Iterative Method for Exponentially Weighted Rolling Regression

Taro Kanatani*†‡

Graduate School of Economics, Kyoto University
Yoshida Honmachi, Sakyo-Ku, Kyoto 606-8501, Japan

*E-mail: taro@e02.mbox.media.kyoto-u.ac.jp
†Tel: +81 75 753-7197
‡Fax: +81 75 753-3492

Abstract

This note proposes an iterative method for exponentially weighted rolling regression (EWRR), which was proved to be an optimal estimator of volatility by Foster and Nelson (1996). The method accelerates the numerical evaluation of EWRR under certain circumstances. An alternative to usual realized volatility is proposed for its application.

Keywords: Rolling regression; Iterative method; Realized volatility

JEL classification: C15; C22; C63

1 Introduction

Various stylized facts about asset return or its volatility can be expressed in state-space models that consist fundamentally of two stochastic differential equations: the observation equation and the state equation (see, e.g., Ghysels et al., 1996). In cases where a observation process is sampled at shorter and shorter time intervals, its conditional variance at any instant can be approximated more accurately using a simple flat-weight moving average of squared residuals. This fact is the theoretical basis for using the standard (flat-weight) rolling regression of squared residuals as an estimator of volatility in the context of high-frequency data.

Foster and Nelson (1996) proved that exponentially weighted rolling regression (EWRR) minimizes the asymptotic variance of measurement error when the time interval is sufficiently small. However, in its application, flat-weight rolling regression (FWRR) was used because it can be calculated efficiently by the conventional iterative method. This note proposes a similar iterative method for EWRR. An alternative to the usual realized volatility is proposed for its application.

2 Iterative Method

First we review the optimal weighted rolling regression explained in Foster and Nelson (1996).[1] Let ${}_hX_t$ be a locally squared integrable semimartingale

[1]For simplification, we restrict our study to scalar and diffusion processes.

that is adapted to the filtration $\{{}_h\mathscr{F}_t\}$, where $\{{}_h\mathscr{F}_t\}$ is increasing and right continuous. Time is discrete such that $t = 0, h, 2h, \cdots, Nh$, where h and N denote the time interval and the number of available observations, respectively. In this note, we assume that the data generating process (DGP) is described by the following state-space representation:

$$\Delta_h X_t = {}_h\mu_t h + \Delta_h M_t, \qquad E((\Delta_h M_t)^2 | {}_h\mathscr{F}_{t-h}) = {}_h\Omega_t h, \tag{1}$$

$$\Delta_h \Omega_t = {}_h\lambda_t h + \Delta_h M_t^*, \qquad E((\Delta_h M_t^*)^2 | {}_h\mathscr{F}_{t-2h}) = {}_h\Lambda_t h, \tag{2}$$

$$\Delta_h B_t = h^{-1/2}((\Delta_h M_t)^2 - {}_h\Omega_t h), \qquad E((\Delta_h B_t)^2 | {}_h\mathscr{F}_{t-h}) = {}_h\theta_t h, \tag{3}$$

where Δ denotes the first order difference (e.g., $\Delta_h X_t = {}_hX_t - {}_hX_{t-h}$), ${}_hM_t$ and ${}_hM_t^*$ are local martingales with respect to ${}_h\mathscr{F}_{t-h}$ and ${}_h\mathscr{F}_{t-2h}$, ${}_h\mu_t$ and ${}_h\Omega_t$ are ${}_h\mathscr{F}_{t-h}$-measurable, and ${}_h\lambda_t$ and ${}_h\Lambda_t$ are ${}_h\mathscr{F}_{t-2h}$-measurable.

Difference equations in (1) and (2) are called the *observation equation* and *state equation*, respectively. ${}_h\Omega_t$ represents volatility when ${}_hX_t$ is the logarithm of the asset price. In (3), the *sampling error* $\Delta_h B_t$ is defined as the martingale difference. Note that ${}_h\theta_t / {}_h\Omega_t^2$ describes the conditional kurtosis of $\Delta_h M_t$ minus one because

$${}_h\theta_t = h^{-1} E((\Delta_h B_t)^2 | {}_h\mathscr{F}_{t-h}) = E((\Delta_h M_t / \sqrt{h})^4 | {}_h\mathscr{F}_{t-h}) - {}_h\Omega_t^2.$$

The estimator addressed in this study is the rolling regression of squared residuals

$${}_h\hat{\Omega}_t \equiv \sum_{s={}_hT_*(t)}^{{}_hT^*(t)} {}_hw_{s-t} z_s h, \quad z_t \equiv \frac{(\Delta_h X_t - {}_h\hat{\mu}_t h)^2}{h},$$

where ${}_hT_*(t)$ and ${}_hT^*(t)$ are the *start* and *end times* of the rolling regression, $\hat{\mu}_t$ is an estimation of μ_t, and $\sum_t {}_hw_t h = 1$. Furthermore, some additional assumptions on DGP and weight are required for the following asymptotic results.[2]

Foster and Nelson (1996) derived the asymptotic distribution of the measurement error:

$$h^{-1/4}({}_h\hat{\Omega}_t - {}_h\Omega_t)|\mathscr{F}_{T_*} \overset{a}{\sim} N(0, {}_hC_{T_*}),$$

where

$${}_hC_{T_*} = {}_h\theta_{T_*}\sqrt{h}\sum_t {}_hw_t^2 h + \frac{{}_h\Lambda_{T_*}}{\sqrt{h}}\sum_t {}_h\Psi_t^2 h,$$

and

$${}_h\Psi_t = \begin{cases} \displaystyle\sum_{s=t+h,t+2h,\cdots}^{\infty} {}_hw_s h & \text{if } t \geq 0, \\ \displaystyle -\sum_{s=-\infty}^{t} {}_hw_s h & \text{if } t < 0. \end{cases}$$

For discussion in the next section, we display variances of EWRR and backward-looking FWRR:

$${}_hC_{T_*} = \begin{cases} \frac{1}{4}\left({}_h\theta_{T_*} a\sqrt{h} + \frac{{}_h\Lambda_{T_*}}{a\sqrt{h}}\right) & \text{if } {}_hw_{s-t} = \frac{a}{2}e^{-a|s-t|}, \\ \frac{{}_h\theta_{T_*}}{n\sqrt{h}} + \frac{{}_h\Lambda_{T_*} n\sqrt{h}}{3} & \text{if } {}_hw_{s-t} = \frac{1}{nh}\cdot I(\{s \in [t-nh, t]\}), \end{cases}$$

where $I(\cdot)$ denotes the indicator function.[3] These variances are minimized, respectively, when $a = \sqrt{{}_h\Lambda_{T_*}/{}_h\theta_{T_*}h}$ and $n = \sqrt{3{}_h\theta_{T_*}/{}_h\Lambda_{T_*}h}$. Foster and

[2]See Foster and Nelson (1996) to review those assumptions.

[3]These can be verified easily by considering the sums as integrals:

$$\sum_t {}_hw_t^2 h \cong \int_{-\infty}^{\infty} {}_hw_t^2 dt,$$

$$\sum_t {}_h\Psi_t^2 h \cong \int_0^{\infty}\left(\int_t^{\infty} {}_hw_s ds\right)^2 dt + \int_{-\infty}^{0}\left(\int_{-\infty}^{t} {}_hw_s ds\right)^2 dt.$$

Nelson (1996) proved that EWRR setting $a = \sqrt{{}_h\Lambda_{T_*}/{}_h\theta_{T_*}h}$ realizes the smallest variance in all weights. If ${}_hw_t$ is constant over time, FWRRs can be evaluated easily because recursive calculation is possible. For example, the backward-looking FWRR is written by the first-order difference equation:

$$ {}_h\hat{\Omega}_t = {}_h\hat{\Omega}_{t-h} + \frac{1}{nh} \cdot (z_{T^*} - z_{T_*-h}). $$

In fact, Foster and Nelson (1996) used two-sided FWRR in an empirical example and in a Monte Carlo simulation.

We propose a similar iterative method for EWRR. To simplify the notation, we define EWRR as

$$ \mathrm{EWRR}[z|a](t) = \sum_s \frac{a}{2} e^{-a|s-t|} z_s h, $$

and divide EWRR into past and future portions as

$$ \mathrm{EWRR}[z|a](t) = P[z|a](t) + F[z|a](t), \tag{4} $$

where

$$ P[z|a](t) = \sum_{s \le t} \frac{a}{2} e^{a(s-t)} z_s h, \quad \text{and} \quad F[z|a](t) = \sum_{s>t} \frac{a}{2} e^{-a(s-t)} z_s h. $$

Thereby, we can find the iterative rule in each process as

$$ P[z|a](t) = e^{-ah} P[z|a](t-h) + \frac{a}{2} z_t h, \tag{5} $$

$$ F[z|a](t) = e^{-ah} F[z|a](t+h) + \frac{a}{2} e^{-ah} z_{t+h} h. \tag{6} $$

In the same manner as for flat-weight, if the weight function does not change (i.e., a is constant) over time, these recurrence formulas improve the efficiency of numerical evaluation. Using (5) and (6), the two series of $\{P[z|a](t)\}_{t=0,h,2h,\cdots}^{Nh}$, and $\{F[z|a](t)\}_{t=Nh,Nh-h,Nh-2h,\cdots}^{0}$, are calculated forward and backward, respectively. Then EWRR$[z|a](t)$ is completed by (4) at each t. As $N \to \infty$, the theoretical computational time with the method increases at order N, whereas that without the method increases at order N^2.

3 An Application: Comparison with Instantaneous Realized Volatility

We require estimates of ${}_h\theta_{T_*}$ and ${}_h\Lambda_{T_*}$ to use optimal EWRR, but producing such estimates is burdensome. Even under the simplifying assumptions that ${}_h\Lambda_t/{}_h\Omega_t^2$ and ${}_h\theta_t/{}_h\Omega_t^2$ are constant over time, they cannot be estimated accurately, as explained in Foster and Nelson (1996). Instead of seeking the optimal estimator, we propose a practical usage of EWRR.

Realized volatility, which is often used as a proxy for true volatility to measure the performance of forecasting in empirical contexts, is defined as backward-looking FWRR,

$$ {}_h\hat{\Omega}_t = \sum_s \frac{1}{n_r} \cdot I(\{s \in [t - n_r h, t]\}) \cdot z_s, \tag{7} $$

where n_r is constant over time. A researcher must determine window length n_r by some method. In the context of the theoretical approach outlined in Foster and Nelson (1996), the estimator (7) implies that the researcher believes n_r to be the optimal $\sqrt{3{}_h\theta_{T_*}/{}_h\Lambda_{T_*}h}$ over time. That implication is equivalent to setting $\sqrt{{}_h\Lambda_{T_*}/{}_h\theta_{T_*}h} = \sqrt{3}/n_r h$. The variances of the asymptotic measurement error of EWRR$[z|\sqrt{3}/n_r h]$ and backward-looking FWRR (7) are

$$\frac{\sqrt{3}}{4}\left(\frac{{}_h\theta_{T_*}}{n_r\sqrt{h}}+\frac{{}_h\Lambda_{T_*}n_r\sqrt{h}}{3}\right), \quad \text{and} \quad \frac{{}_h\theta_{T_*}}{n_r\sqrt{h}}+\frac{{}_h\Lambda_{T_*}n_r\sqrt{h}}{3}, \tag{8}$$

respectively. Therefore, at any t, EWRR realizes a $\sqrt{3}/4$ smaller measurement error variance than realized volatility. Consequently, we expect that the use of EWRR reduces mean squared error (MSE) by $\sqrt{3}/4$ compared to realized volatility.

To confirm this, we performed a Monte Carlo simulation according to Foster and Nelson (1996). We generated 16,885 observations from the following DGP:

$$\Delta \log \Omega_t = 0.0056 \cdot (-0.4246 - \log \Omega_{t-1}) + \sqrt{0.012} \cdot u_{2t}, \tag{9}$$

$$\Delta M_t = \sqrt{\Omega_t} \cdot u_{1t}, \tag{10}$$

where both u_{1t} and u_{2t} are mutually independent, $u_{1t} \sim$ i.i.d. standardized-t_{12}, and $u_{2t} \sim$ i.i.d.$N(0,1)$.[4]

(9) implies that $\log \Omega_t$ is conditionally homoskedastic. This implication is equivalent to the constancy of Λ_t/Ω_t^2, which is specified by 0.012 in this DGP. In (10), kurtosis of u_{1t} is assumed to be 3.75. This assumption means that $\theta_t/\Omega_t^2 = 2.75$ over time because θ_t/Ω_t^2 is conditional kurtosis of u_{1t} minus one. The constancy of Λ_t/Ω_t^2 and θ_t/Ω_t^2 implies that the optimal n_r is $\sqrt{3 \cdot 2.75/0.012}(\approx 26)$ over time.[5]

[4]The prefix $h(=1)$ is dropped for the remainder of this paper.

[5]According to French et al. (1987), that implication seems to be reasonable in reference to U.S. stock prices.

Table 1 shows the average MSE of realized volatility and EWRR from 600 simulations along with ratios of the two estimators' averages of the MSEs. Both estimators minimize the MSE at optimal n_r. As expected, the ratios are approximately $\sqrt{3}/4(\approx 0.433)$ near the optimal n_r. The ratios separate from 0.433 when n_r is far from 26. A very small n_r violates the assumption that the number of observations in the window must be sufficiently large to hold the asymptotic theory. On the other hand, a very large n_r violates the assumption that the window length must be sufficiently short to maintain the parameter constancy.

Although the simplifying assumptions hold in the above example, (8) suggests that regardless of whether the assumptions hold or not (whether nuisance parameters can be estimated accurately or not), the measurement error variances ratio is always $\sqrt{3}/4$. This relation holds unless not-so-restrictive assumptions on DGP and weight (i.e., Foster and Nelson (1996), Assumptions A–D) are violated. We infer that EWRR$[(Residual)^2|\sqrt{3}/n_r]$ is preferable for use in place of the usual realized volatility with window length n_r in a broad range of situations.

4 Conclusion

Using the iterative method presented herein, EWRR is as tractable as FWRR. Nevertheless, the optimal EWRR of Foster and Nelson (1996) requires estimates of nuisance parameters. Even using simplifying assumptions, estimating those parameters is an onerous problem. This note proposes a practical application of EWRR: an alternative to the usual realized volatility with window length n. EWRR$[(Residual)^2|\sqrt{3}/n]$ realizes a $\sqrt{3}/4$ smaller measurement error variance than the realized volatility. Moreover, that relation does not require overly restrictive assumptions. For that reason, instead of realized volatility, we can use EWRR in a wide range of situations as a more accurate and equally simple estimator.

Acknowledgements

The author would like to thank Kimio Morimune, Masato Kagihara, Ryo Okui and seminar participants at Kyoto University for helpful comments and suggestions. This research was partially supported by a Grant-in-Aid for the 21st Century COE Program "Interfaces for Advanced Economic Analysis" of the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

References

Foster, D. P., Nelson, D. B., 1996. Continuous record asymptotics for rolling sample variance estimators. Econometrica 64, 139–174.

French, K. R., Schwert, G. W., Stambaugh, R. F., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.

Ghysels, E., Harvey, A. C., Renault, E., 1996. Stochastic volatility. In: Maddala, G. S., Rao, C. R. (Eds.), Handbook of Statistics. Vol. 14. North Holland, Amsterdam.

Table 1: Average MSE of realized volatility and EWRR

n_r	1	20	26	50	100
Realized volatility	12.456	0.768	0.724	0.805	1.146
	(108.154)	(0.706)	(0.644)	(0.723)	(1.138)
EWRR	8.557	0.354	0.338	0.395	0.603
	(7.241)	(0.258)	(0.237)	(0.273)	(0.478)
Ratio	0.687	0.461	0.467	0.491	0.526

Note: Realized volatility and EWRR are computed as

$$\frac{1}{n_r}\sum_{i=0}^{n_r-1} z_{t-i} \quad \text{and} \quad \frac{\sqrt{3}}{2n_r}\sum_{s} z_s \exp\left[-\frac{\sqrt{3}}{n_r}|s-t|\right],$$

where z_t is the squared residual at t. All means are computed through 600 replications. Standard deviations are shown in parentheses. The 'Ratio' row shows ratios of the two estimators' averages of MSEs at each n_r.