j^{-1/2}h^{-1}\ln(n_j))$ as follows from a standard uniform convergence argument for kernel density estimation with application of Bernstein's large deviation inequality, we have $\hat{D}_7(j) = \mathrm{O_P}(n_j^{-1}j^{1/2}h^{-4+1/v}\ln(n_j))$. It follows from (A.17) that

$$\hat{\varDelta}_{21}(j) = \mathrm{O_P}(n_j^{-1/2} + n_j^{-1}j^{1/2}h^{-4+1/v}\ln(n_j)). \tag{A.18}$$

Next, we consider the second term $\hat{\varDelta}_{22}(j)$ in (A.16). Using (A.13), we have

$$\begin{aligned}\hat{\varDelta}_{22}(j) = &\int_{\mathbb{I}^2}[\tilde{g}_j(w) - 1]\left\{n_j^{-1}\sum_{t=R+j+1}^{T}\frac{\partial^2 K_h[z_1, Z_t(\bar{\theta}_R)]}{\partial\theta\,\partial\theta'}K_h[z_2, Z_{t-j}(\bar{\theta})]\right\}\mathrm{d}w\\ &+\int_{\mathbb{I}^2}[\tilde{g}_j(w) - 1]\left\{n_j^{-1}\sum_{t=R+j+1}^{T}K_h[z_1, Z_t(\bar{\theta}_R)]\frac{\partial^2 K_h[z_2, Z_{t-j}(\bar{\theta})]}{\partial\theta\,\partial\theta'}\right\}\mathrm{d}w\\ &+2\int_{\mathbb{I}^2}[\tilde{g}_j(w) - 1]\left\{n_j^{-1}\sum_{t=R+j+1}^{T}\frac{\partial K_h[z_1, Z_t(\bar{\theta}_R)]}{\partial\theta}\frac{\partial K_h[z_2, Z_{t-j}(\bar{\theta})]}{\partial\theta'}\right\}\mathrm{d}w\\ \equiv\ &\hat{D}_8(j) + \hat{D}_9(j) + 2\hat{D}_{10}(j).\end{aligned} \tag{A.19}$$

For the first term in (A.19), using (A.15), Lemma A.1, and Assumption A.2, we have $|\hat{D}_8(j)| \leqslant \sup_{w\in\mathbb{I}^2}|\tilde{g}(w) - 1|n_j^{-1}\sum_{t=R+j+1}^{T}\int_0^1 |(\partial^2/\partial\theta\,\partial\theta')K_h[z_1, Z_t(\bar{\theta}_R)]\,\mathrm{d}z_1\int_0^1 K_h[z_2, Z_{t-j}(\bar{\theta}_R)]\,\mathrm{d}z_2 = \mathrm{O_P}(n_j^{-1/2}h^{-3}\ln(n_j))$. Similarly, we can show $\hat{D}_9(j) = \mathrm{O_P}(n_j^{-1/2}h^{-3}\ln(n_j))$. We also have $\hat{D}_{10}(j) = \mathrm{O_P}(n_j^{-1/2}h^{-3}\ln(n_j))$ by (A.7), Lemma A.1, and Assumption A.2. It follows from (A.19) that $\hat{\varDelta}_{22}(j) = \mathrm{O_P}(n_j^{-1/2}h^{-3}\ln(n_j))$. This, (A.16), (A.18), Assumptions A.5 and A.7, and the fact that j is a fixed lag order imply $\hat{\varDelta}_2(j) = \mathrm{O_P}(R^{-1/2}n_j^{-1/2} + R^{-1/2}n_j^{-1}j^{1/2}h^{-4+1/v}\ln(n_j) + R^{-1}n_j^{-1/2}h^{-3}\ln(n_j)) = \mathrm{o_P}(n_j^{-1}h^{-1})$. □

Proof of Theorem A.2. See Hong and Li (2005, Theorem A.2). □

Proof of Theorem 2. Let $\lambda \equiv (\lambda_1, \ldots, \lambda_p)'$ be a $p \times 1$ vector such that $\lambda'\lambda = 1$. Define $\hat{Q}_\lambda \equiv \sum_{c=1}^{p}\lambda_c\hat{Q}(j_c)$, where $j_1, \ldots, j_p$ are distinct integers. Following reasoning analogous to that for Theorem A.1, we can show $\hat{Q}_\lambda = \sum_{c=1}^{p}\lambda_c\tilde{Q}(j_c) + \mathrm{o_P}(1)$ for any given p. Moreover, following reasoning analogous to Hong and Li (2005, Theorem A.2), we can show $\sum_{j=1}^{p}\lambda_j\tilde{Q}(j) \to^{\mathrm{d}} \mathrm{N}(0,1)$ given $\lambda'\lambda = 1$. It follows that $\hat{Q}_\lambda \to^{\mathrm{d}} \mathrm{N}(0,1)$. The desired result for $W(p)$ follows immediately by setting $\lambda = (1/\sqrt{p}, \ldots, 1/\sqrt{p})'$. □

Proof of Theorem 3. Put $M(j) \equiv \int_{\mathbb{I}^2}[g_j(w) - 1]^2\,\mathrm{d}w$, and recall $\hat{M}(j) \equiv \int_{\mathbb{I}^2}[\hat{g}_j(w) - 1]^2\,\mathrm{d}w$. Then

$$\hat{M}(j) - M(j) = \int_{\mathbb{I}^2}[\hat{g}_j(w) - g_j(w)]^2\,\mathrm{d}w + 2\int_{\mathbb{I}^2}[\hat{g}_j(w) - g_j(w)][g_j(w) - 1]\,\mathrm{d}w. \tag{A.20}$$

We shall show $\hat{M}(j) - M(j) \rightarrow^p 0$. Note that $\int_{\mathbb{I}^2}[\hat{g}_j(w) - g_j(w)]^2\,\mathrm{d}w \leqslant 2\int_{\mathbb{I}^2}[\hat{g}_j(w) - \tilde{g}_j(w)]^2\mathrm{d}w + 2\int_{\mathbb{I}^2}[\tilde{g}_j(w) - g_j(w)]^2\mathrm{d}w$. For the first term, we have $\int_{\mathbb{I}^2}[\hat{g}_j(w) - \tilde{g}_j(w)]^2\mathrm{d}w \rightarrow^p 0$ following reasoning analogous to that of Theorem A.1. For the second term, from the proof of Hong and Li (2005, proof of Theorem 3), we have $\int_{\mathbb{I}^2}[\tilde{g}_j(w) - g_j(w)]^2\,\mathrm{d}w = \mathrm{O_P}(n_j^{-1}h^{-2} + h^2)$, where the $\mathrm{O}(h^2)$ term is the squared bias given Assumption A.8. It follows that $\int_{\mathbb{I}^2}[\hat{g}_j(w) - g_j(w)]^2\,\mathrm{d}w \rightarrow^p 0$ given Assumption A.7. We thus have $\hat{M}(j) - M(j) \rightarrow^p 0$ by the Cauchy–Schwarz inequality and (A.20).Moreover, given $(n_jh)^{-1}A_h^0 = \mathrm{O}(n_j^{-1}h^{-3}) = \mathrm{o}(1)$, we have $(n_jh)^{-1}\hat{Q}(j) = V_0^{-1/2}M(j) + \mathrm{o_P}(1)$ for any given j. It follows that $\mathrm{P}[\hat{Q}(j) > C_n] \rightarrow 1$ for any $C_n = \mathrm{o}(nh)$ if $M(j) > 0$, which holds when $\{Z_\tau, Z_{\tau-j}\}$ are not independent or U[0,1]. Therefore, $\mathrm{P}[\hat{W}(p) > C_n] \rightarrow 1$ for any fixed $p > 0$ whenever $\mathrm{P}[\hat{Q}(j) > C_n] \rightarrow 1$ at some lag $j \in \{1, \ldots, p\}$. $\square$

References

Andersen, T.G., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance 4, 115–158.

Ang, A., Bekaert, G., 1998. Regime switches in interest rates, NBER Working Paper 6508.

Baillie, R., Bollerslev, T., 1989. The message in daily exchange rates: a conditional variance tale. Journal of Business and Economic Statistics 7, 297–305.

Bates, D., 1996. Jumps and stochastic volatility: exchange rate processes implicit in Deutsche Mark options. Review of Financial Studies 9, 69–107.

Berkowitz, J., 2001. Testing the accuracy of density forecasts, applications to risk management. Journal of Business and Economic Statistics 19, 465–474.

Bollen, N., Gray, S., Whaley, R., 2000. Regime-switching in foreign exchange rates: evidence from currency option prices. Journal of Econometrics 94, 239–276.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Bollerslev, T., 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. Review of Economics and Statistics 69, 542–547.

Boothe, P., Glassman, D., 1987. The statistical distribution of exchange rates: empirical evidence and economic implications. Journal of International Economics 22, 297–319.

Chapman, D., Pearson, N., 2000. Is the short rate drift actually nonlinear? Journal of Finance 55, 355–388.

Clements, M.P., Smith, J., 2000. Evaluating the forecast densities of linear and non-linear models: applications to output growth and unemployment. Journal of Forecasting 19, 255–276.

Corradi, V., Swanson, N., 2005. A test for comparing multiple misspecified conditional interval models. Econometric Theory 21, 991–1016.

Corradi, V., Swanson, N., 2006a. Predictive density evaluation. In: Granger, C.W.J., Elliot, G., Timmerman, A. (Eds.), Handbook of Economic Forecasting. Elsevier, Amsterdam, pp. 197–284.

Corradi, V., Swanson, N., 2006b. Predictive density and conditional confidence interval accuracy tests. Journal of Econometrics, 135, 187–228.

Corradi, V., Swanson, N., 2007. Nonparametric bootstrap procedures for predictive inference based on recursive estimation schemes. International Economic Review, forthcoming.

Cox, D.R., Snell, E.J., 1968. A general definition of residuals. Journal of the Royal Statistical Society B 30, 248–265.

Dawid, A.P., 1984. Statistical theory: the prequential approach. Journal of the Royal Statistical Society A 147, 278–292.

Diebold, F.X., 1988. Empirical Modeling of Exchange Rate Dynamics. Springer, New York.

Diebold, F.X., Mariano, R., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–265.

Diebold, F.X., Nason, J.A., 1990. Nonparametric exchange rate prediction. Journal of International Economics 28, 315–332.
Diebold, F.X., Gunther, T.A., Tay, A.S., 1998. Evaluating density forecasts with applications to financial risk management. International Economic Review 39, 863–883.
Diebold, F.X., Hahn, J., Tay, A.S., 1999. Multivariate density forecast evaluation and calibration in financial risk management: high-frequency returns of foreign exchange. Review of Economics and Statistics 81, 661–673.
Duffie, D., Pan, J., 1997. An overview of value at risk. Journal of Derivatives 4, 13–32.
Engel, C., 1994. Can the Markov switching model forecast exchange rates?. Journal of International Economics 36, 151–165.
Engel, C., Hamilton, J., 1990. Long swings in the dollar: are they in the data and do markets know it? American Economic Review 80, 689–713.
Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. Econometrica 50, 987–1008.
Engle, R.F., Bollerslev, T., 1986. Measuring the persistence in conditional variances. Econometric Reviews 5, 1–50.
Giacomini, R., White, H., 2003. Tests of conditional predictive ability. Working Paper, University of California, San Diego.
Gourieroux, C., Monfort, A., Renault, E., Trognon, A., 1987. Generalized residuals. Journal of Econometrics 34, 5–32.
Granger, C., 1999. The evaluation of econometric models and of forecasts. Invited Lecture in Far Eastern Meeting of the Econometric Society, Singapore.
Granger, C., Pesaran, M.H., 2000. A decision theoretic approach to forecasting evaluation. In: Chan, W.S., Li, W.K., Tong, H. (Eds.), Statistics and Finance: An Interface. Imperial College Press, London, pp. 261–278.
Gray, S., 1996. Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics 42, 27–62.
Hall, P., Heyde, C., 1980. Martingale limit theory and its application. Academic Press, New York.
Hamilton, J.D., 1989. A new approach to the econometric analysis of nonstationary time series and the business cycle. Econometrica 57, 357–384.
Hamilton, J.D., Susmel, R., 1994. Autoregressive conditional heteroskedasticity and changes in regime. Journal of Econometrics 17, 143–173.
Hansen, B.E., 1994. Autoregressive conditional density estimation. International Economic Review 35, 705–730.
Harrison, P.J., Stevens, C., 1976. Bayesian forecasting. Journal of the Royal Statistical Society B 38, 205–247.
Hong, Y., 1996. Consistent testing for serial correlation of unknown form. Econometrica 64, 837–864.
Hong, Y., 2001. Evaluation of out-of-sample density forecasts with application to stock prices. Working Paper, Cornell University.
Hong, Y., Li, H., 2005. Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. Review of Financial Studies 18, 37–84.
Hong, Y., Li, H., Zhao, F., 2006. Can the random walk model be beaten in out-of-sample density forecasts? Evidence from intraday foreign exchange rates. Working Paper, Cornell University.
Hsieh, D., 1988. Statistical properties of daily exchange rates. Journal of International Economics 24, 129–145.
Hsieh, D., 1989. Modeling heteroskedasticity in daily exchange rates. Journal of Business and Economic Statistics 7, 307–317.
Jorion, P., 1988. On jump processes in the foreign exchange and in the stock markets. Review of Financial Studies 1, 427–455.
Jorion, P., 2000. Value at risk: the new benchmark for managing financial risk. McGraw-Hill, New York.
J.P., Morgan, 1996. RiskMetrics—Technical Document, fourth ed. Morgan Guaranty Trust Company, New York.
Lancaster, T., 1990. The Econometric Analysis of Transition Data. Cambridge, Cambridge University Press.
Meese, R., Rogoff, K., 1983. Empirical exchange rate models of the seventies: do they fit out-of-sample?. Journal of International Economics 14, 3–24.
Meese, R., Rose, A., 1990. Nonlinear, nonparametric, nonessential exchange rate estimation. American Economic Review 80, 192–196.
Nelson, D., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59, 347–370.
Putnam, B., Quintana, J., 1994. New Bayesian statistical approaches to estimating and evaluating models of exchange rates determination. In: 1994 Proceedings of the Section on Bayesian Statistical Science. American Statistical Association, Alexandria.

Quintana, J., Putnam, B., 1996. Debating currency markets efficiency using dynamic multiple-factor models. American Statistical Association Proceedings of the Section on Bayesian Statistical Science, 55–60.

Scott, D.W., 1992. Multivariate Density Estimation, Theory, Practice, and Visualization. Wiley, New York.

Tay, A., Wallis, K., 2000. Density forecasting: a survey. Journal of Forecasting 19, 235–254.

West, M., Harrison, P.J., 1997. Bayesian Forecasting and Dynamics Models. Springer, New York, second edition.

White, H., 1984. Asymptotic Theory for Econometricians. Academic Press, San Diego.

Wolff, C.C.P., 1985. Exchange rate models, parameter variation and innovations: a study on the forecasting performance of empirical models of exchange rate determination. Unpublished Doctoral Dissertation, Graduate School of Business, University of Chicago.

Wolff, C.C.P., 1987. Forward foreign exchange rates, expected spot rates, and premia: a signal-extracting approach. Journal of Finance XLII, 395–406.

Zellner, A., 1971. Introduction to Bayesian Inference in Econometrics. Wiley, New York.

Zellner, A., Hong, C., Min, C., 1991. Forecasting turning points in international output growth rates using Bayesian exponentially weighted autoregression, time-varying parameter, and pooling techniques. Journal of Econometrics 49, 275–304.