

Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management:  
High-Frequency Returns on Foreign Exchange

Author(s): Francis X. Diebold, Jinyong Hahn and Anthony S. Tay

Source: *The Review of Economics and Statistics*, Nov., 1999, Vol. 81, No. 4 (Nov., 1999),  
pp. 661-673

Published by: The MIT Press

Stable URL: <https://www.jstor.org/stable/2646715>

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



The MIT Press is collaborating with JSTOR to digitize, preserve and extend access to *The Review of Economics and Statistics*

JSTOR

# MULTIVARIATE DENSITY FORECAST EVALUATION AND CALIBRATION IN FINANCIAL RISK MANAGEMENT: HIGH-FREQUENCY RETURNS ON FOREIGN EXCHANGE

Francis X. Diebold, Jinyong Hahn, and Anthony S. Tay\*

*Abstract*—We provide a framework for evaluating and improving multivariate density forecasts. Among other things, the multivariate framework lets us evaluate the adequacy of density forecasts involving cross-variable interactions, such as time-varying conditional correlations. We also provide conditions under which a technique of density forecast “calibration” can be used to improve deficient density forecasts, and we show how the calibration method can be used to generate good density forecasts from econometric models, even when the conditional density is unknown. Finally, motivated by recent advances in financial risk management, we provide a detailed application to multivariate high-frequency exchange rate density forecasts.

## I. Introduction

THE FORECASTING literature has traditionally focused primarily on point forecasts. Recently, however, attention has shifted to interval forecasts (Chatfield, 1993) and density forecasts (Diebold, Gunther, et al., 1998). The reasons for the recent interest in interval and density forecasts are both methodological and substantive. On the methodological side, recent years have seen the development of powerful models of time-varying conditional variances and densities (Bollerslev et al., 1994; Ghysels et al., 1996; Hansen, 1994; Morvai et al., 1997). On the substantive side, density forecasts and summary statistics derived from density forecasts have emerged as a key part of the explosively growing field of financial risk management (Duffie & Pan, 1997; Jorion, 1997), as well as in more traditional areas such as macroeconomic inflation forecasting (Britton et al., 1998). Moreover, explicit use of predictive densities has long been a prominent feature of the Bayesian forecasting literature (Harrison & Stevens, 1976; West & Harrison, 1997), and recent advances in Markov-chain Monte Carlo (Gelman et al., 1995) have increased the pace of progress. The closely related “prequential” Bayesian literature (Dawid, 1984) also features density forecasts prominently.

Interest in forecasts of various sorts creates a derived demand for methods of evaluating forecasts. In parallel with the historical emphasis on point forecasts, most literature has focused on the evaluation of point forecasts (Diebold & Lopez, 1996), but recent interest in interval and density

forecasts has spurred development of methods for their evaluation (Christoffersen, 1998; Diebold, Gunther, et al., 1998; Wallis, 1999). Diebold, Gunther, et al., in particular, motivate and approach the problem of density forecast evaluation from a risk-management perspective, drawing upon an integral transform dating at least to Rosenblatt (1952), used creatively by Smith (1985) and Shephard (1994), and extended by Seillier-Moiseiwitsch (1993).

Here we extend the density forecast evaluation literature in three ways. First, in contrast to the extant literature, we focus on the multivariate case, which lets us evaluate the adequacy of density forecasts involving cross-variable interactions, such as time-varying conditional correlations, which are crucial in financial settings. Second, we provide conditions under which a technique of density forecast “calibration” can be used to improve deficient density forecasts. Finally, we provide a detailed application to the evaluation of density forecasts of multivariate high-frequency exchange rates, which is of direct substantive interest in addition to illustrating the implementation of our methods.

## II. Density Forecast Evaluation and Calibration

We begin with a brief summary of certain key univariate evaluation results from Diebold, Gunther, et al., (1998), in order to establish ideas and fix notation. We then describe methods of univariate calibration, after which we extend both the evaluation and calibration methods to the multivariate case.

### A. Univariate Evaluation and Calibration

Let  $\{y_t\}_{t=1}^m$  represent a series of realizations generated from the series of conditional densities  $\{f(y_t|\Omega_{t-1})\}_{t=1}^m$ . Diebold, Gunther, et al. (1998) show that, if a series of one-step-ahead density forecasts  $\{p_{t-1}(y_t)\}_{t=1}^m$  coincides with  $\{f(y_t|\Omega_{t-1})\}_{t=1}^m$ , then, assuming a nonzero Jacobian with continuous partial derivatives, the series of probability integral transforms of  $\{y_t\}_{t=1}^m$  with respect to  $\{p_{t-1}(y_t)\}_{t=1}^m$  is i.i.d.  $U(0, 1)$ . That is,

$$\{z_t\}_{t=1}^m = \left\{ \int_{-\infty}^{y_t} p_{t-1}(u) du \right\}_{t=1}^m \text{ i.i.d. } U(0, 1).$$

Thus, to assess whether a series of density forecasts coincides with the corresponding series of true conditional densities, we need only assess whether an observed series is i.i.d.  $U(0, 1)$ .

Received for publication September 22, 1998. Revision accepted for publication June 22, 1999.

\* NYU, University of Pennsylvania, and NBER; University of Michigan; and National University of Singapore, respectively.

Helpful comments were provided by two referees and a coeditor, as well as by Jose Campa, Enrique Sentana, Neil Shephard, Charlie Thomas, and Mike West, and seminar and conference participants at Stern, Wharton, Olsen & Associates, CEMFI, and Erasmus University Rotterdam. All remaining inadequacies are ours alone. For support, we thank the National Science Foundation, the University of Pennsylvania Research Foundation, and the National University of Singapore.

*The Review of Economics and Statistics*, November 1999, 81(4): 661–673

© 1999 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology

If a series of forecasts is found to be suboptimal, an important practical question is how such forecasts might be improved. In the point forecast case, for example, we can regress the  $y$ 's on the  $\hat{y}$ 's (the predicted values) and potentially use the estimated relationship to construct an improved point forecast. Such a regression is sometimes called a Mincer-Zarnowitz regression, after Mincer and Zarnowitz (1969). In this paper, we will use an analogous procedure, which we call *density forecast "calibration,"* for improving density forecasts that produce an i.i.d. but non-uniform  $z$  series.<sup>1</sup>

Suppose that we are in period  $m$  and possess a density forecast of  $y_{m+1}$ . From our earlier discussion,

$$f(y_{m+1}|\Omega_m) = p_m(y_{m+1})q_{m+1}(z_{m+1}).$$

But, if  $z$  is i.i.d., then we can drop the subscript on  $q$  and write

$$f(y_{m+1}|\Omega_m) = p_m(y_{m+1})q(z_{m+1}).$$

Thus, if we knew  $q(z_{m+1})$ , we would know the actual conditional distribution  $f(y_{m+1}|\Omega_m)$ .<sup>2</sup> Because  $q(z_{m+1})$  is unknown, we use an estimate  $\hat{q}(z_{m+1})$  formed using  $\{z_i\}_{i=1}^m$  to construct an estimate  $\hat{f}(y_{m+1}|\Omega_m)$ . In small samples, of course, there is no guarantee that the "improved" forecast will actually be superior to the original, because it is based on an estimate of  $q$  rather than the true  $q$ , and the estimate could be very poor.<sup>3</sup> The practical efficacy of our improvement methods is an empirical matter, which will be assessed shortly.

### B. Multivariate Evaluation

The principles that govern the univariate techniques discussed thus far extend readily to the multivariate case. Suppose that  $y_t$  is now an  $N$  operator 1 vector, and that we have a series of  $m$  multivariate forecasts and their corresponding multivariate realizations. Suppose further that we are able to factor each period's joint forecast into the product of the conditionals,

$$\begin{aligned} p_{t-1}(y_{1t}, y_{2t}, \dots, y_{Nt}) \\ = p_{t-1}(y_{Nt}|y_{N-1,t}, \dots, y_{1t}) \cdots p_{t-1}(y_{2t}|y_{1t})p_{t-1}(y_{1t}). \end{aligned}$$

<sup>1</sup> The basic idea of univariate density forecast calibration traces at least to Smith (1985) and Fackler and King (1990). Shortly, however, we will extend the calibration idea to the multivariate case. Moreover, we will emphasize that density forecast calibration is not universally applicable and develop a sufficient condition for its application in both the univariate and multivariate cases.

<sup>2</sup> Note that  $z$  is uniquely determined by  $y$ , so that, for any given value of  $y$ , we can always compute  $f(y)$  as  $p(y)q(z)$ .

<sup>3</sup> This caveat, of course, applies to any sort of empirical forecast adjustment, such as an empirical bias correction or Mincer-Zarnowitz correction. Fortunately, sample sizes are often large in financial applications.

Then, for each period, we can transform each element of the multivariate observation  $(y_{1t}, y_{2t}, \dots, y_{Nt})$  by its corresponding conditional distribution. This procedure produces a set of  $N$   $z$  series that will be i.i.d.  $U(0, 1)$  individually, and also when taken as a whole, if the multivariate density forecasts are correct. The proof of this assertion is obtained by simply arranging the multivariate series as a series of univariate observations comprising  $\{\dots, y_{1t}, y_{2t}, \dots, y_{Nt}, \dots\}$ .<sup>4</sup> The result for the univariate case can then be applied to show that the resulting  $Nm$  vector  $z$  is i.i.d.  $U(0, 1)$ , and, hence, the  $z$  series corresponding to any particular series of conditional forecasts is also i.i.d.  $U(0, 1)$ .

Note, moreover, that  $N!$   $z$  series can be produced, depending on how the joint density forecasts are factored, giving us a wealth of information with which to evaluate the forecasts. To take the bivariate case as an example, we can decompose the forecasts in two ways:

$$(i) p_{t-1}(y_{1t}, y_{2t}) = p_{t-1}(y_{1t})p_{t-1}(y_{2t}|y_{1t}), \quad \text{and}$$

$$(ii) p_{t-1}(y_{1t}, y_{2t}) = p_{t-1}(y_{2t})p_{t-1}(y_{1t}|y_{2t}),$$

$t = 1, \dots, m$ , and label the respective  $z$  series  $z_1, z_{2|1}, z_2$ , and  $z_{1|2}$ , where we obtain  $z_i$  by taking the probability integral transform of  $y_i$  with respect to  $p_{t-1}(y_i)$ , and we obtain  $z_{i|j}$  by transforming  $y_i$  with respect to  $p_{t-1}(y_i|y_j)$ . Good multivariate forecasts will produce  $z_1, z_{2|1}, z_2$ , and  $z_{1|2}$  that are each i.i.d.  $U(0, 1)$ , as well as combined series  $\{\dots, z_{1,t}, z_{2|1,t}, \dots\}$  and  $\{\dots, z_{2,t}, z_{1|2,t}, \dots\}$  that are also i.i.d.  $U(0, 1)$ .

In many cases of economic and financial interest, the dimension  $N$  of the set of variables being forecast is low. A classic example is density forecasting of returns across different aggregate asset classes, such as equity, foreign exchange, and fixed income. In such cases, it is straightforward to examine and learn from each of the  $N!$  sets of  $z$  series. Sometimes, however,  $N$  may be large, as (for example) when a density forecast is generated for each of the stocks in a broad-based market index, such as the S&P 500, in which case methods such as those of Clements and Smith (1998) may be used to aggregate the information in the individual  $z$  series.

### C. Multivariate Calibration

In parallel with the key formula underlying our discussion of univariate calibration, in the multivariate case we have that

$$\begin{aligned} f(y_{1,m+1}, y_{2,m+1}, \dots, y_{N,m+1}|\Omega_m) \\ = p_m(y_{1,m+1}, y_{2,m+1}, \dots, y_{N,m+1}) \\ \times q(z_{1,m+1}, z_{2,m+1}, \dots, z_{N,m+1}). \end{aligned}$$

<sup>4</sup> That is, we begin with  $m$  observations on an  $N$ -variate variable, and we convert them to a univariate series with  $Nm$  observations.

Moreover, factoring both of the right-side densities, we can write the formula in a way that precisely parallels our multivariate evaluation framework,

$$f(y_{1,m+1}, y_{2,m+1}, \dots, y_{N,m+1} | \Omega_m) = \prod_{i=1}^N (p_m(y_{i,m+1} | y_{i-1,m+1}, \dots, y_{1,m+1}) \times q(z_{i,m+1} | z_{i-1,m+1}, \dots, z_{1,m+1}))$$

As before, an estimate of  $q$  may be used to implement the calibration empirically, yielding

$$\hat{f}(y_{1,m+1}, y_{2,m+1}, \dots, y_{N,m+1} | \Omega_m) = p_m(y_{1,m+1}, y_{2,m+1}, \dots, y_{N,m+1}) \times \hat{q}(z_{1,m+1}, z_{2,m+1}, \dots, z_{N,m+1}),$$

or

$$\hat{f}(y_{1,m+1}, y_{2,m+1}, \dots, y_{N,m+1} | \Omega_m) = \prod_{i=1}^N (p_m(y_{i,m+1} | y_{i-1,m+1}, \dots, y_{1,m+1}) \times \hat{q}(z_{i,m+1} | z_{i-1,m+1}, \dots, z_{1,m+1})).$$

Note that the multivariate calibrating density is the same (in population), regardless of which of the possible  $N!$  factorizations is used.

### III. More on Density Forecast Calibration

Here we elaborate on various aspects of density forecast calibration. First, we give a sufficient condition for an i.i.d. integral transform series, which is required for application of the calibration method. Second, we discuss subtleties associated with estimation of the density of an integral transform, which, by construction, has compact support on the unit interval. Finally, we show how the calibration method can be used to generate good density forecasts even when the conditional density is unknown. We use the generic notation “ $z$ ” to denote an integral transform; it is a scalar integral transform in the univariate case and a vector of integral transforms in the multivariate case.

#### A. Conditions Producing an i.i.d. $z$ Series

In empirical work, the legitimacy of the assumption that  $z$  is i.i.d. can be assessed in a number of ways, ranging from examination of correlograms of various powers of  $z$  to formal tests such as those discussed in Brock et al. (1991). It is nevertheless desirable to characterize theoretically the conditions under which  $z$  will be i.i.d. in order to deepen our

understanding of whether and when we can reasonably hope for an i.i.d.  $z$  series. Here we establish a sufficient condition: If the one-step-ahead density  $f(y_t | \Omega_{t-1})$  belongs to a location-scale family, and, if the forecast  $p_{t-1}(y_t)$  adequately captures dynamics (in a sense to be made precise shortly) but is mistakenly assumed to be in another location-scale family, then  $z$  will be i.i.d.

More precisely, suppose that  $f(y_t | \Omega_{t-1})$  belongs to a location-scale family; that is, the random variable  $[y_t - \mu(\Omega_{t-1})] / \sigma(\Omega_{t-1})$  is i.i.d. with unknown density  $f(\cdot)$ .<sup>5</sup> It then follows that

$$f(y_t | \Omega_{t-1}) = \frac{1}{\sigma(\Omega_{t-1})} f\left(\frac{y_t - \mu(\Omega_{t-1})}{\sigma(\Omega_{t-1})}\right).$$

Now suppose that we misspecify the model by another location-scale family,  $p_{t-1}(y_t)$ , in which we specify  $\mu(\Omega_{t-1})$  and  $\sigma(\Omega_{t-1})$  correctly, in spite of the fact that we use an incorrect conditional density. Thus, we issue the density forecast

$$p_{t-1}(y_t) = \frac{1}{\sigma(\Omega_{t-1})} p\left(\frac{y_t - \mu(\Omega_{t-1})}{\sigma(\Omega_{t-1})}\right).$$

The probability integral transform of the realization with respect to the forecast is therefore

$$z_t = \int_{-\infty}^{y_t} p_{t-1}(y_t) dy_t = \int_{-\infty}^{y_t} \frac{1}{\sigma(\Omega_{t-1})} p\left(\frac{y_t - \mu(\Omega_{t-1})}{\sigma(\Omega_{t-1})}\right) dy_t.$$

Now make the change of variable

$$u_t = \frac{y_t - \mu(\Omega_{t-1})}{\sigma(\Omega_{t-1})}.$$

which yields

$$z_t = \int_{-\infty}^{u_t} \frac{1}{\sigma(\Omega_{t-1})} p(u_t) \sigma(\Omega_{t-1}) du_t = P(u_t).$$

But this means that  $z$  is i.i.d., as claimed, because  $z_t = P(u_t)$  and  $u_t$  is i.i.d.

#### B. Estimation of $q(z)$

Estimation of  $q(z)$  requires care, because  $z$  has support only on the unit interval. One could use a global smoother on

<sup>5</sup> Note that a random variable studentized with respect to the correct conditional mean and standard deviation is not necessarily i.i.d. It could, for example, have nonconstant conditional skewness or kurtosis. The location-scale family assumption rules out such possibilities.

the unit interval, such as a simple two-parameter beta distribution, but the great workhorse beta family is unfortunately not flexible enough to accommodate the multimodal shapes of  $q(z)$  that arise routinely, as documented in Diebold, Gunther, et al. (1998).

One could perhaps stay in the global smoothing framework by using some richer parametric family or a series estimator, as discussed for example by Härdle (1991). In keeping with much of the recent literature, however, we prefer to take a local smoothing approach, as for example with a kernel density estimator. But it has long been recognized that standard kernel estimation of densities with bounded support suffers from a “boundary problem,” that produces biased density estimates near the boundaries. Near the boundary, approximately half of the kernel mass falls outside the range of the data (assuming a symmetric kernel), producing kernel density estimates with expectations approximately equal to half the true underlying density.

Let us elaborate. For i.i.d. data  $z_1, \dots, z_T$  with sufficiently smooth density  $q(z)$  on  $[0, 1]$ , the kernel estimator  $\hat{q}(z)$  of  $q(z)$  is

$$\hat{q}(z) = \frac{1}{Tb} \sum_{i=1}^T K\left(\frac{z - z_i}{b}\right),$$

where  $K(\cdot)$  is a symmetric and sufficiently smooth kernel of choice and  $b > 0$  is bandwidth. It can be shown that, for  $z$  in the interior of  $[0, 1]$ , the kernel density estimates are unbiased:

$$E[\hat{q}(z)] = q(z) + O(b^2).$$

On the other hand, for  $z$  at the boundaries of  $[0, 1]$ , the kernel density estimates are biased, with expectations approximately equal to only half of the true underlying density.<sup>6</sup>

$$E[\hat{q}(0)] = \frac{q(0)}{2} + O(b)$$

$$E[\hat{q}(1)] = \frac{q(1)}{2} + O(b).$$

It can be shown that similar bias occurs at points near the boundaries.

There are a number of local smoothing approaches that address the boundary problem. One is Müller’s (1993) modified kernel density estimator, which is designed to have less bias near boundaries. Another, fiendishly simple and tailor-made for the problem at hand, is the standard histo-

gram—a special type of kernel density estimator in which the kernel never exceeds the boundaries of the distribution. One therefore expects that histogram density estimates will not suffer from the problems associated with kernels that exceed the boundaries. Some experimentation indeed revealed it to be superior in the present context, so we make extensive use of histograms as visual estimates of  $q(z)$  in our subsequent empirical work.

Finally, we can bypass the density estimation problem altogether by casting our evaluation and calibration methods in terms of cdf’s rather than densities, and using empirical cdf’s instead of estimated densities. Although empirical cdf’s are not as easy to interpret as estimated densities, they have a number of attractive features: They are guaranteed to be 0 for  $z = 0$  and 1 for  $z = 1$ , and there is no need for bandwidth selection. Hence, we also make extensive use of empirical cdf’s for doing the calibration transformations in our subsequent empirical work.

### C. *Generating Density Forecasts when the Conditional Density is Not Specified*

The calibration method can also be used to construct density forecasts for situations in which the forecasting model does not specify a conditional density. Consider, for example, the problem of one-step-ahead density forecasting in models with unknown one-step-ahead conditional density and time-varying volatility. We exploit the Bollerslev-Wooldridge (1992) result that GARCH volatility parameters are consistently estimated by maximum likelihood even when the conditional density is misspecified. We maximize the likelihood, which we assume to be Gaussian (incorrectly, in general). We then take the estimated volatility parameters and use them to make conditionally Gaussian density forecasts (again incorrectly in general), and we compute the probability integral transforms of the realizations with respect to those forecasts. If the forecasts using the incorrect conditional density nevertheless capture the dynamics, then, under the conditions given earlier, the series of integral transforms will be i.i.d. but not  $U(0, 1)$ , and we can calibrate them using the series of integral transforms.

Our approach extends readily to  $h$ -step-ahead density forecasting, which in financial contexts is equivalent to one-step-ahead forecasting of  $h$ -period returns.<sup>7</sup> As is well known, even under heroic assumptions such as conditionally Gaussian one-period returns,  $h$ -period returns will not be conditionally Gaussian and have no known closed form. Several earlier papers provide approximations that bear on the problem: Baillie and Bollerslev (1992) study prediction of GARCH processes with Gaussian one-step-ahead conditional density and provide Cornish-Fisher approximations to the  $h$ -step density, while Duan et al. (1997) provide Edgeworth approximations to the  $h$ -step density under similar

<sup>6</sup> See Härdle (1990, p. 130–131), for example, for related discussion in the context of kernel regression, as opposed to density estimation.

<sup>7</sup> See, for example, Duan (1995) and Duan et al. (1997).

conditions. Our methods effectively provide empirical implementation of the Cornish-Fisher and Edgeworth approximations to the  $h$ -step density.

Our approach is in the same spirit as Engle and González-Rivera (1991), who also exploit the Bollerslev-Wooldridge result, using standard nonparametric techniques to estimate the density of returns standardized by estimated GARCH volatilities.<sup>8</sup> Our procedure, however, may address more satisfactorily the well-known “bumpy-tail” problem in nonparametric density estimation, which is an important issue in risk-management contexts. The bumpy-tail problem may be overcome in principle by using larger bandwidths in the tails; our approach is based on the idea that, as long as the density forecasts used in the integral transforms are sufficiently close to the true densities, the integral transforms implicitly provide empirically reasonable variable bandwidths.<sup>9</sup>

The empirical application to which we now turn illustrates all of the ideas developed thus far: We begin by evaluating a series of standard multivariate density forecasts that are revealed to be poor by virtue of associated i.i.d. but non-uniform integral transform series, and we use calibration methods to transform the poor density forecasts into good ones.

#### IV. Evaluating and Calibrating Multivariate Density Forecasts of High-Frequency Returns on Foreign Exchange

Here we present a detailed application of our methods of multivariate density forecast evaluation and calibration to a bivariate system of asset returns. In particular, we study high-frequency DM/\$ and YEN/\$ exchange rate returns, and we generate density forecasts from a forecasting model in the spirit of JP Morgan’s RiskMetrics (JP Morgan, 1996), which is a popular benchmark in the risk-management industry.

We use our multivariate density forecast evaluation tools to assess the adequacy of the RiskMetrics approximation to the dynamics in high-frequency exchange-rate returns, as well as the adequacy of the conditional normality assumption, after which we attempt to improve the forecasts using calibration methods. We begin with a description of the data and a discussion of the statistical properties of the returns

<sup>8</sup> It would be of interest to compare the empirical performance of our approach with that of Engle and González-Rivera (1991) in the univariate one-step-ahead context, and with the Baillie and Bollerslev (1992) and Duan et al. (1997) approaches for  $h$ -step-ahead forecasting, but doing so is beyond the scope of this paper.

<sup>9</sup> Our methods effectively replace the bumpy-tail problem associated with estimation of densities with infinite support (standardized returns) with the boundary-bias problem associated with estimation of densities with compact support (integral transforms), which we handle by using histogram density estimators.

series. We then describe the forecasting model, after which we evaluate and calibrate the density forecasts that it produces.

##### A. Data

The data, kindly provided by Olsen and Associates, are indicative bid and ask quotes posted by banks, spanning the period from January 1, 1996, to December 31, 1996. The data are organized around a grid of half-hour intervals; Olsen provides the quotes nearest the half-hour time stamps. Foreign-exchange trading occurs around the clock during weekdays, but trading is very thin during weekends, so we remove them, as is customary.<sup>10</sup> We consider the weekend to be the period from Friday 21:30 GMT to Sunday 21:00 GMT, as this period appears to correspond most closely to the weekly blocks of zero returns. Thus, a trading week spans Sunday 21:30 GMT to Friday 21:00 GMT, and each of the five trading days therein spans 21:30 GMT on one day to 21:00 GMT the next day. This layout implies that we have one partial week of data, followed by 51 full weeks, followed by another partial week of data. Each full week of data has  $5 \times 48 = 240$  observations. The first partial week has 233 observations (just less than five days of data), and the sample ends with a partial week of 102 observations (about two days), for a total of 12,575 observations.

##### B. Computing Returns

We calculate exchange-rate returns in standard fashion, as for example in Andersen and Bollerslev (1997). We first compute bid and ask prices at each grid point by linearly interpolating the nearest previous and subsequent quotes, and we average the log bid and log ask prices to get a “log price.” We then calculate returns as the changes in log prices between consecutive grid points. It will sometimes be necessary to refer to the return at a particular time of a particular day. In such cases, we will use  $r_{\tau,i}$  to represent the return at time  $\tau$  of day  $i$ .

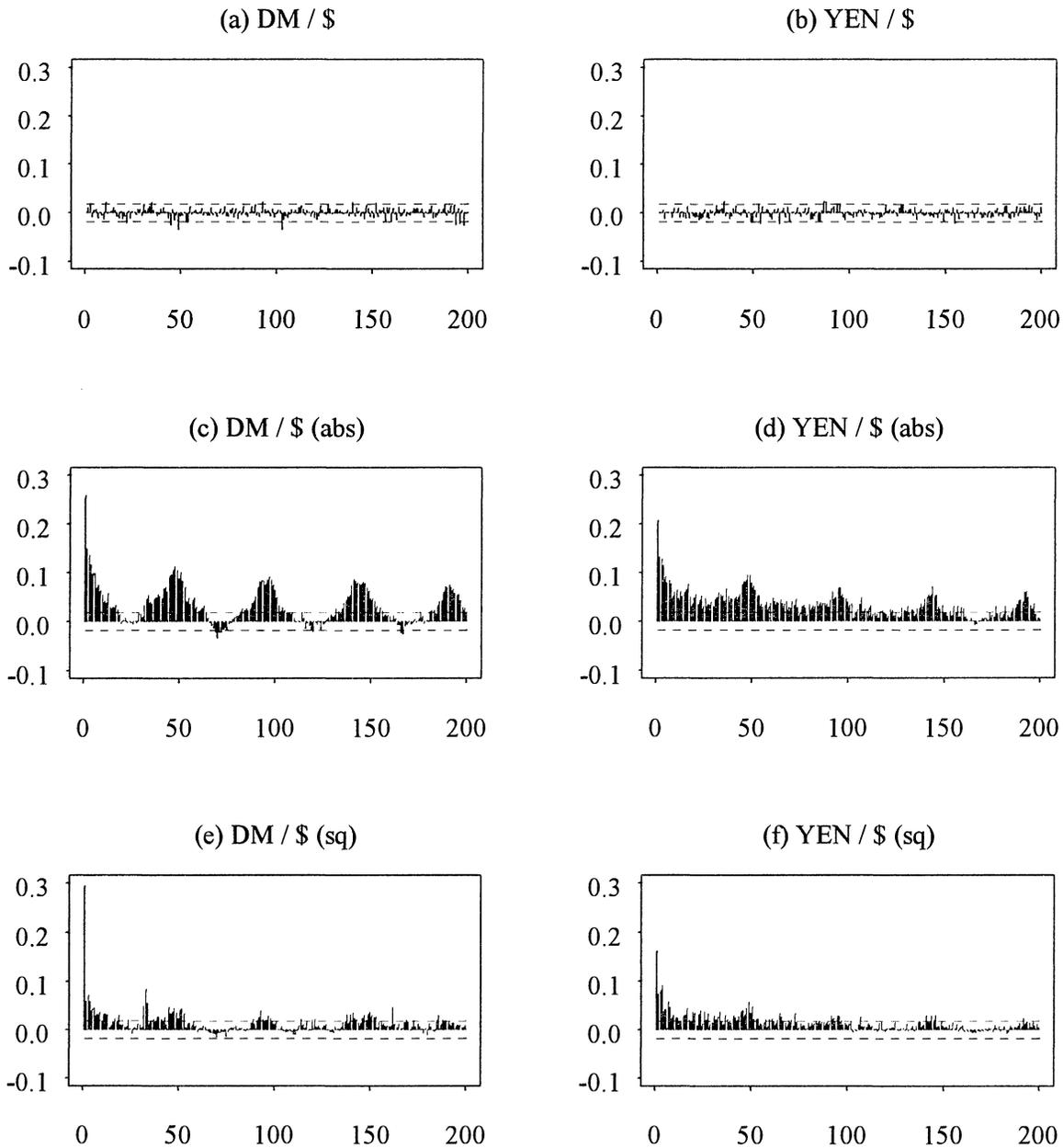
The returns exhibit the MA(1) conditional mean dynamics commonly found in asset returns. Because our focus in this paper is on volatility dynamics and their relation to density forecasting, we follow standard practice and remove the MA(1) dynamics by fitting MA(1) models and treating the residuals as our returns series. Hereafter,  $r_t$  will refer to returns series with the MA(1) component removed.

##### C. Properties of $r_t$ , $r_t^2$ and $|r_t|$

There are strong intra-day calendar effects in both Yen/\$ and DM/\$ returns; figure 1 displays the first 200 sample autocorrelations of  $r_t$ ,  $r_t^2$  and  $|r_t|$ . The calendar effects are present in the conditional variances, not the conditional means; hence, the autocorrelations of  $r_t$  show no patterns,

<sup>10</sup> See, for example, Andersen and Bollerslev (1997).

FIGURE 1—SAMPLE AUTOCORRELATION FUNCTIONS OF RETURNS, ABSOLUTE RETURNS, AND SQUARED RETURNS



while those of  $r_t^2$  and  $|r_t|$  show distinct patterns (especially those of  $|r_t|$ ). Calendar effects in volatility occur because trading is more active at certain times of day than at others. For instance, trading is much less active during the Japanese lunch hour, and much more active when the U.S. markets open for trading.

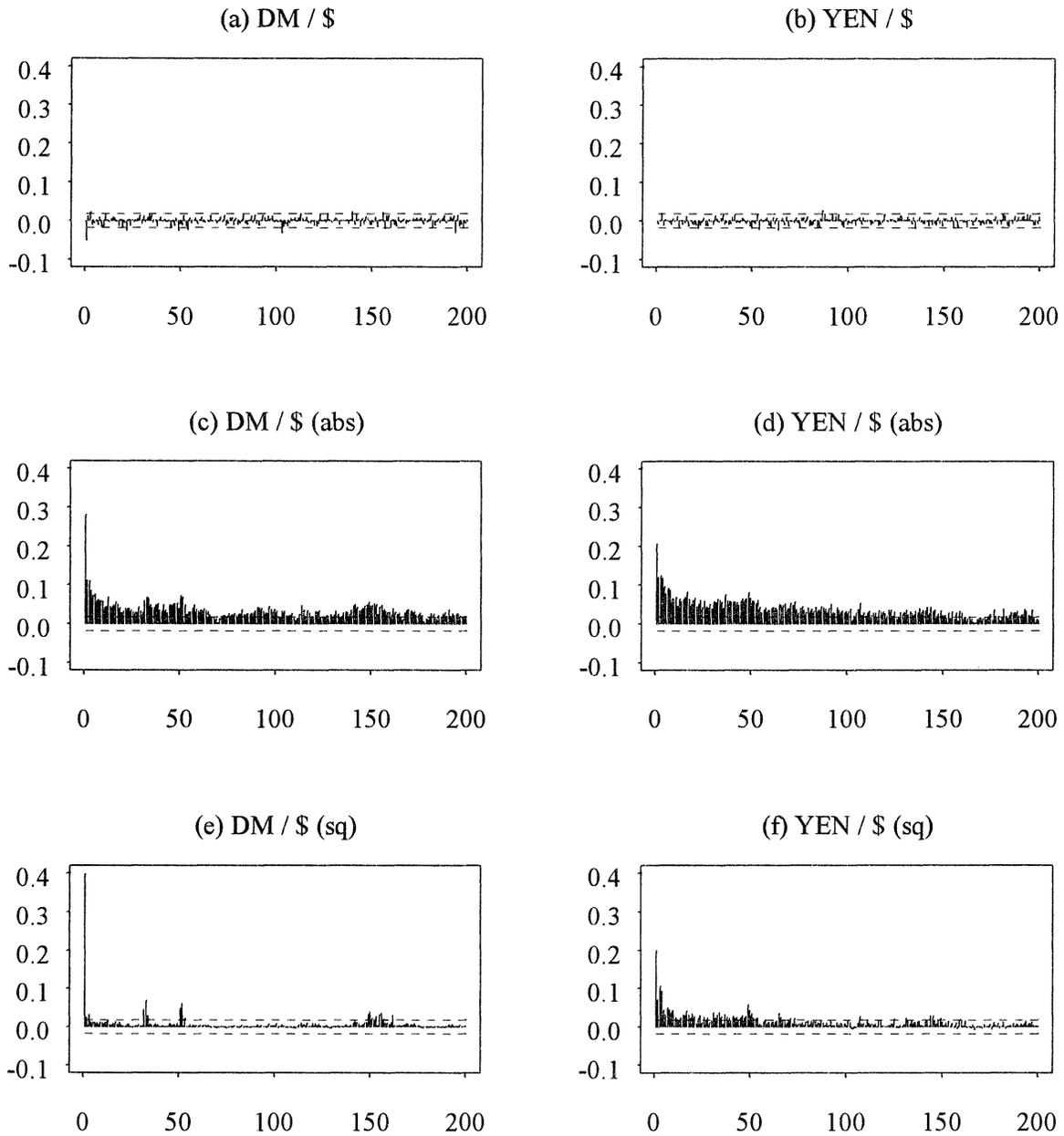
As with the conditional mean of returns, intra-day calendar effects in volatility are not of primary concern to us; hence we remove them. Taking  $r_t = s_{\tau,i} Z_t$  where  $Z_t$  represents the “nonseasonal” portion of the process, we remove volatility calendar effects by fitting time-of-day dummies to  $2 \log |r_t| = 2 \log s_{\tau,i} + 2 \log Z_t$ . We use the estimated time-of-day dummies, suitably normalized so that

$s_{\tau,i}$  summed over the entire sample equals 1, to standardize the returns.<sup>11</sup>

Figure 2 displays the autocorrelations of the levels, squared, and absolute standardized returns of both currencies. The standardization removes the intraday calendar effects in volatility and, as in Andersen and Bollerslev (1997), reveals a feature of the data that was not obvious

<sup>11</sup>  $F$  tests of no time-of-day effects confirm that strong calendar effects are present in absolute returns; the  $F(47, 12527)$  statistics are 17.34 and 8.927 for the DM/\$ and Yen/\$, respectively, with corresponding 0  $p$ -values.

FIGURE 2—SAMPLE AUTOCORRELATION FUNCTIONS OF RETURNS, ABSOLUTE RETURNS, AND SQUARED RETURNS (INTRA-DAY VOLATILITY CALENDAR EFFECTS REMOVED)



before standardization: The autocorrelations of absolute returns decay very slowly.

*D. Construction of Multivariate Density Forecasts*

We construct density forecasts of the two exchange rates using the exponential smoothing approach of RiskMetrics, which assumes that the returns are generated from a multivariate normal distribution, yielding the bivariate density forecasts

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11,t|t-1} & \sigma_{12,t|t-1} \\ \sigma_{21,t|t-1} & \sigma_{22,t|t-1} \end{pmatrix} \right),$$

where

$$\sigma_{11,t|t-1} = \lambda \sigma_{11,t-1|t-2} + (1 - \lambda) y_{1,t-1}^2$$

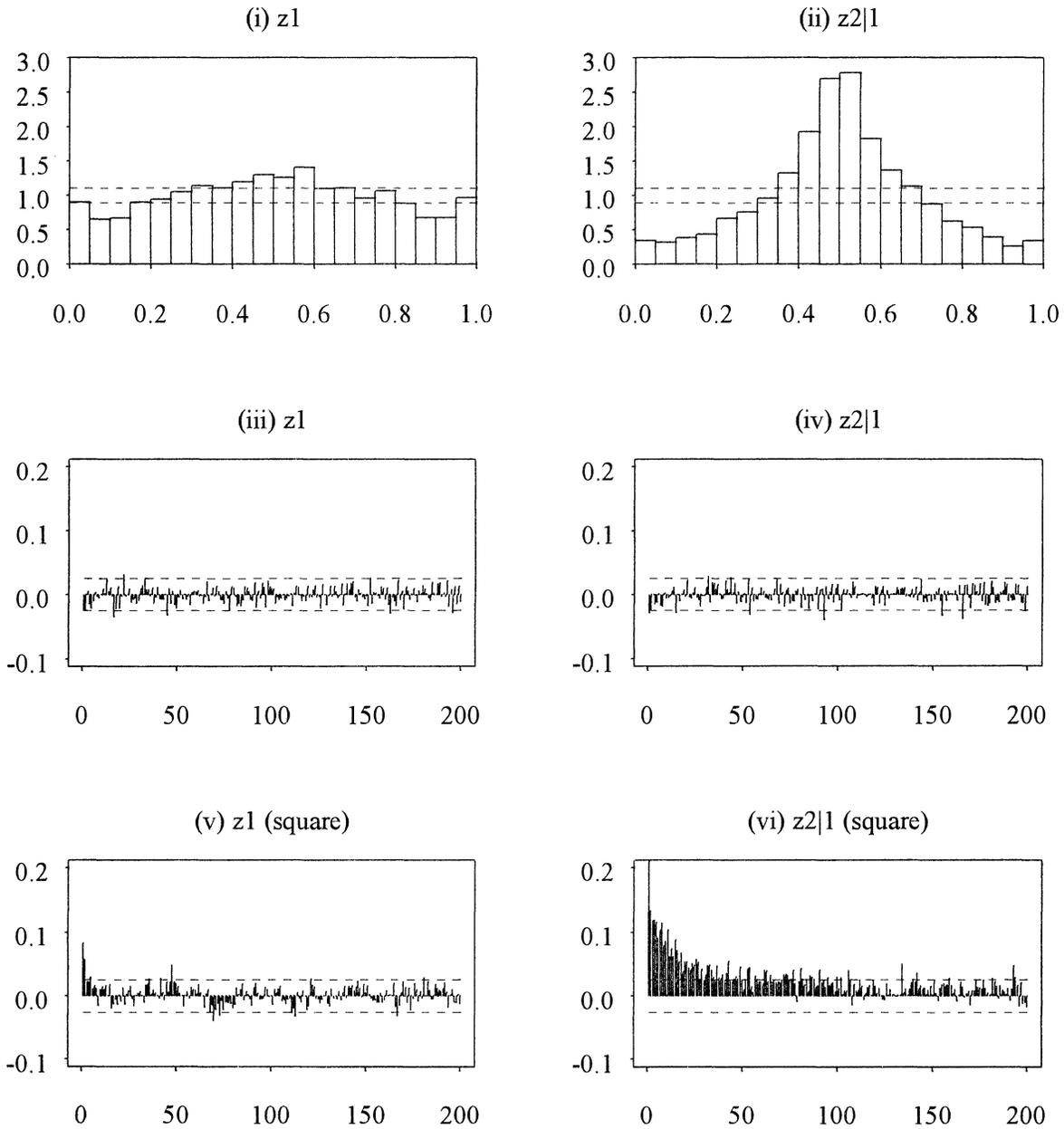
$$\sigma_{22,t|t-1} = \lambda \sigma_{22,t-1|t-2} + (1 - \lambda) y_{2,t-1}^2$$

$$\sigma_{12,t|t-1} = \lambda \sigma_{12,t-1|t-2} + (1 - \lambda) y_{1,t-1} y_{2,t-1},$$

and  $\lambda$  is the decay factor.<sup>12</sup>

<sup>12</sup> Note that, following RiskMetrics, we apply the same decay factor to each of the variances and covariances, which ensures that the variance-covariance matrix is positive definite.

FIGURE 3—HISTOGRAMS AND CORRELOGRAMS OF Z (DECAY FACTOR = 0.95, FORECAST SAMPLE)



*E. Density Forecast Evaluation*

As discussed earlier, we can evaluate the bivariate density forecasts generated by the RiskMetrics approach by decomposing the forecasts in two ways:

$$(i) p(y_{1t}, y_{2t}) = p(y_{1t})p(y_{2t}|y_{1t}) \quad \text{and}$$

$$(ii) p(y_{1t}, y_{2t}) = p(y_{2t})p(y_{1t}|y_{2t}),$$

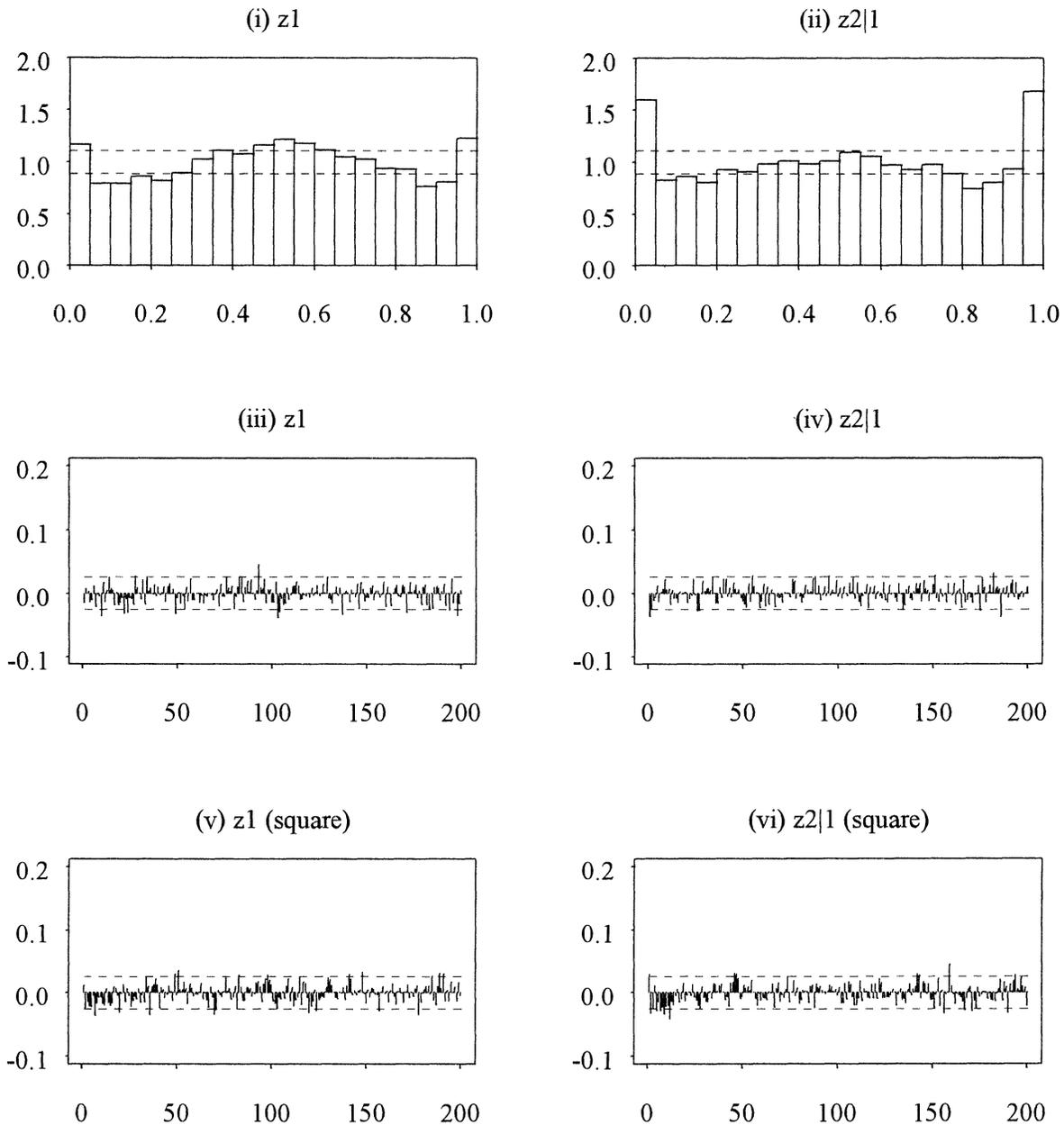
$t = 1, \dots, T$ . Label the respective  $z$  series  $z_1, z_{2|1}, z_2,$  and  $z_{1|2}$ , where we obtain  $z_i$  by taking the probability integral transform of  $y_i$  with respect to  $p(y_i)$ , and we obtain  $z_{i|j}$  by transforming  $y_i$  with respect to  $p(y_i|y_j)$ . Good multivariate

forecasts will produce  $z_1, z_{2|1}, z_2,$  and  $z_{1|2}$  that are each i.i.d.  $U(0, 1)$ . In the following, we will analyze only  $z_1$  and  $z_{2|1}$ , because of space limitations.<sup>13</sup>

To illustrate the multivariate density forecast evaluation procedures, we split the sample in two, the first running from January 1, 00:30 GMT, to June 30, 21:00 GMT (6,234 observations) and the second running from June 30, 21:30 GMT, to December 31, 23:30 GMT (6,431 observations); we call them the “estimation sample” and the “forecast sample,” respectively. We first evaluate forecasts generated using a decay factor of  $\lambda = 0.95$ , which is typical of

<sup>13</sup> See the working paper version of this article (Diebold, Hahn, et al. (1998) for analysis of  $z_2$  and  $z_{1|2}$ .)

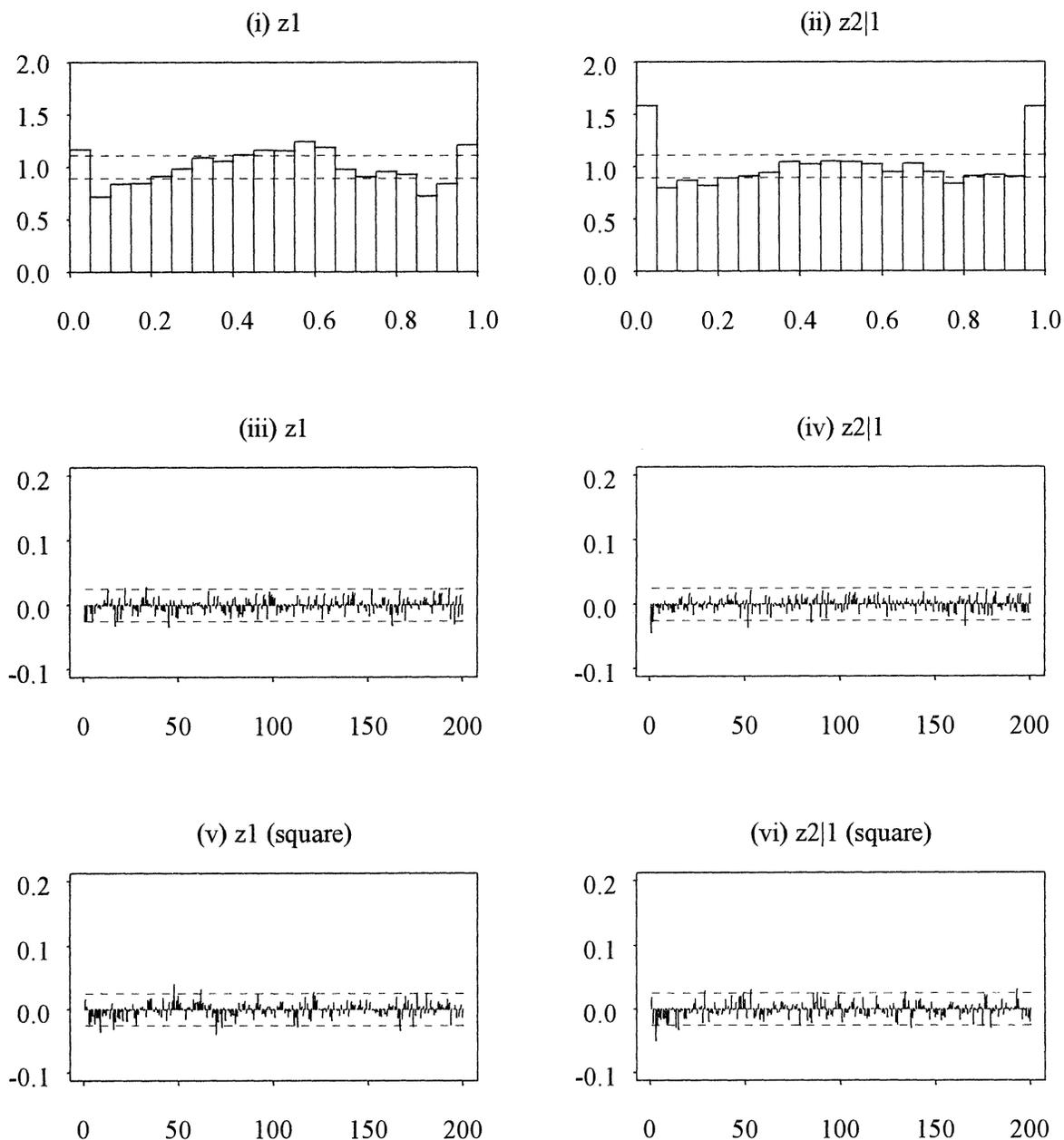
FIGURE 4—HISTOGRAMS AND CORRELOGRAMS OF  $z$  (DECAY FACTOR = 0.83, ESTIMATION SAMPLE)



RiskMetrics implementations. As a referee pointed out, it should be clear that setting  $\lambda = 0.95$  is inappropriate, because, in most GARCH(1, 1) applications, the estimated GARCH parameter is in the neighborhood of 0.82, and sum of the estimated ARCH and GARCH parameters is near 1. Our analysis provides further evidence to this effect. Figure 3 displays the histograms and correlograms of  $z_1$  and  $z_{2|1}$ ; the histograms show clearly that the normality assumption is inappropriate, and the correlograms show clearly that the decay factor  $\lambda = 0.95$  does not produce forecasts that capture the dynamics in return volatilities and correlations. Although the correlograms of levels of  $z_1$  and  $z_{2|1}$  look fine, the correlograms of squares of  $z_1$  and  $z_{2|1}$  indicate strong serial dependence (especially the correlogram of the square

of  $z_{2|1}$ , which highlights the value of the extra information obtained from multivariate, as opposed to univariate, evaluation). The positive serial correlation suggests that choosing  $\lambda = 0.95$  produces volatilities that adapt too slowly. Thus, it appears that a smaller decay factor would be appropriate.

We next turn to forecasts generated with a decay factor chosen based upon the estimation sample to produce integral transforms that visually appear closest to i.i.d., as assessed in our usual way, via correlograms of  $z$  and its powers. As expected, a smaller decay factor ( $\lambda = 0.83$ ) turns out to be optimal. Figure 4 displays the histograms and correlograms of  $z_1$  and  $z_{2|1}$  and their squares, based on the estimation sample. The correlograms indicate that the forecasts capture adequately the dynamics in return volatilities and correla-

FIGURE 5—HISTOGRAMS AND CORRELOGRAMS OF  $z$  (DECAY FACTOR = 0.83, FORECAST SAMPLE)

tions; that is,  $z_1$  and  $z_{2|1}$  and their squares appear serially uncorrelated. The histograms, in contrast, show clearly that the normality assumption is a poor one.

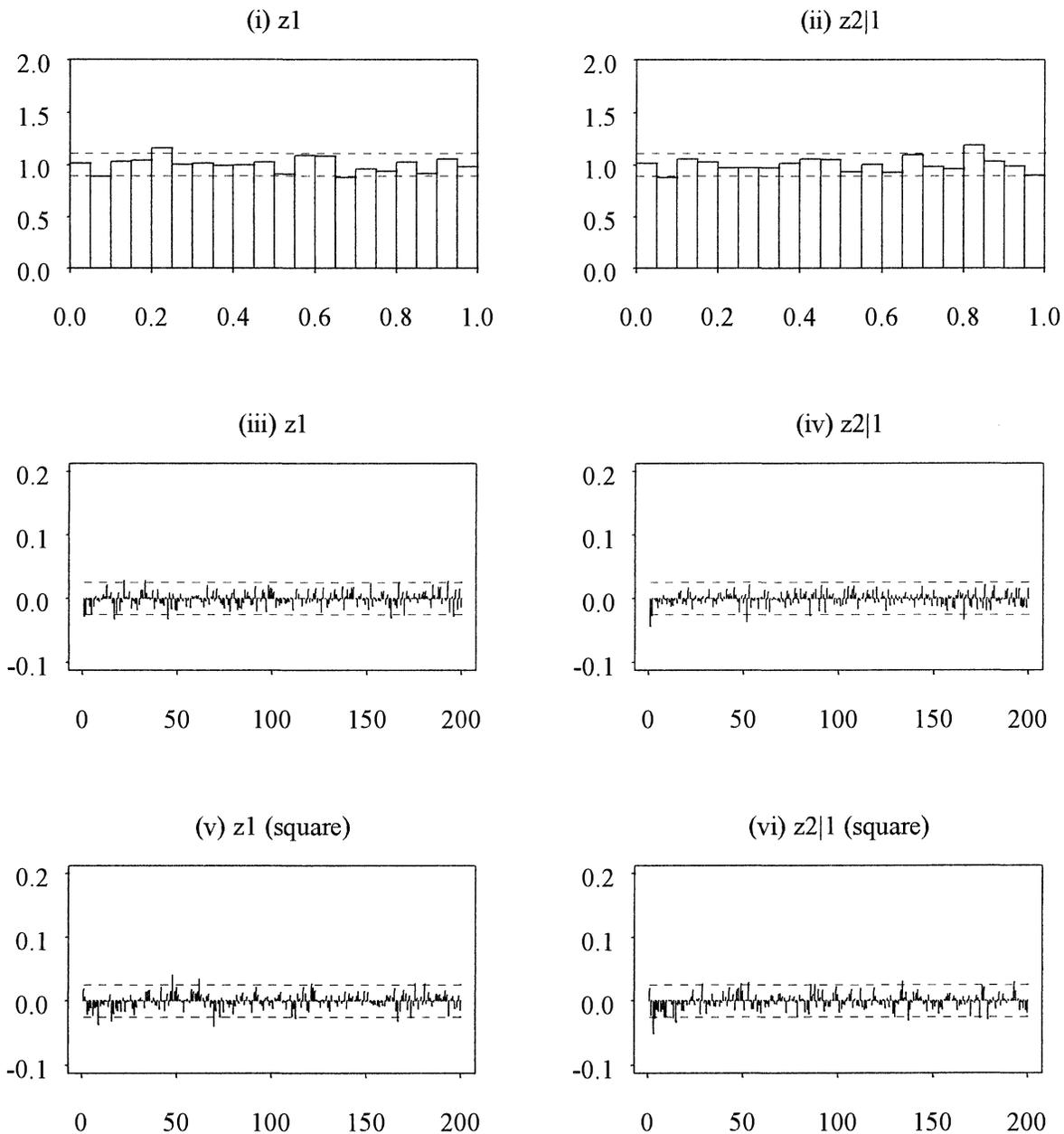
The smaller decay factor of 0.83 also produces much better out-of-sample joint forecasts of the two exchange-rate returns. Figure 5 displays the histograms and correlograms of  $z_1$  and  $z_{2|1}$  and their squares, based on the forecast sample, corresponding to density forecasts produced using  $\lambda = 0.83$ . The correlograms of squares all show clear improvement and indicate little, if any, serial correlation. The histograms, on the other hand, again show clearly that the normality assumption is inadequate.

#### F. Density Forecast Calibration

We now attempt to improve the  $\lambda = .83$  RiskMetrics forecasts, which seemed to capture dynamics adequately but which were plagued by an inappropriate normality assumption, by calibrating. Recall that we use the transformation

$$\begin{aligned} \hat{f}(y_{1,m+1}, y_{2,m+1}, \dots, y_{N,m+1} | \Omega_m) \\ &= p_m(y_{1,m+1}, y_{2,m+1}, \dots, y_{N,m+1}) \\ &\times \hat{q}(z_{1,m+1}, z_{2,m+1}, \dots, z_{N,m+1}), \end{aligned}$$

FIGURE 6—HISTOGRAMS AND CORRELOGRAMS OF  $z$  (DECAY FACTOR = 0.83, FORECAST SAMPLE, CALIBRATED)



or equivalently in terms of empirical cdf's,

$$\hat{F}(y_{1,m+1}, \dots, y_{N,m+1} | \Omega_m) = \hat{Q}(P_m(y_{1,m+1}, \dots, y_{N,m+1}))$$

We obtain  $\hat{Q}(\cdot)$  from the  $z$  series based on the estimation sample, and we use it to do the calibration in the forecast sample. Although we favor the presentation of histograms for density forecast evaluation, calibration is facilitated by using the empirical cdf form. We present the histograms and correlograms of the four  $z$  series corresponding to the calibrated forecasts in figure 6; clearly the histograms of the calibrated forecasts are substantially improved relative to

their noncalibrated counterparts, and the correlograms remain good (that is, they are not affected by the calibration.)

In closing this section, we hasten to add that our calibration methods will not improve certain aspects of a nonoptimal density forecast. In particular, we would not have been able to improve on the dynamics of the forecasts if that was indeed the defect of the original forecasts: A calibration of the forecasts generated using  $\lambda = 0.95$  would not have produced forecasts with improved dynamics. In general, if the diagnostics provide a clear indication of how the underlying structure of a forecasting model can be changed, we should certainly explore that possibility. There

are, however, cases where this may not be feasible. The forecasts that one is dealing with, for example, may have been generated by a third party, and the underlying methodology used may not be known by the forecast user who nonetheless wishes to calibrate the forecasts. Another example is the case in which density forecasts are generated from surveys, see Diebold, et al. (1999). In our example, the exponential smoothing used to capture varying variances and covariances is easy to implement if we are willing to assume multivariate normality. Our analysis of such forecasts, however, suggests that a fatter-tailed distribution, multivariate  $t$  for exdistribution would have been more appropriate. However, the implementation of the exponential smoothing in that context is not entirely straightforward, because the  $t$ -distribution is not described in terms of the variance-covariance matrix, but rather by scale parameters, which would require estimating the appropriate degree of freedom in addition to finding the appropriate decay factor.

#### V. Concluding Remarks and Directions for Future Research

We have proposed a framework for multivariate density forecast evaluation and calibration. The multivariate forecast evaluation procedure is a generalization of the univariate procedure proposed in Diebold, Gunther, et al. (1998) and shares its constructive nature and ease of implementation. We illustrated the power of the procedure to detect and remove defects in bivariate exchange-rate density forecasts generated by a popular method.

An interesting direction for future research involves using recursive techniques for real-time monitoring for breakdown of density forecast adequacy. Real-time monitoring using CUSUM techniques is a simple matter in the univariate case, because, under the adequacy hypothesis, the  $z$  series is i.i.d.  $U(0, 1)$ , which is free of nuisance parameters. Appropriate boundary-crossing probabilities for the CUSUM of the  $z$  series can be computed, as in Chu et al. (1996), using results on boundary-crossing probabilities of sample sums such as those of Robbins and Siegmund (1970). Multivariate CUSUM schemes are also possible, as with the multivariate profile charts of Fuchs and Benjamini (1994).

Finally, research along the lines of West et al. (1993) and Diebold and Mariano (1995) should prove fruitful in developing loss-function based forecast evaluation procedures which would take into account the economic significance of deviations from optimality when evaluating density forecasts.

#### REFERENCES

- Andersen, T. G., and T. Bollerslev, "Intraday Periodicity and Volatility Persistence in Financial Markets," *Journal of Empirical Finance* 4 (1997), 115–158.

- Baillie, R. T., and T. Bollerslev, "Prediction in Dynamic Models with Time Dependent Conditional Variances," *Journal of Econometrics* 52 (1992), 91–114.
- Bollerslev, T., R. F. Engle, and D. Nelson, "ARCH Models," in R. F. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, Vol. 4, (Amsterdam: North-Holland, 1994).
- Bollerslev, T., and J. M. Wooldridge, "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances," *Econometric Reviews* 11 (1992), 143–179.
- Britton, E., P. G. Fisher, and J. D. Whitley, "The Inflation Report Projections: Understanding the Fan Chart," *Bank of England Quarterly Bulletin* 38 (1998), 30–37.
- Brock, W. A., D. A. Hsieh, and B. LeBaron, *Nonlinear Dynamics, Chaos, and Instability*. (Cambridge, MA: MIT Press, 1991).
- Chatfield, C., "Calculating Interval Forecasts," *Journal of Business and Economics Statistics* 11 (1993), 121–135.
- Christoffersen, P. F., "Evaluating Interval Forecasts," *International Economic Review* 39 (1998), 841–862.
- Chu, C.-S. J., M. Stinchcombe, and H. White, "Monitoring Structural Change," *Econometrica* 64 (1996), 1045–1065.
- Clements, M. P., and J. Smith, "Evaluating the Forecast Densities of Linear and Non-Linear Models: Applications to Output Growth and Unemployment," manuscript, University of Warwick (1998).
- Dawid, A. P., "Statistical Theory: The Prequential Approach," *Journal of the Royal Statistical Society, Series A*, 147 (1989), 278–292.
- Diebold, F. X., T. Gunther, and A. S. Tay, "Evaluating Density Forecasts, with Applications to Financial Risk Management," *International Economic Review* 39 (1998), 863–883.
- Diebold, F. X., J. Hahn, and A. S. Tay, "Real-Time Multivariate Density Forecast Evaluation and Calibration: Monitoring the Risk of High-Frequency Returns on Foreign Exchange," NBER Working Paper 6845, <http://www.nber.org> (1998).
- Diebold, F. X., and J. Lopez, "Forecast Evaluation and Combination," in G. S. Maddala and C. R. Rao (eds.), *Statistical Methods in Finance (Handbook of Statistics, Vol 14)*. (Amsterdam: North-Holland 1996), 241–268.
- Diebold, F. X., and R. S. Mariano, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13 (1995), 253–263.
- Diebold, F. X., A. S. Tay, and K. F. Wallis, "Evaluating Density Forecasts of Inflation: The Survey of Professional Forecasters," in R. Engle and H. White (eds.), *Cointegration, Casuality and Forecasting: A Festschrift in Honor of C.W.J. Granger*. (Oxford: Oxford University Press, 1999), 76–90.
- Duan, J.-C. "The GARCH Options Pricing Model," *Mathematical Finance* 5 (1995), 13–32.
- Duan, J.-C., G. Gauthier, and J.-G. Simonato, "An Analytical Approximation for the GARCH Option Pricing Model," manuscript, Hong Kong University of Science and Technology (1997).
- Duffie, D., and J. Pan, "An Overview of Value at Risk," *Journal of Derivatives* 4 (1997), 7–49.
- Engle, R. F., and G. González-Rivera, "Semiparametric ARCH Models," *Journal of Business and Economic Statistics* 9 (1991), 345–359.
- Fackler, P. L., and R. P. King, "Calibration of Options-Based Probability Assessments in Agricultural Commodity Markets," *American Journal of Agricultural Economics* 72 (1990), 73–83.
- Fuchs, C., and Y. Benjamini, "Multivariate Profile Charts for Statistical Process Control," *Technometrics* 36 (1994), 182–195.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis* (London: Chapman and Hall, 1995).
- Ghysels, E., A. Harvey, and E. Renault, "Stochastic Volatility," in G.S. Maddala and C.R. Rao (eds.), *Statistical Methods in Finance (Handbook of Statistics, Vol. 14)* (Amsterdam: North-Holland, 1996).
- Hansen, B. E., "Autoregressive Conditional Density Estimation," *International Economic Review* 35 (1994), 705–730.
- Härdle, W., *Applied Nonparametric Regression* (Cambridge: Cambridge University Press, 1990).
- *Smoothing Techniques* (New York: Springer-Verlag, 1991).
- Harrison, P. J., and C. Stevens, "Bayesian Forecasting" (with discussion), *Journal of the Royal Statistical Society B* 38 (1976), 205–247.

- Jorion, P., *Value at Risk* (San Diego: Academic Press, 1997).
- JP Morgan, "RiskMetrics—Technical Document," (4th ed.) (New York, 1996).
- Mincer, J., and V. Zarnowitz, "The Evaluation of Economic Forecasts," in J. Mincer (ed.), *Economic Forecasts and Expectations* (New York: National Bureau of Economic Research, 1969).
- Morvai, G., S. J. Yakowitz, and P. Algoet, "Weakly Convergent Nonparametric Forecasting of Stationary Time Series," *IEEE Transactions on Information Theory* 43 (1997), 483–498.
- Müller, H.-G., "On the Boundary Kernel Method for Nonparametric Curve Estimation Near End Points," *Scandinavian Journal of Statistics* 20 (1993), 313–328.
- Robbins, H., and D. Siegmund, "Boundary Crossing Probabilities for the Wiener Process and Sample Sums," *Annals of Mathematical Statistics* 41 (1970), 1410–1429.
- Rosenblatt, M., "Remarks on a Multivariate Transformation," *Annals of Mathematical Statistics* 23 (1952), 470–472.
- Seillier-Moiseiwitsch, F., "Sequential Probability Forecasts and the Probability Integral Transform," *International Statistical Review* 61 (1993), 395–408.
- Shephard, N., "Partial Non-Gaussian State Space," *Biometrika* 81 (1994), 115–131.
- Silverman, B. W., *Density Estimation for Statistics and Data Analysis* (New York: Chapman and Hall, 1986).
- Smith, J. Q., "Diagnostic Checks of Non-Standard Time Series Models," *Journal of Forecasting* 4 (1985), 283–291.
- Wallis, K. F., "Asymmetric Density Forecasts of Inflation and the Bank of England's Fan Chart," *National Institute Economic Review* 167 (1999), 106–112.
- West, K. D., H. J. Edison, and D. Cho, "A Utility-Based Comparison of Some Models of Exchange Rate Volatility," *Journal of International Economics* 35 (1993), 23–45.
- West, M., and P. J. Harrison, *Bayesian Forecasting and Dynamic Models* (2nd ed.) (New York: Springer-Verlag, 1997).