Exchange Rate Returns Standardized by Realized Volatility are (Nearly) Gaussian*

Torben G. Andersen  
Northwestern University, U.S.A.

Tim Bollerslev  
Duke University and NBER, U.S.A.

Francis X. Diebold  
University of Pennsylvania and NBER, U.S.A.

Paul Labys  
University of Pennsylvania, U.S.A.

It is well known that high-frequency asset returns are fat-tailed relative to the Gaussian distribution, and that the fat tails are typically reduced but not eliminated when returns are standardized by volatilities estimated from popular ARCH and stochastic volatility models. We consider two major dollar exchange rates, and we show that returns standardized instead by the realized volatilities of Andersen, Bollerslev, Diebold and Labys (2000a) are very nearly Gaussian. We perform both univariate and multivariate analyses, and we trace the differing effects of the different standardizations to differences in information sets (JEL C10, C22, C32, G15, G12).

Keywords: high-frequency data, integrated volatility, realized volatility, risk management.

*This work was supported by the National Science Foundation. We are grateful to Olsen and Associates for making available the intraday exchange rate quotations. For useful comments we thank David Backus, David Bates, Rob Engle, John Geweke, and seminar participants at New York University. Parts of this paper were written while Diebold visited the Stern School of Business, New York University, whose hospitality is gratefully acknowledged.

(Multinational Finance Journal, 2000, vol. 4, no. 3&4, pp. 159–179)  
©Multinational Finance Society, a nonprofit corporation. All rights reserved.
I. Introduction

The prescriptions of modern financial risk management hinge critically on the associated characterization of the distribution of future returns [cf., Diebold, Gunther, and Tay (1998), and Diebold, Hahn, and Tay (1999)]. Because volatility persistence renders speculative returns temporally dependent [e.g., Bollerslev, Chou, and Kroner (1992)], it is the conditional return distribution, not the unconditional distribution, that is of relevance for risk management. This is especially true in high-frequency situations, such as monitoring and managing the risk associated with the day-to-day operations of a trading desk, where volatility clustering is a well recognized fact of life.

Unconditional distributions of exchange rate returns are routinely found to be symmetric but highly leptokurtic. Standardized daily or weekly returns from ARCH and related stochastic volatility models also appear symmetric but leptokurtic; that is, the distributions are not only unconditionally, but also conditionally leptokurtic, although less so than unconditionally. Hence a sizable literature explicitly attempts to model the fat-tailed conditional distributions, including, for example, Bollerslev (1987), Engle and Gonzalez-Rivera (1991), and Hansen (1994).

Let us make the discussion more precise. Assuming that return dynamics operate only through the conditional variance, a standard decomposition of the return innovation is \( r_i = \sigma_i \varepsilon_i \) where \( \sigma_i \) refers to the time-\( t \) conditional standard deviation, and \( \varepsilon_i^{\text{iid}} \sim (0,1) \). Thus, given \( \sigma_i \), it would be straightforward to back out \( \varepsilon_i \) and assess its distributional properties. Of course, \( \sigma_i \) is not directly observable. When using an estimate of \( \sigma_i \), the distributions of the resulting standardized returns are typically found to be fat-tailed, or leptokurtic.¹

The main focus of the present paper is similar – we are also concerned with the shape of the distributions of standardized returns. However, there is an important distinction: our volatility measure is fundamentally different from the ARCH and related estimators that have featured prominently in the literature, and hence our estimates of the conditional distribution differ as well. In particular, we rely on so-

¹ This result has motivated the practical use of various "fudge-factors" relative to the standard normal quantiles in the construction of Value-at-Risk type statistics.
called realized volatility measures constructed from high-frequency intraday returns, as previously analyzed by Schwert (1990), Hsieh (1991), Andersen and Bollerslev (1998), and Andersen, Bollerslev, Diebold and Labys (2001a), Andersen, Bollerslev, Diebold and Ebens (2001), among others.

We proceed to study ten years of high-frequency returns for the Deutschmark - U.S. Dollar (DM/$) and Japanese Yen - U.S. Dollar (Yen/$) exchange rates. In order to establish a proper benchmark, in section 2, we provide a characterization of the distribution of the daily unstandardized returns. In section 3, we characterize the distribution of the daily returns when standardized by univariate realized volatility measures, and, in section 4, we characterize the distribution of the returns when standardized by realized volatilities in a multivariate fashion. For comparison, in section 5, we examine the distribution of returns standardized by GARCH(1,1) volatilities, along with the distribution of returns standardized by one-day-ahead volatility forecasts from a simple ARMA(1,1) model fit directly to realized volatility. We conclude in section 6.

II. Unstandardized Returns

Our empirical analysis is based on 10-year time series of 5-minute DM/$ and Yen/$ returns from December 1, 1986 through December 1, 1996. The data were kindly supplied by Olsen & Associates. After omitting weekend and other holiday non-trading periods, as detailed in Andersen, Bollerslev, Diebold, and Labys (2001a), we are left with a total of $T=2,445$ complete days, each of which consists of 288 5-minute returns. From these we proceed to construct time series of continuously compounded 30-minute and daily returns.

We begin our analysis with a summary of the distributions for the unstandardized, or raw, daily DM/$ and Yen/$ returns. The results appear in table 1 and figures 1 through 3. Consistent with the extant literature, the s-shaped quantile-quantile plots for the two marginal distributions in the top panel of figure 1 indicate that both returns are symmetric but fat-tailed relative to the normal distribution. The statistics reported in the first panel of table 1 confirm that impression: the sample skewness is near 0 for both series, but the sample kurtoses are well above the normal value of 3.
<table>
<thead>
<tr>
<th></th>
<th>Un-Standardized</th>
<th>σ(RV) Standardized</th>
<th>P(RV) Standardized</th>
<th>σ((GARCH)) Standardized</th>
<th>Yen/$</th>
<th>DM/$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>-0.007</td>
<td>0.007</td>
<td>0.007</td>
<td>-0.007</td>
<td>0.016</td>
<td>0.013</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>-0.01</td>
<td>0.017</td>
<td>0.017</td>
<td>-0.01</td>
<td>0.016</td>
<td>0.016</td>
</tr>
<tr>
<td><strong>Maximum</strong></td>
<td>3.099</td>
<td>5.445</td>
<td>5.954</td>
<td>5.782</td>
<td>7.4</td>
<td>0.11</td>
</tr>
<tr>
<td><strong>S.d.Dev.</strong></td>
<td>7.1</td>
<td>7.05</td>
<td>1.000</td>
<td>1.073</td>
<td>-1.39</td>
<td>-1.008</td>
</tr>
<tr>
<td><strong>Skewness</strong></td>
<td>0.033</td>
<td>0.052</td>
<td>0.015</td>
<td>-0.027</td>
<td>0.002</td>
<td>0.061</td>
</tr>
<tr>
<td><strong>Kurtosis</strong></td>
<td>5.395</td>
<td>7.357</td>
<td>2.406</td>
<td>2.414</td>
<td>0.661</td>
<td>0.676</td>
</tr>
<tr>
<td><strong>Correlation</strong></td>
<td>0.659</td>
<td>0.661</td>
<td>0.661</td>
<td>0.661</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1.— Quantile plots daily exchange rate returns.
Figure 2.— Scatterplots of daily exchange rate returns.
Figure 3.—Sample autocorrelation functions of daily exchange rate returns.
Figure 3.— (Continued)
Figure 3.— (Continued)
Turning to the joint distribution, not surprisingly, the two rates show considerable dependence, with a sample correlation of .66. This high degree of dependence is further underscored by the bivariate scatterplot in the top panel of figure 2, which also clearly illustrates the marginal fat tails in terms of the many outliers relative to the tight ellipsoid expected under bivariate normality.

Finally, we consider the conditional distribution of the unstandardized returns, as summarized by the autocorrelations for each of the two daily squared return series and the cross product of the two rates. The relevant correlograms to a displacement of 100 days, along with the Bartlett standard errors, appear in the top panel of figure 3. Again, directly in line with existing evidence in the literature, the results indicate highly persistent conditional variance and covariance dynamics.

III. Univariate Standardization by Realized Volatility

In the absence of any short-run predictability in the mean, which is a good approximation for the two exchange rates analyzed here, the univariate return series are naturally decomposed as $r_t = \sigma_t \varepsilon_t$, where $\varepsilon_t \sim (0,1)$, and $\sigma_t$ is the time-\(t\) conditional standard deviation. On rearranging this decomposition, we obtain the $\sigma$-standardized return,

$$\varepsilon_t = \frac{r_t}{\sigma_t},$$

on whose distribution and dependence structure we now focus.

In practice, of course, $\sigma_t$ is unknown and must be estimated.\(^2\) Many volatility models have been proposed in the literature. However, as formally shown by Andersen, Bollerslev, Diebold and Labys (2001a), in a continuous time setting the ex post volatility over a day may be estimated to any desired degree of accuracy by summing sufficiently high-frequency returns within the day. Following this analysis we shall refer to the corresponding measures as realized volatilities.

In order to define formally our daily realized volatilities, let the two series of 30-minute DM/$ and Yen/$ returns be denoted by $\Delta \log D_{(48),t}$

\[2. \text{ In an abuse of notation, we will continue to use } \sigma_t \text{ to denote an estimate of the volatility, as the meaning will be clear from context.}\]
and \( \Delta \log Y_{(48),t} \), respectively, where \( t = 1/48, 2/48, ..., 2,445 \), and "48" refers to the 48 30-minute intervals in the 24-hour trading day. From these \( 48 \times 2,445 = 117,360 \) 30-minute returns, we estimate the daily variances by simply summing the 48 squared returns within each day. That is,

\[
\sigma^2_{D_t} (RV) = \sum_{j=1, \ldots, 48} \left( \Delta \log D_{(48),t-1+j/48} \right)^2,
\]

\[
\sigma^2_{Y_t} (RV) = \sum_{j=1, \ldots, 48} \left( \Delta \log Y_{(48),t-1+j/48} \right)^2,
\]

where \( t = 1, 2, ..., 2,445 \), and \( RV \) stands for "realized volatility." Our choice of 30-minute returns is motivated by examination of the volatility signature plot of Andersen, Bollerslev, Diebold and Labys (2000), which records average realized volatility as a function of underlying sampling frequency. In the present application, average DM/$ and Yen/$ realized volatility remain stable as underlying sampling frequency increases up to approximately 30-minute returns, as shown in figure 4.3 This suggests that in the present context 30-minute sampling provides a reasonable balance between the salient market microstructure frictions at the very highest sampling frequencies, on the one hand, and the accuracy of the continuous record asymptotics underlying the estimators, on the other.

We now proceed to examine the \( \sigma(RV) \)-standardized returns for each of the two currency series.4 The quantile-quantile plots in the middle panel of figure 1 look radically different from those in the top panel. In particular, they are now nearly linear, indicating that a Gaussian distribution affords a close approximation to each of the marginal

---

3. The quantity "K" on the horizontal axis of the volatility signature plots is the number of 5-minute blocks in each return interval. Hence, for example, \( K=6 \) corresponds to \( 6 \times 5=30 \) minute returns.

4. The reader will notice that here and throughout we do not report results of formal tests of normality. This is intentional. All such tests strongly reject normality, because even small deviations from normality are easily detected in the very large samples available here. But such tests are not constructive, in the sense that they convey little information regarding the "size" and "shape" of the deviation from normality. We find quantile-quantile plots, along with standard skewness and kurtosis coefficients, much more informative. Hence we focus on them.
Figure 4.—Volatility signature plots daily exchange rate returns.

distributions.
The diagnostic statistics in the second panel of table 1 confirm that impression: the distributions of the \( \sigma(RV) \) standardized daily returns are remarkably close to a standard normal. The means are near zero, the standard deviations are close to one, the skewnesses coefficients are close to zero, and the coefficients of kurtosis are near three.\(^5\) If

\(^5\) In an independent study, Bollen and Inder (1999) have recently observed that the distribution of \( \sigma(RV) \)-standardized daily S&P500 futures returns also appears approximately
anything, the distributions appear slightly thin-tailed, or platykurtic. Interpreting the realized volatility as an ideal measure of the rate of information flow to the market, these findings are therefore consistent with the distributional assumptions underlying the Mixture-of-Distributions-Hypothesis (MDH) as originally advocated by Clark (1973); see also Tauchen and Pitts (1983), Taylor (1986), and Andersen (1996).

Proceeding to the joint unconditional distribution of the $\sigma(RV)$-standardized returns, not surprisingly, we see from the second panels of table 1 and figure 2 that the correlation remains high. Interestingly, however, the outliers in the joint density have been largely eliminated. As for the conditional distribution, the correlograms in figure 3 for the squares and the cross product of the daily $\sigma(RV)$-standardized returns indicate the absence of any remaining conditional variance dynamics for the DM/$ rate, and a great reduction in the conditional variance dynamics for the Yen/$ rate. Meanwhile, the autocorrelations for the cross product of the standardized returns decay more slowly than the autocorrelations for the product of the raw returns. Thus, although the univariate standardization has largely eliminated the conditional variance dynamics, it has actually magnified the conditional covariance dynamics. Elimination of both requires a multivariate standardization, to which we now turn.

IV. Multivariate Standardization by Realized Volatility

With a slight abuse of notation, the multivariate case is conveniently written as,

$$r_t = P_t \varepsilon_t,$$

where both $r_t$ and $\varepsilon_t^{iid} \sim (0, I)$ are now $N \times 1$ vectors, and $P_t$ refers to the $N \times N$ matrix square-root of the time-$t$ conditional covariance matrix for the raw returns, $\Sigma_t$, so that in particular $P_t' P_t = \Sigma_t$. Of course, the

Gaussian.
matrix square-root operator is not unique. For concreteness, we rely here on the unique $N \times N$ lower-triangular Cholesky factorization. The corresponding $P$-standardized return vector is then readily defined as

$$
\varepsilon_i = P_i^{-1} r_i,
$$

which, in general, will differ from the corresponding vector of stacked univariate $\sigma$-standardized returns. In particular, we have

$$
\Sigma_i = \begin{bmatrix} \sigma_{DD}^2 & \sigma_{DY}^2 \\ \sigma_{DY}^2 & \sigma_{YY}^2 \end{bmatrix} = P_i' P_i' = \begin{bmatrix} p_{11t} & 0 \\ p_{21t} & p_{22t} \end{bmatrix} \begin{bmatrix} p_{11t} & p_{21t} \\ 0 & p_{22t} \end{bmatrix}
$$

$$
= \begin{bmatrix} p_{11t}^2 & p_{21t} p_{11t} \\ p_{11t} p_{21t} & p_{21t}^2 + p_{22t}^2 \end{bmatrix},
$$

where we have arbitrarily arranged the bivariate returns as (DM/$, Yen/$). Upon matching terms, it follows that

$$
p_{11t} = \sigma_{DD}^1, \quad p_{21t} = \frac{\sigma_{DY}}{\sigma_{DD}^1}, \quad p_{22t} = \sqrt{\frac{\sigma_{YY}^2 - \sigma_{DY}^2}{\sigma_{DD}^1}},
$$

so that

$$
P_i^{-1} = \frac{1}{p_{11t} p_{22t}} \begin{bmatrix} p_{22t} & 0 \\ -p_{21t} & p_{11t} \end{bmatrix} = \begin{bmatrix} \frac{1}{p_{11t}} & 0 \\ -\frac{p_{21t}}{p_{11t}} & \frac{1}{p_{22t}} \end{bmatrix} \begin{bmatrix} 1 \\ p_{11t} \\ -p_{21t} \\ p_{11t} p_{22t} \end{bmatrix}.$$
\[
\begin{bmatrix}
\frac{1}{\sigma_{DDt}} & 0 \\
-\frac{\sigma_{DYt}}{\sigma_{DDt}^2} \times \frac{1}{\sqrt{\sigma_{YYt}^2 - \sigma_{DYt}^2 / \sigma_{DDt}^2}} & \frac{1}{\sqrt{\sigma_{YYt}^2 - \sigma_{DYt}^2 / \sigma_{DDt}^2}}
\end{bmatrix}
\]

Hence, \( P \)-standardization of a time-\( t \) return vector is equivalent to element-by-element, or univariate, \( \sigma \)-standardization only in the special and counterfactual case of \( \sigma_{DYt} = 0 \). The \( P \)-standardization simply \( \sigma \)-standardizes the return placed first in the ordering, whereas it substitutes a linear combination of the two unstandardized returns for the second return.\(^6\)

Of course, the \( P \) matrix involves both exchange rate variances and their covariance. Analogous to our realized variance estimator, the \textit{realized covariance} is readily defined as the sum of the intra-day cross products:

\[
\sigma_{DY}(RV) \equiv \sum_{j=1, \ldots, 48} \Delta \log D_{(48), t-l+j/48} \times \Delta \log Y_{(48), t-l+j/48}
\]

Armed with these realized variances and covariances, we now proceed to construct and examine -standardized returns.

We report the results in the third panel of table 1 and figures 1-3.\(^7\) The differences, as expected, arise primarily in the multivariate dimensions of the distribution. The sample correlation between the bivariate \( P(RV) \)-standardized returns, as reported in table 1, is greatly reduced from .66 to only .08. Moreover, the scatterplot reported in the third panel of figure 2 now appears spherical, confirming the negligible correlation. Importantly, the correlogram for the cross products of the daily \( P(RV) \)-standardized returns, reported in the third panel of figure 3,

\(^6\) This mirrors the dependence on the ordering of the variables in the analysis of vector autoregressions identified by a Wold Causal Chain.

\(^7\) We have ordered the bivariate returns as (DM/$, Yen/$). Although in general the ordering can affect the results, it is inconsequential in the present application.
confirms that the conditional covariance dynamics have been eliminated. The differences, however, are not exclusively in terms of the multivariate features. In particular, the $P(RV)$-standardization also produces an improved correlogram for the Yen/$\$$ returns relative to that of the $\sigma(RV)$-standardized returns.

V. Standardization by GARCH(1,1) Volatility and by Forecasts of Realized Volatility

Numerous parametric volatility models have been suggested in the literature for best capturing the conditional temporal dependencies in $\sigma_t$. The most commonly used specification is the simple univariate GARCH(1,1) model, and we follow standard practice by utilizing this as an illustrative benchmark for each of the two rates. That is, we posit that

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2.$$

We refer to the associated estimates of the conditional standard deviations as $\sigma(GARCH)$, with the $\sigma(GARCH)$-standardized daily returns defined accordingly.

Consistent with the prior literature, the summary statistics in the fourth panel of table 1 show that standardization by $\sigma(GARCH)$ reduces, but does not eliminate, the excess kurtosis. In particular, the sample kurtosis for the DM/$\$$ drops from 5.4 to 4.8, while the Yen/$\$$ kurtosis is reduced from 7.4 to 5.4. Thus, in each case, significant excess kurtosis remains after the standardization.

It is natural to ask why such different results obtain for the $\sigma(RV)$-standardized versus the $\sigma(GARCH)$-standardized returns. Of course, in general, we would expect different measures for $\sigma_t$ to affect the properties of the standardized returns. However, in this case, there is a specific aspect of the calculations that makes an obvious difference: $\sigma_t(RV)$ is an estimate of the volatility for the day-$t$ returns conditional on the continuous (or high-frequency discrete intraday) sample path of stochastic volatility up to and including day $t$, whereas $\sigma_t(GARCH)$ is
an estimate of the volatility of day-\( t \) returns conditional on the discrete sample path of returns up to but not including day \( t \).

To further underscore the importance of this difference, we next calculate \( \sigma_t(RVF) \) as a one-day-ahead forecast of the realized volatility made at day \( t-1 \), where the forecast is obtained by linear projection of the realized volatility on its own past. That is, we fit an ARMA model to realized volatility and use the fitted model to make forecasts. This approach is much closer in spirit to the \( \sigma(GARCH) \) estimator analyzed above, and we therefore conjecture that standardization by \( \sigma_t(RVF) \) will reduce, but not eliminate, the excess kurtosis.\(^8\)

For ease of comparison to the GARCH(1,1) case, we shall rely on a simple univariate ARMA(1,1) structure for modeling the realized volatilities.\(^9\) Also, in direct analogy to the GARCH(1,1) case, the model is estimated over the full ten-year sample.\(^10\) From these estimates, we proceed with the creation of standard 1-day-ahead forecasts, from which we obtain our \( \sigma_t(RVF) \) series, and corresponding \( \sigma(RVF) \)-standardized returns.

The diagnostic statistics in the last panel of table 1 show that the distributions of the \( \sigma(RVF) \)-standardized returns and the \( \sigma(GARCH) \)-standardized returns are fairly similar. In particular, both exhibit fat tails relative to the normal. Figure 5 clearly reveals the reason behind this divergence between the \( \sigma(RV) \)-standardized returns, which to a first

---

8. It is tempting to conjecture that \( \sigma(RVF) \)-standardized returns will be less fat-tailed than \( \sigma(GARCH) \)-standardized returns, because \( \sigma_{t-1}(RV) \) should provide a superior measure of recent volatility relative to \( \sigma_{t-1}(GARCH) \), which is operative in the GARCH(1,1) recursion. However, there is generally no simple relation between forecasts based on more relative to less information and the resulting amount of excess kurtosis of the corresponding standardized returns, as explained for example in Nelson (1996).

9. The univariate \( \sigma(RV) \) forecasting exercise reported here is highly stylized. A more detailed analysis based on a multivariate model is undertaken in Andersen, Bollerslev, Diebold and Labys (2001b). A related approach was recently pursued by Taylor and Xu (1997) in analyzing the informational content in high-frequency exchange rates and implied volatilities from options.

10. Assuming no structural breaks during our 10-year sample, and that the dynamics remained unchanged, we can justify the use of one-day-ahead volatility forecasts based on full-sample parameter estimates. Estimation with the full sample also has the obvious advantage that it avoids the early-on instability associated with recursive estimation.
Figure 5.—Time series of alternative volatility measures daily exchange rate returns.
approximation appear Gaussian, and the leptokurtic $\sigma_{(RVF)}$- and $\sigma_{(GARCH)}$-standardized returns. The $\sigma_{(RVF)}$ and $\sigma_{(GARCH)}$ volatility series are both one-day-ahead forecasts, and so are smoother than the object being forecast, which is effectively the $\sigma_{(RV)}$ volatility series. Hence, standardization by $\sigma_{(RVF)}$ or $\sigma_{(GARCH)}$ is insufficient to eliminate the excess kurtosis, whereas standardization by the $\sigma_{(RV)}$ is able to accomplish that goal.

VI. Summary and Concluding Remarks

It is well known that daily asset returns are fat-tailed relative to the Gaussian distribution, and that the fat tails are typically reduced but not eliminated when returns are standardized by volatilities estimated from popular models such as GARCH. We have considered two major dollar exchange rates, and we have shown that returns standardized instead by the realized volatilities of Andersen, Bollerslev, Diebold and Labys (2000a) are very nearly Gaussian. We performed both univariate and multivariate analyses, and we traced the different effects of the different standardizations to differences in information sets, thereby extending the independent work of Zhou (1996).

Our analysis is important because it helps set the stage for improved high-dimensional volatility modeling and out-of-sample forecasting, which in turn hold promise for the development of better decision making in practical situations of risk management, portfolio allocation, and asset pricing. The lognormal-normal mixture model developed in Andersen, Bollerslev, Diebold and Labys (2001b), for example, relies heavily on both the lognormality of realized volatility, as documented in Andersen, Bollerslev, Diebold and Labys (2001a) and Andersen, Bollerslev, Diebold and Ebens (2001), and the normality of returns standardized by realized volatility, as documented here.

References


