Dynamic Integration of Time- and State-Domain Methods for Volatility Estimation

Jianqing FAN, Yingying FAN, and Jiancheng JIANG

Time- and state-domain methods are two common approaches to nonparametric prediction. Whereas the former uses data predominantly from recent history, the latter relies mainly on historical information. Combining these two pieces of valuable information is an interesting challenge in statistics. We surmount this problem by dynamically integrating information from both the time and state domains. The estimators from these two domains are optimally combined based on a data-driven weighting strategy, which provides a more efficient estimator of volatility. Asymptotic normality is separately established for the time domain, the state domain, and the integrated estimators. By comparing the efficiency of the estimators, we demonstrate that the proposed integrated estimator uniformly dominates the other two estimators. The proposed dynamic integration approach is also applicable to other estimation problems in time series. Extensive simulations are conducted to demonstrate that the newly proposed procedure outperforms some popular ones, such as the RiskMetrics and historical simulation approaches, among others. In addition, empirical studies convincingly endorse our integration method.

KEY WORDS: Bayes; Dynamical integration; Smoothing; State domain; Time domain; Volatility.

1. INTRODUCTION

When forecasting a future event or making an investment decision, two pieces of useful information are frequently consulted. On the one hand, based on the recent history, a form of local average, such as the moving average in the time domain, can be used to forecast a future event. This approach uses the continuity of a function and ignores the information in the remote history, which is related to the current through stationarity. On the other hand, a future event can be forecast based on state-domain modeling techniques (see Fan and Yao 2003 for details). For example, forecasting the volatility of bond yields with the current rate 6.47% involves computing the standard deviation based on the historical information with yields of around 6.47%. This approach relies on the stationarity of the yields and depends predominately on historical data, and ignores the importance of recent data.

In general, the foregoing two pieces of information are weakly dependent. For example, consider the weekly data on the yields of 3-month Treasury Bills presented in Figure 1. Suppose that the current time is January 4, 1991 and the interest rate is 6.47%. Based on the weighted squared differences in the past 52 weeks (1 year), for example, the volatility may be estimated. This corresponds to the time-domain smoothing, using the small vertical stretch of data in Figure 1(a). Figure 1(b) computes the squared differences of the past year's data and depicts the associated exponential weights. The estimated volatility (conditional variance) is indicated by the dashed horizontal bar. Let the resulting estimator be $\hat{\sigma}^2_{t,\text{time}}$. On the other hand, in financial activities, we do consult historical information to make better decisions. The current interest rate is 6.47%. The volatility of the yields may be examined when the interest rate is around 6.47%, say 6.47% ± .25%. This corresponds to using the part of the data indicated by the horizontal bar. Figure 1(c)

plots the squared differences $(X_t - X_{t-1})^2$ against X_{t-1}, with X_{t-1} restricted to the interval 6.47% ± .25%. Applying the local kernel weight to the squared differences results in a state-domain estimator $\hat{\sigma}^2_{t,\text{state}}$, indicated by the horizontal bar in Figure 1(c). Clearly, as shown in Figure 1(a), except in the 3-week period immediately before January 4, 1991, the last period with an interest rate within 6.47% ± .25% is the period from May 15, 1988 and July 22, 1988. Thus, the time- and state-domain estimators use two weakly dependent components of the time series, because these two components are 136 weeks apart in time; see the horizontal and vertical bars of Figure 1(a).

It is important to combine the estimators from the time domain and the state domain separately, because this combination allows us to use more sampling information and to give better estimates. In fact, compared with the state-domain estimator, our new integrated estimator will put more emphasis (weight) on recent data; in contrast with the time-domain estimator, it will use historical data to improve the efficiency. These will also be demonstrated mathematically.

Both state- and time-domain smoothing have been popularly studied in the literature. Many authors have contributed their works on the first topic; the survey papers by Cai and Hong (2003) and Fan (2005) provide overviews. Other works include, for example, drift and volatility estimation for short rate by Chapman and Pearson (2000); density estimation for spatial linear processes by Hallin, Lu, and Tran (2001); diffusion estimation by Bandi and Phillips (2003); Fan and Zhang (2003), and Aït-Sahalia and Mykland (2004); kernel estimation and testing for continuous-time financial models by Arapis and Gao (2004); and tests for diffusion model by Chen, Gao, and Tang (2007) and Hong and Li (2005). On the other hand, there is also a large literature on time-domain smoothing, including work by Hall and Hart (1990), Robinson (1997), Gijbels, Pope, and Wand (1999), Hardle, Herwartz, and Spokoiny (2002), Fan and Gu (2003), Mercurio and Spokoiny (2004), and Aït-Sahalia, Mykland, and Zhang (2005), among others. The accuracies of both time- and state-domain estimators for the volatility depend on the information contained in the recent data and historical data. Brenner, Harjes, and Kroner (1996) proposed an interesting model on term interest rate that has some flavor of

Figure 1. Illustration of Time-Domain and State-Domain Estimation. (a) The yields of 3-month Treasury Bills from 1954 to 2004. The vertical bar indicates localization in time, and the horizontal bar represents localization in state. (b) Illustration of time-domain smoothing. Squared differences are plotted against its time index, and the exponential weights are used to compute the local average. (c) Illustration of state-domain smoothing. Squared differences are plotted against the level of interest rates, restricted to the interval 6.47% ± .25% indicated by the horizontal bar in (a). The Epanechnikov kernel is used for computing the local average.

combining the time- and state-domain information in a parametric form. However, there is no formal work in the literature on efficiently integrating the time- and state-domain estimators.

In this article we realize the existence of two pieces of weakly independent information and introduce a general approach for integrating the time- and state-domain estimators. The integrated estimator borrows the strengths of both the time- and state-domain estimators with aggregated information from the data, and thus dominates both of the estimators when the data-generating process is a continuous stationary diffusion process. In particular, when the time domain is far more informative than the state domain, the integrated estimator will basically become the time-domain estimator. On the other hand, if at a particular time when the performance of the state-domain estimator dominates that of the time-domain estimator, the procedure opts automatically for the state-domain estimator.

Our strategy for integration is to introduce a dynamic weighting scheme, $0 \leq w_t \leq 1$, to combine the two weakly dependent estimators. Define the resulting integrated estimator as

$$\hat{\sigma}_t^2 = w_t \hat{\sigma}_{t,\text{time}}^2 + (1 - w_t)\hat{\sigma}_{t,\text{state}}^2. \tag{1}$$

The question is how to choose the dynamic weight w_t to optimize the performance. A reasonable approach is to minimize the variance of the combined estimator, leading to the dynamic

optimal weight

$$w_t = \frac{\text{var}(\hat{\sigma}_{t,\text{state}}^2)}{\text{var}(\hat{\sigma}_{t,\text{time}}^2) + \text{var}(\hat{\sigma}_{t,\text{state}}^2)}. \tag{2}$$

Estimation of the unknown variances in (2) is introduced in Section 3. Another approach to integration is to use the Bayesian approach, which considers the historical information the prior. We explore this idea in Section 4.

To appreciate the intuition behind our approach and how it works, consider the diffusion process

$$dr_t = \mu(r_t)\,dt + \sigma(r_t)\,dW_t, \tag{3}$$

where W_t is a Wiener process on $[0, \infty)$. This diffusion process is frequently used to model asset price and the yields of bonds, which are fundamental to fixed income securities, financial markets, consumer spending, corporate earnings, asset pricing, and inflation. The volatilities of interest rates are fundamentally important to value the prices of bonds and contract debt obligations (CDOs), the derivatives of bonds, and their associated risk management (e.g., hedging).

The family of models in (4) includes famous ones such as the Vasicek (1977) model, the CIR model (Cox, Ingersoll, and Ross 1985), and the CKLS model (Chan, Karolyi, Longstaff, and Sanders 1992). Suppose that at time t we have historic data $\{r_{t_i}\}_{i=0}^N$ from process (3) with a sampling interval Δ. Our aim

is to estimate the volatility $\sigma_t^2 \equiv \sigma^2(r_t)$. Let $Y_i = \Delta^{-1/2}(r_{t_{i+1}} - r_{t_i})$. Then, for model (3), the Euler approximation scheme is

$$Y_i \approx \mu(r_{t_i})\Delta^{1/2} + \sigma(r_{t_i})\varepsilon_i, \tag{4}$$

where $\varepsilon_i \overset{\text{iid}}{\sim} N(0, 1)$ for $i = 0, \ldots, N - 1$. Fan and Zhang (2003) studied the impact of the order of difference on statistical estimation. They found that although a higher order can possibly reduce approximation errors, it substantially increases the variances of data. They recommended the Euler scheme (4) for most practical situations.

To facilitate the derivation of mathematical theory, we focus on the estimation of volatility in model (3) to illustrate how to deal with the problem of dynamic integration. Asymptotic normality of the proposed estimator is established under this model, and extensive simulations are conducted. This theoretically and empirically demonstrates the dominant performance of the integrated estimation. Our method focuses on only the estimation of volatility, but it can be adapted to other estimation problems, such as the value at risk studied by Duffie and Pan (1997), the drift estimation for diffusion considered by Spokoiny (2000), and conditional moments, conditional correlation, conditional distribution, and derivative pricing. It is also applicable to other estimation problems in time series, such as forecasting the mean function. Further studies along these lines are beyond the scope of the current investigation.

2. ESTIMATION OF VOLATILITY

From here on, for theoretical derivations, we assume that the data $\{r_{t_i}\}$ are sampled from the diffusion model (3). As demonstrated by Stanton (1997) and Fan and Zhang (2003), the drift term in (4) contributes to the volatility in the order of $o(1)$ under certain conditions when $\Delta \to 0$ (e.g., Stanton 1997). Thus, for simplicity, we sometimes ignore the drift term when estimating the volatility.

2.1 Time-Domain Estimator

A popular version of the time-domain estimator of the volatility is the moving average estimator,

$$\hat{\sigma}_{MA,t}^2 = n^{-1} \sum_{i=t-n}^{t-1} Y_i^2, \tag{5}$$

where n is the size of the moving window. This estimator ignores the drift component and uses local n data points. An extension of the moving average estimator is the exponential smoothing estimation given by

$$\begin{aligned}\hat{\sigma}_{ES,t}^2 &= (1 - \lambda)Y_{t-1}^2 + \lambda\hat{\sigma}_{ES,t-1}^2 \\ &= (1 - \lambda)\{Y_{t-1}^2 + \lambda Y_{t-2}^2 + \lambda^2 Y_{t-3}^2 + \cdots\},\end{aligned} \tag{6}$$

where λ is a smoothing parameter between 0 and 1 that controls the size of the local neighborhood. This estimator is closely related to that from the GARCH(1, 1) model (see Fan, Jiang, Zhang, and Zhou 2003 for more details). The RiskMetrics of J. P. Morgan (1996), which is used for measuring the risk, called value at risk (VaR), of financial assets, recommends $\lambda = .94$ and $\lambda = .97$ for calculating the VaR of the daily and monthly returns.

The exponential smoothing estimator in (6) is a weighted sum of the squared returns before time t. Because the weight decays exponentially, it essentially uses recent data. A slightly modified version that explicitly uses only n data points before time t is

$$\hat{\sigma}_{ES,t}^2 = \frac{1 - \lambda}{1 - \lambda^n} \sum_{i=1}^{n} Y_{t-i}^2 \lambda^{i-1}, \tag{7}$$

where the smoothing parameter λ depends on n. When $\lambda \to 1$, it becomes the moving average estimator (5).

Theorem 1. Suppose that $\sigma_t^2 > 0$. Under conditions (C1) and (C2), if $n \to \infty$ and $n\Delta \to 0$, then $\hat{\sigma}_{ES,t}^2 - \sigma_t^2 \to 0$, almost surely. Moreover, if the limit $c = \lim_{n \to \infty} n(1 - \lambda)$ exists and $n\Delta^{(2p-1)/(4p-1)} \to 0$ where p is as specified in condition (C2), then

$$\sqrt{n}[\hat{\sigma}_{ES,t}^2 - \sigma_t^2]/s_{1,t} \overset{\mathcal{D}}{\longrightarrow} \mathcal{N}(0, 1),$$

where $s_{1,t}^2 = c\,\sigma_t^4 \frac{e^c+1}{e^c-1}$.

Theorem 1 has very interesting implications. We can compute the variance as if the data were independent. Indeed, if the data in (7) were independent and locally homogeneous, then

$$\text{var}(\hat{\sigma}_{ES,t}^2) \approx \frac{(1 - \lambda)^2}{(1 - \lambda^n)^2} 2\sigma_t^4 \sum_{i=1}^{n} \lambda^{2(i-1)} \approx \frac{1}{n} s_{1,t}^2.$$

This is indeed the asymptotic variance given in Theorem 1.

2.2 Estimation in State Domain

To obtain the nonparametric estimation of the functions $f(x) = \Delta^{1/2}\mu(x)$ and $\sigma^2(x)$ in (4), we use the local linear smoother studied by Ruppert, Wand, Holst, and Hössjer (1997) and Fan and Yao (1998). The local linear technique is chosen for its several nice properties, including asymptotic minimax efficiency and design adaptation. Furthermore, it automatically corrects edge effects and facilitates bandwidth selections (Fan and Yao 2003).

Note that the historical data used in the state-domain estimation at time t are $\{(r_{t_i}, Y_i), i = 0, \ldots, N - 1\}$. Let $\hat{f}(x) = \hat{\alpha}_1$ be the local linear estimator that solves the weighted least squares problem

$$(\hat{\alpha}_1, \hat{\alpha}_2) = \arg\min_{\alpha_1, \alpha_2} \sum_{i=0}^{N-1} [Y_i - \alpha_1 - \alpha_2(r_{t_i} - x)]^2 K_{h_1}(r_{t_i} - x),$$

where $K(\cdot)$ is a kernel function and $h_1 > 0$ is a bandwidth. Denote the squared residuals by $\hat{R}_i = \{Y_i - \hat{f}(r_{t_i})\}^2$. Then the local linear estimator of $\sigma^2(x)$ is $\hat{\sigma}_S^2(x) = \hat{\beta}_0$ given by

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{\beta_0, \beta_1} \sum_{i=0}^{N-1} \{\hat{R}_i - \beta_0 - \beta_1(r_{t_i} - x)\}^2 W_h(r_{t_i} - x), \tag{8}$$

with a kernel function W and a bandwidth h. Fan and Yao (1998) gave strategies for bandwidth selection. Stanton (1997) and Fan and Zhang (2003) showed that Y_i^2 instead of $\hat{R}_i$ in (8) also can be used for the estimation of $\sigma^2(x)$.

The asymptotic bias and variance of $\hat{\sigma}_S^2(x)$ have been given by Fan and Zhang (2003, thm. 4). Set $v_j = \int u^j W^2(u)\,du$ for $j = 0, 1, 2$, and let $p(\cdot)$ be the invariant density function of the

Markov process $\{r_s\}$ from (3). We then have the following result.

Theorem 2. Set $s_2^2(x) = 2v_0\sigma^4(x)/p(x)$. Let x be in the interior of the support of $p(\cdot)$. Suppose that the second derivatives of $\mu(\cdot)$ and $\sigma^2(\cdot)$ exist in a neighborhood of x. Under conditions (C3)–(C6), $\sqrt{Nh}[\hat{\sigma}_S^2(x) - \sigma^2(x)]/s_2(x) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$.

3. DYNAMIC INTEGRATION OF TIME– AND STATE–DOMAIN ESTIMATORS

In this section we first show how the optimal dynamic weights in (2) can be estimated, and then prove that the time-domain and state-domain estimators are indeed asymptotically independent.

3.1 Estimation of Dynamic Weights

For the exponential smoothing estimator in (7), we can apply the asymptotic formula given in Theorem 1 to get an estimate of its asymptotic variance. But because the estimator is a weighted average of Y_{t-i}^2, we also can obtain its variance directly by assuming that $Y_{t-j} \sim \mathrm{N}(0, \sigma_t^2)$ for small j. Indeed, with the foregoing local homogeneous model, we have

$$\mathrm{var}(\hat{\sigma}_{ES,t}^2)$$

$$\approx \frac{(1-\lambda)^2}{(1-\lambda^n)^2} 2\sigma_t^4 \sum_{i=1}^n \sum_{j=1}^n \lambda^{i+j-2} \rho_t(|i-j|)$$

$$= \frac{2(1-\lambda)^2 \sigma_t^4}{(1-\lambda^n)^2} \left\{ n + 2 \sum_{k=1}^{n-1} \frac{\rho_t(k)\lambda^k(1-\lambda^{2(n-k)})}{1-\lambda^2} \right\}, \quad (9)$$

where $\rho_t(k) = \mathrm{cor}(Y_t^2, Y_{t-k}^2)$ is the autocorrelation of the series $\{Y_{t-k}^2\}$. The autocorrelation can be estimated from the data in history. Note that due to the locality of the exponential smoothing, only $\rho_t(k)$'s with the first 30 lags, say, contribute to the variance calculation.

We now turn to estimate the variance of $\hat{\sigma}_{S,t}^2 = \hat{\sigma}_S^2(r_t)$. Details of this have been given by Fan and Yao (1998, 2003, sec. 6.2). Let

$$V_j(x) = \sum_{i=1}^{N-1} (r_{t_i} - x)^j W\big((r_{t_i} - x)/h_1\big) \quad (10)$$

and $\xi_i(x) = W(\frac{r_{t_i}-x}{h_1})\{V_2(x) - (r_{t_i} - x)V_1(x)\}/\{V_0(x)V_2(x) - V_1(x)^2\}$. Then the local linear estimator can be expressed as $\hat{\sigma}_S^2(x) = \sum_{i=1}^{N-1} \xi_i(x)\hat{R}_i$, and its variance can be approximated as

$$\mathrm{var}(\hat{\sigma}_S^2(x)) \approx 2\sigma^4(x) \sum_{i=1}^{N-1} \xi_i^2(x). \quad (11)$$

Substituting (9) and (11) into (2), we propose to combine the time-domain and state-domain estimators with the dynamic weight

$$\hat{w}_t = \frac{\hat{\sigma}_{S,t}^4 \sum_{i=1}^{N-1} \xi_i^2(r_t)}{\hat{\sigma}_{S,t}^4 \sum_{i=1}^{N-1} \xi_i^2(r_t) + c_t \hat{\sigma}_{ES,t}^4}, \quad (12)$$

where $c_t = \frac{(1-\lambda)^2}{(1-\lambda^n)^2}\{n + 2\sum_{k=1}^{n-1} \rho_t(k)\lambda^k(1-\lambda^{2(n-k)})/(1-\lambda^2)\}$. For practical implementation, we truncate the series $\{\rho_t(k)\}_{k=1}^{t-1}$

in the summation as $\{\rho_t(k)\}_{k=1}^{30}$. This results in the dynamically integrated estimator

$$\hat{\sigma}_{I,t}^2 = \hat{w}_t\hat{\sigma}_{ES,t}^2 + (1 - \hat{w}_t)\hat{\sigma}_{S,t}^2, \quad (13)$$

where $\hat{\sigma}_{S,t}^2 = \hat{\sigma}_S^2(r_t)$. The function $\hat{\sigma}_S^2(\cdot)$ uses historical data up to the time t, and we need to update this function as time evolves. Fortunately, we need know only the function value at the point r_t, which significantly reduces the computational cost. The computational cost can be further reduced if we update the estimated function $\hat{\sigma}_{S,t}^2$ on a prescribed time schedule (e.g., once every 2 months for weekly data).

Finally, we note that in the choice of weight, only the variance of the estimated volatility is considered, not the mean squared error (MSE). This is mainly to facilitate the choice of dynamic weights. Because the smoothing parameters in $\hat{\sigma}_{ES,t}^2$ and $\hat{\sigma}_S^2(x)$ have been tuned to optimize their performance separately, their bias and variance trade-offs have been considered indirectly. Thus, controlling the variance of the integrated estimator $\hat{\sigma}_{I,t}^2$ also controls, to some extent, the bias of the estimator.

3.2 Sampling Properties

The fundamental component in the choice of dynamic weights is the asymptotic independence between the time- and state-domain estimators. The following theorem shows that the time- and state-domain estimators are indeed asymptotically independent. To facilitate the expression of notations, we present the result at the current time t_N.

Theorem 3. Suppose that the second derivatives of $\mu(\cdot)$ and $\sigma^2(\cdot)$ exist in a neighborhood of r_{t_N}. Under conditions (C1) and (C3)–(C6), if condition (C2) holds at time point $t = t_N$, then

(a) Asymptotic independence:

$$\left[\frac{\sqrt{n}(\hat{\sigma}_{ES,t_N}^2 - \sigma_{t_N}^2)}{s_{1,t_N}}, \frac{\sqrt{Nh}(\hat{\sigma}_{S,t_N}^2 - \sigma_{t_N}^2)}{s_{2,t_N}} \right]^T \xrightarrow{\mathcal{D}} \mathcal{N}(0, I_2),$$

where s_{1,t_N} and $s_{2,t_N} = s_2(r_{t_N})$ are as given in Theorems 1 and 2.

(b) Asymptotic normality of $\hat{\sigma}_{t_N}^2$ with w_{t_N} in (2): If the limit $d = \lim_{N\to\infty} n/[Nh]$ exists, then $\sqrt{Nh/\omega}[\hat{\sigma}_{t_N}^2 - \sigma_{t_N}^2] \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$, where $\omega = w_{t_N}^2 s_{1,t_N}^2/d + (1 - w_{t_N})^2 s_{2,t_N}^2$.

(c) Asymptotic normality of $\hat{\sigma}_{I,t_N}^2$ with an estimated weight $\hat{w}_{t_N}$: If the weight $\hat{w}_{t_N}$ converges to the weight w_{t_N} in probability and the limit $d = \lim_{N\to\infty} n/[Nh]$ exists, then $\sqrt{Nh/\omega} \times (\hat{\sigma}_{I,t_N}^2 - \sigma_{t_N}^2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$.

Because the estimators $\hat{\sigma}_{S,t}^2$ and $\hat{\sigma}_{ES,t}^2$ are consistent,

$$\hat{w}_t \approx \frac{\sum_{i=1}^{t-1} \xi_i^2(r_t)}{\sum_{i=1}^{t-1} \xi_i^2(r_t) + c_t} \approx \frac{\mathrm{var}(\hat{\sigma}_S^2(r_t))}{\mathrm{var}(\hat{\sigma}_S^2(r_t)) + \mathrm{var}(\hat{\sigma}_{ES,t}^2(r_t))} = w_t,$$

and hence the condition in Theorem 3(c) holds. Note that the theoretically optimal weight minimizing the variance in Theorem 3(b) is

$$w_{t_N,\mathrm{opt}} = \frac{ds_{2,t_N}^2}{s_{1,t_N}^2 + ds_{2,t_N}^2} \approx \frac{\mathrm{var}(\hat{\sigma}_S^2(r_t))}{\mathrm{var}(\hat{\sigma}_S^2(r_t)) + \mathrm{var}(\hat{\sigma}_{ES,t}^2(r_t))} \approx w_{t_N}.$$

It follows that the integrated estimator $\hat{\sigma}^2_{I,t_N}$ is optimal in the sense that it achieves the minimum variance among all of the weighted estimators in (1).

When $0 < d < \infty$ in Theorem 3, the effective sample sizes in both time and state domains are comparable; thus neither the time-domain estimator nor the state-domain estimator dominates. From Theorem 3, based on the optimal weight, the asymptotic relative efficiencies of $\hat{\sigma}^2_{I,t_N}$ with respect to $\hat{\sigma}^2_{S,t_N}$ and $\hat{\sigma}^2_{ES,t_N}$ are

$$\text{eff}(\hat{\sigma}^2_{I,t_N}, \hat{\sigma}^2_{S,t_N}) = 1 + ds^2_{2,t_N}/s^2_{1,t_N} \qquad \text{and}$$

$$\text{eff}(\hat{\sigma}^2_{I,t_N}, \hat{\sigma}^2_{ES,t_N}) = 1 + s^2_{1,t_N}/(ds^2_{2,t_N}),$$

which are > 1. This demonstrates that the integrated estimator $\hat{\sigma}^2_{I,t_N}$ is more efficient than the time-domain and state-domain estimators.

4. BAYESIAN INTEGRATION OF VOLATILITY ESTIMATES

Another possible approach is to consider the historical information as the prior and to incorporate it into the estimation of volatility using the Bayesian framework. We now explore such an approach.

4.1 Bayesian Estimation of Volatility

The Bayesian approach is to consider the recent data $Y_{t-n}, \ldots, Y_{t-1}$ as an independent sample from N$(0, \sigma^2)$ [see (4)] and to consider the historical information being summarized in a prior. To incorporate the historical information, we assume that the variance σ^2 follows an inverse-gamma distribution with parameters a and b, which has the density function

$$f(\sigma^2) = b^a \Gamma^{-1}(a)\{\sigma^2\}^{-(a+1)}\exp(-b/\sigma^2).$$

Write $\sigma^2 \sim \text{IG}(a, b)$. It is well known that

$$\text{E}(\sigma^2) = \frac{b}{a-1},$$

$$\text{var}(\sigma^2) = \frac{b^2}{(a-1)^2(a-2)}, \qquad \text{and} \qquad (14)$$

$$\text{mode}(\sigma^2) = \frac{b}{a+1}.$$

The hyperparameters a and b are estimated from historical data using the state-domain estimators.

It can be easily shown that given $\mathbf{Y} = (Y_{t-n}, \ldots, Y_{t-1})$, the posterior density of σ^2 is IG(a^*, b^*), where $a^* = a + \frac{n}{2}$, and $b^* = \frac{1}{2}\sum_{i=1}^n Y^2_{t-i} + b$. From (14), the Bayesian mean of σ^2 is

$$\hat{\sigma}^2 = \frac{b^*}{a^* - 1} = \sum_{i=1}^n \frac{Y^2_{t-i} + 2b}{2(a-1) + n}.$$

This Bayesian estimator can be easily written as

$$\hat{\sigma}^2_B = \frac{n}{n + 2(a-1)}\hat{\sigma}^2_{MA,t} + \frac{2(a-1)}{n + 2(a-1)}\hat{\sigma}^2_P, \qquad (15)$$

where $\hat{\sigma}^2_{MA,t}$ is the moving average estimator given by (5) and $\hat{\sigma}^2_P = b/(a-1)$ is the prior mean, determined from the historical data. This combines the estimate based on the data and prior knowledge.

The Bayesian estimator (15) uses the local average of n data points. To incorporate the exponential smoothing estimator (6), we consider it the local average of $n^* = \sum_{i=1}^n \lambda^{i-1} = \frac{1-\lambda^n}{1-\lambda}$ data points. This leads to the following integrated estimator:

$$\hat{\sigma}^2_{B,t} = \frac{n^*}{n^* + 2(a-1)}\hat{\sigma}^2_{ES,t} + \frac{2(a-1)}{2(a-1) + n^*}\hat{\sigma}^2_P$$

$$= \frac{1-\lambda^n}{1-\lambda^n + 2(a-1)(1-\lambda)}\hat{\sigma}^2_{ES,t}$$

$$+ \frac{2(a-1)(1-\lambda)}{1-\lambda^n + 2(a-1)(1-\lambda)}\hat{\sigma}^2_P. \qquad (16)$$

In particular, when $\lambda \to 1$, the estimator (16) reduces to (15).

4.2 Estimation of Prior Parameters

A reasonable source for the prior information in (16) is the historical data up to time t. Thus, the hyperparameters a and b should depend on t and can be used to match with the moments from the historical information. Using the approximation model (4), we have

$$E\big[(Y_t - \hat{f}(r_t))^2 \mid r_t\big] \approx \sigma^2(r_t) \qquad \text{and}$$

$$\text{var}\big[(Y_t - \hat{f}(r_t))^2 \mid r_t\big] \approx 2\sigma^4(r_t). \qquad (17)$$

These can be estimated from the historical data up to time t, namely the state-domain estimator $\hat{\sigma}^2_S(r_t)$. Because we have assumed that the prior distribution for σ^2_t is IG(a_t, b_t), by (17) and the method of moments, we would get the following estimation equations by matching the moments from the prior distribution with the moment estimation from the historical estimate:

$$E(\sigma^2_t) = \hat{\sigma}^2_S(r_t) \qquad \text{and} \qquad \text{var}(\sigma^2_t) = 2\hat{\sigma}^4_S(r_t).$$

This, together with (14), leads to

$$\frac{b_t}{a_t - 1} = \hat{\sigma}^2_S(r_t) \qquad \text{and} \qquad \frac{b^2_t}{(a_t - 1)^2(a_t - 2)} = 2\hat{\sigma}^4_S(r_t),$$

where E represents from the expectation with respect to the prior. Solving the foregoing equations, we obtain

$$\hat{a}_t = 2.5 \qquad \text{and} \qquad \hat{b}_t = 1.5\hat{\sigma}^2_S(r_t).$$

Substituting this into (16), we obtain the estimator

$$\hat{\sigma}^2_{B,t} = \frac{1-\lambda^n}{1-\lambda^n + 3(1-\lambda)}\hat{\sigma}^2_{ES,t} + \frac{3(1-\lambda)}{1-\lambda^n + 3(1-\lambda)}\hat{\sigma}^2_{S,t}. \qquad (18)$$

Unfortunately, the weights in (18) are static and do not depend on the time t. Thus, the Bayesian method that we use does not produce a satisfactory answer to this problem. However, other implementations of the Bayesian method may yield the dynamic weights.

5. SIMULATION STUDIES AND EXAMPLES

To facilitate the presentation, we use the simple abbreviations in Table 1 to denote seven volatility estimation methods. Details on the first three methods have been given by Fan and Gu (2003). In particular, the first method is to estimate volatility using the standard deviation of the yields in the past year, and the RiskMetrics method is based on the exponential smoothing with $\lambda = .94$. The semiparametric method of Fan and Gu (2003) is an extension of a local model used in the exponential

Table 1. Seven Volatility Estimators

Hist: The historical method
RiskM: The RiskMetrics method of J. P. Morgan
Semi: The semiparametric estimator (SEV) of Fan and Gu (2003)
NonBay: The nonparametric Bayesian method in (18)
Integ: The integration method of time and state domains in (13)
GARCH: The maximum likelihood method based on a GARCH(1, 1) model
StaDo: The state-domain estimation method in Section 2.2

smoothing, with the smoothing parameter determined by minimizing the prediction error. It includes exponential smoothing with λ selected by data as a specific example. For our integrated methods, the parameter λ is taken as .94.

The following five measures are used to assess the performance of different procedures for estimating the volatility. Other related measures also can be used (see Davé and Stahl 1997).

Measure 1: Exceedence Ratio Against Confidence Level. This measure counts the number of the events for which the loss of an asset exceeds the loss predicted by the normal model at a given confidence α. It was given by Fan and Gu (2003) and computed as

$$\mathrm{ER}(\hat{\sigma}_t^2) = m^{-1} \sum_{i=T+1}^{T+m} I(Y_i < \Phi^{-1}(\alpha)\hat{\sigma}_i), \qquad (19)$$

where $\Phi^{-1}(\alpha)$ is the α-quantile of the standard normal distribution and m is the size of the out-sample. This gives an indication of how effectively the volatility estimator can be used for prediction.

As noted by Fan and Gu (2003), one shortcoming of the measure is its large Monte Carlo error. Unless the postsample size m is sufficiently large, this measure has difficulty by differentiating the performance of various estimators due to the presence of large error margins. Note that the ER depends strongly on the assumption of normality. In our simulation study, we use the true α-quantile of the error distribution instead of $\Phi^{-1}(\alpha)$ in (19) to compute the ER. For real data analysis, we use the α-quantile of the last 250 residuals for the in-sample data.

Measure 2: Mean Absolute Deviation Error. To motivate this measure, we first consider the MSEs

$$\mathrm{PE}(\hat{\sigma}_t^2) = m^{-1} \sum_{i=T+1}^{T+m} (Y_i^2 - \hat{\sigma}_i^2)^2.$$

The expected value can be decomposed as

$$E(\mathrm{PE}) = m^{-1} \sum_{i=T+1}^{T+m} E(\sigma_i^2 - \hat{\sigma}_i^2)^2 + m^{-1} \sum_{i=T+1}^{T+m} E(Y_i^2 - \sigma_i^2)^2$$

$$+ m^{-1} \sum_{i=T+1}^{T+m} E(Y_i^2 - \sigma_i^2)(\sigma_i^2 - \hat{\sigma}_i^2). \quad (20)$$

Because $\hat{\sigma}_i^2$ is predicable at time t_i and $Y_i^2 - \sigma_i^2$ is a martingale difference, the third term is 0 [the drift is assumed to be negligible in the forgoing formulation of PE; otherwise, the drift term should be removed first. In both cases, it has an affection $O(\Delta^2)$]. Therefore, we need consider only the first two

terms. The first term reflects the effectiveness of the estimated volatility, whereas the second term is the size of the stochastic error and independent of estimators. As in all statistical prediction problems, the second term is usually of an order of magnitude larger than the first term. Thus a small improvement in PE could mean substantial improvement over the estimated volatility. However, due to the well-known fact that financial time series contain outliers, the MSE is not a robust measure. Therefore, we used the mean absolute deviation error (MADE), $MADE(\hat{\sigma}_t^2) = m^{-1} \sum_{i=T+1}^{T+m} |Y_i^2 - \hat{\sigma}_i^2|$.

Measure 3: Square-Root Absolute Deviation Error. An alternative to MADE is the Square-root absolute deviation error (RADE), defined as

$$RADE(\hat{\sigma}_t^2) = m^{-1} \sum_{i=T+1}^{T+m} \big| |Y_i| - \sqrt{2/\pi}\hat{\sigma}_i \big|.$$

The constant factor comes from the fact that $E|\varepsilon_t| = \sqrt{2/\pi}$ for $\varepsilon_t \sim N(0, 1)$. If the underlying error distribution deviates from normality, then this measure is not robust.

Measure 4: Ideal Mean Absolute Deviation Error and Others. To assess the estimation of the volatility in simulations, we also can use the ideal mean absolute deviation error (IMADE),

$$IMADE = m^{-1} \sum_{i=T+1}^{T+m} |\hat{\sigma}_i^2 - \sigma_i^2|,$$

and the ideal square root absolute deviation error (IRADE),

$$IRADE = m^{-1} \sum_{i=T+1}^{T+m} |\hat{\sigma}_i - \sigma_i|.$$

Another intuitive measure is the relative IMADE (RIMADE),

$$RIMADE = m^{-1} \sum_{i=T+1}^{T+m} \frac{|\hat{\sigma}_i^2 - \sigma_i^2|}{\sigma_i^2}.$$

This measure may create some outliers when the true volatility is quite small at certain time points. For real data analysis, both measures are not applicable.

5.1 Simulations

To assess the performance of the five estimation methods listed in Table 1, we compute the average and standard deviation of each of the four measures over 600 simulations. Generally speaking, the smaller the average (or the standard deviation), the better the estimation approach. We also compute the "score" of an estimator, which is the percentage of times among 600 simulations that the estimator outperforms the average of the 7 methods in terms of an effectiveness measure. To be more specific, for example, consider RiskMetrics using MADE as an effectiveness measure. Let m_i be the MADE of the RiskMetrics estimator at the ith simulation, and let $\bar{m}_i$ be the average of the MADEs for the five estimators at the ith simulation. Then the "score" of the RiskMetrics approach in terms of the MADE is defined as $\frac{1}{600} \sum_{i=1}^{600} I(m_i < \bar{m}_i)$. Obviously, estimators with higher scores are preferred. In addition, we define the "relative loss" of an estimator $\hat{\sigma}_t^2$ relative to $\hat{\sigma}_{I,t}^2$, in terms of MADEs, as

$$RLOSS(\hat{\sigma}_t^2, \hat{\sigma}_{I,t}^2) = \overline{MADE}(\hat{\sigma}_t^2)/\overline{MADE}(\hat{\sigma}_{I,t}^2) - 1,$$

where $\overline{MADE}(\hat{\sigma}_t^2)$ is the average of $MADE(\hat{\sigma}_t^2)$ among simulations.

Example 1. In this example we compare the proposed methods with the maximum likelihood estimation approach for a parametric generalized autoregressive conditional heteroscedasticity [GARCH(1, 1)] model. This will give an assessment of how well our methods perform compared with the most efficient estimation method.

Because the GARCH model does not have a state variable, r_{t_i}, but the diffusion model does, the diffusion model cannot be used to fit the data simulated from the GARCH model. A sensible approach is to use the diffusion limit of the GARCH(1, 1) model to generate data. There are several interesting works on reconciling the two modeling approaches along this line; for example, Nelson (1990) established the continuous-time diffusion limit for the discrete ARCH model, Duan (1997) proposed an augmented GARCH model to unify various parametric GARCH models and derived its diffusion limit, and Wang (2002) studied the asymptotic relationship between GARCH models and diffusions. Here we use the result on the diffusion limit of the GARCH(1, 1) model of Wang (2002).

Specifically, we first generate data from the GARCH(1, 1) model,

$$\begin{cases} Y_t = \sigma_t \varepsilon_t \\ \sigma_t^2 = c_0 + a_0 \sigma_{t-1}^2 + b_0 Y_{t-1}^2, \end{cases}$$

where the ε_t's are from the standard normal distribution and the true parameters $c_0 = 4.98 \times 10^{-7}$, $a_0 = .9289615$, and $b_0 = .0411574$ are from the fitted values of the parameters for the daily exchange rate of the Australian dollar with the U.S. dollar from January 3, 1994 to December 29, 2000. The GARCH(1, 1) model is then fitted by maximum likelihood estimation.

Second, we use the diffusion limit (see Wang 2002, sec. 2.3) of the foregoing GARCH model,

$$\begin{cases} dr_t = \sigma_t \, dW_{1t}, \\ d\sigma_t^2 = (\beta_0 + \beta_1 \sigma_t^2) \, dt + \beta_2 \sigma_t^2 \, dW_{2t}, \end{cases}$$

where $W_{1,t}$ and $W_{2,t}$ are two independent standard Wiener process and the parameters are computed as $(\beta_0, \beta_1, \beta_2) = (2.5896 \times 10^{-5}, -1.5538, .4197)$. To simulate the diffusion process σ_t^2, we use the discrete-time order 1.0 strong approximation scheme of Kloeden, Platen, Schurz, and Sørensen (1996).

For each of the two models, we generate 600 series of daily data, each with length 1,200. For each simulation, we set the first 900 observations as the "in-sample" data and the last 300 observations as the "out-sample" data. The results, summarized in Table 2, show that all estimators have acceptable ER values close to .05. The integrated estimator has minimum RIMADE, IMADE, MADE, IRADE, and RADE on average among all of the nonparametric estimators. Compared with the most efficient maximum likelihood estimator (MLE) of the GARCH model, the integrated estimator has smaller MADE but larger RIMADE, IMADE, and IRADE. The RADEs of the MLE and the integrated estimator are very close, demonstrating that the integrated estimator is very impressive in forecasting the volatility of this example.

Note that the dynamic weights for the integrated estimator are estimated by minimizing the variance of the combined estimator in (1). As one anonymous referee pointed out, one may choose to minimize the MSE instead, which seems to be difficult and computationally intensive in current situations. As argued in Section 3.1, the biases and variances trade-off have been considered indirectly in the time-domain and state-domain estimation methods. The bias of the integrated estimator was controlled before the weights were chosen.

To investigate whether the foregoing intuition is correct, here we opt for the suggestion from the referee to visually display how the integrated estimator dominates the state-domain/time-domain estimators over time through the dynamic weights, and compare the biases of the state-domain estimator and the integrated estimator. The weights and biases are shown in Figure 2.

Table 2. Comparisons of Several Volatility Estimation Methods

Measure	Empirical formula	Hist	RiskM	Semi	NonBay	Integ	GARCH	StaDo
RIMADE	Score (%)	40.73	28.05	64.94	50.92	73.79	76.29	40.90
	Average	.205	.204	.183	.187	.178	.169	.204
	Standard ($\times 10^{-2}$)	6.55	3.31	4.50	3.08	3.97	3.67	5.67
	Relative loss (%)	15.44	14.44	2.66	4.97	0	−4.76	14.75
IMADE	Score (%)	61.77	25.54	53.42	46.08	72.12	75.46	40.90
	Average ($\times 10^{-5}$)	.332	.367	.330	.336	.318	.304	.362
	Standard ($\times 10^{-6}$)	1.118	.744	.877	.690	.851	.800	1.113
	Relative loss (%)	4.40	15.44	3.73	5.81	0	−4.47	13.90
MADE	Score (%)	28.21	49.58	49.75	56.59	88.15	57.43	79.63
	Average ($\times 10^{-4}$)	.164	.153	.153	.152	.149	.152	.149
	Standard ($\times 10^{-6}$)	2.314	2.182	2.049	2.105	1.869	1.852	1.876
	Relative loss (%)	10.16	2.84	2.93	2.09	0	2.07	.11
IRADE	Score (%)	62.77	24.04	51.09	47.08	70.62	77.13	40.40
	Average ($\times 10^{-3}$)	.403	.448	.407	.410	.391	.369	.446
	Standard ($\times 10^{-3}$)	.121	.075	.092	.070	.093	.085	.131
	Relative loss (%)	3.17	14.75	4.11	4.98	0	−5.53	14.22
RADE	Score (%)	30.55	50.42	43.41	59.10	84.64	58.93	79.13
	Average ($\times 10^{-2}$)	.20	.19	.19	.19	.19	.19	.19
	Standard ($\times 10^{-4}$)	1.53	1.45	1.41	1.42	1.34	1.33	1.37
	Relative loss (%)	5.13	1.42	1.65	.96	0	1.06	−.25
ER	Average ($\times 10^{-2}$)	4.95	5.45	5.38	5.28	5.25	5.16	5.00
	Standard ($\times 10^{-2}$)	1.57	1.13	1.28	1.12	1.36	1.25	1.49
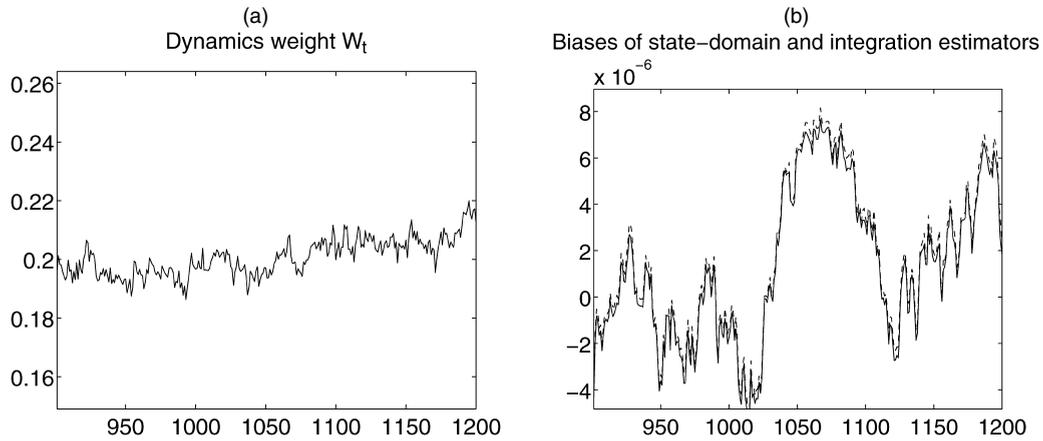
Figure 2. The Dynamic Weights for Prediction for Example 1 (a) and the Biases for Prediction for the State-Domain Estimator ($- - -$) and the Integration Estimator (—) (b).

It is seen that the bias is actually controlled, because both of the estimators have very close biases. The weights are around .2, and through the weights the integrated method improves both the RiskMetrics method and the state-domain approach.

Example 2. To simulate the interest rate data, we consider the Cox–Ingersoll–Ross (CIR) model,

$$dr_t = \kappa(\theta - r_t)\,dt + \sigma r_t^{1/2}\,dW_t, \qquad t \geq t_0,$$

where the spot rate, r_t, moves around a central location or long-run equilibrium level $\theta = .08571$ at speed $\kappa = .21459$. The σ is set at .07830. These parameters values given by Chapman and Pearson (2000), satisfy the condition $2\kappa\theta \geq \sigma^2$, so that the process r_t is stationary and positive. The model has been studied by Chapman and Pearson (2000) and Fan and Zhang (2003).

We simulated weekly data (600 runs) from the CIR model. For each simulation, we designated the first 900 observations

as the "in-sample" data and the last 300 observations the "out-sample" data. The results, summarized in Table 3, show that the integrated method performs closely to the state-domain method. This is mainly because the dynamic weights shown in Figure 3 are very small and the integration estimator is very close to the state-domain estimator, supporting our statement at the end of the third paragraph in Section 1. Table 3 shows that the integrated estimator uniformly dominates the other estimators on average due to its highest score and lowest averaged RIMADE, IMADE, MADE, IRADE, and RADE. The improvement in RIMADE, IMADE, and IRADE is about 100%. This demonstrates that our integrated volatility method better captures the volatility dynamics. The historical simulation method performs poorly due to misspecification of the function of the volatility parameter. The results here show the advantage of aggregating the information of the time and state domains. Note that all estimators have reasonable ER values at level .05, and that the ER value of the integrated estimator is closest to .05.

Table 3. Comparisons of Several Volatility Estimation Methods

Measure	Empirical formula	Hist	RiskM	Semi	NonBay	Integ	GARCH	StaDo
RIMADE	Score (%)	5.01	18.20	21.04	36.56	99.33	24.04	99.17
	Average ($\times 10^{-1}$)	2.67	1.90	1.88	1.64	.61	2.02	.62
	Standard ($\times 10^{-1}$)	1.48	.35	.81	.31	.39	1.30	.43
	Relative loss (%)	337.32	210.74	207.98	168.21	0	230.12	1.12
IMADE	Score (%)	7.68	13.19	14.69	29.72	99.67	21.20	99.50
	Average ($\times 10^{-6}$)	2.37	2.08	1.91	1.79	.64	1.96	.63
	Standard ($\times 10^{-6}$)	1.00	.78	.69	.68	.45	.91	.46
	Relative loss (%)	271.62	225.58	199.05	181.07	0	206.96	−1.63
MADE	Score (%)	38.23	52.09	57.76	56.09	71.45	53.92	69.45
	Average ($\times 10^{-5}$)	1.02	.93	.93	.93	.91	.94	.92
	Standard ($\times 10^{-6}$)	3.49	3.33	3.21	3.30	3.17	3.10	3.17
	Relative loss (%)	11.00	2.12	2.21	1.62	0	2.29	.21
IRADE	Score (%)	8.18	12.52	14.02	30.72	99.83	21.04	99.50
	Average ($\times 10^{-4}$)	3.76	3.22	3.07	2.77	.99	3.16	.98
	Standard ($\times 10^{-4}$)	1.40	.75	.90	.66	.60	1.92	.62
	Relative loss (%)	278.55	223.97	208.98	178.92	0	218.21	−1.11
RADE	Score (%)	40.40	52.09	50.42	55.93	73.79	48.75	72.79
	Average ($\times 10^{-2}$)	.15	.15	.15	.15	.15	.15	.15
	Standard ($\times 10^{-4}$)	2.72	2.67	2.57	2.65	2.58	2.51	2.57
	Relative loss (%)	6.08	1.23	1.60	.92	0	1.89	.06
ER	Average	.056	.055	.054	.053	.049	.053	.049
	Standard	.023	.011	.013	.011	.013	.036	.013
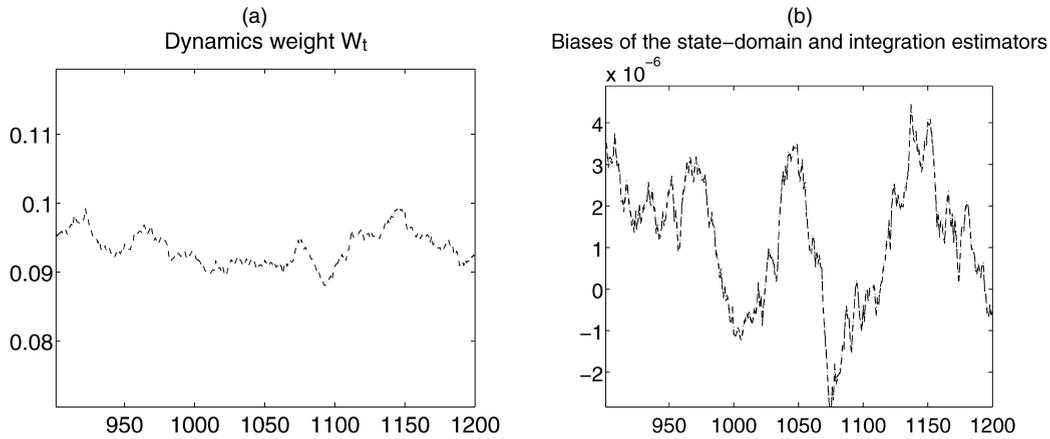
Figure 3. The Dynamic Weights for Prediction for Example 2 (a) and the Biases for Prediction for the State-Domain Estimator (− − −) and for the Integration Estimator (—) (b).

Again, Figure 3(b) displays the biases for the state domain and integration estimators. It can be seen that the state-domain and integration estimators have almost overlaid biases, demonstrating that the bias of the integrated estimator is taken care of even though we choose the dynamic weights by minimizing the variance.

Example 3. To assess the performance of the proposed methods under a nonstationary situation, we now consider the geometric Brownian (GBM),

$$dr_t = \mu r_t \, dt + \sigma r_t \, dW_t,$$

where W_t is a standard one-dimensional Brownian motion. This is a nonstationary process against which we check whether our method continues to apply. Note that the celebrated Black–Scholes option price formula is derived from Osborne's assumption that the stock price follows the GBM model. By Itô's formula, we have

$$\log r_t - \log r_0 = (\mu - \sigma^2/2)t + \sigma^2 W_t.$$

We set $\mu = .03$ and $\sigma = .26$ in our simulations. With the Brownian motion simulated from independent Gaussian increments, we can generate the samples for the GBM. We simulate 600 times with $\Delta = 1/252$. For each simulation, we generate 1,000 observations and use the first two-thirds of the observations as in-sample data and the remaining observations as out-sample data.

Because of the nonstationarity of the process, the simulation results are somewhat unstable with a few uncommon realizations appearing. Therefore, we use a robust method to summarize the results. For each measure, we compute the relative loss and the median, along with the median absolute deviation (MAD) for scale estimation. The MAD for sample $\{a_i\}_{i=1}^s$ is defined as $MAD(a) = median\{|a_i - median(a)|, i = 1, \ldots, s\}$. It is easy to see that $MAD(a)$ provides a robust estimator of the standard deviation of $\{a_i\}$. Table 4 reports the results. The table shows that the integrated estimator performs quite well. All estimators have acceptable ER values. The proposed esti-

Table 4. Robust Comparisons of Several Volatility Estimation Methods

Measure	Empirical formula	Hist	RiskM	Semi	NonBay	Integ	GARCH	StaDo
RIMADE	Score (%)	7.33	43.00	35.83	72.83	61.33	27.33	52.33
	Median ($\times 10^{-1}$)	2.254	1.561	1.770	1.490	1.382	2.036	1.443
	MAD	.035	.012	.024	.012	.034	.036	.037
	Relative loss (%)	63.10	12.96	28.05	7.81	0	47.32	4.44
IMADE	Score (%)	13.33	43.33	26.33	73.33	65.17	25.50	53.00
	Median ($\times 10^{-7}$)	.217	.174	.193	.163	.146	.265	.154
	MAD ($\times 10^{-8}$)	.566	.474	.486	.445	.322	.817	.357
	Relative loss (%)	48.64	19.13	32.11	12.05	0	81.45	5.61
MADE	Score (%)	39.00	39.50	26.50	49.33	52.00	42.33	51.33
	Median ($\times 10^{-6}$)	.115	.111	.111	.110	.109	.111	.109
	MAD ($\times 10^{-7}$)	.236	.244	.240	.244	.231	.230	.230
	Relative loss (%)	5.92	1.43	1.81	1.23	0	1.74	.29
IRADE	Score (%)	14.17	43.17	25.00	73.17	64.67	26.83	53.00
	Median ($\times 10^{-4}$)	.346	.267	.310	.255	.251	.378	.266
	MAD ($\times 10^{-5}$)	.599	.406	.488	.389	.566	.847	.644
	Relative loss (%)	37.87	6.49	23.58	1.49	0	50.37	5.91
RADE	Score (%)	39.00	37.00	24.33	51.17	61.50	29.50	57.50
	Median ($\times 10^{-4}$)	1.581	1.509	1.515	1.508	1.504	1.661	1.503
	MAD ($\times 10^{-4}$)	.185	.201	.198	.200	.193	.257	.192
	Relative loss (%)	5.09	.30	.73	.23	0	10.42	−.07
ER	Median	.054	.051	.054	.051	.048	.057	.0480
	MAD	.010	.003	.004	.003	.006	.007	.006
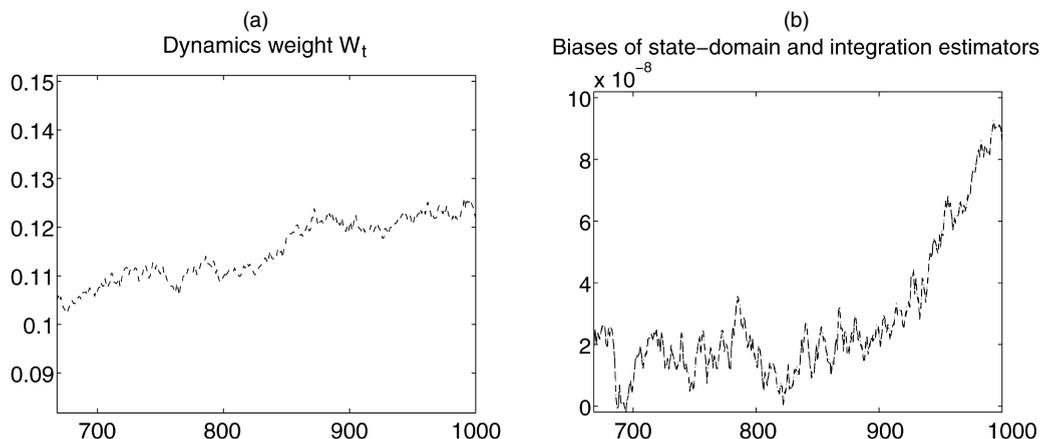
Figure 4. The Dynamic Weights for Prediction for Example 3 (a) and the Biases for Prediction; for the State-Domain Estimator (− − −) and the Integration Estimator (—) (b).

mators have the smallest measure values in median or robust scale. This shows that our integrated method continues to perform better than the others for this nonstationary case.

Again, Figure 4(b) displays the biases for the state domain and integration estimators. It can be seen that the state-domain and integration estimators have almost overlaid biases, demonstrating that the bias of the integrated estimator is taken care of even though we choose the dynamic weights by minimizing the variance.

5.2 Real Examples

In this section we apply the integrated volatility estimation methods and other methods to the analysis of real financial data.

5.2.1 Treasury Bond. Here we consider the weekly returns of three Treasury Bonds with terms of 1, 5, and 10 years. The in-sample data are for January 4, 1974–December 30, 1994, and out-sample data are for January 6, 1995–August 8, 2003. The total sample size is 1,545, and the in-sample size is 1,096. The results are reported in Table 5.

From Table 5, for the 1-year Treasury Notes, the integrated estimator is of the smallest MADE and the smallest RADE, reflecting that the integrated estimation method of the volatility is the best among the seven methods. For the bonds with 5- or 10-year maturities, the seven estimators have close MADEs and RADEs, where the historical simulation method is better than the RiskMetrics in terms of MADE and RADE and the integrated estimation approach has the smallest MADEs. In addition, the integrated estimator has ER values closest to

.05. This demonstrates the advantage of using state-domain information, which can improve the time-domain prediction of the volatilities in the interest dynamics. The nonparametric Bayesian method does not perform as well as the integrated estimator because it uses fixed weights and the inverse-gamma prior, which may not be true in practice.

5.2.2 Exchange Rate. We analyze the daily exchange rate of several foreign currencies against the U.S. dollar. The data are for January 3, 1994–August 1, 2003. The in-sample data consist of the observations before January 1, 2001, and the out-sample data consist of the remaining observations. The results, reported in Table 6, show that the integrated estimator has the smallest MADEs and moderate RADEs for the exchange rates. Note that the RADE measure is effective only when the data seem to be from a normal distribution and thus cannot calibrate the accuracy of volatility forecast for data from an unknown distribution. For this reason, the RADE might not be a good measure of performance for this example. Both the integrated volatility estimation and GARCH perform nicely for this example. They outperform other nonparametric methods.

Figure 5 reports the dynamic weights for the daily exchange rate of the U.K. pound to the U.S. dollar. Because the RiskM estimator is far more informative than the StaDo estimator, our integrated estimator becomes basically the time-domain estimator by automatically choosing large dynamic weights. This exemplifies our statement in the paragraph immediately before (1) in Section 1.

Table 5. Comparisons of Several Volatility Estimation Methods

Term	Measure ($\times 10^{-2}$)	Hist	RiskM	Semi	NonBay	Integ	GARCH	StaDo
1 year	MADE	1.044	.787	.787	.794	.732	.893	.914
	RADE	5.257	4.231	4.256	4.225	4.105	4.402	4.551
	ER	2.237	2.009	2.013	1.562	3.795	0	7.366
5 years	MADE	1.207	1.253	1.296	1.278	1.201	1.245	1.638
	RADE	5.315	5.494	5.630	5.563	5.571	5.285	6.543
	ER	.671	1.339	1.566	1.112	5.804	0	8.929
10 years	MADE	1.041	1.093	1.103	1.111	1.018	1.046	1.366
	RADE	4.939	5.235	5.296	5.280	5.152	4.936	6.008
	ER	1.112	1.563	1.790	1.334	4.911	0	6.920

Table 6. Comparison of Several Volatility Estimation Methods

Currency	Measure	Hist	RiskM	Semi	NonBay	Integ	GARCH	StaDo
U.K.	MADE ($\times 10^{-4}$)	.614	.519	.536	.519	.493	.506	.609
	RADE ($\times 10^{-3}$)	3.991	3.424	3.513	3.440	3.492	3.369	3.900
	ER	.009	.014	.019	.015	.039	.005	.052
Australia	MADE ($\times 10^{-4}$)	.172	.132	.135	.135	.126	.132	.164
	RADE ($\times 10^{-3}$)	1.986	1.775	1.830	1.798	1.761	1.768	2.033
	ER	.054	.025	.026	.022	.042	.018	.045
Japan	MADE ($\times 10^{-1}$)	5.554	5.232	5.444	5.439	5.064	5.084	6.987
	RADE ($\times 10^{-1}$)	3.596	3.546	3.622	3.588	3.559	3.448	4.077
	ER	.012	.011	.019	.012	.028	.003	.032

6. CONCLUSIONS

We have proposed a dynamically integrated method and a Bayesian method to aggregate the information from the time domain and the state domain. We studied the performance comparisons both empirically and theoretically. We showed that the proposed integrated method effectively aggregates the information from both the time and state domains and has advantages over some previous methods; for example, it is powerful in forecasting volatilities for the yields of bonds and for exchange rates. Our study also revealed that proper use of information from both the time domain and the state domain makes volatility forecasting more accurate. Our method exploits both the continuity in the time domain and the stationarity in the state domain, and can be applied to situations in which these two conditions hold approximately.

APPENDIX: CONDITIONS AND PROOFS OF THEOREMS

We state technical conditions for the proof of our results:

(C1) (Global Lipschitz condition.) There exists a constant $k_0 \geq 0$ such that

$$|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \leq k_0|x - y| \quad \text{for all } x, y \in \mathbb{R}.$$

(C2) Given time point $t > 0$, there exists a constant $L > 0$ such that

$$E|\mu(r_s)|^{4(p+\delta)} \leq L \quad \text{and} \quad E|\sigma(r_s)|^{4(p+\delta)} \leq L$$

for any $s \in [t - \eta, t]$, where η is some positive constant, p is a positive integer, and δ is some small positive constant.

Dynamics weight W$_t$

Figure 5. Dynamic Weights for the Daily Exchange Rate of the U.K. Pound and the U.S. Dollar.

(C3) The discrete observations $\{r_{t_i}\}_{i=0}^{N}$ satisfy the stationarity conditions of Banon (1978). Furthermore, the G_2 condition of Rosenblatt (1970) holds for the transition operator.

(C4) The conditional density $p_\ell(y|x)$ of $r_{t_i+\ell}$ given r_{t_i} is continuous in the arguments (y, x) and is bounded by a constant independent of ℓ.

(C5) The kernel W is a bounded, symmetric probability density function with compact support, say $[-1, 1]$.

(C6) $(N - n)h \to \infty$, $(N - n)h^5 \to 0$ and $(N)h\Delta \to 0$.

Condition (C1) ensures that (3) has a unique strong solution, $\{r_t, t \geq 0\}$, continuous with probability 1, if the initial value r_0 satisfies that $P(|r_0| < \infty) = 1$ (see thm. 4.6 of Liptser and Shiryayev 2001). Condition (C2) indicates that, given time point $t > 0$, there is a time interval $[t - \eta, t]$ on which the drift and the volatility have finite $4(p + \delta)$th moments, which is needed for deriving asymptotic normality of the time-domain estimate $\hat{\sigma}_{ES,t}^2$. Conditions (C3)–(C5) are similar to those of Fan and Zhang (2003). Condition (C6) is assumed for bandwidths where undersmoothing is uses to yield zero bias in the asymptotic normality of the state-domain estimate.

Throughout the proof, we denote by M a generic positive constant and use μ_s and σ_s to represent $\mu(r_s)$ and $\sigma(r_s)$.

Proposition A.1. Under conditions (C1) and (C2), we have, for almost all sample paths,

$$|\sigma_s^2 - \sigma_u^2| \leq K|s - u|^{(2p-1)/(4p)} \tag{A.1}$$

for any $s, u \in [t - \eta, t]$, where the coefficient K satisfies $E[K^{2(p+\delta)}] < \infty$.

Proof. First, we show that the process $\{r_s\}$ is locally Hölder-continuous with order $q = (p - 1)/(2p)$ and coefficient K_1, such that $E[K_1^{4(p+\delta)}] < \infty$, that is,

$$|r_s - r_u| \leq K_1|s - u|^q, \quad \forall s, u \in [t - \eta, t]. \tag{A.2}$$

In fact, by Jensen's inequality and martingale moment inequalities (Karatzas and Shreve 1991, sec. 3.3.D, p. 163), we have

$$E|r_u - r_s|^{4(p+\delta)}$$

$$\leq M\left(E\left|\int_s^u \mu_v\, dv\right|^{4(p+\delta)} + E\left|\int_s^u \sigma_v\, dW_v\right|^{4(p+\delta)}\right)$$

$$\leq M(u - s)^{4(p+\delta)-1}\int_s^u E|\mu_v|^{4(p+\delta)}\, dv$$

$$\quad + M(u - s)^{2(p+\delta)-1}\int_s^u E|\sigma_v|^{4(p+\delta)}\, dv$$

$$\leq M(u - s)^{2(p+\delta)}.$$

Then, by theorem 2.1 of Revuz and Yor (1999, p. 26),

$$E\left[\left(\sup_{t \neq s}\{|r_s - r_u|/|s - u|^\alpha\}\right)^{4(p+\delta)}\right] < \infty \tag{A.3}$$

for any $\alpha \in [0, \frac{2(p+\delta)-1}{4(p+\delta)})$. Let $\alpha = \frac{p-1}{2p}$ and $K_1 = \sup_{s \neq u}\{|r_s - r_u|/|s-u|^{(p-1)/(2p)}\}$. Then $E[K_1^{4(p+\delta)} < \infty]$, and the inequality (A.2) holds.

Second, by condition (C1), we have $|\sigma_s^2 - \sigma_u^2| \leq k_0(\sigma_s + \sigma_u)|r_s - r_u|$. This, combined with (A.2), leads to

$$|\sigma_s^2 - \sigma_u^2| \leq k_0(\sigma_s + \sigma_u)K_1|s-u|^q \equiv K|s-u|^q.$$

Then, by Hölder's inequality,

$$E[K^{2(p+\delta)}] \leq ME[(\sigma_s K_1)^{2(p+\delta)}]$$
$$\leq M\sqrt{E[\sigma_s^{4(p+\delta)}]E[K_1^{4(p+\delta)}]} < \infty.$$

Proof of Theorem 1

Let $Z_{i,s} = (r_s - r_{t_i})^2$. Applying Itô's formula to $Z_{i,s}$, we obtain

$$dZ_{i,s} = 2\left(\int_{t_i}^s \mu_u \, du + \int_{t_i}^s \sigma_u \, dW_u\right)(\mu_s \, ds + \sigma_s \, dW_s) + \sigma_s^2 \, ds$$

$$= 2\left[\left(\int_{t_i}^s \mu_u \, du + \int_{t_i}^s \sigma_u \, dW_u\right)\mu_s \, ds + \sigma_s\left(\int_{t_i}^s \mu_u \, du\right) dW_s\right]$$

$$+ 2\left(\int_{t_i}^s \sigma_u \, dW_u\right)\sigma_s \, dW_s + \sigma_s^2 \, ds.$$

Then Y_i^2 can be decomposed as $Y_i^2 = 2a_i + 2b_i + \bar{\sigma}_i^2$, where

$$a_i = \Delta^{-1}\left[\int_{t_i}^{t_{i+1}} \mu_s \, ds \int_{t_i}^s \mu_u \, du + \int_{t_i}^{t_{i+1}} \mu_s \, ds \int_{t_i}^s \sigma_u \, dW_u\right.$$

$$\left. + \int_{t_i}^{t_{i+1}} \sigma_s \, dW_s \int_{t_i}^s \mu_u \, du\right],$$

$$b_i = \Delta^{-1} \int_{t_i}^{t_{i+1}} \int_{t_i}^s \sigma_u \, dW_u \sigma_s \, dW_s,$$

and

$$\bar{\sigma}_i^2 = \Delta^{-1} \int_{t_i}^{t_{i+1}} \sigma_s^2 \, ds.$$

Therefore, $\hat{\sigma}_{ES,t}^2$ can be written as

$$\hat{\sigma}_{ES,t}^2 = 2\frac{1-\lambda}{1-\lambda^n}\sum_{i=t-n}^{t-1} \lambda^{t-i-1} a_i + 2\frac{1-\lambda}{1-\lambda^n}\sum_{i=t-n}^{t-1} \lambda^{t-i-1} b_i$$

$$+ \frac{1-\lambda}{1-\lambda^n}\sum_{i=t-n}^{t-1} \lambda^{t-i-1}\bar{\sigma}_i^2$$

$$\equiv A_{n,\Delta} + B_{n,\Delta} + C_{n,\Delta}.$$

By Proposition A.1, as $n\Delta \to 0$, $|C_{n,\Delta} - \sigma_t^2| \leq K(n\Delta)^q$, where $q = (2p-1)/(4p)$. This, combined with Lemmas A.1 and A.2, completes the proof of the theorem.

Lemma A.1. If condition (C2) is satisfied, then $E[A_{n,\Delta}^2] = O(\Delta)$.

Proof. Simple algebra gives the result. In fact,

$$E(a_i^2) \leq 3E\left[\Delta^{-1}\int_{t_i}^{t_{i+1}} \mu_s \, ds \int_{t_i}^s \mu_u \, du\right]^2$$

$$+ 3E\left[\Delta^{-1}\int_{t_i}^{t_{i+1}} \mu_s \, ds \int_{t_i}^s \sigma_u \, dW_u\right]^2$$

$$+ 3E\left[\Delta^{-1}\int_{t_i}^{t_{i+1}} \sigma_s \, dW_s \int_{t_i}^s \mu_u \, du\right]^2$$

$$\equiv I_1(\Delta) + I_2(\Delta) + I_3(\Delta).$$

Applying Jensen's inequality, we obtain that

$$I_1(\Delta) = O(\Delta^{-1})E\left[\int_{t_i}^{t_{i+1}} \int_{t_i}^s \mu_s^2\mu_u^2 \, du \, ds\right]$$

$$= O(\Delta^{-1})\int_{t_i}^{t_{i+1}} \int_{t_i}^s E(\mu_u^4 + \mu_s^4) \, du \, ds = O(\Delta).$$

By Jensen's inequality, Hölder's inequality, and martingale moment inequalities, we have

$$I_2(\Delta) = O(\Delta^{-1})\int_{t_i}^{t_{i+1}} E\left(\mu_s \int_{t_i}^s \sigma_u \, dW_u\right)^2 ds$$

$$= O(\Delta^{-1})\int_{t_i}^{t_{i+1}} \left\{E[\mu_s]^4 E\left[\int_{t_i}^{t_{i+1}} \sigma_u \, dW_u\right]^4\right\}^{1/2} ds$$

$$= O(\Delta).$$

Similarly, $I_3(\Delta) = O(\Delta)$. Therefore, $E(a_i^2) = O(\Delta)$. Then, by the Cauchy–Schwartz inequality, and noting that $n(1-\lambda) = O(1)$, we obtain that

$$E[A_{n,\Delta}^2] \leq n\left(\frac{1-\lambda}{1-\lambda^n}\right)^2 \sum_{i=1}^n \lambda^{2(n-i)} E(a_i^2) = O(\Delta).$$

Lemma A.2. Under conditions (C1) and (C2), if $n \to \infty$ and $n\Delta \to 0$, then

$$s_{1,t}^{-1}\sqrt{n}B_{n,\Delta} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{A.4}$$

Proof. Note that $b_j = \sigma_t^2\Delta^{-1}\int_{t_j}^{t_{j+1}}(W_s - W_{t_j}) \, dW_s + \epsilon_j$, where

$$\epsilon_j = \Delta^{-1}\int_{t_j}^{t_{j+1}}(\sigma_s - \sigma_t)\left[\int_{t_j}^s \sigma_u \, dW_u\right] dW_s$$

$$+ \Delta^{-1}\sigma_t\int_{t_j}^{t_{j+1}}\left[\int_{t_j}^s (\sigma_u - \sigma_t) \, dW_u\right] dW_s.$$

By the central limit theorem for martingales (see Hall and Heyde 1980, cor. 3.1), it suffices to show that

$$V_n^2 \equiv E[s_{1,t}^{-2}nB_{n,\Delta}^2] \to 1 \tag{A.5}$$

and that the following Lyapunov condition holds:

$$\sum_{i=t-n}^{t-1} E\left(\sqrt{n}\frac{1-\lambda}{1-\lambda^n}\lambda^{t-i-1}b_i\right)^4 \to 0. \tag{A.6}$$

Note that

$$\frac{\Delta^2}{2}E(\epsilon_j^2) \leq E\left\{\int_{t_j}^{t_{j+1}}(\sigma_s - \sigma_t)\left[\int_{t_j}^s \sigma_u \, dW_u\right] dW_s\right\}^2$$

$$+ \sigma_t^2 E\left\{\int_{t_j}^{t_{j+1}}\left[\int_{t_j}^s (\sigma_u - \sigma_t) \, dW_u\right] dW_s\right\}^2$$

$$\equiv L_{n1} + L_{n2}. \tag{A.7}$$

By Jensen's inequality, Hölder's inequality, and moment inequalities for martingale, we have

$$L_{n1} \leq \int_{t_j}^{t_{j+1}} E\left\{(\sigma_s - \sigma_t)^2\left[\int_{t_j}^s \sigma_u \, dW_u\right]^2\right\} ds$$

$$\leq \int_{t_j}^{t_{j+1}} \left\{E(\sigma_s - \sigma_t)^4 E\left[\int_{t_j}^s \sigma_u \, dW_u\right]^4\right\}^{1/2} ds$$

$$\leq \int_{t_j}^{t_{j+1}} \left\{E[k_0 K_1(n\Delta)^q]^4 36\Delta \int_{t_j}^s E(\sigma_u^4) \, du\right\}^{1/2} ds$$

$$\leq M(n\Delta)^{2q}\Delta^2, \tag{A.8}$$

where K_1 is defined in Proposition A.1. Similarly,

$$L_{n2} \le M(n\Delta)^{2q}\Delta^2. \qquad (A.9)$$

By (A.7), (A.8), and (A.9), we have $E(\epsilon_j^2) \le M(n\Delta)^{2q}$; therefore, $E[\sigma_t^{-4}b_j^2] = \frac{1}{2} + O((n\Delta)^q)$. By the theory of stochastic calculus, simple algebra gives that $E(b_j) = 0$ and $E(b_i b_j) = 0$ for $i \ne j$. It follows that

$$V_n^2 = E(s_{1,t}^{-2} n B_{n,\Delta}^2) = \sum_{i=t-n}^{t-1} E\left(2 s_{1,t}\sqrt{n}\frac{1-\lambda}{1-\lambda^n}\lambda^{t-i-1}b_i\right)^2 \to 1;$$

that is, (A.5) holds. For (A.6), it suffices to prove that $E(b_j^4)$ is bounded, which holds from condition (C2) and by applying the moment inequalities for martingales to b_j^4.

Proof of Theorem 2

The proof is completed along the same lines as given by Fan and Zhang (2003).

Proof of Theorem 3

By Fan and Yao (1998), the volatility estimator $\hat{\sigma}_{S,t_N}^2$ behaves as if the instantaneous return function f were known. Thus, without loss of generality, we assume that $f(x) = 0$ and hence $\hat{R}_i = Y_i^2$. Let $\mathbf{Y} = (Y_0^2, \ldots, Y_{N-1}^2)^T$ and $\mathbf{W} = \mathrm{diag}\{W_h(r_{t_0} - r_{t_N}), \ldots, W_h(r_{t_{N-1}} - r_{t_N})\}$, and let $\mathbf{X}$ be a $(N - t_0) \times 2$ matrix with each of the elements in the first column being 1 and the second column being $(r_{t_0} - r_{t_N}, \ldots, r_{t_{N-1}} - r_{t_N})'$. Let $m_i = E[Y_i^2 | r_{t_i}]$, $\mathbf{m} = (m_0, \ldots, m_{N-1})^T$, and $\mathbf{e}_1 = (1, 0)^T$. Define $\mathbf{S}_N = \mathbf{X}^T\mathbf{W}\mathbf{X}$ and $\mathbf{T}_N = \mathbf{X}^T\mathbf{W}\mathbf{Y}$. Then it can be written that $\hat{\sigma}_{S,t_N}^2 = \mathbf{e}_1^T\mathbf{S}_N^{-1}\mathbf{T}_N$ (see Fan and Yao, 2003). Hence

$$\hat{\sigma}_{S,t_N}^2 - \sigma_{t_N}^2 = \mathbf{e}_1^T\mathbf{S}_N^{-1}\mathbf{X}^T\mathbf{W}\{\mathbf{m} - \mathbf{X}\boldsymbol{\beta}_N\} + \mathbf{e}_1^T\mathbf{S}_N^{-1}\mathbf{X}^T\mathbf{W}(\mathbf{Y} - \mathbf{m})$$

$$\equiv \mathbf{e}_1^T\mathbf{b} + \mathbf{e}_1^T\mathbf{t}, \qquad (A.10)$$

where $\boldsymbol{\beta}_N = (m(r_{t_N}), m'(r_{t_N}))^T$ with $m(r_x) = E[Y_1^2 | r_{t_1} = x]$. Following Fan and Zhang (2003), the bias vector $\mathbf{b}$ converges in probability to a vector $\bar{\mathbf{b}}$ with $\bar{\mathbf{b}} = O(h^2) = o(1/\sqrt{(N)h})$. In what follows, we show that the centralized vector $\mathbf{t}$ is asymptotically normal.

In fact, write $\mathbf{u} = (N)^{-1}\mathbf{H}^{-1}\mathbf{X}^T\mathbf{W}(\mathbf{Y} - \mathbf{m})$, where $\mathbf{H} = \mathrm{diag}\{1, h\}$. Then, following Fan and Zhang (2003), the vector $\mathbf{t}$ can be written as

$$\mathbf{t} = p^{-1}(r_{t_N})\mathbf{H}^{-1}\mathbf{S}^{-1}\mathbf{u}(1 + o_p(1)), \qquad (A.11)$$

where $\mathbf{S} = (\mu_{i+j-2})_{i,j=1,2}$ with $\mu_j = \int u^j W(u)\,du$ and $p(x)$ is the invariant density of r_t defined in Section 2.2. For any constant vector $\mathbf{c}$, define

$$Q_N = \mathbf{c}^T\mathbf{u} = \frac{1}{N}\sum_{i=0}^{N-1}\{Y_i^2 - m_i\}C_h(r_{t_i} - r_{t_N}),$$

where $C_h(\cdot) = 1/hC(\cdot/h)$ with $C(x) = c_0 W(x) + c_1 x W(x)$. Applying the "big-block" and "small-block" arguments of Fan and Yao (2003, thm. 6.3, p. 238), we obtain

$$\theta^{-1}(r_{t_N})\sqrt{(N)h}Q_N \xrightarrow{\mathrm{D}} \mathrm{N}(0,1), \qquad (A.12)$$

where $\theta^2(r_{t_N}) = 2p(r_{t_N})\sigma^4(r_{t_N})\int_{-\infty}^{+\infty} C^2(u)\,du$. In what follows, we decompose Q_N into two parts, Q_N' and Q_N'', which satisfy the following:

(a) $NhE[\theta^{-1}(r_{t_N})Q_N']^2 \le (h/N)(h^{-1}a_N(1 + o(1)) + (N) \times o(h^{-1})) \to 0$.

(b) Q_N'' is identically distributed as Q_N and is asymptotically independent of $\hat{\sigma}_{ES,t_N}^2$.

Define $Q_N' = \frac{1}{N}\sum_{i=0}^{a_N}\{Y_i^2 - E[Y_i^2|r_{t_i}]\}C_h(r_{t_i} - r_{t_N})$ and $Q_N'' = Q_N - Q_N'$, where a_N is a positive integer satisfying $a_N = o(N)$ and $a_N\Delta \to \infty$. Obviously, $a_N \gg n$. Let $\vartheta_{N,i} = (Y_i^2 - m_i)C_h(r_{t_i} - r_{t_N})$. Then, following Fan and Zhang (2003),

$$\mathrm{var}[\theta^{-1}(r_{t_N})\vartheta_{N,1}] = h^{-1}(1 + o(1))$$

and

$$\sum_{\ell=1}^{N-2}|\mathrm{cov}(\vartheta_{N,1}, \vartheta_{N,\ell+1})| = o(h^{-1}),$$

which yields the result in (a). This, combined with (A.12), (a), and the definition of Q_N'' leads to

$$\theta^{-1}(r_{t_N})\sqrt{Nh}Q_N'' \xrightarrow{\mathrm{D}} \mathrm{N}(0,1). \qquad (A.13)$$

Note that the stationarity conditions of Banon (1978) and the G_2 condition of Rosenblatt (1970) on the transition operator imply that the ρ_t mixing coefficient $\rho_t(\ell)$ of $\{r_{t_i}\}$ decays exponentially, and that the strong mixing coefficient $\alpha(\ell) \le \rho_t(\ell)$, it follows that

$$\left|E\exp\{i\xi(Q_N'' + \hat{\sigma}_{ES,t_N}^2)\} - E\exp\{i\xi(Q_N'')\}E\exp\{(i\xi\hat{\sigma}_{ES,t_N}^2)\}\right|$$

$$\le 32\alpha((a_N - n)\Delta) \to 0,$$

for any $\xi \in \mathbb{R}$. Using the theorem of Volkonskii and Rozanov (1959), we get the asymptotic independence of $\hat{\sigma}_{ES,t_N}^2$ and Q_N''.

By (a), $\sqrt{Nh}Q_N'$ is asymptotically negligible. Then, by Theorem 1,

$$d_1\theta^{-1}(r_{t_N})\sqrt{Nh}Q_N + d_2 V_2^{-1/2}\sqrt{n}[\hat{\sigma}_{ES,t_N}^2 - \sigma^2(r_{t_N})]$$

$$\xrightarrow{\mathcal{D}} \mathcal{N}(0, d_1^2 + d_2^2),$$

for any real d_1 and d_2, where $V_2 = \frac{e^c+1}{e^c-1}\sigma^4(r_{t_N})$. Because Q_N is a linear transformation of $\mathbf{u}$,

$$\mathbf{V}^{-1/2}\begin{bmatrix}\sqrt{Nh}\mathbf{u} \\ \sqrt{n}[\hat{\sigma}_{ES,t_N}^2 - \sigma^2(r_{t_N})]\end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, I_3),$$

where $\mathbf{V} = \mathrm{blockdiag}\{V_1, V_2\}$ with $V_1 = 2\sigma^4(r_{t_N})p(r_{t_N})\mathbf{S}^*$, where $\mathbf{S}^* = (v_{i+j-2})_{i,j=1,2}$ with $v_j = \int u^j W^2(u)\,du$. This, combined with (A.11), gives the joint asymptotic normality of $\mathbf{t}$ and $\hat{\sigma}_{ES,t_N}^2$. Note that $\mathbf{b} = o_p(1/\sqrt{Nh})$, and it follows that

$$\Sigma^{-1/2}\begin{pmatrix}\sqrt{Nh}[\hat{\sigma}_{S,t_N}^2 - \sigma^2(r_{t_N})] \\ \sqrt{n}[\hat{\sigma}_{ES,t_N}^2 - \sigma^2(r_{t_N})]\end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, I_2),$$

where $\Sigma = \mathrm{diag}\{2\sigma^4(r_{t_N})v_0/p(r_{t_N}), V_2\}$. Note that $\hat{\sigma}_{S,t_N}^2$ and $\hat{\sigma}_{ES,t_N}^2$ are asymptotically independent; it follows that the asymptotical normality in (b) holds.

The result in (c) follows from (b) and the fact that $\hat{w}_{t_N} \xrightarrow{P} w_{t_N}$ and $\hat{\sigma}_{I,t_N}^2 = \hat{w}_{t_N}\hat{\sigma}_{ES,t_N}^2 + (1 - \hat{w}_{t_N})\hat{\sigma}_{S,t_N}^2$. It follows that

$$\hat{\sigma}_{I,t_N}^2 - \sigma_{t_N}^2$$

$$= \hat{\sigma}_{t_N}^2 - \sigma_{t_N}^2 + (\hat{w}_{t_N} - w_{t_N})[(\hat{\sigma}_{ES,t_N}^2 - \sigma_{t_N}^2) - (\hat{\sigma}_{S,t_N}^2 - \sigma_{t_N}^2)]$$

$$\equiv L_{n1} + (\hat{w}_{t_N} - w_{t_N})(L_{n2} - L_{n3}).$$

Because L_{n1}, L_{n2}, and L_{n3} are of the same order, the second and third terms are dominated by the first term, L_{n1}. Then $\hat{\sigma}_{I,t_N}^2 - \sigma_{t_N}^2 = (\hat{\sigma}_{t_N}^2 - \sigma_{t_N}^2)(1 + o_p(1))$, and thus $\hat{\sigma}_{I,t_N}^2 - \sigma_{t_N}^2$ shares the same asymptotical normality as $\hat{\sigma}_{t_N}^2 - \sigma_{t_N}^2$.

REFERENCES

Aït-Sahalia, Y., and Mykland, P. (2004), "Estimators of Diffusions With Randomly Spaced Discrete Observations: A General Theory," *The Annals of Statistics*, 32, 2186–2222.

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2005), "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," *Review of Financial Studies*, 18, 351–416.

Arapis, M., and Gao, J. (2004), "Nonparametric Kernel Estimation and Testing in Continuous-Time Financial Econometrics," manuscript.

Bandi, F., and Phillips (2003), "Fully Nonparametric Estimation of Scalar Diffusion Models," *Econometrica*, 71, 241–283.

Banon, G. (1978), "Nonparametric Identification for Diffusion Processes," *SIAM Journal of Control Optimization*, 16, 380–395.

Barndoff-Neilsen, O. E., and Shephard, N. (2001), "Non-Gaussian Ornstein–Uhlenbeck–Based Models and Some of Their Uses in Financial Economics" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 63, 167–241.

——— (2002), "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society*, Ser. B, 64, 253–280.

Bollerslev, T., and Zhou, H. (2002), "Estimating Stochastic Volatility Diffusion Using Conditional Moments of Integrated Volatility," *Journal of Econometrics*, 109, 33–65.

Brenner, R. J., Harjes, R. H., and Kroner, K. F. (1996), "Another Look at Models of the Short-Term Interest Rate," *Journal of Financial and Quantitative Analysis*, 31, 85–107.

Cai, Z., and Hong, Y. (2003), "Nonparametric Methods in Continuous-Time Finance: A Selective Review," in *Recent Advances and Trends in Nonparametric Statistics*, eds. M. G. Akritas and D. M. Politis, Amsterdam: Elsevier, pp. 283–302.

Chan, K. C., Karolyi, A. G., Longstaff, F. A., and Sanders, A. B. (1992), "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *Journal of Finance*, 47, 1209–1227.

Chapman, D. A., and Pearson, N. D. (2000), "Is the Short Rate Drift Actually Nonlinear?" *Journal of Finance*, 55, 355–388.

Chen, S. X., Gao, J., and Tang, C. Y. (2007), "A Test for Model Specification of Diffusion Processes," *The Annals of Statistics*, to appear.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985), "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–467.

Davé, R. D., and Stahl, G. (1997), "On the Accuracy of VaR Estimates Based on the Variance–Covariance Approach," working paper, Olshen and Associates.

Duan, J. C. (1997), "Augumented GARCH(p, q) Process and Its Diffusion Limit," *Journal of Econometrics*, 79, 97–127.

Duffie, D., and Pan, J. (1997), "An Overview of Value at Risk," *The Journal of Derivatives*, 4, 7–49.

Fan, J. (2005), "A Selective Overview of Nonparametric Methods in Financial Econometrics" (with discussion), *Statistical Science*, 20, 317–357.

Fan, J., and Gu, J. (2003), "Semiparametric Estimation of Value-at-Risk," *Econometrics Journal*, 6, 261–290.

Fan, J., Jiang, J., Zhang, C., and Zhou, Z. (2003), "Time-Dependent Diffusion Models for Term Structure Dynamics and the Stock Price Volatility," *Statistica Sinica*, 13, 965–992.

Fan, J., and Yao, Q. (1998), "Efficient Estimation of Conditional Variance Functions in Stochastic Regression," *Biometrika*, 85, 645–660.

——— (2003), *Nonlinear Time Series: Nonparametric and Parametric Methods*, New York: Springer-Verlag.

Fan, J., and Zhang, C. M. (2003), "A Reexamination of Diffusion Estimators With Applications to Financial Model Validation," *Journal of the American Statistical Association*, 98, 118–134.

Genon-Catalot, V., Jeantheau, T., and Laredo, C. (1999), "Parameter Estimation for Discretely Observed Stochastic Volatility Models," *Bernoulli*, 5, 855–872.

Gijbels, I., Pope, A., and Wand, M. P. (1999), "Understanding Exponential Smoothing via Kernel Regression," *Journal of the Royal Statistical Society*, Ser. B, 61, 39–50.

Hall, P., and Hart, J. D. (1990), "Nonparametric Regression With Long-Range Dependence," *Stochastic Processes and Their Applications*, 36, 339–351.

Hall, P., and Heyde, C. (1980), *Martingale Limit Theorem and Its Applications*, New York: Academic Press.

Hallin, M., Lu, Z., and Tran, L. T. (2001), "Density Estimation for Spatial Linear Processes," *Bernoulli*, 7, 657–668.

Härdle, W., Herwartz, H., and Spokoiny, V. (2002), "Time-Inhomogeneous Multiple Volatility Modelling," *Journal of Finance and Econometrics*, 1, 55–95.

Hong, Y., and Li, H. (2005), "Nonparametric Specification Testing for Continuous-Time Models With Applications to Term Structure of Interest," *Review of Financial Studies*, 18, 37–84.

Karatzas, I., and Shreve, S. (1991), *Brownian Motion and Stochastic Calculus* (2nd ed.), New York: Springer-Verlag.

Kloeden, D. E., Platen, E., Schurz, H., and Sørensen, M. (1996), "On Effects of Discretization on Estimators of Drift Parameters for Diffusion Processes," *Journal of Applied Probability*, 33, 1061–1076.

Liptser, R. S., and Shiryayev, A. N. (2001), *Statistics of Random Processes I, II* (2nd ed.), New York: Springer-Verlag.

Mercurio, D., and Spokoiny, V. (2004), "Statistical Inference for Time-Inhomogeneous Volatility Models," *The Annals of Statistics*, 32, 577–602.

Morgan, J. P. (1996), *RiskMetrics Technical Document* (4th ed.), New York: Author.

Nelson, D. B. (1990), "ARCH Models as Diffusion Approximations," *Journal of Econometrics*, 45, 7–38.

Revuz, D., and Yor, M. (1999), *Continuous Martingales and Brownian Motion*, New York: Springer-Verlag.

Robinson, P. M. (1997), "Large-Sample Inference for Nonparametric Regression With Dependent Errors," *The Annals of Statistics*, 25, 2054–2083.

Rosenblatt, M. (1970), "Density Estimates and Markov Sequences," in *Nonparametric Techniques in Statistical Inference*, ed. M. L. Puri, Cambridge, U.K.: Cambridge University Press, pp. 199–213.

Ruppert, D., Wand, M. P., Holst, U., and Hössjer, O. (1997), "Local Polynomial Variance Function Estimation," *Technometrics*, 39, 262–273.

Spokoiny, V. (2000), "Drift Estimation for Nonparametric Diffusion," *The Annals of Statistics*, 28, 815–836.

Stanton, R. (1997), "A Nonparametric Models of Term Structure Dynamics and the Market Price of Interest Rate Risk," *Journal of Finance*, LII, 1973–2002.

Vasicek, O. (1977), "An Equilibrium Characterization of Term Structure," *Journal of Financial Economics*, 5, 177–188.

Volkonskii, V. A., and Rozanov, Yu. A. (1959), "Some Limit Theorems for Random Functions, Part I," *Theory of Probability and Its Applications*, 4, 178–197.

Wang, Y. (2002), "Asymptotic Nonequivalence of GARCH Models and Diffusions," *The Annals of Statistics*, 30, 754–783.