Optimal Threshold Selection for Realized Volatility Forecasts in the Presence of Jumps

Benjamin E. Fissel\textsuperscript{1,2} and Yixiao Sun\textsuperscript{1}

November 24, 2010

(1) Email: bfissel@ucsd.edu and yisun@ucsd.edu.
Department of Economics
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0508

(2) Email: Ben.Fissel@noaa.gov
NOAA, Alaska Fisheries Science Center
Economics and Social Sciences Research Program
7600 Sand Point Way NE
Seattle, WA 98115

Abstract: When estimating and forecasting realized volatility in the presence of jumps, a form of bias-variance tradeoff is present in the selection of the truncation threshold. We propose an optimal method for threshold selection that minimizes the out-of-sample forecasting loss. The use of a forecasting framework is fundamentally different from the testing framework in the literature. We find that \textit{a priori} large truncation thresholds may not be optimal from a forecasting perspective and smaller thresholds should be used. An extensive simulation study and an empirical application to S&P 500 futures demonstrate the effectiveness of the proposed method.

Keywords: Bipower Variation, Forecasting, Jump Diffusion, Quadratic Variation, Realized Volatility.

JEL Classification: C01, C14, C52, C53
1 Introduction

Accurate measurement of volatility is of paramount importance in the world of finance where volatility is risk. Furthermore, the volatility of asset returns has been shown to possess long memory meaning that volatility and hence risk are forecastable. Accurate measurement of volatility will increase the forecasting performance. Measuring and forecasting volatility has an extensive history and has been widely researched. Many different parametric and nonparametric methods are available to practitioners. A good summary of the literature can be found in Andersen et al. (2007).

A popular method of measuring volatility is through realized volatility. The theoretical foundation for realized volatility is set in the context of continuous time diffusion processes. The use of diffusion processes to model asset returns is widespread due to the tractability of the model and the breadth of research and literature on Brownian motion and diffusion processes. From an economic perspective, the diffusion process can be motivated through market participants interacting under no arbitrage. For pure diffusion processes, realized volatility can be reliably estimated by the so-called Quadratic Variation (QV) estimator (Andersen et al. (2000)).

There are, however, aspects of the price process that the pure diffusion process does not fully capture. In particular, we observe too many large movements in asset returns and the tails are too fat for the price process to be consistent with a pure continuous time diffusion process, see Andersen et al. (2001). This has led researchers to propose a number of different yet related alternatives to the pure diffusion process. One popular alternative dating back to perhaps Merton (1976) is incorporating jumps into the price process. This ultimately leads to what has become known as the jump diffusion process. This type of process augments the pure diffusion process with an additive jump component producing occasional large discrete movements in the process. The jump diffusion process can be parameterized to better characterize the return process and display large movements and fat tails that the pure diffusion process lacks. However, the additional flexibility comes at the cost of tractability. Given that we can only sample the continuous time process on discrete intervals, it is difficult to know whether the movement between two observations was the result of normal variation of the diffusion process or a jump occurred somewhere within the interval.

In the context of jump diffusion processes, the QV estimator is biased by the jumps. The bias induced by the jumps can potentially be quite large. The Bipower Variation (BV) estimator, an alternative measure of the daily volatility, is asymptotically robust to the presence of jumps but is not as accurate as the QV estimator. That is to say, the BV estimator is asymptotically unbiased but has a larger variance. An improved estimator of the daily volatility proposed by Barndorff-Nielsen and Shephard (2004) and explored by Andersen et al. (2007) uses the QV estimator on days when no jumps occurred and the BV estimator on jump days. We call this the switching estimator. However, as mentioned before, identifying jump days is difficult given that we only observe the process on discrete intervals.

Work started by Barndorff-Nielsen and Shephard (2002, 2003, 2004, 2006), and continued by Huang and Tauchen (2005) and Andersen et al. (2007) shows that a
Hausman type $Z$-statistic can effectively identify the jump days. More specifically, the $Z$-statistic is equal to $(QV - BV)/\sigma$, the normalized difference between the BV and QV estimators. Under the null of no jumps, the $Z$-statistic is asymptotically standard normal while under the alternative of having jumps the $Z$-statistic diverges to infinity. If the $Z$-statistic is larger than a critical value or truncation threshold, we conclude that jumps have occurred. Through simulations, Huang and Tauchen (2005) show that the test based on the $Z$-statistic displays good size and power properties. Under realistic parameterizations of the diffusion process and using 2.32 (or 99% critical value) as the threshold, the $Z$-test correctly classifies 70% of the jump days (Table 3, p. 477). To date, the critical value or the truncation threshold has been somewhat arbitrarily chosen. It typically ranges from the 0.95 to 0.9999 quantiles of the standard normal distribution, reflecting the belief that jumps are rare but their sizes are large.

In this paper we take a different perspective on the selection of the truncation threshold and argue that from a forecasting perspective we can select the threshold to minimize prediction error. Our objective is not to identify the jump days per se. Instead, we focus on the forecasting problem, which is fundamentally different from the testing problem considered in the literature. From the testing perspective, misclassification of small jumps is as bad as misclassification of large jumps as both contribute equally to the type II error of the test. However, from the forecasting perspective, misclassification of small jumps may not matter much while misclassification of large jumps can be disastrous.

We provide a conceptual framework for optimal threshold determination. We assume that the daily realized volatility or its logarithm follows an AR(1) process and we do not directly observe this process but we have two noisy measures: one from the QV estimate and the other from the BV estimate. The switching estimator is equal to the (log)QV estimator if the $Z$-statistic is less than a given truncation threshold and is equal to the (log)BV estimator otherwise. This simple framework is not meant to capture reality perfectly but it captures the essence of the forecasting problem and allows us to see clearly the offsetting forces behind the optimal threshold determination. We find that the optimal threshold decreases with the jump probability while it decreases with the jump size initially and then increases with it as the jump size becomes large.

To support the conceptual framework, we simulate a one-factor stochastic volatility jump-diffusion process. Following the same parametrization set forth in Huang and Tauchen (2005), we demonstrate that the forecast loss is U-shaped in the truncation threshold. For practical use, we employ the $hv$-block cross-validation method of Racine (2000) to optimally select the threshold. The $hv$-block cross-validation is a nonparametric method for estimating the out-of-sample forecasting loss. We do not use the simple leave-one-out cross validation here because the observations are not iid. Another complication is that in practice we do not observe the true realized volatility process, the cross validation target. We show that we can use the BV estimate as a plug-in target for the true unknown realized volatility. The BV estimator is an appropriate target to use because it is consistent regardless of the presence of
the jumps. Simulations confirm that we can select the optimal truncation threshold using either the true volatility as the target or the BV estimate as a proxy.

The paper will proceed as follows. Section 2 will formalize much of the discussion above by presenting realized volatility, the asymptotic theory resulting in the $Z$-statistic, and the switching estimator. Section 3 will present a simple framework for truncation threshold selection and show that under a mean squared error criterion an optimal threshold exists. Section 4 will present an extensive simulation study, which lends support to the effectiveness of our optimal rule for threshold determination. Section 5 will apply our method to S&P 500 futures. The final section will conclude.

2 Background on Realized Volatility

A rigorous exposition of realized volatility has been presented in a number of papers such as Barndorff-Nielsen and Shephard (2002) and Andersen et al. (2000). It is not the intent of this paper to cover this theory thoroughly, rather what is presented here is a brief summary and the form of the estimators considered in this paper.

The theoretical justification for using realized volatility as a measure of variance is set in the context of continuous-time diffusion processes. The use of continuous-time diffusion processes to model asset prices is ubiquitous in financial econometrics. Following the notation of Andersen et al. (2007), we assume that the log price process satisfies the stochastic differential equation:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + \kappa(t)dq(t), \quad t \in (0,T)$$

(1)

where $\mu(t)$ is a continuous process with locally bounded variation, $\sigma(t)$ is a strictly positive volatility process, and $W(t)$ is a standard Brownian motion. The term $\kappa(t)dq(t)$ captures the jump component of the process, $q(t)$ is a counting process where $P[dq(t) = 1] = \lambda(t)dt$ for a possibly time varying intensity parameter $\lambda(t)$ and $\kappa(t)$ is the jump size of the process. The daily realized volatility of the diffusion process is defined to be the integrated volatility process over a trading day, i.e.

$$V_t = \int_{t-1}^t \sigma^2(s)ds.$$  

(2)

We focus on $V_t$, the volatility induced by the diffusion component, because there is ample empirical evidence that $V_t$ is serially correlated and thus predictable while the jump component is hard to predict.

Let

$$r(s+1) = p(s+1) - p(s)$$

be the return over the period from $s$ to $s+1$. Define the quadratic variation estimator to be

$$QV_t := QV_t(\Delta) = \sum_{j=1}^{1/\Delta} r_{t-1+j\Delta}^2$$

(3)

where $\Delta$ is the length of the sampling interval. To clarify, consider data measured in 5-minute increments from 8:30am to 4:30pm for a total of 97 observations per
day. If we wanted to measure the daily variation (i.e. \( t \to t + 1 \) represents a day), then we could set \( 1/\Delta = 97 \). Under the assumptions given in a series of papers by Barndorff-Nielsen and Shephard (2002, 2003, 2004, 2006), we have

\[
QV_t(\Delta) \to_p \int_{t-1}^{t} \sigma^2(s) ds + \sum_{t-1<s\leq t} \kappa^2(s) \text{ as } \Delta \to 0. \tag{4}
\]

It follows that in the presence of jumps the QV estimator is biased for the realized volatility \( V_t \) with the jump terms inducing the bias.

An analogous measure of volatility has been presented by Barndorff-Nielsen and Shephard (2003) known as realized bipower variation. It is defined as

\[
BV_t := BV_t(\Delta) = \mu_1 \left( \frac{M}{M - 1} \right)^{1/\Delta} \sum_{j=2}^{1/\Delta} |r_{t-1+j\Delta}| |r_{t-1+(j-1)\Delta}| \tag{5}
\]

where \( M = 1/\Delta \) and \( \mu_x = E(|Z|^x) \) for \( Z \sim N(0, 1) \) hereafter. The Bipower Variation (BV) estimator is consistent for the realized volatility:

\[
BV_t(\Delta) \to_p \int_{t-1}^{t} \sigma^2(s) ds \text{ as } \Delta \to 0. \tag{6}
\]

However, unlike the QV estimator, it is robust to the presence of jumps in the price process.

In the absence of jumps, the consistency results in (4) and (6) can be strengthened to

\[
\Delta^{-1/2} \Gamma^{-1}_t \left( \frac{QV_t - \int_{t-1}^{t} \sigma^2(s) ds}{BV_t - \int_{t-1}^{t} \sigma^2(s) ds} \right) \to^d N \left( \left( \begin{array}{c} 0 \\ 0 \end{array} \right), \left( \begin{array}{cc} 2 & 2 \\ 2 & 2.609 \end{array} \right) \right) \tag{7}
\]

where \( \Gamma^2_t = \int_{t-1}^{t} \sigma^4(s) ds \) is the integrated quarticity. It follows that a normalized difference between \( QV_t \) and \( BV_t \) has a stable distribution in the absence of jumps while it diverges in the presence of jumps. For details see Barndorff-Nielsen and Shephard (2004, 2006). This suggests that jump detection can be based on a Hausman type Z-statistic defined by

\[
Z_t := Z_t(\Delta) = \frac{QV_t(\Delta) - BV_t(\Delta)}{\sqrt{\Delta} \sqrt{IQ_t(\Delta)}} \Rightarrow N(0, 1), \text{ as } \Delta \to 0 \tag{8}
\]

where

\[
IQ_t(\Delta) = \mu_{-3}^{4/3} \left( \frac{M}{M - 2} \right)^{1/\Delta} \sum_{j=3}^{1/\Delta} |r_{t+j\Delta}|^{4/3} |r_{t+(j-1)\Delta}|^{4/3} |r_{t+(j-2)\Delta}|^{4/3} \tag{9}
\]

is an estimator of the integrated quarticity commonly called the tri-power quarticity.

Let \( z_\alpha = \Phi^{-1}(\alpha) \) be the \( 100 \times \alpha \% \) quantile of the standard normal distribution. For a given \( \alpha \), the days that \( Z_t(\Delta) > z_\alpha \) are considered to be days in which a jump occurred. Huang and Tauchen (2005) show that this identification strategy with
\( z_\alpha = 2.32 \) can correctly detect approximately 70\% of the jump days. Based on this identification strategy, we introduce a switching estimator of the realized volatility:

\[
\hat{V}_t = QV_t I\{Z_t < z_\alpha\} + BV_t I\{Z_t \geq z_\alpha\}. \tag{9}
\]

The use of this switching estimator is studied by Barndorff-Nielsen and Shephard (2004), Andersen et al. (2007) and Huang and Tauchen (2005). The switching estimator is a function of \( z_\alpha \), a critical value from the standard normal distribution. Andersen et al. (2007) find that using the switching estimator increases forecasting performance. Given the positive performance of the switching estimator, a natural next step is to refine the estimator by selecting the threshold parameter appropriately. In the rest of paper, we refer to both \( z_\alpha \) and \( \alpha \) as the truncation threshold. This should not cause any confusion.

### 3 Threshold Selection for Volatility Forecast

The selection of the truncation threshold for the switching estimator to date has been somewhat arbitrary. Previous values have been selected based on the belief that jumps are rare but their sizes are large. This leads researchers to consider threshold levels \( \alpha \) in the range from 0.95 to 0.9999. For example, Andersen et al. (2007) use 0.999. In this paper, we propose a data-driven method to optimally select the truncation threshold. The proposed optimal threshold minimizes the forecasting loss and reflects an optimal trade-off that is analogous to the standard bias-variance trade-off.

To instill some intuition, consider first a very low truncation threshold. With a low truncation threshold a large number of days are “identified” as having jumps and consequently the BV estimator is frequently used. Due to the larger variance of the BV estimator, we can decrease our prediction error by using the QV estimator more often at the cost of incurring some jump-induced bias. Conversely, consider a very large truncation threshold. In this scenario we use the QV estimator very often, thereby introducing a large amount of bias due to the presence of jumps. By using the BV estimator more often, we can reduce the bias at the cost of some variance inflation. These two opposing forces lead to a level where the marginal change in variance from using the QV estimator is equal to the marginal change in bias from using the BV estimator.

#### 3.1 Model and Forecast Log-volatility

To provide a rigorous analysis, we first have to decide whether we want to model volatility itself or its logarithm. From a modeling perspective, it seems natural to model the logarithm of the volatility process as we do not have to impose any positiveness constraint. In this subsection, we focus on modeling and forecasting log-volatility.

We assume that the true log realized volatility follows a simple AR(1) process

\[
v_{t+1} = \rho_0 + \rho_1 v_t + u_{t+1}
\]
where $v_t = \log V_t$ and $u_{t+1}$ is a white noise idiosyncratic error term. We do not observe $v_t$ directly but we have two error-ridden measures: $bv = \log(BV)$ and $qv = \log(QV)$. According to (7) and using the Delta method, these two estimators can be represented asymptotically as

$$bv_t = v_t + \frac{\epsilon_t + \sigma_e \epsilon_t}{\exp(v_t)}, \quad qv_t = v_t + \frac{\epsilon_t + \sigma_e \lambda_t}{\exp(v_t)}$$

where

$$\sigma_e = \sqrt{\Delta \Gamma},$$

$\epsilon_t$ is the irreducible estimator error of the QV estimator in the absence of jumps, $\sigma_e \epsilon_t$ is the additional estimator error incurred by the less efficient BV estimator, and $\lambda_t$ is the bias term induced by the jumps. $\epsilon_t, \epsilon_t$ and $\lambda_t$ are mutually independent with

$$\begin{pmatrix} \epsilon_t \\ \epsilon_t \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & 1 \end{pmatrix} \right),$$

and

$$\lambda_t = \begin{cases} \xi, & \text{with probability } p \\ 0, & \text{with probability } 1 - p \end{cases}$$

The normality assumption for $(\epsilon_t, \epsilon_t)$ is compatible with Barndorff-Nielsen and Shephard (2004, 2006) in an asymptotic sense. The binary specification for $\lambda_t$ gives a simple model of jumps: for a given trading day, jumps happen with probability $p$ and the jump size is $\sigma_e \xi$. For simplicity we assume that the jump size is a constant. It is straightforward to generalize the current analysis to random jump sizes.

Note that the variance of $\epsilon_t$ is of order $\Delta$ and decays to zero as the sampling interval approaches zero. The variance of the extra estimator error $\sigma_e \epsilon_t$ is by definition $\sigma_e^2$, which is also of order $\Delta$. The parametrization of jumps in (10) implicitly assumes that the jump size is comparable to the extra estimation error. This is analogous to the parametrization in a local power analysis where the alternatives are assumed to lie in a shrinking neighborhood of the null hypothesis.

We assume that $\rho_0, \rho_1$ and $\sigma_e$ are known parameters. This is not realistic but can be justified by an asymptotic argument. Suppose these parameters can be estimated at the parametric rate. If the length of the time series used in estimating $\rho_0, \rho_1$ and $\sigma_e$ is of larger order than the number of sampling points per trading day, then the estimation errors for $\rho_0, \rho_1$ and $\sigma_e$ are of smaller order than the squared loss $L$ defined in equation (11) below. As a result, the estimation error can be ignored for our purpose here.

As in the previous section, we can define the switching estimator based on the log-volatility:

$$\hat{v}_t = \mathbb{I} \{ Z_t < z_\alpha \} qv_t + \mathbb{I} \{ Z_t \geq z_\alpha \} bv_t$$

$$= v_t + \frac{\epsilon_t}{\exp(v_t)} + \frac{\sigma_e}{\exp(v_t)} \left[ \mathbb{I}_\alpha \lambda_t + (1 - \mathbb{I}_\alpha) \epsilon_t \right]$$
where
\[ Z_t = \frac{\exp(v_t)(qv_t - bv_t)}{\sigma_e} \quad \text{and} \quad \mathbb{I}_\alpha = \mathbb{I}\{Z_t \leq \alpha\} = \mathbb{I}\{\lambda_t - \epsilon_t \leq \alpha\}. \]

Suppose we want to forecast \( v_{t+1} \) conditional on the information set \( \mathfrak{F}_t = \sigma(bv_t, qv_t, \ldots) \). Then the best forecast under the squared error loss is \( \rho_0 + \rho_1 v_t \) if \( v_t \) is known. In practice, \( v_t \) is not known and a plug-in version of the best forecast is
\[ \hat{v}_{t+1} = \rho_0 + \rho_1 \hat{v}_t. \]

We focus on the forecasts of the above form in the rest of the paper. The corresponding forecasting loss is
\[ L = E \left[ (v_{t+1} - \hat{v}_{t+1})^2 \mid \mathfrak{F}_t \right]. \tag{11} \]
Assuming that \( u_{t+1} \) is uncorrelated with \( e_t \) and the jumps \( \lambda_t \), the above loss can be rewritten as
\[ L(z_\alpha) = E \left[ u_{t+1}^2 \mid \mathfrak{F}_t \right] + \rho_0^2 E \left[ (v_t - \hat{v}_t)^2 \mid \mathfrak{F}_t \right]. \tag{12} \]
So in this simple setting, minimizing the forecasting loss is equivalent to minimizing the mean squared error of the switching estimator \( \hat{v}_t \). The MSE of \( \hat{v}_t \) is proportional to \( \hat{L}(z_\alpha) \) defined as:
\[ \hat{L}(z_\alpha) = E \left[ \mathbb{I}_\alpha \lambda_t + (1 - \mathbb{I}_\alpha) e_t \right]^2 = E \left[ \mathbb{I}_\alpha \lambda_t^2 \right] + E \left[ (1 - \mathbb{I}_\alpha) e_t^2 \right]. \tag{13} \]

The above loss function is based on log-volatility. Suppose the ultimate goal of modelling the log volatility is to forecast the volatility itself. Then according to the simple AR(1) model, the best forecast under the quadratic loss is
\[ E(V_{t+1} \mid \mathfrak{F}_t) = E(\exp(v_{t+1}) \mid \mathfrak{F}_t) = E(\exp(\rho_0 + \rho_1 v_t + u_{t+1}) \mid \mathfrak{F}_t) = CE[\exp(\rho_0 + \rho_1 v_t) \mid \mathfrak{F}_t], \]
where \( C = E[\exp(u_{t+1}) \mid \mathfrak{F}_t] \). A plug-in forecast is then given by
\[ \hat{V}_{t+1} = C \exp(\rho_0 + \rho_1 \hat{v}_t) \]
and the associated forecast error is
\[ \hat{V}_{t+1} - \hat{V}_{t+1} = \exp(\rho_0 + \rho_1 v_t) \exp(u_{t+1}) - \exp(\rho_0 + \rho_1 \hat{v}_t) E[\exp(u_{t+1}) \mid \mathfrak{F}_t] \]
\[ = \exp(\rho_0 + \rho_1 v_t) \{ \exp(u_{t+1}) - E[\exp(u_{t+1}) \mid \mathfrak{F}_t] \}
+ [\exp(\rho_0 + \rho_1 v_t) - \exp(\rho_0 + \rho_1 \hat{v}_t)] E[\exp(u_{t+1}) \mid \mathfrak{F}_t] := I_1 + I_2. \]

In the above decomposition, the first part \( I_1 \) reflects the unforecastable component and the second part \( I_2 \) captures the forecast error due to the inaccuracy of \( \hat{v}_t \) as a measure of \( v_t \). To the first order, \( I_2 \) is proportional to \( (v_t - \hat{v}_t) \). As a result, the optimal \( \alpha \) that minimizes \( E \left[ (v_{t+1} - \hat{v}_{t+1})^2 \mid \mathfrak{F}_t \right] \) also minimizes the dominating term in
\[ E \left[ (\hat{V}_{t+1} - \hat{V}_{t+1})^2 \mid \mathfrak{F}_t \right]. \]

8
3.2 Model and Forecast Volatility

From the modelling perspective, it is natural to focus on log-volatility. However, from the forecasting point of view, it is sometimes more convenient and often more straightforward to model volatility itself. In this subsection, we briefly discuss this practice.

We assume that the true volatility process follows an AR(1) model:

$$V_{t+1} = \tilde{\rho}_0 + \tilde{\rho}_1 V_t + u_{t+1}$$  \hspace{1cm} (14)

where $u_{t+1}$ is mean zero error term that is uncorrelated with $V_t$. The simple linear process is meant to be the best linear approximation to the true data generating process for $V_t$. We do not observe $V_t$ directly but we have two error-ridden measures: the BV and QV estimates. According to (7), these two estimators can be represented asymptotically as

$$BV_t = V_t + \epsilon_t + \sigma_e \epsilon_t$$  \hspace{1cm} (15)
$$QV_t = V_t + \epsilon_t + \sigma_e \lambda_t$$  \hspace{1cm} (16)

where $\epsilon_t, \lambda_t, \epsilon_t$ are defined as before. To ensure positivity of $BV_t, QV_t$ and $V_{t+1}$, it is technically necessary to truncate the tails of $u_{t+1}, \epsilon_t$ and $\epsilon_t$. This, however, is such a low probability event for small $\Delta$ that the truncation and renormalization can effectively be ignored in practice.

With the above parametrization, we can rewrite the switching estimator in Section 2 as

$$\hat{V}_t = \mathbb{I} \{ Z_t < z_\alpha \} QV_t + \mathbb{I} \{ Z_t \geq z_\alpha \} BV_t$$

$$= V_t + \epsilon_t + \sigma_e \mathbb{I}_a \lambda_t + (1 - \mathbb{I}_a) \epsilon_t \right\} .$$

The term $V_t + \epsilon_t$ in $\hat{V}_t$ can be regarded as the best possible target.

Following the same argument in the previous subsection, it is easy to show that, to minimize the quadratic forecasting loss is to minimize the mean squared error of the switching estimator $\hat{V}_t$. Again using the same argument, we can show that the mean squared error of $\hat{V}_t$ is proportional to the loss given in (13).

3.3 Optimal Threshold Determination

We have shown that the best forecast is achieved by using a threshold that minimizes $\hat{L}(z_\alpha)$ as given in (13). We proceed to compute $E \left[ \mathbb{I}_a \lambda_t^2 \right]$ and $E \left[ (1 - \mathbb{I}_a) \epsilon_t^2 \right]$. First,

$$E \left[ \mathbb{I}_a \lambda_t^2 \right] = E \left[ \mathbb{I} \{ \lambda_t - \epsilon_t < z_\alpha \} \lambda_t \right] \lambda_t^2 = E \left[ \Phi(z_\alpha - \lambda_t) \lambda_t^2 \right] .$$  \hspace{1cm} (17)

Second,

$$E \left[ (1 - \mathbb{I}_a) \epsilon_t^2 \right] = E \left[ \mathbb{I} \{ \lambda_t - \epsilon_t \geq z_\alpha \} \epsilon_t^2 \right]$$
$$= E \left[ \epsilon_t^2 | \epsilon_t \leq \lambda_t - z_\alpha \right] P(\epsilon_t \leq \lambda_t - z_\alpha) = \Phi(\lambda_t - z_\alpha) - (\lambda_t - z_\alpha) \phi(\lambda_t - z_\alpha) .$$  \hspace{1cm} (18)
Bringing together the two terms in (17) and (18), we get

\[
\hat{L}(z_\alpha) = E \left[ (1 - \Phi(\lambda_t - z_\alpha)) \lambda_t^2 \right] + E \left[ \Phi(\lambda_t - z_\alpha) - (\lambda_t - z_\alpha) \phi(\lambda_t - z_\alpha) \right]
\]

\[
= p \left[ 1 - \Phi(\xi - z_\alpha) \right] \xi^2 + p \left[ \Phi(\xi - z_\alpha) - (\xi - z_\alpha) \phi(\xi - z_\alpha) \right] + (1 - p) \left[ \Phi(-z_\alpha) + z_\alpha \phi(z_\alpha) \right].
\]

For each \((\xi, p)\) combination, we can graph \(\hat{L}(z_\alpha)\), the rescaled loss function against \(z_\alpha\) or \(\alpha\). Figure 1 presents such graphs for a few \((\xi, p)\) combinations. For most of the \((\xi, p)\) combinations in the figure, the loss function is U-shaped. The figure clearly demonstrates that there is an opportunity to select the truncation threshold optimally. When the jump size is large, the risk of using an overly large threshold is very high.

For each given \((\xi, p)\) combination, we minimize \(\hat{L}(z_\alpha)\) with respect to \(z_\alpha\) and obtain the optimal threshold \(z_\alpha^* = z_\alpha^*(\xi, p)\). For some \((\xi, p)\) combinations, \(\hat{L}(z_\alpha)\) does not have a minimum on the real line but due to continuity, it always has a minimum on a compact set. In the following, we set the compact set to be \([0,4]\). We take zero as the lower bound because the optimal \(z_\alpha^*\) is nonnegative except in the extreme case when both the jump size and the jump probability are very large. We take 4 as the upper bound because it is large enough relative to critical values from the standard normal distribution. Higher upper bound will have only a very small effect on the loss function \(\hat{L}(z_\alpha)\). Figure 2 presents a three-dimensional graph of the optimal threshold function \(z_\alpha^* = z_\alpha^*(\xi, p)\).

A few observations are in order. First, the optimal \(z_\alpha^*\) is located on the boundary of the compact set \([0,4]\) for some \((\xi, p)\) configurations. When the jump probability \(p\) is very small, the optimal threshold hits the upper bound of the compact set. Intuitively, when the jumps do not happen very often, we should use a threshold that is as large as possible and thus employ the efficient QV estimator as often as possible. Similarly, when the jump size \(\xi\) is very small, the optimal threshold \(z_\alpha^*\) should be maximized in order to take advantage of the small contamination from the jumps. When the jump size \(\xi\) is very large, the optimal threshold \(z_\alpha^*\) also hits the upper bound. In this case, in the presence of jumps, the robust BV estimator and the efficient QV estimator are clearly distinguishable. A large threshold that is close to the jump size will be optimal. Such a large threshold is likely to hit the upper bound.

Second, for each fixed \(\xi\), the optimal \(z_\alpha^*\) decreases with \(p\). This is intuitive. When the probability of jumps is higher, we decrease the threshold so that we use the robust BV estimator more often. Figure 3 presents \(z_\alpha^*(\xi, p)\) as a function of \(p\) for a range of \(\xi\) values. When \(\xi\) is very small such as \(\xi = 1\), the optimal threshold takes only two boundary values: \(z_\alpha^* = 4\) when \(p\) is small while \(z_\alpha^* = 0\) when \(p\) is large. For other value of \(\xi\), the optimal threshold starts at \(z_\alpha^* = 4\) and then decreases as \(p\) increases and finally reaches zero when \(p\) approaches 1.

Third, for each given \(p\), the optimal threshold is not a monotonic function of \(\xi\). See Figure 4. It is well expected that the optimal threshold hits the upper bound 4 when the jump size is quite small. It is perhaps surprising that the optimal threshold increases with the jump sizes when the jump size is large enough. Note that the selection indicator \(I_{e_\alpha} = I \{ \lambda_t - e_t < z_\alpha \}\) depends on both \(z_\alpha\) and \(\lambda_t\). For each given
$z_\alpha$, the probability that $I_\alpha = 1$ decreases with the jump size. To some degree, the selection indicator automatically guards against the use of the QV estimator when the jump size is large. Figure 4 shows that this automatic mechanism is conservative and it is advantageous to increase the optimal threshold in response to an increase in the jump size when the jump size is already large.

4 Simulation Analysis

The simple framework in Section 3 clearly shows that there is an opportunity to optimally determine the truncation threshold. In this section, we consider the truncation threshold determination under a more realistic situation. We simulate a jump diffusion process that mimics main features of a financial time series. The simulation experiment provides us with a controlled setting to examine the offsetting forces underlying optimal threshold determination. More importantly, for the simulation, the true volatility process and the jumps are known. We define the term realized variation ($RV_t$) to be the sample analogue of the true volatility process without jumps given by equation (2). The difference between $RV_t$ and $V_t$ is the discretization error, which is unavoidable in simulating a continuous time process numerically. For jump diffusion processes, $RV_t$ is unknown in practice. This setting allows us to analyze the switching estimator relative to $RV_t$ in order to measure its forecasting performance for different levels of truncation threshold.

The strategy for analyzing threshold determination in the simulation setting consists of a number of steps. First, we simulate a representative one-factor stochastic volatility model. Second, we record the summary statistics and present some plots showing that the simulated process resembles the typical time series one would observe in the financial market. Third, we use Corsi’s forecasting model (Corsi (2009)) to forecast the one-step-ahead volatility. This forecasting model is parameterized by the truncation threshold. It is estimated for each threshold value $\alpha$ in a discretized subset of $[5, 1)$. Finally, using a true hold-out sample, we compare the forecasting performance of the optimally selected threshold with that of an ad hoc threshold of 0.999.

To make the selection of threshold operational, we have to deal with two additional problems. The first is a target to measure the prediction loss since in practice the true realized variation $RV_t$ is unknown. The second is an algorithm to produce the prediction error curve that isn’t subject to overfitting. These two problems are considered in turn. We propose using the bipower variation $BV_t$, which is observable in practice, as the plug-in target. By comparing $BV_t$ with the true variation $RV_t$, we show that using $BV_t$ leads to nearly the same threshold choice. To address the second problem, we suggest using the $hv$-block cross-validation method of Racine (2000) to obtain feasible estimates of the out-of-sample prediction errors.
4.1 Simulation Set-up

The process simulated is the jump diffusion process with single factor stochastic volatility (JD1F), which is virtually identical to the process analyzed by Huang and Tauchen (2005). More specifically, we assume that the log price satisfies

\[ dp(t) = \mu \, dt + \exp(\beta_0 + \beta_1 v(t)) \, dW_p(t) + \kappa(t) \, dq(t), \]  
\[ dv(t) = \gamma v(t) dt + dW_v(t), \]  

where the stochastic volatility is modeled as exponential volatility. Nested in the exponential function is another diffusion process \( v(t) \). Equation (20) characterizes the dynamics of \( v(t) \) and motivates the simplified AR(1) setting in Section 3. The two Brownian motion processes \( W_v(t) \) and \( W_p(t) \) are correlated, which induces a correlation between the pricing process and the volatility process. The jump process is modeled as a compound Poisson normal jump process.

We set the drift parameter \( \mu = 0.0004 \), the volatility parameters \( \gamma = -0.1, \beta_0 = 0, \beta_1 = 0.125 \), and \( \text{Corr}(W_v(t), W_p(t)) = \rho_{vp} = -0.62 \). The jump process \( \kappa(t) dq(t) \) has intensity parameter \( \lambda = 0.118 \) and a jump magnitude \( \tau = 1.5 \) where \( P[dq(t) = 1] = \lambda dt \) and \( \kappa \sim N(0, \tau^2) \). These model parameters reflect empirical regularities in the financial sector and are taken mostly from Huang and Tauchen (2005). The only difference is that the drift term we use is smaller. In the absence of the volatility component and jumps, the drift parameter we choose produces approximately 10% yearly return. The drift component is largely inconsequential in the estimation of volatility. For more details on the one-factor stochastic volatility model, see Huang and Tauchen (2005).

The jump diffusion process is simulated continuously in that it does not account for breaks at the end of a day. A total of 5000 days are simulated and split into an in-sample period of 4000 days and an out-of-sample period of 1000 days. The simulation algorithm uses the methods outlined in Chapter 6 of Cont and Tankov (2004). The simulation is produced on a fine scale which we call seconds \( dt = (3600 \times 6.5)^{-1} \), then sampled at a larger scale. We sample at 5-minute intervals, which amounts to sampling every 300th observation. We refer to a day as 6.5 simulated hours of trading. Our ultimate desire is to get a daily measure of variation and predict its future value.

Visually, the simulated prices can be seen in Figure 5. The top panel shows the log prices \( p(t) \), the middle panel the daily returns \( r(t) = p(t) - p(t - 1) \), while in the bottom panel we see the true daily jump process \( \sum_{t-1<s\leq t} \kappa(s) \), which is unknown in practice. The log price process exhibits some variation around the trend but generally it is monotonically increasing over time. The persistence in the variation of returns is clearly demonstrated by the oscillation of the width of the daily returns series. Furthermore, by comparing the return plot to the jump plot below it, we see that large returns are generally induced by the jump process. There are many small jumps; Large jumps are rare, reflecting the use of a normal distribution as the compounding distribution. Large rare jumps are consistent with the common belief about jumps.

We can, in some respects, link this continuous process back to the simplified conceptual framework of Section 3 through parameter choice. Considering first the
volatility, if $\rho_1 = 1 + \gamma$, then the simplified process is a discretized version of the continuous process. The relation to the compound jump process is not as exact; the probability $p$ is related to the jump intensity $\lambda$, and the jump variance $\tau^2$ is related to $\sigma^2 e$ from the simplified process. There are notable differences between this model and the simplified model in Section 3. First, the simplified model does not specify a price process. Second, the jump process is greatly simplified. Third, because we have to sample the continuous-time model on discreet intervals, the bipower variation is still slightly contaminated by jumps. Section 3 models the limiting behavior of the bipower variation, which is free from jumps. Later we will see that the influence of jumps on the bipower variation is quite small. The model in Section 3 is simplified to make the problem analytically tractable while retaining the salient features of the problem. In total, these differences will be inconsequential to the results of Section 3.

For the simulated price process, we compute the daily return $D_{ly} R_t$, daily $QV_t$ according to equation (3), daily $BV_t$ according to equation (5), and $RV_t$ as the sample analogue of equation (2). $QV_t$ and $BV_t$ are computed using 5-minute returns, $RV_t$ is computed using seconds data. In addition, we compute a modified $Z_t$ statistic:

$$Z_t(\Delta) = \Delta^{-1/2} \frac{1 - \frac{BV_t(\Delta)}{QV_t(\Delta)}}{\sqrt{(\mu_1^4 + 2\mu_1^2 - 5) \int_t^{t+1} \max[1, \frac{QV_t(\Delta)}{BV_t(\Delta)}]}}$$

(21)

The modified $Z$-statistic and the original $Z$-statistic from equation (8) are asymptotically equivalent. Barndorff-Nielsen and Shephard (2004) show that the modified $Z$-statistic has better finite sample performance than the original $Z$-statistic. We use this modified $Z$-statistic for all computations in this paper. Figure 6 displays the histogram of the modified $Z$-statistic. Comparing this to the standard normal density, the excess mass is seen in the right tail of the histogram. Days that have a $Z$-statistic of 3 or greater are almost certainly jump days. Based on this observation, we would expect the optimal threshold to be less that 3.

Table 1 presents summary statistics for the simulated jump diffusion process. The statistics of the simulated process are similar to the statistics one would obtain from an empirical time series. Comparing these statistics to those in Table 3 for the S&P 500 futures, we find a notable difference — a smaller Ljung-Box statistic for $QV_t$ for the simulated process. This is likely caused by the purely independent jump process in the simulation. While worthy of note, the difference is not large. The null of no autocorrelation at 10 lags is soundly rejected for both time series. All together, the moments of the simulated process match those of the S&P 500 futures quite well. Given this, we may conclude that the model provides a realistic framework for optimal threshold determination.

Figure 7 shows the daily variation estimates for the simulated JD1F based on 5-minute returns. The top two panels show the daily $QV_t$ and $BV_t$. For comparison, the bottom panel shows the true realized variation $RV_t$, which is not polluted by the jumps and is unknown in practice. Comparing the three panels and taking note of the scale on the vertical axis, we see that $BV_t$ has approximately the same magnitude
as \( RV_t \) and is not nearly as highly polluted by the jumps as \( QV_t \). The switching estimator essentially switches between \( QV_t \) and \( BV_t \) to approximate \( RV_t \). A plot such as this might lead one to believe that the bipower variation would always be the optimal estimator. Both \( QV_t \) and \( BV_t \) seem to capture the underlying trend, the largest difference occurring only on jump days. While theory stands contrary to this perspective, we can also note that if \( BV_t \) were always optimal, then the selected truncation threshold would be \( \alpha = 0.5 \) with \( z_\alpha = 0 \). We find in the next section that a much higher threshold is optimal, thus it pays to use \( QV_t \) sometimes.

### 4.2 In-Sample Prediction

To predict the realized volatility, we use the simple HAR-RV model. This model is due to Corsi (2009) and has been used by Andersen et al. (2007) among others. Its popularity is due in part to its easy implementation as compared to the more complicated fractionally differenced models sometimes used. Furthermore, the Corsi model can capture slow autocorrelation decay, a feature of volatility that has been shown in numerous papers. Overall the model has enjoyed empirical success and popularity as a forecasting algorithm. The simple model (14) in Section 3 is a simplified version of the HAR-RV model.

To implement the HAR-RV procedure, define

\[
V_{t-h,t}^\alpha = h^{-1}(V_{t-h+1}^\alpha + \ldots + V_t^\alpha)
\]

where

\[
V_t^\alpha = \mathbb{I}\{Z_t < z_\alpha\} QV_t + \mathbb{I}\{Z_t \geq z_\alpha\} BV_t
\]

is the switching estimator presented before. The subscript \( \alpha \) on \( V_t^\alpha \) indicates the dependence of \( V_t^\alpha \) on the truncation threshold \( z_\alpha = \Phi^{-1}(\alpha) \). The HAR-RV model is then given by

\[
V_{t+1} = \beta_0 + \beta_1 V_t^\alpha + \beta_2 V_{t-5,t}^\alpha + \beta_3 V_{t-22,t}^\alpha + \epsilon_{t+1}.
\] (22)

The parameters of the HAR-RV model \( (\beta_0, \beta_1, \beta_2, \beta_3) \) can now be estimated simply using OLS. The estimation is carried out only over the 4000 in-sample days. Since the interest is in values \( \alpha \in [0.5, 1) \), the set is discretized and estimation is repeated for each value of \( \alpha \) in the discretized set. Having obtained parameter estimates, we can use the model to forecast and then measure the forecast error for each set of parameters \( \{\alpha, \beta_0(\alpha), \beta_1(\alpha), \beta_2(\alpha), \beta_3(\alpha)\}, \alpha \in [0.5 : 1) \) where :: indicates that the set has been discretized. The exact discretization is up to the users’ preferences for accuracy versus computational time. In what follows the set is discretized as \([0.5 : 0.005 : 0.95], [0.9505 : 0.0005 : 0.9950], [0.9951 : 0.0001 : 0.9999]\). We use the different increments in order to visually improve the plot since our plots use \( z_\alpha \) on the horizontal axis. It should also be noted that in our experience thresholds larger than 0.90 tend to be selected as ‘best’, thus cutting up the interval more finely at higher thresholds will produce more accurate results.

Figure 8 plots the prediction error curves versus the threshold \( z_\alpha \) for the simulated jump diffusion process. The top panel uses \( BV_t \) as the dependent variable in (22)
and the target for prediction while the bottom panel uses $RV_t$. The threshold that minimizes the prediction error when $RV_t$ is used as the target is $z_\alpha = 1.41$ or $\alpha = 0.92$. The optimal threshold when $BV_t$ is used as the target is $z_\alpha = 1.78$ or $\alpha = 0.96$. The difference is small.

The bottom panel plots the prediction error curve determined by using the true volatility $RV_{t+1}$ as the prediction target. In this case, the mean squared prediction error is measured by

$$PE_{TV}(\alpha) = E(\widetilde{V}_{t+1}^\alpha - RV_{t+1})^2$$

(23)

where $\widetilde{V}_{t+1}^\alpha$ indicates forecast from the HAR-RV model and $RV_{t+1}$ is the discrete analogue to $\int_t^{t+1} \sigma^2(s)ds$. $RV_{t+1}$ is the ideal target but in practice we do not have a reasonable measurement of $RV_{t+1}$ when jumps are present. Hence the bottom panel is more of theoretical interest.

The bottom panel of Figure 8 shows that the prediction error curve is U-shaped in the truncation threshold and the minimizing value of $\alpha$ is an interior point. This is consistent with Figure 1. For very large truncation thresholds, say $z_\alpha > 3$, there is a very rapid increase in prediction error due to the bias induced by the jumps. This result is useful to practitioners, where the conventional wisdom is to use a large threshold, sometimes as high as $\alpha = 0.999$ as in Andersen et al. (2007). Given the rapid loss of predictive ability for large $z_\alpha$, a rule of thumb would be to use smaller truncation thresholds such as $\alpha = 0.95$ or even $\alpha = 0.9$ to ensure that we stay away from this bad region of bias. This recommendation is supported by the observation in both the theoretical analysis and the simulation study that the marginal increase in prediction error from the introduction of additional variance to the switching estimator has a smaller impact on predictive loss than the introduction of a large jump-induced bias.

The top panel of Figure 8 gives the feasible prediction error curve with the bipower variation as a target. Mathematically the curve is an approximate measure of

$$PE_{BV}(\alpha) = E(\widetilde{V}_{t+1}^\alpha - BV_{t+1})^2$$

Recall that as $\Delta \to 0$, $BV_{t+1} \to \int_t^{t+1} \sigma^2(s)ds$ so in the limit bipower variation is the appropriate plug-in estimator. Other feasible plug-in targets such as $QV_t$ and the switching estimator itself $V_{t+1}$ are considered. The plots of the prediction error using these plug-in targets do not produce curves that capture the shape of the prediction error curve as accurately as $BV_{t+1}$. This is expected as they are not consistent and therefore will not capture the variation of $RV_t$ well. Using the bipower variation as a target provides a feasible estimator of the prediction error curve. It performs well from an empirical perspective as displayed by Figure 8 and is theoretically appealing because it is consistent for $\int_t^{t+1} \sigma^2(s)ds$.

A natural question is: even if $BV_t$ is the best feasible plug-in target, how big is the difference between using $BV_t$ and $RV_t$? The additional error incurred by using the bipower variation as a plug-in is minimal. Recall, the thresholds that minimize the prediction error when $RV_t$ and $BV_t$ are used as the targets are $z_\alpha = 1.41$ and $z_\alpha = 1.78$ respectively. Approximately 4.5% of the Z-statistics fall between 1.41 and 1.78 and
have been “misclassified”. However, the prediction error curve has a relatively flat bottom around the true threshold, resulting in very little loss in predictive ability from using $BV_t$ as opposed to $RV_t$.

4.3 Out-of-Sample Forecasting

In this subsection the forecast performance of the HAR-RV model is considered over the hold-out-sample. Recall that all of the parameter estimation in the previous section is carried out on the 4000 in-sample days. Based on this analysis we determine that the optimal $\alpha$ is 0.92, which corresponds to $z_\alpha = 1.41$. Using this threshold and its corresponding parameter estimates, we employ the HAR-RV model in (22) to forecast 1000 out-of-sample days that are held out from the original estimation phase. The mean squared prediction error (MSPE) is presented in Table 2. All statistics in this section are computed using the true realized volatility $RV_t$ as the target of estimation and forecasting. We can also use $BV_{t+1}$ as the dependent variable in the forecasting model and select the optimal threshold accordingly. As shown before, the optimal thresholds based on $BV_{t+1}$ or $RV_{t+1}$ are close to each other. So it does not matter much how the optimal threshold is determined.

The question is, how much do we gain in an out-of-sample sense by selecting the optimal threshold as compared to some ad hoc selection? To answer this, we compute and compare the out-of-sample MSPE for the optimally selected threshold with that of an ad hoc threshold of $\alpha = 0.999$ or $z_\alpha = 3.09$. The threshold of $z_\alpha = 3.09$ is not without precedent. This threshold is considered in both Huang and Tauchen (2005) and Andersen et al. (2007). We find that there is a 10% reduction in MSPE from using the optimal threshold compared to the ad hoc threshold. To obtain some idea of the accuracy of this percentage gain, we construct confidence intervals using block bootstrap and recompute the percentage gain for each bootstrap iteration. The block bootstrap is a commonly used bootstrap method designed for dependent data. Politis and White (2004) provide a good explanation of the procedure along with guidance on the selection of the block size. We follow their guidance and perform 1000 bootstrap iterations. 90% and 85% confidence intervals are reported in Table 2. The percentage gain is marginally significant, zero is contained in the 90% confidence interval but not in the 85% confidence interval.

4.4 $h$-Block Cross-Validation for Selecting the Threshold

In Subsection 4.2, we select the optimal threshold on the basis of in-sample prediction errors. In general, the in-sample MSPE is an optimistic estimate of its out-of-sample version. The threshold that optimizes the in-sample performance criterion may not necessarily optimize the corresponding out-of-sample version. Although we have shown that in the present context the optimal threshold determined by the in-sample criterion delivers superior out-of-sample performance, in this subsection, we explore the possibility of selecting the optimal threshold that minimizes a direct estimator of the out-of-sample forecasting errors.
A feasible and robust method for estimating out-of-sample forecast error for independent data is through cross validation. A version of cross validation known as $hv$-block cross-validation is meant to handle dependent data. $hv$-block cross-validation, an extension to leave-$v$-out cross validation, accounts for serial dependence by buffering the out-of-sample validation set by $h$ observations. So the $hv$-block cross-validation combines leave-$v$-out cross validation with $h$-block cross validation.

We prefer $hv$-block cross-validation in part because of the straightforward intuition of the cross validation method and because of its nonparametric nature. These features make it particularly suitable for evaluating nonparametric estimators of the realized volatility.

The $hv$-block cross-validation method can be attributed to Racine (2000) and is essentially a combination of the results on linear cross validation by Shao (1993) and Burman et al. (1994). Shao (1993) shows that for cross validation to select the correct model with probability 1 in the limit, a dominatingly large validation set is required. If $T$ is our sample size and $v$ is our validation set size so that $T - v = n_c$ is the training set size, then the result of Shao (1993) requires that $T \to \infty, n_c \to \infty$, and $T/v \to 1$ for consistent linear model selection with iid data.

Burman et al. (1994) propose an $h$-block cross-validation method, which extends cross validation in a straightforward way to dependent processes. The idea is to reduce the training set by removing the $h$ observations preceding and following the observations in the testing or validation set. When $h \to \infty$, the training set is effectively independent from the testing or validation set. Burman et al. (1994) consider a validation set size of 1, as is the case in the leave-one-out cross validation. Racine (2000) brings these two strains of literature together into what he calls $hv$-block cross-validation. He uses a validation set size large enough to achieve consistent model selection, while using the $h$-block buffering to achieve independence between the validation sets and the training sets. Through simulations Racine (2000) shows that the $hv$-block cross-validation method selects the correct model with high probability even for highly dependent processes. Racine (2000) suggests selecting $v$ such that $v = T - 2h - 1/2(1 - 2h)^{1-\varepsilon}$ and $h = 0.05T$. In our simulation where $T = 5000$ we let $\varepsilon = .01$ so that $v = 2431$ and $h = 250$.

Operationally, the $hv$-block cross-validation statistic can be written as

$$CV_{hv}(\alpha) = \frac{1}{(T - v)v} \sum_{i=1}^{T-v} ||V_{(i,v)}^{\alpha} - RV_{(i,v)}||^2$$

where $||a||^2 = a'a$, $RV_{(i,v)}$ is the true realized volatility from observation $i$ to observation $i + v - 1$. $V_{(i,v)}^{\alpha}(-i;h,v)$ is the forecast over observations $i$ to $i + v - 1$ obtained from a model estimated on a training set, which excludes the validation set, and the $2h$ buffer observations surrounding the training set.

The bottom panel of Figure 9 shows the $hv$-block cross-validated curve. This curve is obtained by simulating a single 5000-day time series, then the $hv$-block cross-validated prediction error is computed by (24) for each threshold in the discretized set $\alpha \in [.5 :: 1]$. The same discretization as Section 4.2 is used.
To evaluate the quality of the $h_v$-block cross-validated error curve, we carry out a Monte Carlo simulation. We simulate the JD1F model from Section 4.1 500 times over 5000 days. To determine the out-of-sample prediction error, we divide each replication into an out-of-sample set and an in-sample set. The out-of-sample set size is the last 1000 days of the simulation and the in-sample set size is the first 4000 days. The same discretized set of thresholds as Section 4.2 is used again. For each threshold the switching estimator is constructed and the parameters of the HAR-RV model given by equation (22) are estimated using the in-sample data. Using these estimates, we forecast over the out-of-sample data. From this forecast we determine the out-of-sample prediction error at that threshold. This process is done for each threshold and then repeated 500 times. Averaging over the 500 replications produces the out-of-sample prediction error curve seen in the top panel of Figure 9.

Comparing the two panels of Figure 9, we can see some noise in the $h_v$-block cross-validated curve. However, the shape is largely preserved so that the trough of the pseudo true curve in the top panel is approximately the same as the trough of the $h_v$-block cross-validated curve. As long as the $h_v$-block cross-validation algorithm chooses a minimum that is within the trough of the true prediction error curve, there will be little loss of predictive ability, as the prediction curve has a relatively flat bottom. Perhaps more importantly, the $h_v$-block cross-validated curve produces the same steep increase at large thresholds. The Monte Carlo analysis shows that $h_v$-block cross-validated curve produces a curve that approximates the true prediction error curve well.

5 Empirical Analysis

The theoretical analysis of Section 3 has shown that an optimal threshold exists. Section 4 presents a simulation analysis and gives a feasible procedure to select the optimal threshold. In this section we perform a threshold selection analysis of the S&P 500 futures market. Data from this market has been used in the work of Andersen et al. (2007) and Huang and Tauchen (2005).

5.1 Data Description

The data analyzed is tick-by-tick data from Jan. 1, 1990 through Dec. 31, 2007. The data was purchased from the Chicago Mercantile Exchange. The method of cleaning and reconstructing the data mimics that of Andersen, Bollerslev, Diebold and Vega (2007). Floor trading hours for the CME run from 9:30am EST to 4:30pm EST. Since the east coast market opens at 8:30am EST, there are a substantial number of trades and consequently initial opening day volatility from electronic trades that occur between 8:30am and 9:30am EST. To account for this additional volatility, we augment the CME floor observations with the Globex data base from 8:30-9:30am EST. We construct five-minute returns by using the observations nearest to the end of the minute interval. Five-minute intervals are the smallest time interval one can use without running into serious microstructure noise problems. Even at this frequency
microstructure noise bias can be a problem. An adjustment for microstructure noise is made by using the one-off variation estimators, which is explained later in Sec 5.2. For five-minute periods that have no observation, we linearly interpolate between the prices of the nearest proceeding and succeeding observations. Days where very few observations are observed are discarded as light trading days. Light trading days are defined as days where there are less than 50 five-minute intervals with at least one observation. For a trading day spanning 8:30am to 4:30pm EST, there are 97 five-minute intervals. In total there are 4489 days after dropping light trading days.

There are four S&P 500 futures contracts that expire every year. These contracts expire on the last day of March, June, Sept. and Dec. Each contract is most heavily traded in the few months preceding its expiration. To concatenate these multiple contracts, we follow the methods employed in Andersen, Bollerslev, Diebold and Vega (2007). We always use the futures contract that is closest to expiration, switching to the next contract 5 trading days before its expiration. This creates a more or less continuous time series from the multiple futures contracts. Plots depicting the constructed daily log prices and daily returns can be found in Figure 10.

5.2 Methodology

We use the same method as in Section 4. Using $hv$-block cross-validation and the HAR-RV forecasting model, we compute the mean squared prediction error for each value in the discretized subset of $[0.5, 1)$ and select the optimal threshold as the value of $\alpha$ that minimizes the prediction error.

Many of the statistics necessary for threshold selection are computed exactly as stated in Section 2. Returns are calculated from log prices as $r(t) = p(t) - p(t-1)$. The quadratic variation estimator ($QV_t$) is calculated in the standard fashion according to equation (3). However for the statistics involving adjacent observations, as is the case with $BV_t$ and $IQ_t$, a microstructure noise adjustment is needed. Market microstructure noise causes a spurious correlation in adjacent returns. To reduce the microstructure noise bias, Andersen et al. (2007) propose one-off estimators of the $BV_t$ and $IQ_t$:

\[
BV_t(\Delta) = \mu_1 \left( \frac{M}{M-3} \right) \sum_{j=2}^{1/\Delta} |r_{t-1+j*\Delta}||r_{t-1+(j-2)*\Delta}|
\]

\[
IQ_t(\Delta) = \mu_4^{3/3} \left( \frac{M}{M-5} \right) \sum_{j=3}^{1/\Delta} |r_{t-1+j*\Delta}^{3/3}||r_{t-1+(j-1)*\Delta}^{3/3}||r_{t-1+(j-2)*\Delta}^{3/3}|
\]

Huang and Tauchen (2005) show that use of the one-off estimators results in a marked reduction in bias. Furthermore, they are both consistent for their respective statistics. For the empirical analysis of the S&P 500 returns we calculate $BV_t$ and $IQ_t$ according to the one-off estimators.

Summary statistics for the S&P 500 Futures data set are given in Table 3. The data is split into an in-sample set where we will carry out all estimation and an out-of-sample set used purely for assessing the optimal threshold. The in-sample
period extends 12 years from Jan. 1, 1990 through Dec. 31, 2002. We can see from these statistics that $QV_t$ is highly leptokurtotic. In contrast, $BV_t$ is not nearly as heavy tailed due to its robustness to jumps. $BV_t$ also has a larger Ljung-Box test statistic indicating that it exhibits more dependence. This greater dependence is also a result of $BV_t$’s robustness to jumps, which to some degree are thought to be serially independent and independent of the diffusion process. Hence, it makes sense that a volatility process purged of independent jumps displays more dependence. $BV_t$ and $QV_t$ can also be compared through their plots in Figure 11. The bias reduction feature of $BV_t$ is not as obvious from the plots here as it is for the simulated JD1F in Section 4.1. However, it is still present nonetheless.

The histogram of the $Z$-statistic in Figure 12 indicates the presence of jumps as we observe an abnormally large number of observations in the right tail of the distribution. Comparing the $Z$-statistic histogram to the standard normal density, we see that there are many more medium sized jumps for this time series than for the simulated JD1F. The presence of many medium size jumps makes threshold selection even more critical. As is previously shown through simulation, the introduction of bias from jumps has the largest marginal impact on prediction error. Thus, the gains from selecting a larger threshold may be substantial here. However, it is interesting to note that the threshold selected is $z_\alpha = 1.598$, which is approximately where we see the histogram of the $Z$-statistic begin to protrude from the normal density.

### 5.3 Empirical Results

The important distinction between the observed data and the simulated data is that we observe only $QV_t$ and $BV_t$ but not $RV_t$. This has the consequence of forcing us to use $BV_t$ as the target of the loss function. Figure 13 displays the prediction error curve for the switching estimator under the MSPE criterion. This prediction error curve is obtained through $hv$-block cross-validation where $h$ and $v$ are 161 and 750 respectively. The prediction error curve is largely U-shaped in the threshold and increases rapidly with the truncation threshold when the truncation threshold is large. These qualitative observations are consistent with our simulation results. We see that the prediction error curve reaches a minimum at 0.945 or $z_\alpha = 1.598$.

Table 4 displays results from one-step-ahead forecasts over the out-of-sample period of Jan. 2003 to Dec. 2007. All parameter estimates for the model are carried out on the in-sample period of Jan. 1990 through Dec. 2002. Under the MSPE criterion, we observe 4% reduction in out-of-sample predictive ability relative to the ad hoc threshold level of 0.999. As before, we employ block bootstrap to construct the CI for the percentage gain. The 95% CI is [0.0238 0.0716] which does not contain 0. This result shows that the optimal threshold significantly out performs the ad hoc threshold.
6 Conclusion

This paper proposes a new practical method to select the truncation threshold in jump determination that is optimal from a forecasting perspective. The selected threshold minimizes an out-of-sample forecasting loss. A rigorous theoretical analysis shows that the prediction error curve is convex in the truncation threshold so that a minimizing threshold exists. To find this optimal threshold, we use the bipower variation estimate as the validation target and employ $h_v$-block cross-validation to obtain a robust and nonparametric estimator of the out-of-sample forecasting loss. An extensive simulation study and an empirical application show that using optimal threshold can reduce the forecasting loss by 5% to 10% as compared to ad-hoc thresholds previously used in the literature.

The idea of the paper can be used to select the truncated threshold for other purposes. For example, we may be interested in forming optimal portfolios on the basis of realized variances and covariances. In this case, an economic loss should be employed to determine the optimality of the truncation threshold. We leave this for future research.
References


Table 1: Summary Statistics of the Simulated One-factor Jump-diffusion Process. Daily Returns ($DlyRet_t$) are computed using the end-of-day observations. The quadratic variation ($QV_t$) and bipower variation ($BV_t$) are computed using 5 min returns. The realized variation ($RV_t$) is computed using seconds data. LB10 is the Ljung-Box test statistic for autocorrelation of order 10.

<table>
<thead>
<tr>
<th></th>
<th>$DlyRet_t$</th>
<th>$QV_t$</th>
<th>$BV_t$</th>
<th>$RV_t$</th>
<th>$Z_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.075</td>
<td>1.475</td>
<td>1.193</td>
<td>1.163</td>
<td>0.529</td>
</tr>
<tr>
<td>Median</td>
<td>0.043</td>
<td>1.040</td>
<td>1.005</td>
<td>0.991</td>
<td>0.116</td>
</tr>
<tr>
<td>St Dev</td>
<td>1.194</td>
<td>1.827</td>
<td>0.757</td>
<td>0.676</td>
<td>2.020</td>
</tr>
<tr>
<td>Skew</td>
<td>0.201</td>
<td>8.332</td>
<td>2.079</td>
<td>1.820</td>
<td>2.441</td>
</tr>
<tr>
<td>LB10</td>
<td>10.288</td>
<td>440.261</td>
<td>12297.809</td>
<td>18324.198</td>
<td>6.188</td>
</tr>
</tbody>
</table>

Table 2: Out-of-Sample Forecasting Performances of Different Threshold Selection Methods. Percentage Gain is the percentage gain in MSPE of using the optimal threshold against using the ad hoc threshold. Confidence intervals for the percentage gain is constructed by block bootstrap.

<table>
<thead>
<tr>
<th></th>
<th>Optimal Thresholding</th>
<th>Ad Hoc Thresholding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold Selected</td>
<td>1.410</td>
<td>3.090</td>
</tr>
<tr>
<td>Mean Squared Prediction Error</td>
<td>0.132</td>
<td>0.148</td>
</tr>
<tr>
<td>Percentage Gain</td>
<td>0.106</td>
<td></td>
</tr>
<tr>
<td>90% CI for Percentage Gain</td>
<td>$[-0.0006, 0.2182]$</td>
<td></td>
</tr>
<tr>
<td>85% CI for Percentage Gain</td>
<td>$[0.0037, 0.2019]$</td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Summary Statistics for the S&P 500 Futures. The in-sample period is from Jan.1990 to Dec.02 while the out-of sample period is from Jan.2003 to Dec.2007. Quadratic and bipower variation are computed using 5 min returns. LB10 gives the Ljung-Box Test statistic for autocorrelation up to order 10.

<table>
<thead>
<tr>
<th>Stats</th>
<th>$DlyRet_t$</th>
<th>$QV_t$</th>
<th>$BV_t$</th>
<th>$Z_t$</th>
<th>In-sample Periods, 3253 days</th>
<th>Out-of-sample Periods, 1236 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.009</td>
<td>1.114</td>
<td>1.031</td>
<td>0.707</td>
<td>0.004</td>
<td>0.723</td>
</tr>
<tr>
<td>Median</td>
<td>0.034</td>
<td>0.661</td>
<td>0.616</td>
<td>0.554</td>
<td>0.040</td>
<td>0.461</td>
</tr>
<tr>
<td>St.Dev</td>
<td>1.005</td>
<td>1.657</td>
<td>1.445</td>
<td>1.425</td>
<td>0.741</td>
<td>0.813</td>
</tr>
<tr>
<td>Skew</td>
<td>−0.192</td>
<td>6.998</td>
<td>5.750</td>
<td>1.115</td>
<td>−0.257</td>
<td>4.657</td>
</tr>
<tr>
<td>$LB_{10}$</td>
<td>13.862</td>
<td>5297</td>
<td>8478</td>
<td>101.63</td>
<td>19.835</td>
<td>2532</td>
</tr>
</tbody>
</table>

Table 4: Out-of-Sample Forecasting Performances of Different Threshold Selection Methods. Percentage Gain is the percentage gain in MSPE of using the optimal threshold against using the ad hoc threshold. 95% confidence interval for the percentage gain was constructed using a block bootstrap.

<table>
<thead>
<tr>
<th></th>
<th>Optimal Thresholding</th>
<th>Ad Hoc Thresholding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold Selected</td>
<td>1.598</td>
<td>3.090</td>
</tr>
<tr>
<td>Mean Squared Prediction Error</td>
<td>0.222</td>
<td>0.231</td>
</tr>
<tr>
<td>Percentage Gain</td>
<td>3.97%</td>
<td></td>
</tr>
<tr>
<td>95% CI for Percentage Gain</td>
<td>[0.0238, 0.0716]</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1: Rescaled loss as a function of the truncation threshold $z_\alpha$

Figure 2: Figure of $z_\alpha(\xi, p)$ as a function of jump size $\xi$ and jump probability $p$
Figure 3: Graph of $z^*_\alpha(\xi, p)$ as a function of jump probability $p$ for different jump sizes $\xi$.

Figure 4: Graph of $z^*_\alpha(\xi, p)$ as a function of $\xi$ for different jump probabilities $p$. 

27
Figure 5: Simulated one-factor jump-diffusion Process. Top Panel: end-of-day log prices. Middle Panel: daily returns. Bottom Panel: daily compound jump process (sum of jumps within a day) with Poisson intensity $\lambda = 0.118$ and jump sizes $N(0, 1.5^2)$.

Figure 6: Histogram of the daily modified $Z$-statistic based on the simulated one-factor jump-diffusion process.
Figure 7: Daily variation estimates of the simulated jump diffusion process with one-factor stochastic volatility based on 5-minute returns. Top Panel: daily Quadratic Variation ($QV_t$). Middle Panel: daily Bipower Variation ($BV_t$). Bottom Panel: true daily Realized Variation ($RV_t$)
Figure 8: Mean squared prediction error as a function of the truncation threshold for the simulated jump diffusion process. Top Panel: Mean squared prediction error using the Bipower Variation as the target of the loss function. Bottom Panel: Mean squared prediction error using the true Realized Variation as the target of the loss function.
Figure 9: Top Panel: Monte Carlo estimate of the out-of-sample prediction error for the JD1F model. The in-sample set size is 4000 while the out-of-sample set size is 1000. 500 MC iterations are performed. Bottom Panel: hv-block cross-validated prediction error of a single simulated time series with v=2431 and h=250.
Figure 12: S&P 500 Futures. Histogram of the modified Z-statistic compared to a standard normal density.

Figure 13: S&P 500 Futures. Mean squared prediction error curve as a function of the truncation threshold.