11ª Escola de Séries

Temporais e Econometria

Analysis of High Frequency Financial Time Series: Methods, Models and Software

Eric Zivot

Associate Professor and Gary Waterman Distinguished Scholar, Department of Economics

Adjunct Associate Professor, Department of Finance

University of Washington

August 1, 2005

Agenda

- Lecture 1
  - Introduction to high frequency data
- Lecture 2
  - Realized variance measures: theory
- **Lecture 3**
  - **Realized variance measures: empirical analysis**

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

- Historical data made available for use in three conferences on the statistical analysis of high frequency data: HFDF-1993, HFDF-1996, and HF-2000.
- 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

Author	Series	Sample	Days, T	m
AB 1998	DM/$, Y/$	87-93	260	288
AB 1998	DM/$, Y/$	87-93	260	48
ABDL 2000	DM/$, Y/$	86-96	2,445	288
ABDL 2001	DM/$, Y/$	86-96	2,449	288
ABDL 2003	DM/$, Y/$	86-99	3,045	48
ABDM 2005	DM/$, Y/$	89-99	3,045	48
BNS 2001	DM/$	86-96	2,449	various
BNS 2002	DM/$	86-96	2,449	288

Distribution of RV

- ABDL (2001): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*.
- BNS (2001): "Estimating Quadratic Variation Using Realized Variance," *Journal of Applied Econometrics*.

Summary Statistics for Daily RV Measures, m=228

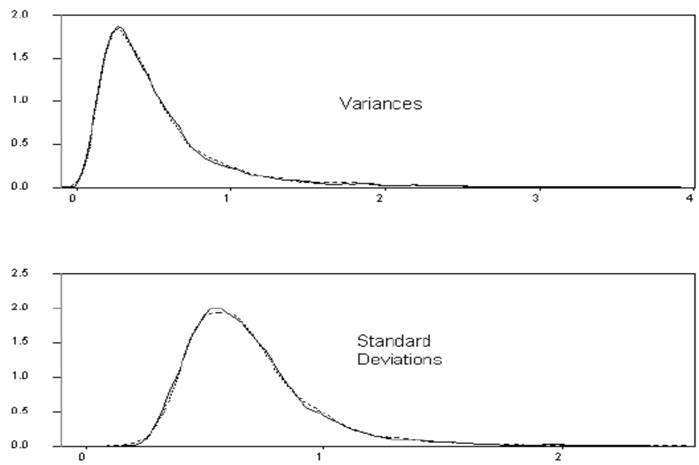
	RV_D	RV_Y	$RVOL_D$	$RVOL_Y$	$RLVOL_D$	$RLVOL_Y$	$RCOV$	$RCOR$
Mean	.529	.538	.679	.684	-.449	-.443	.243	.435
Variance	.234	.272	.067	.070	.120	.123	.073	.028
Skewness	3.71	5.57	1.68	1.87	.345	.264	3.78	-.203
Kurtosis	24.1	66.5	7.78	10.4	3.26	3.53	25.3	2.72

Table 3: Summary statistics for daily RV measures. Source ABDL (2001).

Non-Gaussian

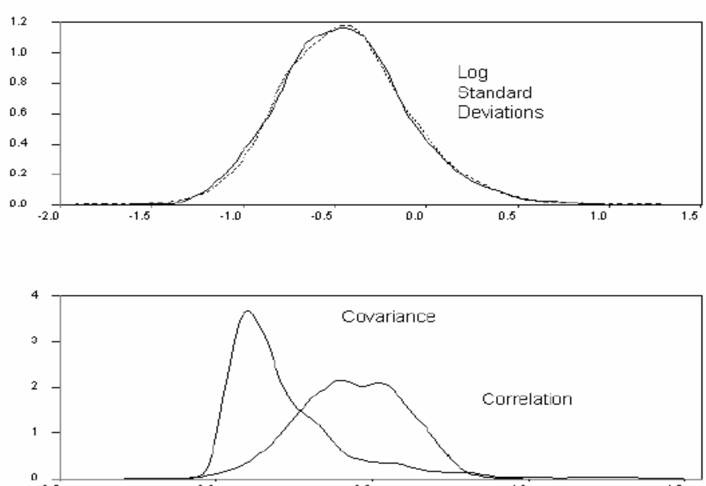
Gaussian

Unconditional Distributions: m=288



Source: ABDL 2001

Unconditional Distributions: m=288

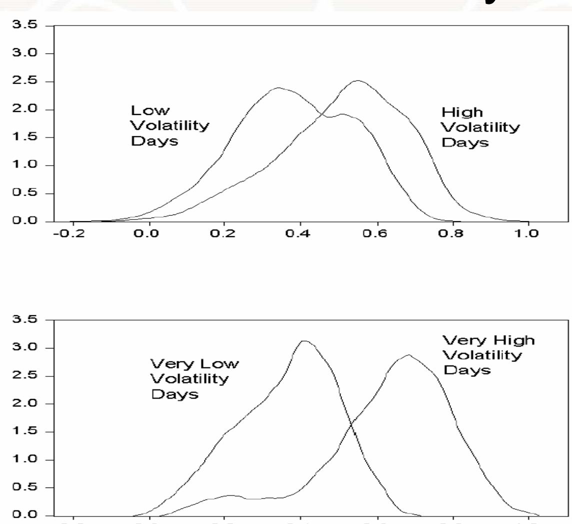


Source: ABDL 2001

Correlation Matrix for Daily RV Measures

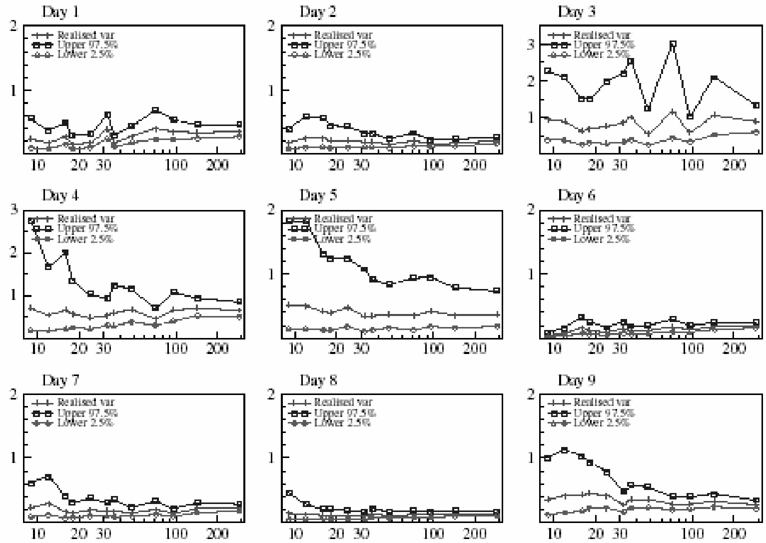
	RV_Y	$RVOL_D$	$RVOL_Y$	$RLVOL_D$	$RLVOL_Y$	$RCOV$	$RCOR$
RV_D	.539	.061	.552	.860	.512	.806	.341
RV_Y	1.00	.546	.945	.514	.825	.757	.234
$RVOL_D$		1.00	.592	.965	.578	.793	.383
$RVOL_Y$			1.00	.589	.959	.760	.281
$RLVOL_D$				1.00	.604	.720	.389
$RLVOL_Y$					1.00	.684	.294
$RCOV$						1.00	.590

"Correlation-in-Volatility" Effect



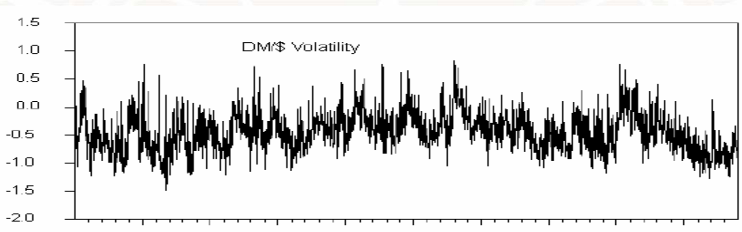
Source: ABDL (2001)

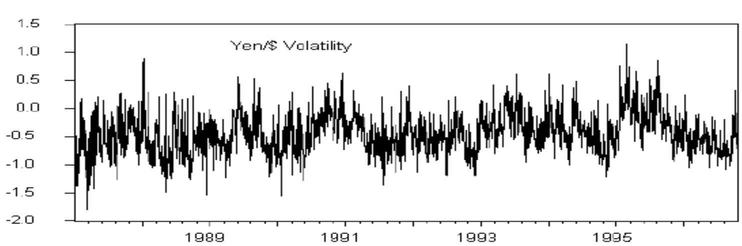
Accuracy of RV Measures: 95% CI from BNS Asymptotic theory



Source: BNS (2002)

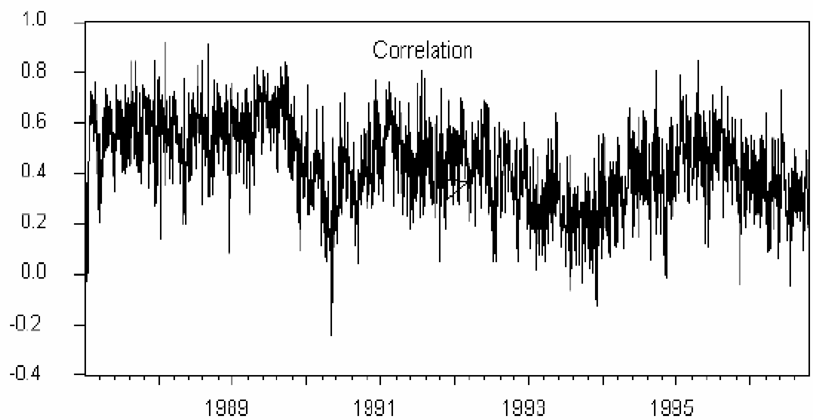
Time Series of Daily RVOL: m=228





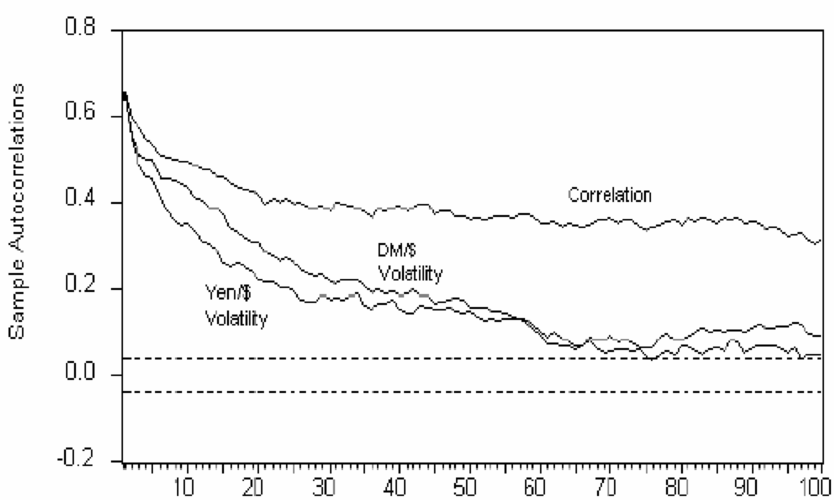
Source: ABDL (2001)

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:

$$\rho_k \to C_\rho \bullet k^{-\alpha}, \text{ as } k \to \infty$$
$$\alpha \in (0,1)$$

Fractionally Differenced Processes

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

$$(1-L)^d (y_t - \mu) = u_t,\ u_t \sim I(0)$$
$$0 < d < 0.5: \text{ stationary long memory}$$
$$0.5 \le d < 1: \text{ nonstationary long memory}$$

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 Estimated of d

	RV_D	RV_Y	$RVOL_D$	$RVOL_Y$	$RLVOL_D$	$RLVOL_Y$	$RCOV$	$RCOR$
$\hat{d}$	.356	.339	.381	.428	.420	.455	.334	.413

Note: Multivariate estimation of common d using ($RLVOL_D$, $RLVOL_Y$, $RLVOL_{DY}$) 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

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

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

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

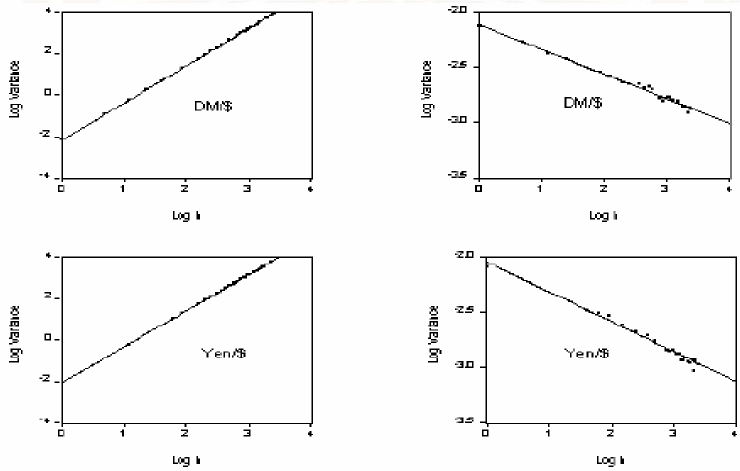
Temporal Aggregation and Estimated of d

GPH Estimates of d

h	RV_D	RV_Y	$RVOL_D$	$RVOL_Y$	$RLVOL_D$	$RLVOL_Y$	$RCOV$	$RCOR$
1	.356	.339	.381	.428	.420	.455	.334	.413
5	.457	.429	.446	.473	.405	.496	.368	.519
10	.511	.490	.470	.501	.515	.507	.436	.494
15	.400	.426	.384	.440	.421	.440	.319	.600
20	.455	.488	.440	.509	.496	.479	.439	.630

Temporal Aggregation and Scaling Laws

RV RLVOL



Source: ABDL (2001)

Distribution of Returns Standardized by RV

- ABDL (2000): “Exchange Rate Returns Standardized by Realized Volatility Are (Nearly) Gaussian,” *Multinational Finance Journal*

Stochastic Volatility Model

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

$$r_t = \sigma_t \varepsilon_t$$
$$\sigma_t = \text{latent volatility}$$
$$\varepsilon_t \sim iid\ (0,1)$$
$$\Rightarrow E[r_t^2] = \sigma_t^2$$

Standardized Returns

- Compute returns standardized by estimates of conditional volatility

$$\hat{\varepsilon}_t = \frac{r_t}{\hat{\sigma}_t}$$
$$\hat{\sigma}_t = RVOL_t,\ m = 288$$
$$\hat{\sigma}_t = \hat{\sigma}_t^{GARCH(1,1)}$$
$$\text{GARCH(1,1): } \sigma_t^2 = w + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2$$

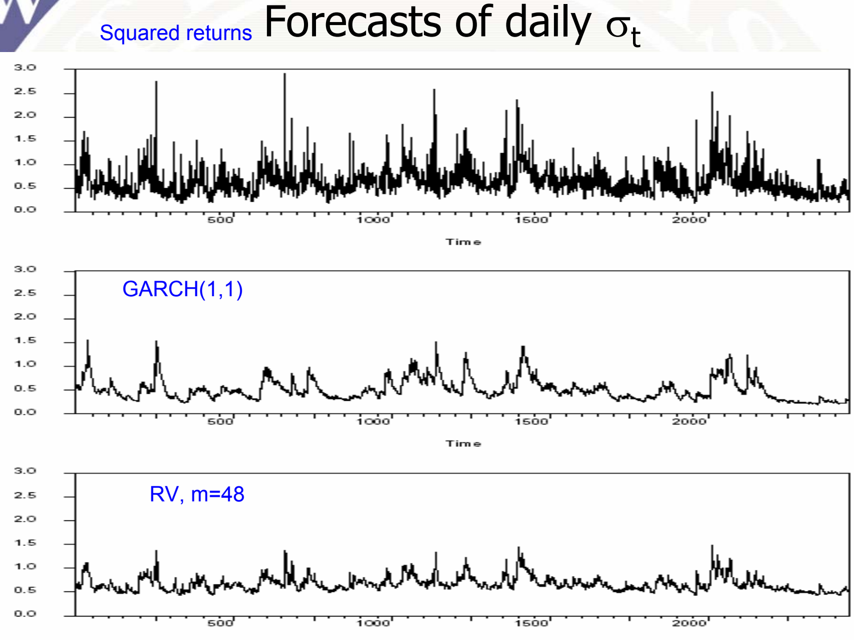
Multivariate Standardized Returns

- Standardized returns based RCOV

$$\begin{pmatrix} \hat{\varepsilon}_{D,t} \\ \hat{\varepsilon}_{Y,t} \end{pmatrix} = RCOV_t^{-1/2} \begin{pmatrix} r_{D,t} \\ r_{Y,t} \end{pmatrix}$$

$$RCOV_t^{1/2} = \text{ Cholesky factor of } RCOV_t$$

Forecasts of daily σ_t



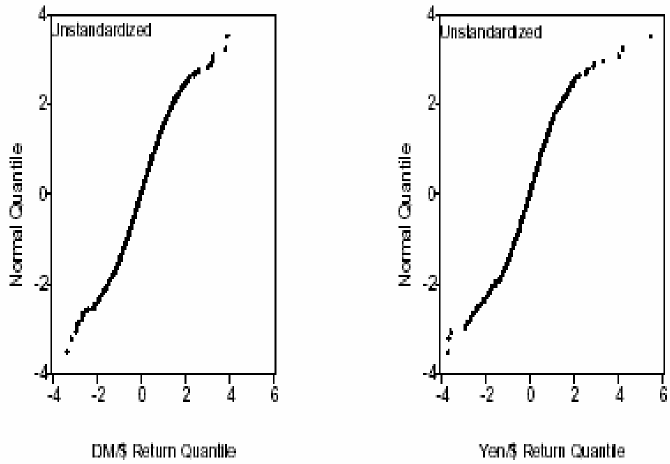
Comparison of Volatility Forecasts

- Squared returns are unbiased but very noisy
- GARCH(1,1) forecasts are smoother than RV forecasts; do not utilize information between time *t*-1 and *t* (exponentially weighted average of past returns)
- RV forecasts make exclusive use of information between time *t*-1 and *t*; better forecast of time *t* volatility

Summary Statistics

	r_t		$\frac{r_t}{\hat{\sigma}_t^{GARCH}}$		$\frac{r_t}{RVOL_t}$	
	DM/$	Y/$	DM/$	Y/$	DM/$	Y/$
Mean	-.007	-.009	-.002	-.011	-.007	.007
Std. Dev.	.710	.705	1.00	1.00	1.01	.984
Skewness	.033	.052	-.027	-.139	.015	.002
Kurtosis	5.40	7.36	4.75	5.41	2.41	2.41
Correlation	.659		.661		.661	

Distribution of Daily Returns

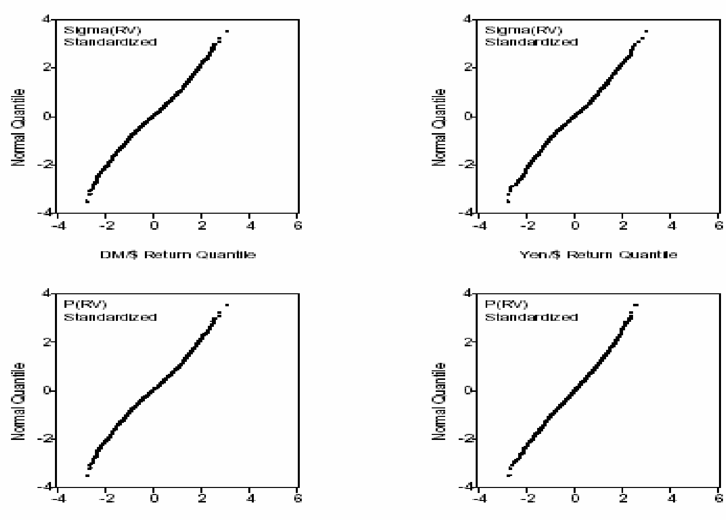


Source: ABDL (2000)

Distribution of Standardized Returns

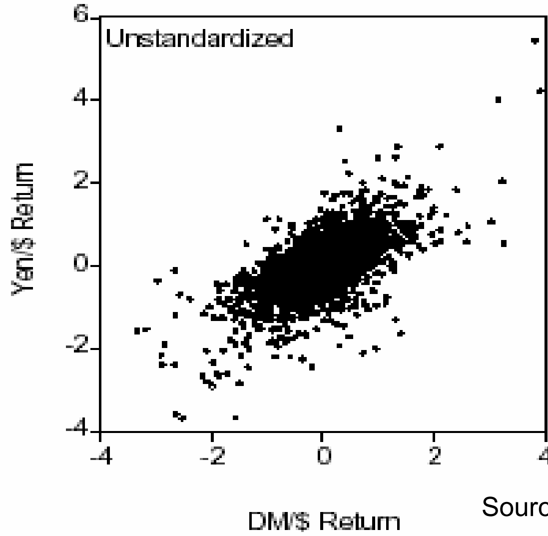
RV

RCOV



Source: ABDL (2000)

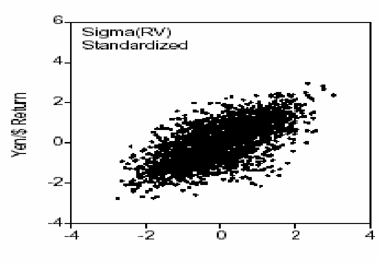
Scatterplot of Daily Returns



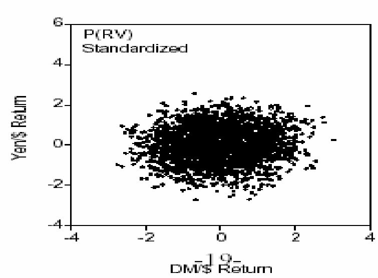
Source: ABDL (2000)

Scatterplot or Standardized Returns

RV

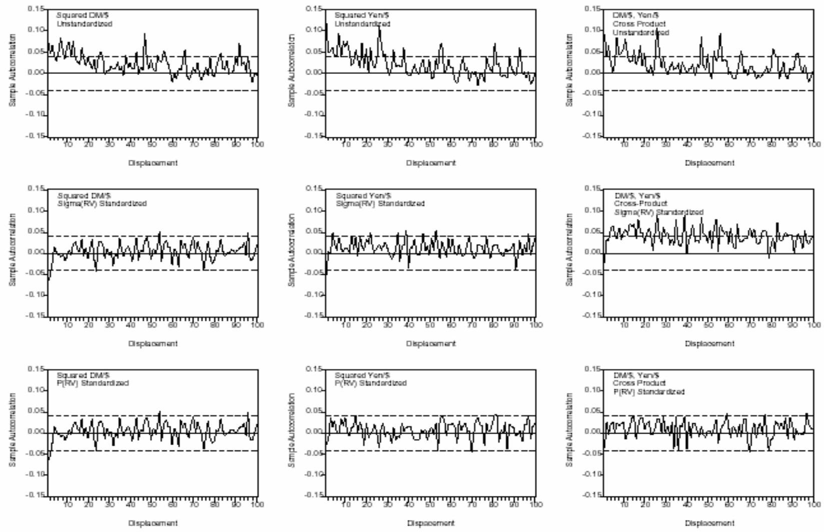


RCOV



Source: ABDL (2000)

SACF of Squared Returns



RAW

RV

RCOV

DM/$ | Yen/$ | DM/$, Yen/$

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
- RV provides superior volatility forecasts

Modeling and Forecasting RV

- ABDL (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*

Traditional Conditional Volatility Models

- Normal GARCH(1,1)

$$r_t = \sigma_t \varepsilon_t,\ \varepsilon_t \sim iid\ N(0,1)$$
$$\sigma_t^2 = w + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2$$

- Log-Normal SV model

$$r_t = \sigma_t \varepsilon_t,\ \varepsilon_t \sim iid\ N(0,1)$$
$$\ln \sigma_t^2 = \delta + \phi \ln \sigma_{t-1}^2 + \sigma_u u_t,\ \mathrm{u}_t \sim iid\ N(0,1)$$
$$E[\varepsilon_t u_t] = 0$$

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

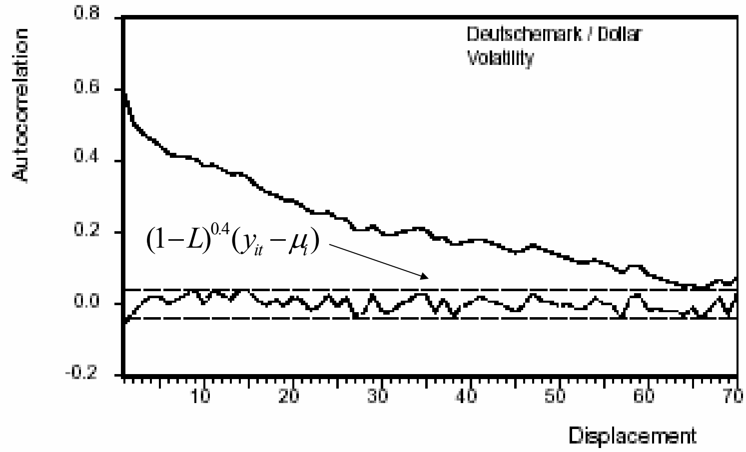
Trivariate System of Exchange Rates

$$y_t = \begin{pmatrix} RLVOL_{D/\$,t} \\ RLVOL_{Y/\$,t} \\ RLVOL_{Y/D,t} \end{pmatrix},\ m = 48$$

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

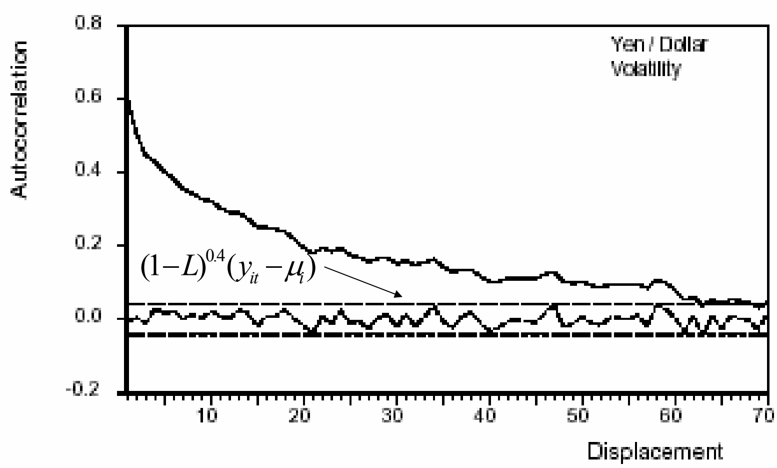
- Fit models for y_t in sample: 12/1/86-12/1/96
- Forecast y_t out-of-sample: 12/2/96 – 6/30/99

SACF of Daily DM/$ RLVOL: m=48



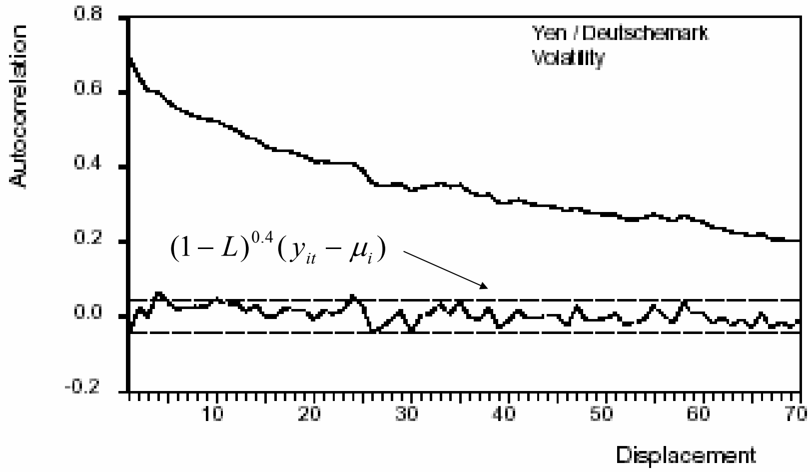
Source: ABDL (2003)

SACF of Daily Yen/$ RLVOL: m=48



Source: ABDL (2003)

SACF of Daily Yen/DM RLVOL: m=48



Source: ABDL (2003)

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 $|r_t|$
- AR-RV: univariate AR(5) fit to $(1-L)^{0.4}RLVOL_{i,t}$
- 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

$$RVOL_{i,t} = b_0 + b_1 R\hat{V}OL_{i,t}^{VAR-RV} + b_2 R\hat{V}OL_{i,t}^{\text{model}} + error_t$$

$$R\hat{V}OL_{i,t}^{VAR-RV} = \text{1-day ahead forecast from RV-VAR}$$

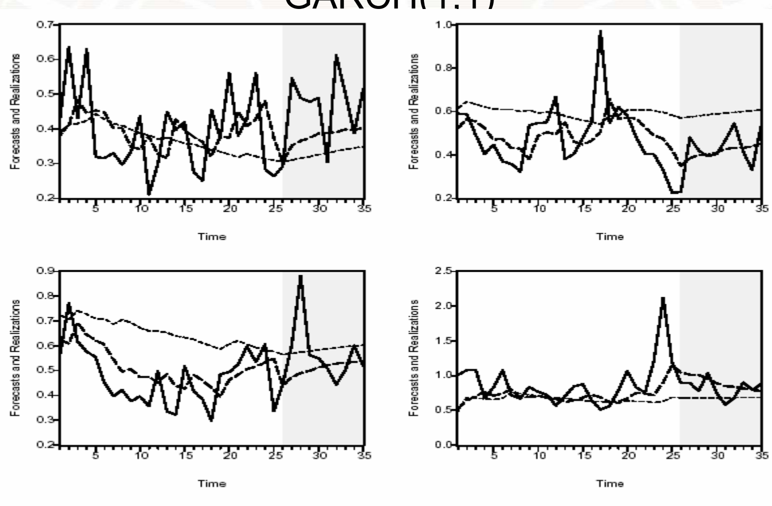
$$R\hat{V}OL_{i,t}^{\text{model}} = \text{1-day ahead forecast from alternative model}$$

$$H_0 : b_0 = 0,\ b_1 = 1,\ b_2 = 0$$

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.

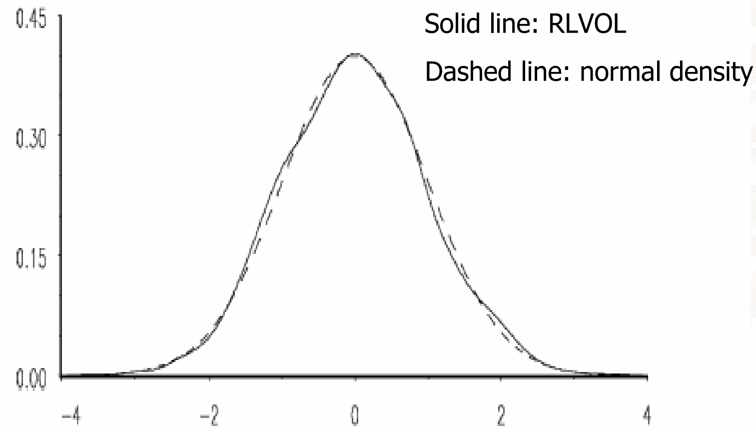
Empirical Analysis of TAQ Data

- Andersen, Bollerslev, Diebold, Ebens (2001): "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics*
  - 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

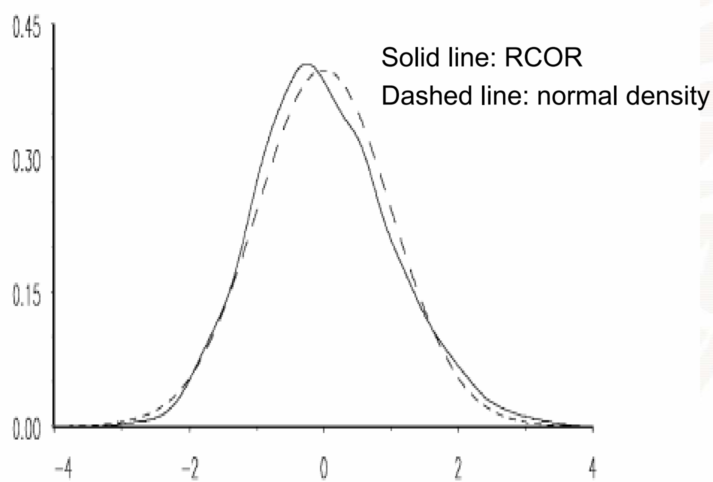
- Results for equity returns are similar to those for FX returns
  - RLVOL, RCOR are approximately Gaussian
  - RV measures exhibit long memory
  - Daily returns standardized by RVOL are nearly Gaussian
- Little evidence of leverage effect
- Evidence of factor structure in multivariate system of RV measures

Distribution of Daily RLVOL: Alcoa



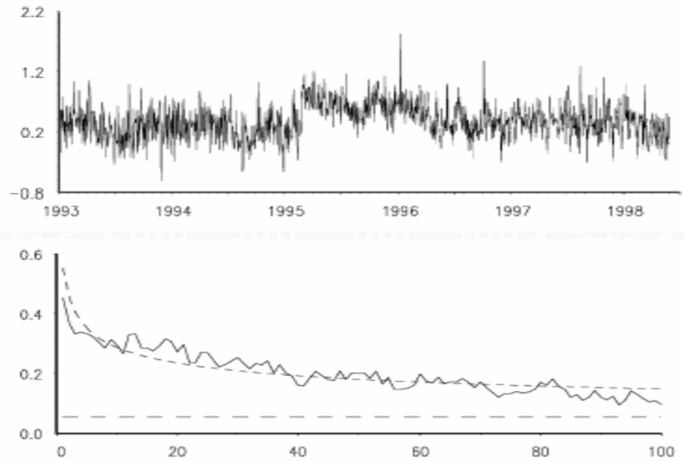
Source: ABDE (2001)

Distribution of Daily RCOR: Alcoa,Exxon



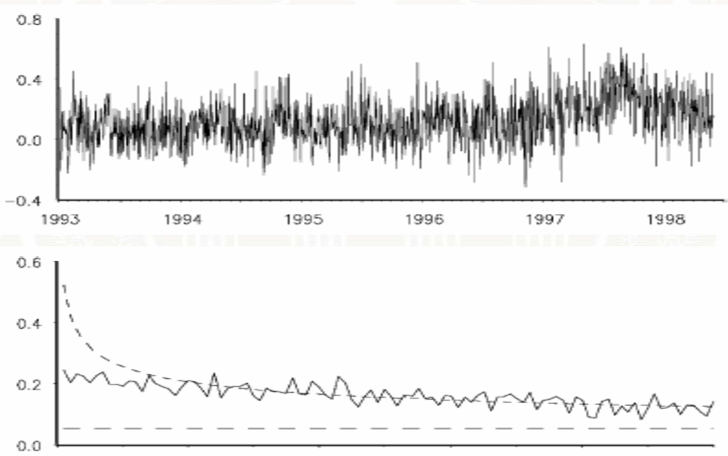
Source: ABDE (2001)

Time Series of Daily RLVOL: Alcoa



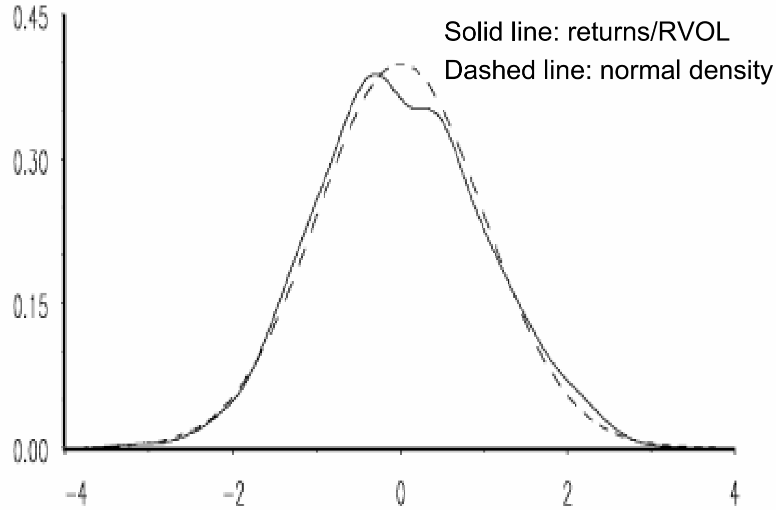
8/8/2005 Source: ABDE (2001)

Time Series of Daily RCOR: Alcoa, Exxon

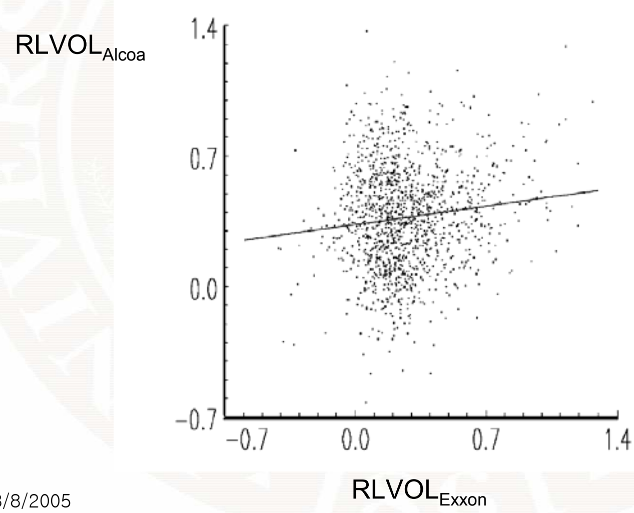


Source: ABDE (2001)

Distribution of Daily Standardized Returns for Alcoa

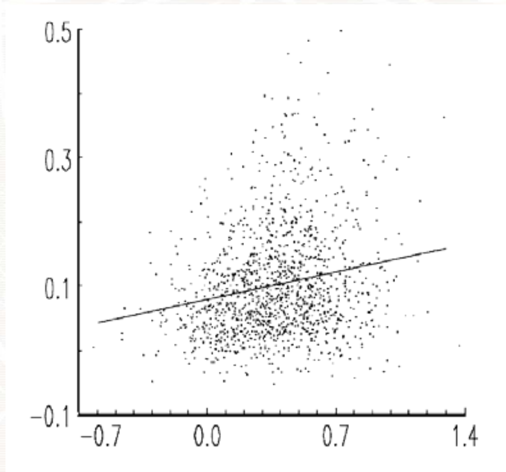


Evidence for Factor Structure



Evidence of Factor Structure

$RCOR_{Alcoa,i}$

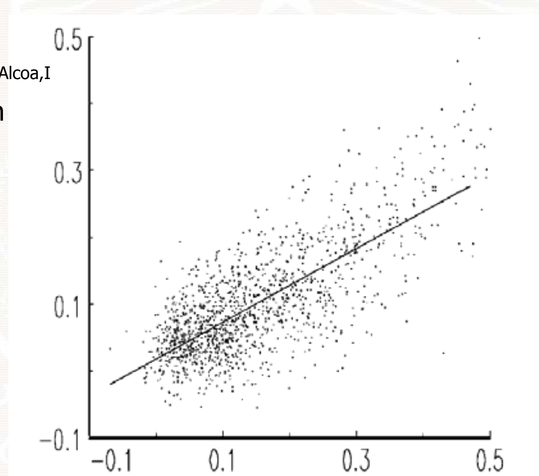


$RLVOL_{Alcoa}$

Evidence of Factor Structure

Average $RCOR_{Alcoa,I}$

i≠Alcoa, Exxon



Average $RCOR_{Exxon,I}$ i≠Alcoa, Exxon

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