Volatility Forecasting with High Frequency Data

Youngjun Jang

Department of Economics
Stanford University
Stanford, CA 94305
jyjglory@gmail.com

Advisor: Professor Peter Hansen

May 27, 2007

ABSTRACT

The daily volatility is typically unobserved but can be estimated using high frequent tick-by-tick data. In this paper, we study the problem of forecasting the unobserved volatility using past values of measured volatility. Specifically, we use daily estimates of volatility based on high frequency data, called realized variance, and construct the optimal linear forecast of future volatility. Utilizing single exponential smoothing, we develop formulae that yield the optimal coefficients for our forecast. We compare the precision of our forecast with those of two popular forecasting models, the HAR regression model and the Local Level model, in terms of mean squared errors. In empirical analysis of the seven DJIA stocks, our model performs better than two competing models in most of the cases.

Keywords: Realized Variance, High-Frequency Data, Instrumental Variables, Local Level Model, HAR model, linear forecasting
1. Introduction

In common terms, volatility refers to ‘the fluctuations observed in some phenomenon.’ In economic terms, it refers to ‘the variability of the random component of a time series.’ Specifically, in financial economics, volatility can be defined as ‘the standard deviation of the random Wiener-driven component in a continuous-time diffusion model’ (Andersen et al. 2005.)

Measuring and forecasting volatility has a wide range of applications. Point forecasting, interval forecasting, probability forecasting, and density forecasting are all good examples of statistical applications. For financial applications, risk management, asset allocation with time-varying covariances, and option valuation with dynamic volatility are representative ones (Andersen et al. 2005.) In their seminal paper “Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts”, Andersen and Bollerslev (1998) showed that volatility models produce good forecasts for the latent volatility factor which is important for many financial applications.

CAPM (Capital Asset Pricing Model) assumes that investors are only concerned with the rate of return and risk, which is represented as a standard deviation of returns. However, the standard deviation of the return might be too simple to represent the risk facing investors. Return volatility is a primary input to option pricing and portfolio allocation problems. Moreover, good forecasts of volatility are essential to implement and evaluate asset pricing, trading, and hedging theories (Andersen and Bollerslev, 1998.) Recently, complicated methods such as Value-at-Risk (VaR) and Expected Shortfall (ES) that utilize volatility forecasting techniques have been used to measure risks more accurately.
Forecasting daily volatility is central for financial risk managements. The Realized Variance (RV), which is a sum of squared intraday returns, is the simplest measure of daily volatility based on high-frequent data. A drawback of this estimator is that it is sensitive to market microstructure noise, such as rounding errors in the intraday prices. So the empirical estimates of daily volatility used in this paper are the kernel-based estimates of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2007). These are closely related to the RV, but robust to market microstructure noise. For simplicity, we will simply refer to these estimates as the RV. Among several methods, the HAR (Heterogeneous Autoregressive model of the Realized Volatility) model and the LL (Local Level) model are widely used to forecast volatility, since they show nice predictive performance empirically and can be easily computed.

In this paper, we will model the latent volatility with an Autoregressive (AR) model. For simplicity, we will mainly focus on an AR(1) model, however an AR(2) model with a unit root will also be used. The problem we seek to solve is to determine future volatility expressed as a linear combination of previous days’ RV, including today’s RV.

Section 2 summarizes the related research. In Section 3, we detail our notation and study several models. In Section 4, seven equities of Dow Jones Industrial Average (DJIA) data that are extracted from Trade and Quote (TAQ) database, sampled from 2001 to 2004, will be used to determine optimal coefficients for models. The precision of a forecast will be measured by the value of the mean squared error (MSE) that is calculated by comparing the daily forecasts in 2005 with the observed volatilities, RVs. In Section 5, we conclude with suggestions for future research and discussion of our results.
2. Literature Review

Since the joint distributional characteristics of asset returns are essential to evaluate the risk-return tradeoff, they are primary factors to price financial instruments. Moreover, they are important to manage risks as they are related to the conditional portfolio return distributions which decide the likelihood of extreme shifts in portfolio value (Andersen et al. 2003). The second moment structure of the conditional distributions is the major element of time-varying characteristics of the distribution. Thus, many researchers have investigated forecasting and modeling return volatility in depth (Andersen et al. 2003).

With the development of computer and network systems, data can be collected for shorter time horizons. This improvement in the availability of data changed the focus from quarterly or monthly modeling to weekly or daily modeling. Further, intraday or tick-by-tick data analyses have also become possible.

Volatility forecasting is especially effective on a high-frequency basis, such as hourly or daily (Christoffersen and Diebold, 2000). However, the standard volatility models used to forecast at the daily level cannot accommodate intraday information in detail, and also models developed for intraday level cannot explain daily data sufficiently well (Andersen et al. 2003). This discrepancy is caused by the presence of market microstructure noise in high-frequency financial data. Market microstructure noise makes standard estimators such as RV unreliable, since it induces autocorrelation in the intraday returns which makes estimators biased (Hansen and Lunde, 2006). In response to this problem, Barndorff-Nielsen et al. (2007) develop the class of realized kernel estimators of quadratic variation. As already mentioned in the introduction, in this paper, we will use these realized kernel-
based estimates and call these estimates as RV for simplicity.

Corsi presents a simple AR-type model of the realized volatility that considers volatilities realized over different time horizons (2004). He terms his model a Heterogeneous Autoregressive model of the Realized Volatility (HAR). The HAR model is based on the HARCH model of Muller et al. (1997) which was inspired by Heterogeneous Market Hypothesis and the asymmetric propagation of volatility between various time horizons. Using 12 years (from December ’89 to July 2001) of tick-by-tick logarithmic middle prices of USD/CHF FX rates, Corsi finds that his simple HAR model successfully reproduces some of the main empirical characteristics of the financial data: fat tails, long memory in the volatility, and distributional properties of realized volatility (2004). Beginning with the HAR model, we will use the next section to discuss theoretical results of several models, including the LL model and an AR(1) model.
3. Theoretical Framework

First of all, we distinguish actual volatility from observed volatility. Actual volatility is a latent variable, while observed volatility is a noisy estimator of actual volatility. For example, the daily integrated variance is actual volatility and the realized variance, RV, is observed volatility (Hansen and Lunde, 2006).

Suppose prices are given by $dp_t = \sigma_t dW_t$ where $W_t$ is a standard Brownian process, $p_t$ is the logarithm of instantaneous price, and $\sigma_t$ is the time-varying volatility. Integrated variance associated with day $t$ is defined by the integral of the instantaneous volatility over the time interval, $\int_{t-1}^t \sigma_s^2 ds$. The standard definition of the RV associated with day $t$, is $RV_t = \Delta \sum_{j=0}^{M-1} r_{t-j\Delta}$, where $\Delta = \frac{1d}{M}$ and $r_{t-j\Delta}$ = intraday returns sampled at time interval $\Delta$ (Corsi, 2004).

Barndorff-Nielsen et al. (2007) define

$$\gamma_h(X_{\delta}) = \sum_{j\delta} (X_{j\delta} - X_{(j-1)\delta})(X_{(j-h)\delta} - X_{(j-h-1)\delta}), \quad h = -M, \ldots, -1, 0, 1, 2, \ldots, M$$ for any process $X$. In this paper, we define the realized kernel $K(X_{\delta}) = \sum_{h=\delta}^{H} w_h \gamma_h(X_{\delta})$ and call this as RV for simplicity. $w_h$ are non-random weights here. See Barndorff-Nielsen et al. (2007) for details.

We use notation $X_t$ to denote log (RV = realized variance = observed volatility) at time $t$, and $Y_t$ to denote log (Integrated variance = actual volatility) at time $t$. The difference between actual volatility and observed volatility can be represented as a measurement error $\eta_t$, where $\{\eta_t\}$ are iid white noise that are not correlated with $X_t$ and
Mathematically, \( Y_t = X_t + \eta_t \).

We want to find optimal linear forecast of \( Y_{t+1} \) given \( X_T, X_{T-1}, \cdots X_1 \). We define the optimality to minimize \( E\left[ (Y_{t+1} - \hat{X}_{t+1})^2 \right] \) where \( \hat{X}_{t+1} = f(X_T, X_{T-1}, \cdots X_1) = \sum_{i=1}^T w_i X_{T+i} \).

Since \( E\left[ (X_{t+1} - \hat{X}_{t+1})^2 \right] = E\left[ (Y_{t+1} + \eta_{t+1} - \hat{X}_{t+1})^2 \right] \)
\( = E(\eta_{t+1}^2) + 2E(\eta_{t+1} (Y_{t+1} - \hat{X}_{t+1})) + E\left[ (Y_{t+1} - \hat{X}_{t+1})^2 \right] = \sigma^2 + E\left[ (Y_{t+1} - \hat{X}_{t+1})^2 \right] \) and \( \sigma^2 \) is a constant, minimizing \( E\left[ (Y_{t+1} - \hat{X}_{t+1})^2 \right] \) is same as minimizing \( E\left[ (X_{t+1} - \hat{X}_{t+1})^2 \right] \).

As \( E\left[ (X_{t+1} - \hat{X}_{t+1})^2 \right] \) is easier to compute, we will focus on minimizing this term from now on.

**I. HAR Model**

Following Corsi (2004), we define weekly and monthly RV at time \( t \) as

**Weekly RV at time t**
\[
RV_t^{(w)} = \frac{1}{5} \left( X_{t+1} + X_{t-2} + \cdots + X_{t-5} \right) = \frac{1}{5} \sum_{i=1}^5 \log(RV_{t-i})
\]

**Monthly RV at time t**
\[
RV_t^{(m)} = \frac{1}{22} \left( X_{t+1} + X_{t-2} + \cdots + X_{t-22} \right) = \frac{1}{22} \sum_{i=1}^{22} \log(RV_{t-i})
\]

Note that one week has five trading days and one month has 22 trading days on average.

Corsi (2004) proposes the HAR model that estimates \( X_t \) from \( X_{t-1} \), \( RV_t^{(w)} \), \( RV_t^{(m)} \), by regressing \( X_t \) on \( X_{t-1}, \frac{1}{5} \sum_{j=1}^5 X_{t-j}, \frac{1}{22} \sum_{j=1}^{22} X_{t-j} \), and 1.

In short, \( \hat{X}_t = w^{(d)} X_{t+1} + w^{(w)} \left( \frac{1}{5} \sum_{i=1}^5 X_{t-i} \right) + w^{(m)} \left( \frac{1}{22} \sum_{i=1}^{22} X_{t-i} \right) + w^{(c)} \cdots \) (1)

(\( w^{(c)} \) is a constant term.)
II. Local Level Model

Local Level (LL) model starts with \( Y_t = Y_{t-1} + \eta_t \) and \( Y_t = Y_{t-1} + \varepsilon_t \), where \( \{\eta_t\} \) and \( \{\varepsilon_t\} \) are iid white noise and independent of each other. Since \( Y_{t+1} - X_t = \sum_{s=1}^{t+1} \varepsilon_s - \eta_t \),

\[
\begin{bmatrix}
Y_{t+1} - X_T \\
Y_{t+1} - X_{T-1} \\
Y_{t+1} - X_1
\end{bmatrix} = \sigma^2_x \begin{bmatrix}
1 & 1 & 1 & \cdots & 1 \\
1 & 2 & 2 & \cdots & 2 \\
1 & 2 & 3 & \cdots & 3 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 2 & 3 & \cdots & T
\end{bmatrix} + \sigma^2_\eta I_T \quad (= A)
\]

We want to find the optimal linear combination of \( X_T, X_{T-1}, \cdots, X_1 \), our goal is to choose optimal \( w = (w_1, \cdots, w_T)' \) to minimize \( \text{var}(\hat{Y}_{T+1} - \hat{Y}_{T+1}(\omega)) = w'Aw \) such that \( w'1 = 1 \) where \( t = (1, \cdots, 1)' \), \( \hat{Y}_{T+1}(\omega) = \sum_{s=1}^{T} w_s X_{T+1-s} \), \( \sum_{t=1}^{T} w_t = 1 \), \( w = (w_1, \cdots, w_T)' \). Using the method of Lagrange multiplier, we derive \( w^* = A^{-1}t' (t'A^{-1}t)^{-1} \) from the first order conditions. Note that \( w^* = (0, \cdots, 0)' \) when \( \sigma^2_\eta = 0 \), as should be the case in the absence of measurement errors.

Define \( \lambda = \frac{\sigma^2_\eta}{\sigma^2_\varepsilon} \) and \( A_\lambda = \frac{\lambda^2}{\sigma^2_\varepsilon} \)

\[
\begin{bmatrix}
1 & 1 & 1 & \cdots & 1 \\
1 & 2 & 2 & \cdots & 2 \\
1 & 2 & 3 & \cdots & 3 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 2 & 3 & \cdots & T
\end{bmatrix} + \lambda I_T \quad \text{Thus, we find the optimal weights given by } \quad w^*_\lambda = A_\lambda^{-1}t' (t'A_\lambda^{-1}t)^{-1}.
\]

Theoretically, for LL model, \( \text{corr}(\Delta X_t, \Delta X_{t-1}) = -\frac{1}{2 + \frac{1}{\lambda}} \) (Harvey, 1993) \( \quad (2) \).

We can estimate \( \hat{\lambda}_{LL} \) by computing \( \text{corr}(\Delta X_t, \Delta X_{t-1}) \) from 2001~2004 data. For
instance, GM (General Motors) stock has a correlation -0.427 and \( \hat{\lambda}_{LL} = 2.906 \). After calculating optimal weights, we estimate
\[
\hat{X}_t = \sum_{h=1}^{N} w_h X_{t-h} \quad \cdots \quad (3)
\]

\( N \) is the number of previous days we use to estimate. For consistency with the HAR model, \( N = 22 \) is used in this paper. Theoretically, as \( T \to \infty \), \( \{ w_h \} \) forms a geometric sequence.

In particular, \( w_{\lambda, h}^* = (1 - p) p^{h-1} \) where \( p = \frac{\hat{\lambda}}{\lambda + 1} \) \( \cdots \quad (4) \) (Hansen, 2007)

### III. Stationary AR(1)

More generally, for \( Y_t = \varphi Y_{t-1} + \varepsilon_t \) (where \( \varphi < 1 \)) we get

\[
A_\varphi = \frac{A}{\sigma_\varepsilon^2} = \begin{pmatrix}
1 & 1 & 1 & 1 & \cdots & 1 \\
1 + \varphi^2 & 1 + \varphi^2 & 1 + \varphi^2 & \cdots & 1 + \varphi^2 \\
1 + \varphi^2 & 1 + \varphi^2 + \varphi^4 & 1 + \varphi^2 + \varphi^4 & \cdots & 1 + \varphi^2 + \varphi^4 \\
1 + \varphi^2 & 1 + \varphi^2 + \varphi^4 & \cdots & 1 - \frac{\varphi^{2T}}{1 - \varphi^2} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots& \frac{\varphi^2}{
\end{pmatrix}

Therefore, we can compute optimal weights if we know, or estimate, values of \( \varphi \) and \( \lambda \).

For LL model, \( \varphi = 1 \). After calculating optimal weights, we estimate \( \hat{X}_t = \sum_{h=1}^{N} w_h X_{t-h} \).

### IV. AR(2) with unit roots

For AR(2) case, we can find optimal weights similarly. Note that this procedure does not require a unit root assumption.

\[
X_t = Y_t + \eta_t, Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varepsilon_t \quad \text{for AR(2).} \quad \text{By iterating, we get}
\]
\[
\begin{align*}
&\text{Var} \begin{pmatrix}
Y_{t+1} - a_1 X_t - b_1 X_{t-1} \\
Y_{t+1} - a_2 X_t - b_2 X_{t-2} \\
\vdots \\
Y_{t+1} - a_{t-1} X_t - b_{t-1} X_1
\end{pmatrix} = \text{Var} \begin{pmatrix}
\varepsilon_{t+1} - a_1 \eta_t - b_1 \eta_{t-1} \\
\varepsilon_{t+1} + a_2 \varepsilon_t - a_2 \eta_{t-1} - b_2 \eta_{t-2} \\
\quad \vdots \\
\varepsilon_{t+1} + a_{t-2} \varepsilon_t + \cdots + a_{t-2} \varepsilon_1 - a_{t-1} \eta_1 - b_{t-1} \eta_t
\end{pmatrix} \\
&= \sigma_x^2 \begin{pmatrix}
c_0 & c_0 & \cdots & c_0 \\
c_0 & c_1 & \cdots & c_1 \\
\vdots & \vdots & \ddots & \vdots \\
c_0 & c_1 & \cdots & c_{T-2}
\end{pmatrix} + \sigma_q^2 \begin{pmatrix}
d_0 & e_0 & 0 & \cdots & 0 \\
0 & d_1 & e_1 & 0 & \vdots \\
0 & 0 & \ddots & \ddots & 0 \\
\vdots & \vdots & \ddots & \ddots & e_{T-2} \\
0 & 0 & \cdots & 0 & d_{T-2}
\end{pmatrix}
\end{align*}
\]

where
\[
c_t = \sum_{k=0}^i a_k^2, d_t = \sum_{k=1}^{i+1} (b_k^2 + c_k^2), e_t = a_{i+2} b_{i+1} \text{ for } a_0 = 1. \text{ This is a special case of AR}(p).
\]

General cases are analyzed in Hansen and Lunde (2006).

**V. Estimation with Instrumental Variable** \( Z_t = X_{t-3} \)

We use a simple AR(1) model: \( X_t = Y_t + \eta_t \) and \( Y_t = \varphi Y_{t-1} + \mu + \varepsilon_t \) (\( \varphi < 1 \)) and use \( X_{t-3} \) as an instrumental variable \( Z_t \) (Since applying \( X_{t-3} \) gives better estimates, we use it instead of \( X_{t-2} \)). Then Ordinary Least Squares estimator is
\[
\hat{\varphi}_{LS} = \frac{\sum (X_{t-1} - \overline{X}_{t-1})(X_t - \overline{X})}{\sum (X_{t-1} - \overline{X}_{t-1})} \quad \ldots \quad (5)
\]
and Instrumental Variable estimator is
\[
\hat{\varphi}_{IV} = \frac{\sum (X_{t-3} - \overline{X}_{t-3})(X_t - \overline{X})}{\sum (X_{t-3} - \overline{X}_{t-3})(X_{t-1} - \overline{X}_{t-1})} \quad \ldots \quad (6)
\]

\[
\frac{\hat{\varphi}_{LS}}{\var{X}_{t-1}} \rightarrow \frac{\cov(X_t, X_{t-1})}{\var(X_{t-1})} = \frac{\cov(Y_t + \eta_t, Y_{t-1} + \eta_{t-1})}{\var(Y_{t-1} + \eta_{t-1})} = \varphi \frac{\var{\eta}_t^2}{\sigma_x^2} = \varphi \frac{1 - \varphi^2}{\sigma_x^2} = \varphi \frac{\var{\eta}_t^2}{\sigma_x^2 + \sigma_\eta^2} = \varphi \frac{1 - \varphi^2}{1 - \varphi^2 + \sigma_\eta^2} = \varphi \frac{1}{1 + \lambda(1 - \varphi^2)}
\]

\[
\varphi_{LS} = \lim(\hat{\varphi}_{LS}) \quad \text{and} \quad \varphi_{IV} = \lim(\hat{\varphi}_{IV}). \text{ Then we get} \quad \lambda_{IV} = \frac{\varphi_{IV} - 1}{\varphi_{IV} - 1} \quad \ldots \quad (7).
\]

Single exponential smoothing is a very popular scheme to produce a smoothed time
series that assigns exponentially decreasing weights on past observations. Let \( a_t \) be an
observation and \( s_t \) be a smoothed value for time \( t \). Starting from \( s_2 = a_1 \), we recursively
compute \( s_t \) from the basic equation of exponential smoothing, \( s_t = \beta a_{t-1} + (1 - \beta) s_{t-1} \) for \( t \geq 3 \) (the expanded equation is \( s_t = \beta \sum_{i=1}^{t-2} (1 - \beta)^{-1} a_{t-i} + (1 - \beta)^{-2} a_1 \)). Note that recent
observations are given relatively more weight in forecasting than the older observations are.
\( \beta \) is called the smoothing constant; \( \varphi \) is the smoothing constant in our model.

Utilizing single exponential smoothing, we substitute \( \mu \) with \((1 - \varphi)\delta\). Instead of
\( \delta = X_{t-1} (s_{t-1} \text{ in the basic equation}), \) we set \( \delta = \frac{1}{T} \sum_{r=1}^{T} X_r \) in our model. Intuitively, \( \delta \) is
an ‘average’ of all past values, instead of a single past observation or a single smoothed value.
We have \( Y_t = \varphi Y_{t-1} + \mu + \epsilon_t = \varphi Y_{t-1} + (1 - \varphi)\delta + \epsilon_t \Leftrightarrow (Y_t - \delta) = \varphi (Y_{t-1} - \delta) + \epsilon_t \).
Since we have estimates of \( \varphi \) and \( \lambda \), we can achieve optimal weights \( w^* = (w_1, \cdots, w_T)' \).

Then, \( (\hat{X}_t - \delta) = \sum_{h=1}^{N} w_h (X_{t-h} - \delta) = \sum_{h=1}^{N} w_h X_{t-h} - \sum_{h=1}^{N} w_h \delta \).

Therefore, \( \hat{X}_t = \sum_{h=1}^{N} w_h X_{t-h} + \left(1 - \sum_{h=1}^{N} w_h \right) \delta \quad \cdots \quad (8) \)

Note that \( \sum_{h=1}^{N} \varphi^{-h} w_h = 1 \), by the definition of the exponential smoothing.

Initially, we set \( \delta = \frac{1}{T} \sum_{r=1}^{T} X_r \). Here, \( \delta \) represents an average of all previous
values available. Yet, it is plausible to define \( \delta \) as an average of the recent values. For
example, we can take an average of past month \( (\delta = \frac{1}{N} \sum_{h=1}^{N} X_{t-h} \text{ where } N = 22) \), past
quarter \( (\delta = \frac{1}{3N} \sum_{h=1}^{3N} X_{t-h}) \), or past year \( (\delta = \frac{1}{250} \sum_{h=1}^{250} X_{t-h}) \). We will examine all suggested
δ to find which δ provides best estimates, i.e., estimates that give minimum mean squared error (MSE).

Instead of estimating λ from $\lambda_{IV} = \frac{\phi_{LS}}{1-\phi_{IV}}$, we can estimate λ and φ by the prediction based criterion (Hansen and Lunde, 2006). We define $Q(\lambda, \phi) = \frac{1}{T} \sum_{t=1}^{T} (X_t - \hat{Y}_t)^2$, where $\hat{Y}_t = \sum_{h=1}^{N} w_h X_{t-h}$ and $T$ = the last day of the available data, December 31, 2004 in this case. Then $Q$ is a sample MSE function; we estimate $(\hat{\lambda}_{Q,IV}, \hat{\phi}_{Q,IV}) = \arg \min_{\lambda, \phi} Q(\lambda, \phi) \ldots$ (9).

We also test this new set of estimators $(\hat{\lambda}_{Q,IV}, \hat{\phi}_{Q,IV})$; keep $\varphi_{IV}$, but use $\hat{\lambda}_{Q,IV}$ instead of $\lambda_{IV}$. For comparison, we also apply this method to the LL model and estimate $\hat{\lambda}_{Q,LL}$ instead of $\hat{\lambda}_{LL}$ from (2), $\text{corr}(\Delta X_t, \Delta X_{t-1}) = -\frac{1}{2 + \frac{1}{\hat{\lambda}}}$.

VI. Estimation of AR(2)

For this section, we assume $\varphi_1 + \varphi_2 = 1$. As this unit root condition will eliminate the constant in the model, we can use $Q(\lambda, \phi) = \frac{1}{T} \sum_{t=1}^{T} (X_t - \hat{Y}_t)^2$ where $\hat{Y}_t = \sum_{h=1}^{N} w_h X_{t-h}$ as we did in the previous section (Without the unit root condition, the formula for $\hat{Y}_t$ becomes $\hat{Y}_t = \sum_{h=1}^{N} w_h X_{t-h} + g(\hat{\mu})$.) We are especially interested in the values of $\varphi_2$; if $\varphi_2$ is close to zero, it is likely that applying an AR(2) model by adding an additional term from an AR(1) model does not improve forecasting significantly.
VII. Possible Extension: Multiple Instrumental Variables

We start from an AR(1) model, \( X_t = Y_t + \eta_t \) and \( Y_t = \varphi Y_{t-1} + \mu + \varepsilon_t \). We used \( Z_t = X_{t-3} \) in part \( V \) as an instrumental variable. In this section, we consider multiple instrumental variables, \( X_{t-2}, X_{t-3}, \ldots, X_{t-k} \). We can estimate \( \varphi \) by using the two stage least squares (TSLS) estimator. Let \( Z_t = (1, X_{t-2}, X_{t-3}, \ldots, X_{t-k})' \).

\[
X_t = Y_t + \eta_t = \varphi Y_{t-1} + \mu + \varepsilon_t + \eta_t = \varphi X_{t-1} + \mu + (\varepsilon_t + \eta_t - \varphi \eta_{t-1}) \quad \cdots \quad (\ast)
\]

For the first-stage regression of TSLS, we regress \( X_{t-1} \) on \( Z_t \). We get

\[
X_{t-1} = \Pi' Z_t + \nu_t, \quad \text{where} \quad \Pi = (\pi_0, \pi_1, \ldots, \pi_{k-1})'.
\]

By OLS regression, \( \hat{\Pi}' = (Z_t, Z_t)^{-1} (Z_t X_{t-1}) \). We denote \( \hat{X}_{t-1} \) as the predicted value from this regression. After replacing \( X_{t-1} \) with \( \hat{X}_{t-1} \) in (\ast), we run the second-stage regression of TSLS. Finally, we get an estimate of \( \varphi \).

\[
\hat{\varphi}_{TSLS} = \frac{\sum (\hat{X}_{t-1} - \overline{X}_{t-1})(X_t - \overline{X})}{\sum (\hat{X}_{t-1} - \overline{X}_{t-1})^2}
\]

By utilizing more instrumental variables, we can improve the precision of the resulting estimator. For this paper, we do not take advantage of this approach to estimation.
4. Empirical Analysis

We analyze empirical data of stock returns for the 7 equities of DJIA. Since tick size was reduced from 1/16 of 1 dollar to 1 cent on January 29, 2001, we choose the sample period from February 1, 2001 to December 31, 2005. The data were extracted from Trade and Quote (TAQ) database and days with less than 5 hours of trading were removed from the sample.

To test models, we estimate coefficients for (1), (3), (8) using data from February 1, 2001 to December 31, 2004. After estimating coefficients, we forecast year 2005 $X_t$ with three models and compute MSE with the actual year 2005 data. Note that $\delta$ at (8) is the only variable that can be changed during the forecasting procedures.

I. Estimating coefficients

We use Alcoa (AA) case for an example. For HAR model, coefficients of (1) are given by the OLS regression.

$$\hat{X}_t = 0.204 X_{t-1} + 0.571 \left( \frac{1}{5} \sum_{i=1}^{5} X_{t-i} \right) + 0.137 \left( \frac{1}{22} \sum_{i=1}^{22} X_{t-i} \right) + 0.097,$$ from (1).

Since $corr(\Delta X_t, \Delta X_{t-1}) = -0.4285 = -\frac{1}{2 + \frac{1}{\lambda}}$ from (2), $\hat{\lambda}_{LL} = 2.997$. After calculating with our formulae, we get $\hat{\lambda}_{Q,LL} = 3.622$ from (9), $\hat{\phi}_{LS} = 0.6115$ from (5), $\hat{\phi}_{IV} = 0.9036$ from (6), $\hat{\lambda}_{IV} = 2.6033$ from (7), $\hat{\lambda}_{Q,IV} = 9.286$ from (9). The big difference between $\hat{\phi}_{LS}$ and $\hat{\phi}_{IV}$ indicates that the use of the instrumental variable is essential; $X_{t-1}$ is much more likely to be correlated with the noise at time $t$ than $X_{t-3}$ is. From computed $\phi$ and $\lambda$, we compute optimal weights for the IV model and the LL model.
Graph 1 summarizes this result.

*Graph 1. Optimal weights of AA (Alcoa) for three models*

The graph shows that for all models, coefficients decay exponentially as the number of lagged periods increases. Generally speaking, recent data counts much more. For the LL model, theoretically calculated weights using (4) coincide with the derived optimal weights. For the IV model, the ratio \( w_i/\bar{w}_{i+1} = 0.5229 \) is consistent by single exponential smoothing. Lastly, a sum of weights is 1 for the LL model, while \(<1\) for the IV model due to the \( \delta \) term.
II. MSE comparisons

For convenience, let IV(0) be the IV model: 
\[ \hat{X}_t = \sum_{h=1}^{N} w_h X_{t-h} + \left(1 - \sum_{h=1}^{N} w_h\right) \delta \]
with
\[ \delta = \frac{1}{T} \sum_{t=1}^{T} X_t \] (all past data). Similarly, the IV(m) model uses
\[ \delta = \frac{1}{N} \sum_{h=1}^{N} X_{t-h} \] where
\[ N = 22 \] (past month), the IV(q) model uses
\[ \delta = \frac{1}{3N} \sum_{h=1}^{3N} X_{t-h} \] (past quarter), and IV(y) uses
\[ \delta = \frac{1}{250} \sum_{h=1}^{250} X_{t-h} \] (past year).

Mean Squared Error (MSE) of AA are shown at Table 1. Better results (smaller MSEs) are underlined.

Table 1. Mean Squared Errors of AA (Alcoa)

<table>
<thead>
<tr>
<th>LL Q,LL or ( \hat{\lambda}_{LL} )</th>
<th>HAR Q,LL or ( \hat{\lambda}_{HAR} )</th>
<th>IV(0)</th>
<th>IV (m)</th>
<th>IV (q)</th>
<th>IV (y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\lambda}<em>{LL} ) or ( \hat{\lambda}</em>{IV} )</td>
<td>0.201</td>
<td>0.190</td>
<td><strong>0.203</strong></td>
<td>0.191</td>
<td>0.190</td>
</tr>
<tr>
<td>( \hat{\lambda}<em>{Q,LL} ) or ( \hat{\lambda}</em>{Q,IV} )</td>
<td><strong>0.198</strong></td>
<td>0.190</td>
<td>0.217</td>
<td><strong>0.184</strong></td>
<td><strong>0.184</strong></td>
</tr>
</tbody>
</table>

From Table 1, we observe that using \( \hat{\lambda}_{Q,LL} \) or \( \hat{\lambda}_{Q} \) provides better (smaller) MSE for most of the cases. Also, IV(m) and IV(q) shows better results than IV(0), and IV(y) provides the best results among all IV’s. These results hold for all seven stocks except GM. It is likely that there is a certain ‘optimal’ length of the period to calculate \( \delta \) that gives the best estimates.

MSE for seven stocks for all forecasting models are shown at Table 2. As before, underlined numbers are the smallest MSEs.
### Table 2. Mean Squared Errors of seven stocks

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>BA</th>
<th>CAT</th>
<th>DIS</th>
<th>GE</th>
<th>GM</th>
<th>IBM</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAR</td>
<td>0.190</td>
<td>0.189</td>
<td>0.218</td>
<td>0.218</td>
<td>0.204</td>
<td>0.464</td>
<td>0.211</td>
</tr>
<tr>
<td>LL</td>
<td>$\hat{\lambda}_{Q,LL}$</td>
<td>0.198</td>
<td>0.188</td>
<td>0.239</td>
<td>0.235</td>
<td>0.217</td>
<td>0.486</td>
</tr>
<tr>
<td></td>
<td>$\hat{\lambda}_{LL}$</td>
<td>0.201</td>
<td>0.190</td>
<td>0.239</td>
<td>0.236</td>
<td>0.219</td>
<td>0.486</td>
</tr>
<tr>
<td>IV(0)</td>
<td>$\hat{\lambda}_{Q,IV}$</td>
<td>0.217</td>
<td>0.255</td>
<td>0.237</td>
<td>0.254</td>
<td>0.262</td>
<td>0.489</td>
</tr>
<tr>
<td></td>
<td>$\lambda_{IV}$</td>
<td>0.203</td>
<td>0.215</td>
<td>0.225</td>
<td>0.244</td>
<td>0.238</td>
<td>0.468</td>
</tr>
<tr>
<td>IV(m)</td>
<td>$\hat{\lambda}_{Q,IV}$</td>
<td>0.184</td>
<td>0.182</td>
<td>0.223</td>
<td>0.223</td>
<td>0.207</td>
<td>0.486</td>
</tr>
<tr>
<td></td>
<td>$\lambda_{IV}$</td>
<td>0.191</td>
<td>0.193</td>
<td>0.222</td>
<td>0.233</td>
<td>0.221</td>
<td>0.471</td>
</tr>
<tr>
<td>IV(q)</td>
<td>$\hat{\lambda}_{Q,IV}$</td>
<td>0.184</td>
<td>0.182</td>
<td>0.218</td>
<td>0.220</td>
<td>0.204</td>
<td>0.494</td>
</tr>
<tr>
<td></td>
<td>$\lambda_{IV}$</td>
<td>0.190</td>
<td>0.192</td>
<td>0.219</td>
<td>0.231</td>
<td>0.219</td>
<td>0.471</td>
</tr>
<tr>
<td>IV(y)</td>
<td>$\hat{\lambda}_{Q,IV}$</td>
<td>0.180</td>
<td>0.181</td>
<td>0.214</td>
<td>0.218</td>
<td>0.202</td>
<td>0.505</td>
</tr>
<tr>
<td></td>
<td>$\lambda_{IV}$</td>
<td>0.188</td>
<td>0.191</td>
<td>0.217</td>
<td>0.230</td>
<td>0.218</td>
<td>0.473</td>
</tr>
</tbody>
</table>

GM is an unusual case. MSE of GM is more than twice of other stocks’ MSE.

Except GM, IV(y) provides lower or similar MSE compared to the LL or HAR model.
III. Forecasts

Graph 2 plots actual year 2005 values of $X_i$ of Alcoa (AA) stock and forecasts from the model. Other stocks show similar patterns.

*Graph 2. Actual data and model forecasts for 2005, AA (Alcoa)*
Graph 3 shows that the range of $X_t$ is larger for the GM case.

Graph 3. Actual data and model forecasts for 2005, GM (General Motors)
Graph 1 shows that the LL model has higher weights on recent data than other models have; this fact explains why the LL model has the biggest fluctuations among three models in graph 2 and 3. On the other hand, the IV model has the smallest fluctuations due to the existence of the $\delta$ term. The $\delta$ term, the average of past values (for certain period; one year is used in this case), acts like a ‘buffer’ and reduces variations.

**IV. $\hat{\lambda}_{LL}, \hat{\lambda}_{Q,LL}, \hat{\lambda}_{Q,IV}, \hat{\lambda}_{IV}$**

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>BA</th>
<th>CAT</th>
<th>DIS</th>
<th>GE</th>
<th>GM</th>
<th>IBM</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_{IV}$</td>
<td>2.603</td>
<td>1.224</td>
<td>2.017</td>
<td>2.358</td>
<td>1.662</td>
<td>2.225</td>
<td>1.375</td>
<td>1.923</td>
</tr>
<tr>
<td>Ratio ($\hat{\lambda}<em>{Q,IV} / \lambda</em>{IV}$)</td>
<td>3.57</td>
<td>5.47</td>
<td>4.59</td>
<td>3.61</td>
<td>5.30</td>
<td>4.70</td>
<td>4.90</td>
<td>4.59</td>
</tr>
<tr>
<td>$\hat{\lambda}_{LL}$</td>
<td>2.997</td>
<td>3.261</td>
<td>1.508</td>
<td>3.333</td>
<td>3.548</td>
<td>3.135</td>
<td>2.233</td>
<td>2.859</td>
</tr>
<tr>
<td>$\hat{\lambda}_{Q,LL}$</td>
<td>3.622</td>
<td>4.081</td>
<td>1.569</td>
<td>3.640</td>
<td>4.063</td>
<td>3.152</td>
<td>3.004</td>
<td>3.304</td>
</tr>
<tr>
<td>Ratio ($\hat{\lambda}<em>{Q,LL} / \hat{\lambda}</em>{LL}$)</td>
<td>1.21</td>
<td>1.25</td>
<td>1.04</td>
<td>1.09</td>
<td>1.14</td>
<td>1.01</td>
<td>1.35</td>
<td>1.16</td>
</tr>
</tbody>
</table>

Overall, $\hat{\lambda}_{Q,IV}$ is about 4.59 times of $\lambda_{IV}$ and $\hat{\lambda}_{Q,LL}$ is about 1.16 times of $\hat{\lambda}_{LL}$. For the LL model, lambdas from the prediction based criterion, $\hat{\lambda}_{Q,LL}$, are similar to lambdas from the OLS regression, $\hat{\lambda}_{LL}$. Yet, for the IV model, $\hat{\lambda}_{Q,IV}$ are more than 4 times of $\lambda_{IV}$ in general.

However, in both models, ratios are fairly close to averages. This result suggests that there are certain relationships between lambdas derived from the OLS regression and
lambdas obtained from the prediction based criterion. In this case, $\hat{\lambda}_{Q,LL}$ (or $\hat{\lambda}_{Q,IV}$) could be estimated from $\hat{\lambda}_{LL}$ (or $\hat{\lambda}_{IV}$), if we need them since they provide better forecasts.

V. Values of $\varphi$ from an AR(2) with unit roots

Table 4. Values of $\varphi_1$ and $\varphi_2$ from AR(2) under the unit root condition

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>BA</th>
<th>CAT</th>
<th>DIS</th>
<th>GE</th>
<th>GM</th>
<th>IBM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varphi_1$</td>
<td>1.0013</td>
<td>0.9989</td>
<td>0.9994</td>
<td>1.0016</td>
<td>1.0026</td>
<td>1.0059</td>
<td>1.0003</td>
</tr>
<tr>
<td>$\varphi_2$</td>
<td>-0.0013</td>
<td>0.0011</td>
<td>0.0007</td>
<td>-0.0016</td>
<td>-0.0026</td>
<td>-0.0059</td>
<td>-0.0003</td>
</tr>
</tbody>
</table>

$(\varphi_1, \varphi_2) \approx (1,0)$ for all seven stocks. The $Y_{t-2}$ term has a coefficient very close to zero. In other words, the $Y_{t-2}$ term appears to contribute little for the prediction power. Therefore, using an AR(1) model provides almost equally effective forecasting results and requires less computations, compared to an AR(2) model under the unit root condition.
5. Summary and Discussion

As we have seen in Table 2: Mean Squared Errors of seven stocks, the IV(y) model provides better results than other models (including the HAR model and the LL model) do, except for the GM case. Also, from the same table, we observe that IV(y) model works better than IV(0). This indicates that using only recent data to compute \( \delta \) (the average of past data), instead of all available past data, gives better estimates. One possible explanation could be the volatility clustering in the empirical data.

However, IV(y) model also works better than IV(m) and IV(q) even though these two models use much limited recent data (only past month or past quarter, which are shorter periods compared to the past year that is used in the IV(y) model.) This comparison between IV(y) and other IV models implies that choosing “proper” length of the past period to get \( \delta \) results in better estimates. Questions including what are the ‘proper’ lengths and what factors determine them should be answered by future research.

Another key point of Table 2 is that using \( \hat{\lambda}_{Q,IV} \) provides better estimates than using \( \lambda_{IV} \). Also, for the LL model, using \( \hat{\lambda}_{Q,LL} \) works better than using \( \hat{\lambda}_{LL} \). Table 3 shows that \( \hat{\lambda}_{Q,IV} \) is about 4.59 times of \( \lambda_{IV} \) and \( \hat{\lambda}_{Q,LL} \) is about 1.16 times of \( \hat{\lambda}_{LL} \). These ratios are fairly consistent among stocks; ratio \( \hat{\lambda}_{Q,IV} / \lambda_{IV} \) varies from 3.57 to 5.47 and Ratio \( \hat{\lambda}_{Q,LL} / \hat{\lambda}_{LL} \) varies from 1.01 to 1.35. It is possible that there are certain theoretical relationships between \( \hat{\lambda}_{Q,IV} \) and \( \lambda_{IV} \) or \( \hat{\lambda}_{Q,LL} \) and \( \hat{\lambda}_{LL} \) that define ratios.

Graphs that plot actual data and estimates of AA and GM show that GM stock has more volatile RV. For GM, the range of \( X_t \) is -2 to 4, approximately; for AA and other stocks, the range of \( X_t \) is -1 to 2, approximately. Also, GM has the largest MSE among
seven stocks; the MSE of GM is more than twice of those of other stocks. Testing more stocks or testing longer periods could help us to confirm that the GM case is an unusual case.

Table 4 shows that \((\phi_1, \phi_2) \approx (1, 0)\) for all stocks. The fact that coefficients of the additional terms, \(\phi_2\), are almost zero suggests that adding the second term does not improve forecasts significantly. Using the instrumental variable \(Z_t = X_{t-3}\), we find that our simple AR(1) model captures almost all the dynamics that is important for forecasting. However, studying a non-unit root AR(2) model case could be an interesting future topic to pursue.

Admittedly, since this paper focuses on finding better forecasting model for the specific case, using 2001~2004 data to forecast year 2005 for seven stocks, there is no guarantee that the IV(y) model performs better than the HAR model and the LL model in other periods or for other stocks. Testing the proposed IV(y) model on other data sets could be helpful to support the effectiveness of the model. The most important achievement of this paper is that it has utilized a simple approach to estimate future volatilities.
Reference


