Data Sources

• Much of the published empirical analysis of RV has been based on high frequency data from two sources:
  – Olsen and Associates proprietary FX data set for foreign exchange
    • www.olsendata.com
  – The NYSE Trades and Quotation (TAQ) data for equity
    • www.nyse.com/taq

Olsen FX Data

• The HFDF-2000 data is the most commonly used data set
  – spot exchange rates sampled every 5 minutes for the $, DM, CHF, BP, Yen over the period December 1, 1986 through June 30, 1999.
  – All interbank bid/ask indicative quotes for the exchange rates displayed on the Reuters FXFX screen.
  – Highly liquid market: 2000-4000 observations per day per currency
  – Outlier filtered log-price at each 5-minute tick is interpolated from the average of bid and ask quotes for the two closest ticks, and 5-minute cc return is difference in the log-price.
Olsen FX Data

- Data cleaning prior to computation of RV measures:
  - 5-minute return data is restricted to eliminate non-trading periods, weekends, holidays, and lapses of the Reuters data feed.
  - The slow weekend period from Friday 21:05 GMT until Sunday 21:00 GMT is eliminated from the sample.
  - Holidays removed: Christmas (December 24-26), New Year's (December 31-January 2), July 4th, Good Friday, Easter Monday, Memorial Day, Labor Day, and Thanksgiving and the day after.
  - Days that contain long strings of zero or constant returns (caused by data feed problems) are eliminated.

Empirical Analysis of FX Returns

<table>
<thead>
<tr>
<th>Author</th>
<th>Series</th>
<th>Sample</th>
<th>Days, T</th>
<th>m</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB 1998</td>
<td>DM/$, Y/$</td>
<td>87-93</td>
<td>260</td>
<td>288</td>
</tr>
<tr>
<td>AB 1998</td>
<td>DM/$, Y/$</td>
<td>87-93</td>
<td>260</td>
<td>48</td>
</tr>
<tr>
<td>ABDL 2000</td>
<td>DM/$, Y/$</td>
<td>86-96</td>
<td>2,445</td>
<td>48</td>
</tr>
<tr>
<td>ABDL 2001</td>
<td>DM/$, Y/$</td>
<td>86-96</td>
<td>2,449</td>
<td>288</td>
</tr>
<tr>
<td>ABDL 2003</td>
<td>DM/$, Y/$</td>
<td>86-99</td>
<td>3,045</td>
<td>48</td>
</tr>
<tr>
<td>ABDM 2005</td>
<td>DM/$, Y/$</td>
<td>89-99</td>
<td>3,045</td>
<td>48</td>
</tr>
<tr>
<td>BNS 2001</td>
<td>DM/$</td>
<td>86-96</td>
<td>2,449</td>
<td>various</td>
</tr>
<tr>
<td>BNS 2002</td>
<td>DM/$</td>
<td>86-96</td>
<td>2,449</td>
<td>288</td>
</tr>
</tbody>
</table>
Distribution of RV


Summary Statistics for Daily RV Measures, m=228

<table>
<thead>
<tr>
<th></th>
<th>RV_D</th>
<th>RV_Y</th>
<th>RVOL_D</th>
<th>RVOL_Y</th>
<th>RLVOL_D</th>
<th>RLVOL_Y</th>
<th>RCOV</th>
<th>RCOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>.529</td>
<td>.588</td>
<td>.679</td>
<td>.684</td>
<td>-.449</td>
<td>-.443</td>
<td>.243</td>
<td>.435</td>
</tr>
<tr>
<td>Variance</td>
<td>.234</td>
<td>.272</td>
<td>.067</td>
<td>.070</td>
<td>.120</td>
<td>.123</td>
<td>.073</td>
<td>.028</td>
</tr>
<tr>
<td>Skewness</td>
<td>3.71</td>
<td>5.57</td>
<td>1.68</td>
<td>1.87</td>
<td>.345</td>
<td>.264</td>
<td>3.78</td>
<td>-.203</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>24.1</td>
<td>66.5</td>
<td>7.78</td>
<td>10.4</td>
<td>3.26</td>
<td>3.53</td>
<td>25.3</td>
<td>2.72</td>
</tr>
</tbody>
</table>

Table 3: Summary statistics for daily RV measures. Source ABDL (2001).
Unconditional Distributions: $m=288$

Source: ABDL 2001
**Correlation Matrix for Daily RV Measures**

<table>
<thead>
<tr>
<th></th>
<th>RV_D</th>
<th>RV_L</th>
<th>RVOL_D</th>
<th>RVOL_L</th>
<th>RLVL_D</th>
<th>RLVL_L</th>
<th>RCOV</th>
<th>RCOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>RV_D</td>
<td>.539</td>
<td>.061</td>
<td>.525</td>
<td>.860</td>
<td>.512</td>
<td>.806</td>
<td>.341</td>
<td></td>
</tr>
<tr>
<td>RV_L</td>
<td>1.00</td>
<td>.546</td>
<td>.945</td>
<td>.514</td>
<td>.825</td>
<td>.757</td>
<td>.234</td>
<td></td>
</tr>
<tr>
<td>RVOL_D</td>
<td>1.00</td>
<td>.592</td>
<td>.965</td>
<td>.578</td>
<td>.793</td>
<td>.383</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RVOL_L</td>
<td>1.00</td>
<td>.589</td>
<td>.959</td>
<td>.760</td>
<td>.281</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLVL_D</td>
<td>1.00</td>
<td>.604</td>
<td>.720</td>
<td>.389</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLVL_L</td>
<td>1.00</td>
<td>.684</td>
<td>.294</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCOV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>.590</td>
</tr>
</tbody>
</table>

**“Correlation-in-Volatility” Effect**

Source: ABDL (2001)
Accuracy of RV Measures: 95% CI from BNS
Asymptotic Theory as Functions of m

Time Series of Daily RVOL: m=228

Source: BNS (2002)

Source: ABDL (2001)
Time Series of Daily RCOR: m=228

Source: ABDL (2001)

SACF of Daily RV Measures: m=228

Source: ABDL (2001)
Long Memory Behavior of RV Measures

A stationary process $y_t$ has long memory, or long range dependence, if its autocorrelation function decays slowly at a hyperbolic rate:

\[
\rho_k \rightarrow C\rho \cdot k^{-\alpha}, \text{ as } k \rightarrow \infty
\]

\[
\alpha \in (0,1)
\]

Fractionally Differenced Processes

- A long memory process $y_t$ can be modeled parametrically by extending an integrated process to a fractionally integrated process:

\[
(1 - L)^d (y_t - \mu) = u_t, \quad u_t \sim I(0)
\]

0 < $d < 0.5$: stationary long memory

0.5 ≤ $d < 1$: nonstationary long memory
Estimating $d$

- Nonparametric estimation
  - Geweke-Porter-Hudak (GPH) log-periodogram regression
  - Local Whittle estimator
  - Phillips-Kim modified GPH estimator
  - Andrews-Guggenberger biased corrected GPH estimator

- Parametric estimation
  - ARFIMA($p,d,q$) model with normal errors

GPH Estimates of $d$

<table>
<thead>
<tr>
<th></th>
<th>$R_{V_D}$</th>
<th>$R_{V_Y}$</th>
<th>$R_{VOL_D}$</th>
<th>$R_{VOL_Y}$</th>
<th>$R_{RLVOL_D}$</th>
<th>$R_{RLVOL_Y}$</th>
<th>$R_{RCOV}$</th>
<th>$R_{RCOR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d$</td>
<td>0.356</td>
<td>0.339</td>
<td>0.381</td>
<td>0.428</td>
<td>0.420</td>
<td>0.455</td>
<td>0.334</td>
<td>0.413</td>
</tr>
</tbody>
</table>

Note: Multivariate estimate of common $d$ using ($R_{RLVOL_{D}}, R_{RLVOL_{Y}}, R_{RLVOL_{DY}}$) is 0.4
Temporal Aggregation and Scaling Laws

- The fractional differencing parameter $d$ is invariant under temporal aggregation.
- If $x_t$ is fractionally integrated with parameter $d$ then

$$\text{var}([x_t]_h) = c \cdot h^{2d+1}$$

$$[x_t]_h = \sum_{j=1}^{h} x_{(t-1)+j}$$

$$\Rightarrow \ln(\text{var}([x_t]_h)) \propto (2d + 1)\ln(h)$$

Temporal Aggregation and Estimated of $d$

<table>
<thead>
<tr>
<th>$h$</th>
<th>$RV_D$</th>
<th>$RV_Y$</th>
<th>$RVOLS_D$</th>
<th>$RVOLS_Y$</th>
<th>$RLVOLS_D$</th>
<th>$RLVOLS_Y$</th>
<th>$RCOV$</th>
<th>$RCOR$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.356</td>
<td>.339</td>
<td>.381</td>
<td>.428</td>
<td>.420</td>
<td>.455</td>
<td>.334</td>
<td>.413</td>
</tr>
<tr>
<td>5</td>
<td>.457</td>
<td>.429</td>
<td>.446</td>
<td>.473</td>
<td>.405</td>
<td>.496</td>
<td>.368</td>
<td>.519</td>
</tr>
<tr>
<td>10</td>
<td>.511</td>
<td>.490</td>
<td>.470</td>
<td>.501</td>
<td>.515</td>
<td>.507</td>
<td>.496</td>
<td>.494</td>
</tr>
<tr>
<td>15</td>
<td>.400</td>
<td>.426</td>
<td>.384</td>
<td>.440</td>
<td>.421</td>
<td>.440</td>
<td>.319</td>
<td>.600</td>
</tr>
<tr>
<td>20</td>
<td>.455</td>
<td>.488</td>
<td>.440</td>
<td>.509</td>
<td>.496</td>
<td>.479</td>
<td>.439</td>
<td>.630</td>
</tr>
</tbody>
</table>
Temporal Aggregation and Scaling Laws

Source: ABDL (2001)

Distribution of Returns Standardized by RV

Stochastic Volatility Model

• Assume daily returns $r_t$ may be decomposed following a standard conditional volatility model

$$r_t = \sigma_t \varepsilon_t$$

$$\sigma_t = \text{latent volatility}$$

$$\varepsilon_t \sim iid \ (0,1)$$

Standardized Returns

• Compute returns standardized by estimates of conditional volatility

$$\hat{\varepsilon}_t = \frac{r_t}{\hat{\sigma}_t}$$

$$\hat{\sigma}_t = RVOL_t, \ m = 48$$

$$\hat{\sigma}_t = \hat{\sigma}_t^{GARCH(1,1)}$$

GARCH(1,1): $\sigma_t^2 = w + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2$
Multivariate Standardized Returns

• Standardized returns based RCOV

\[
\begin{pmatrix}
\hat{e}_{D,t} \\
\hat{e}_{Y,t}
\end{pmatrix} = RCOV_{t}^{-1/2} \begin{pmatrix}
r_{D,t} \\
r_{Y,t}
\end{pmatrix}
\]

\[RCOV_{t}^{1/2} = \text{Cholesky factor of } RCOV_{t}\]

Comparison of Volatility Forecasts

• Squared returns are unbiased but very noisy

• GARCH(1,1) estimates are smoother than RV estimate; do not utilize information between time \(t-1\) and \(t\) (exponentially weighted average of past returns)

• RV estimates make exclusive use of information between time \(t-1\) and \(t\); better forecast of time \(t\) volatility
### Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>$r_t$</th>
<th>$r_t / \sigma_t$ARCH</th>
<th>$r_t / \sigma_t$RVOL_t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DM/$</td>
<td>Y/$</td>
<td>DM/$</td>
</tr>
<tr>
<td>Mean</td>
<td>.007</td>
<td>.009</td>
<td>-.002</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>.710</td>
<td>.705</td>
<td>1.00</td>
</tr>
<tr>
<td>Skewness</td>
<td>.033</td>
<td>.052</td>
<td>-.027</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>5.40</td>
<td>7.36</td>
<td>4.75</td>
</tr>
<tr>
<td>Correlation</td>
<td>.659</td>
<td>.661</td>
<td>.661</td>
</tr>
</tbody>
</table>

Gaussian!

### Distribution of Daily Returns

- **DM/$ Return Quantile**
- **Y/$ Return Quantile**

Source: ABDL (2000)
Distribution of Standardized Returns

Source: ABDL (2000)

Scatterplot of Daily Returns

Source: ABDL (2000)
Scatterplot or Standardized Returns

Source: ABDL (2000)

SACF of Squared Returns

Source: ABDL (2000)
Returns Standardized by 1-Day-Ahead Forecasts

<table>
<thead>
<tr>
<th></th>
<th>( \frac{\sigma_t}{\sigma_{GARCH}} )</th>
<th>( \frac{r_t}{RVOL_{t-1}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DM/$</td>
<td>Y/$</td>
</tr>
<tr>
<td>Mean</td>
<td>-.002</td>
<td>-.011</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Skewness</td>
<td>-.027</td>
<td>-.139</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4.75</td>
<td>5.41</td>
</tr>
<tr>
<td>Correlation</td>
<td>.661</td>
<td>.661</td>
</tr>
</tbody>
</table>

5/24/2010
Conclusions

• Daily returns standardized by RV measures are nearly Gaussian
• Supports diffusion model for returns
• Alternative to copula methods for characterizing multivariate distributions
• Advantages for value-at-risk computation

Modeling and Forecasting RV

• ABDL (2003): “Modeling and Forecasting Realized Volatility,” *Econometrica*
Traditional Conditional Volatility Models

- Normal GARCH(1,1)
  \[ r_t = \sigma_t \epsilon_t, \epsilon_t \sim iid \ N(0,1) \]
  \[ \sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2 \]

- Log-Normal SV model
  \[ r_t = \sigma_t \epsilon_t, \epsilon_t \sim iid \ N(0,1) \]
  \[ \ln \sigma_t^2 = \delta + \phi \ln \sigma_{t-1}^2 + \sigma_u u_t, u_t \sim iid \ N(0,1) \]
  \[ E[\epsilon u_t] = 0 \]

Advantages of Using RV

- RV provides an observable estimate of latent volatility
- Standard time series models (e.g. ARIMA) may be used to model and forecast RV
- Multivariate time series models may be used model and forecast RCOV, RCOR
Trivariate System of Exchange Rates

\[ y_t = \begin{pmatrix} RLVOL_{D/S,t} \\ RLVOL_{Y/S,t} \\ RLVOL_{Y/D,t} \end{pmatrix}, \quad m = 48 \]

\[ RCOV_{D/S,Y/S} = \frac{1}{2} \left( RV_{D/S,t} + RV_{Y/S,t} - RV_{Y/D,t} \right) \]

- Fit models for \( y_t \) in sample: 12/1/86-12/1/96
- Forecast \( y_t \) out-of-sample: 12/2/96 – 6/30/99

SACF of Daily DM/$ RLVOL: m=48

SACF of Daily Yen/$ RLVOL: m=48

\[(1 - L)^{0.4}(Y_t - \mu_t)\]


SACF of Daily Yen/DM RLVOL: m=48

\[(1 - L)^{0.4}(Y_t - \mu_t)\]

FI-VAR(5) Model (VAR-RV)

\[ \Phi(L)(1 - L)^{0.4} (y_t - \mu) = \varepsilon_t \]
\[ \varepsilon_t \sim iid \ N(0, \Omega) \]
\[ \Phi(L) = I_3 - \Phi_1 L - \cdots - \Phi_5 L^5 \]

Alternative Models

- **VAR-ABS**: VAR(5) fit to |rt|
- **AR-RV**: univariate AR(5) fit to (1-L)^{0.4}RLVL_{i,t}
- **Daily GARCH(1,1)**: normal-GARCH(1,1) fit to daily returns r_{i,t}
- **Daily RiskMetrics**: exponentially weighted moving average model for r_{i,t}^2 with \lambda=0.94
- **Daily FIEGARCH(1,1)**: univariate fractionally integrated exponential GARCH(1,1) fit to r_{i,t}
- **Intra-day FIEGARCH deseason/filter**: univariate fractionally integrated exponential GARCH(1,1) fit to 30-minute filtered and deseasonalized returns r_{i,t+\Delta}.
Forecast Evaluation

\[ RVOL_{i,t} = b_0 + b_1 \hat{RVOL}_{VAR-RV}^{VAR} + b_2 \hat{RVOL}_{model}^{model} + error_t \]

\( \hat{RVOL}_{VAR-RV}^{VAR} \) = 1-day ahead forecast from RV-VAR

\( \hat{RVOL}_{model}^{model} \) = 1-day ahead forecast from alternative model

\( H_0 : b_0 = 0, b_1 = 1, b_2 = 0 \)

Findings

- RV-VAR is consistently best forecasting model in-sample and out-of-sample: highest \( R^2 \) from forecast evaluation regressions.
- Rarely reject \( H_0 : b_0 = 0, b_1 = 1, b_2 = 0 \) for RV-VAR model
- RV-AR is close to RV-VAR
Forecasts of Daily RVOL: VAR-RV vs. GARCH(1,1)

NYSE TAQ Data

- Intra-day trade and quotation information for all securities listed on NYSE, AMEX, and NASDAQ.
- The most active period for equity markets is during the trading hours of the NYSE between 9:30 a.m. EST until 4:00 p.m. EST.
- Not as liquid as FX markets
NYSE TAQ Data

• Equity returns are generally subject to more pronounced market microstructure effects (e.g., negative first order serial correlation caused by bid-ask bounce effects) than FX data. As a result, equity returns are often filtered to remove these microstructure effects prior to the construction of RV measures.

• A common filtering method involves estimating an MA(1) or AR(1) model to the returns, and then constructing the filtered returns as the residuals from the estimated model.

Empirical Analysis of TAQ Data

  – Analyze 30 Dow Jones Industrial Average Stocks over the period 1/2/93 – 5/29/98
  – Restrict analysis to NYSE exchange hours
  – T=1,336; m=79 5-minute returns
Summary of Findings

- Results for equity returns are similar to those for FX returns
  - RLVOL, RCOR are approximately Gaussian
  - RV measures exhibit long memory
  - Daily returns standardized by RVOL are nearly Gaussian
- Little evidence of leverage effect
- Evidence of factor structure in multivariate system of RV measures

Distribution of Daily RLVOL: Alcoa

Solid line: RLVOL
Dashed line: normal density

Source: ABDE (2001)
Distribution of Daily RCOR: Alcoa, Exxon

Solid line: RCOR
Dashed line: normal density

Source: ABDE (2001)

Time Series of Daily RLVOL: Alcoa

Source: ABDE (2001)
Time Series of Daily RCOR: Alcoa, Exxon

Source: ABDE (2001)

Distribution of Daily Standardized Returns for Alcoa

Solid line: returns/RVOL
Dashed line: normal density
Evidence for Factor Structure

Evidence of Factor Structure
Evidence of Factor Structure

Directions for Future Research

- Continued development of methods for exploiting the volatility information in high-frequency data
- Volatility modeling and forecasting in the high-dimensional multivariate environments of practical financial economic relevance