How Accurate are Value-at-Risk Models at Commercial Banks?

Jeremy Berkowitz*
Graduate School of Management
University of California, Irvine

James O'Brien
Division of Research and Statistics
Federal Reserve Board

April 17, 2001

Abstract: In recent years, the trading accounts at large commercial banks have grown substantially and become progressively more diverse and complex. We provide descriptive statistics on the trading revenues from such activities and on the associated Value-at-Risk forecasts internally estimated by banks. For a sample of large bank holding companies, we evaluate the performance of banks' trading risk models by examining the statistical accuracy of the VaR forecasts. Although a substantial literature has examined the statistical and economic meaning of Value-at-Risk models, this article is the first to provide a detailed analysis of the performance of models actually in use.

Keywords: market risk, portfolio models, value-at-risk, volatility

Acknowledgements: We gratefully acknowledge the support and comments of Jim Embersit and Denise Dittrich of the Federal Reserve Board's Division of Supervision and Regulation, Philippe Jorion, Matt Pritsker, Mike Gibson, Hao Zhou, colleagues at the Federal Reserve Board and the New York Fed and an anonymous referee. The opinions expressed do not necessarily represent those of the Federal Reserve Board or its staff.
*Corresponding author, Irvine, CA, 92697-3125, jberkowitz@gsm.uci.edu

In recent years, the trading accounts at large commercial banks have grown rapidly and become progressively more complex. To a large extent, this reflects the sharp growth in the over-the-counter derivatives markets, in which commercial bank are the principal dealers. In order to manage market risks, major trading institutions have developed large scale risk measurement models. While approaches may differ, most models measure and aggregate market risks in current positions at a highly detailed level. The models employ a standard risk metric, Value-at-Risk (VaR), which is a lower tail percentile for the distribution of profit and loss (P&L). VaR models have been sanctioned for determining market risk capital requirements for large banks by U.S. and international banking authorities through the 1996 Market Risk Amendment to the Basle Accord. Spurred by these developments, VaR has become a standard measure of financial market risk that is increasingly used by other financial and even non-financial firms as well.

The general acceptance and use of VaR models has spawned a substantial literature including statistical descriptions of VaR and examinations of different modeling issues and approaches (for a survey and analysis see Jorion (2001)). Yet, because of their proprietary nature, there has been little empirical study of risk models actually in use, their VaR output, or indeed the P&L distributions of trading firms. For the most part, VaR analyses in the public domain have been limited to comparing modeling approaches and implementation procedures using illustrative portfolios (e.g., Beder (1995), Hendricks (1996), Marshall and Seigel (1997), Pritsker (1997)).[1]

In this paper, we provide the first direct evidence on the performance of bank VaR models. We analyze the distribution of historical trading P&L and the *daily* performance of VaR estimates of 6 large U.S. banks. All are large multinational institutions and meet the Basle "large trader" criterion—with trading activity equal to at least at 10 percent of total assets or $1 billion. The banks include the largest US bank derivative dealers and all are in the top 10 in terms of notional amounts outstanding as of year-end 1999. P&L and VaR data series are maintained by the banks to assess compliance with the Basle market risk capital requirements -- they serve as a gauge of the forecast accuracy of the models used for internal risk management. Regulations

[1] Jorion (2000) studies the usefulness of VaR disclosures in banks' annual and quarterly financial reports for forecasting risk.

stipulate that estimates are to be calculated for a 99 percent lower critical value of aggregate trading P&L with a one-day horizon. The forecasts provide a lower bound on aggregate trading P&L that should be breached 1 day in 100.

We evaluate the VaR forecasts in several ways. First, the null hypothesis of a 99 percent coverage rate is tested. Two important findings are that, unconditionally, the VaR estimates tend to be conservative relative to the 99th percentile of P&L. However, at times losses can substantially exceed the VaR and such events tend to be clustered. This suggests that the banks' models, besides a tendency toward conservatism, have difficulty forecasting changes in the volatility of P&L.

In part, the empirical performance of current models reflects difficulties in structural modeling when portfolios are large and complex. Large trading portfolios can have exposures to several thousand market risk factors, with individual positions numbering in the tens of thousands. It is virtually impossible for the models to turn out daily VaRs that measure the joint distribution of all material risks conditional on current information. The models therefore employ approximations to reduce computational burdens and overcome estimation hurdles. Additionally, the precision of the model may reflect risk management objectives and regulatory modeling constraints.

To further assess the performance of the banks' structural models, we compare their VaR forecasts with those from a standard GARCH model of the bank's P&L volatility. The GARCH model is reduced form and attempts no accounting for changes in portfolio composition. In principal, the banks' structural models should deliver superior forecasts. Our results, however, indicate that the bank VaR models do not provide forecasts superior to a simple model of P&L volatility. The GARCH model of P&L generally provides for lower VaRs and is better at predicting changes in volatility. Because of the latter, the GARCH model permits comparable risk coverage with less regulatory capital.

Reduced form forecasts based on time-varying volatility offer a simple alternative to structural models that may warrant further consideration. While the GARCH P&L model used here ignores current trading positions, such models can be adapted to account for changes in portfolio composition if such information is available. At a minimum, the results presented here illustrate that even naive reduced-form, time series models might serve as a useful ingredient in VaR forecasting.

The remainder of the paper is organized as follows. Section I defines the data and describes the distribution of daily P&L and bank VaRs. Section II presents the econometric methodology used to evaluate the performance of the models against the observed P&L. Section III considers some current practices and difficulties in constructing structural models of large complex trading portfolios, which might help to explain the performance of the banks' VaR estimates. A final section provides some general conclusions.

I. Daily Trading P&L and VaR

Daily profit and loss from trading activities and the associated VaR forecasts were collected from 6 large banking institutions subject to the Basle capital standards for trading risk. The trading revenue is based on position values recorded at the close of day and, unless reported otherwise, represents the bank holding company's consolidated trading activities. These activities include trading in interest rate, foreign exchange, and equity assets, liabilities, and derivatives contracts. Trading revenue includes gains and losses from daily marking to market of positions. Also included is fee income net of brokerage expenses related to the purchase and sale of trading instruments, excluding interest income and expenses.

The daily VaR estimates are maintained by the banks for the purpose of forecast evaluation or "back-testing" and are required by regulation to be calculated with the same risk model used for internal measurement of trading risk. The VaRs are for a one-day ahead horizon and a 99 percent confidence level for losses, i.e., the 1% lower tail of the P&L distribution. Because the internal models are based on positions at the close of business preceding the forecast day, they omit intra-day position changes. The bank models also omit net fee income although it is included in reported trading P&L.[2]

Summary statistics are reported in Table 1 for daily P&L and VaR data from January 1998 through March 2000. For these and other statistics reported below, each bank's daily P&L and VaR are divided by the bank's full-sample standard deviation of P&L to protect confidentiality. All banks reported positive average profits over the period. Sizeable differences in average P&L and standard deviation across banks (not shown) correspond to differences in the

[2] Additionally, the internal models are subject to regulatory standards that include a minimum sample period, quarterly parameter updating, and explicit treatment of risk factor correlations and nonlinearity in options (see Basle Committee on Banking Supervision (1996)).

size of trading activity, although column 2 of the Table also indicates significant disparity in mean P&L relative to its variation. In column 4, we report the 99^{th} percent losses of the P&L distributions, the statistic forecasted by VaR. These losses, coming once in 100 days, are quite large and are clustered at about 3 standard deviations below the mean. As a result, the excess kurtosis estimates (relative to the Normal distribution) displayed in column 5 are also large.

The last three columns of Table 1 show summary statistics for the banks' 99^{th} percentile VaRs. For 5 of 6 banks, the average VaR lies outside the lower 99^{th} percentile P&L, with VaRs for four banks ranging from 1.6 to over 3 times their respective 99^{th} percentile P&Ls. At the 99^{th} percentile, P&L would be expected to violate VaR 5 times in 500 trading days. However, only one bank experienced more than three violations. Nonetheless, the magnitudes of P&L losses *in excess of* the VaR are large. For two banks the mean violation is about as large as the mean VaR itself. For one bank, it is more than three times the mean VaR.

Histograms of P&L are presented in Figure 1 for the six banks. In all histograms, daily P&L are de-meaned and divided by their standard deviation. At least 5 of the 6 banks exhibit extreme outliers, with a preponderance of the outliers in the left tail. Both the skewness estimates reported in Table 1 and the histograms in Figure 1 suggest that the portfolio returns tend to be left-skewed.

In Figure 2, we display the time series of each bank's P&L and corresponding one-day ahead 99^{th} percentile VaR forecast (expressed in terms of the standard deviation of that bank's P&L). The plots tend to confirm the conservativeness of the VaR forecasts where violations of VaR are relatively few but large. The plots also show a differences in VaR performances among banks. For banks 1, 2, and 6, VaRs are in the general vicinity of the lower range of their P&Ls, but for banks 3, 4, and 5 this is not the case. The VaRs for these 3 banks also appear to exhibit trends. In particular, banks 4's VaR trends down while bank 5's VaR trends up.

The large losses in Figure 2 are mostly a consequence of high volatility in P&L between August and October 1998, in the wake of the Russian default and turmoil in Asian financial markets. Table 2 (column 1) shows that during this period, average returns are lower, standard deviations of the P&L for most banks are exceptionally large, and the 99^{th} percentiles are blown out. As shown in column 3, almost all violations for the bank VaRs occurred in this period. Figure 3 shows the timing and magnitudes of the violations, again expressed in standard deviations.

These findings suggest that P&L may be correlated across banks, a potential concern to bank supervisors because it raises the specter of systemic risk – the simultaneous realization of large losses at several banks. In the upper panel of Table 3, cross-correlations between banks' daily P&Ls are reported. While uniformly positive, the correlation coefficients for daily P&L are generally low, mostly below .2. Moreover, the daily correlations are low even for the subset of observations August-October 1998. Low correlations may reflect differences in portfolio compositions among banks. That is, even when market disruptions are widespread, shocks across different markets do not necessarily occur on the same calendar day. Additionally, trading firms have some discretion in the exact timing for reporting losses or gains in P&L, especially for inactively traded instruments. When P&L is aggregated over multi-day horizons, these idiosyncrasies may be less important. For example, over 5-day holding periods, the P&L cross-correlations approximately double.

The lower panel of Table 3 displays correlations for daily VaR across banks. The VaR correlations are as often negative as they are positive and no clear pattern of co-movement is evident. Results are qualitatively the same even when the sample is restricted to the August-October 1998 period and they are the same using 5-day average VaRs (not shown). These findings are consistent with different patterns in the bank VaRs displayed in Figure 2 and contrast with the small but positive daily cross-correlations in P&L.

II. Testing Model Performance

In this section we study the forecast accuracy of the bank VaR estimates and their sensitivity to daily portfolio volatility. Denote the portfolio's P&L by r_t, so that each day t the bank forecasts r_{t+1}. The VaR forecast is the quantity $\bar{r}$ such that $pr(r_{t+1} < \bar{r}) = \alpha$ over the next trading day. Here $\alpha = .01$, so that the model predicts a lower bound on losses not to be exceeded with 99% confidence.

A. Forecast Evaluation

The traditional approach to validating such interval forecasts is to compare the targeted violation rate, α, to the observed violation rate. The first column of Table 4 reports the actual rates at which violations occurred for the 6 banks. The average violation rate across banks is less than ½ of one percent. Column 2 reports likelihood ratio (LR) statistics for the null of a 1%

violation rate. The p-values, shown in square brackets, are the probabilities of the likelihood ratio values exceeding the observed value under the 1% null.

These p-values indicate that one of the coverage rates is significantly different from 1% at standard test levels. In addition, the LR test is undefined for one bank which had no violations. Both rejections arise because the frequency of violations is *less* than the desired one percent. Because of the small samples involved, unconditional coverage tests are known to have low power against alternative hypotheses (e.g., Kupiec (1995), Christoffersen (1998), Berkowitz (2000)).

More powerful tests are developed by Christoffersen (1998) who observes that not only should violations occur 1% of the time, but they should also be independent and identically distributed over time. Statistically, the variable defined as

$$I_t = 1 \quad \textit{if violation occurs}$$
$$= 0 \quad \textit{if no violation occurs}$$

should be an iid Bernoulli sequence with parameter α. Likelihood ratio tests of this null are easily constructed. These tests are referred to as conditional coverage and reported in column 3 of Table 4, with p-values shown in square brackets. At conventional significance levels, the VaR forecasts are rejected for two banks. A third bank shows a p-value of .14.

A useful feature of the likelihood framework is the following identity:

$$LR_{cc} = LR_{uc} + LR_{ind.}$$

That is, the conditional coverage test (LR_{cc}) can be decomposed into a test of the unconditional coverage (LR_{uc}), i.e., violation rate of α, plus a test that violations are independent (LR_{ind}). Column 4 reports the results of LR tests for first-order serial dependence.[3] The p-values suggest that for 2 banks, given a violation on one day there is a high probability of a violation the next day (higher than 1%). Similarly, the last column in Table 4 reports the sample autocorrelation, corr(I_t,I_{t-1}), a diagnostic suggested by Christoffersen and Diebold (2000). Monte Carlo p-values indicate two significant first-order autocorrelations. While these results are limited to first-order serial dependence, as noted earlier, almost all of the VaR violations occurred during a single three-month period.

[3] The tests are restricted to first-order dependence, rather than considering higher-order dependence as well, because of the small number of observations.

B. Comparisons with a Benchmark Model

The clustering of violations suggests that the volatility of P&L may be time varying to a degree not captured by the models. To further pursue the potential for predictable volatility, we formulate an alternative VaR model determined from an ARMA(1,1) plus GARCH(1,1) model of portfolio returns. That is, we estimate the following reduced form model of r_t

$$(1) \quad r_t = \mu + \rho r_{t-1} + u_t + \lambda u_{t-1}$$

where u_t is an iid innovation with mean zero and variance σ_t. The volatility process σ_t is described by

$$(2) \quad \sigma_t = \omega + \theta u_{t-1}^2 + \phi \sigma_{t-1}$$

where ω, θ and ϕ are parameters to be estimated. We apply the standard GARCH model where innovations are assumed to be conditionally Normal. Thus the 99% VaR forecast at time t is given by $\hat{r}_{t+1} - 2.33\hat{\sigma}_{t+1}$, where $\hat{r}_{t+1}$ is the predicted value of r_{t+1} from equation (1) and $\hat{\sigma}_{t+1}$ is the estimated volatility from equation (2).[4]

A time-series model of P&L is a natural benchmark for evaluating the banks' VaR models, whose hallmark has been the employment of detailed information on current positions and their exposures to the various market risk factors. The reduced form model cannot account for changes in current positions or relationships between positions and the market risk factors because it is fit to the aggregate returns data. Nonetheless, it is potentially a more tractable approach for capturing trend and time varying volatility in a banks' P&L without the structure that makes large-scale models so complex and unwieldy.

The ARMA and GARCH parameters are estimated each day with data available up to that point. To obtain stable estimates for the initial period, forecasts for days 1 through 165 are in-sample. Rolling out-of-sample forecasts begin after day 165, which is in the third week of

August 1998 except for one bank (where it is May 1998). Out-of-sample estimates are updated daily. Given parameter estimates, we forecast the next day's 99% VaR assuming Normality of the GARCH innovations. The resultant forecasts, both within and out-of sample, are shown in Figure 4 by the solid line, along with the P&L and the internal model forecasts. One-day ahead reduced-form forecasts appear to track the lower tails of P&L remarkably well. Compared to the structural model, the time series model does far better at adjusting to changes in volatility.

Summary statistics and backtests for the GARCH model VaRs are presented in Table 5. The second column shows that the GARCH model successfully removes first-order persistence in banks' P&L volatility. The average GARCH VaRs shown in column 3 are also lower than average bank VaRs, except for bank 6, and the number of violations shown in column 4 average out to about 1 percent. Thus, on average, the GARCH VaRs achieve the targeted violation rate and a 99th percentile VaR coverage. The mean violation rate for the GARCH VaRs also is lower than that of the banks' VaRs.

While more conservative VaRs might be expected to have higher mean violations, they also would be expected to produce smaller aggregate violations and maximum violations, other things equal. [5] However, this is not the case. Aggregate violations (column 4 times column 5) and maximum violations (see below) for the GARCH VaRs are comparable to the bank model VaRs, even though the bank VaRs are more conservative. These results indicate a potentially important advantage for the reduced-form GARCH model. The magnitudes of the banks' VaR forecasts are used to determine regulatory capital requirements, and likely influence banks' internal capital allocations as well. The GARCH VaRs are able to deliver lower required capital levels without producing larger violations. As described below, this reflects the GARCH VaRs greater responsiveness to changes in P&L volatility.

---

[4] Note that it is possible for the reduced form ARMA+GARCH VaR to be positive. This would occur if the conditional mean of the distribution is large so that $\hat{r}_{t+1} > 2.33\hat{\sigma}_{t+1}$.

[5] Under either a normal distribution or heavy-tailed distributions, such as the t distribution, the conditional expected value of lower tail returns is increasing in the lower critical tail value, while the unconditional, aggregate, and maximum expected values are inversely related to the lower critical tail value.

Formal back-tests of the GARCH models are presented in the bottom panel of Table 5.[6] The backtest results provide little basis for distinguishing between the GARCH and bank VaR modeling approaches. In terms of coverage, one GARCH VaR model is rejected at standard significance levels. Even though the GARCH VaRs on average have a 1 percent violation rate and the bank models less than a ½ percent violation rate, the rejection rate is the same for both sets of models. Results for independence of violations also are similar between the two modeling approaches. For the GARCH VaRs two banks are rejected for independence in violations.

Despite the comparability of the backtests, the GARCH models' greater responsiveness to changes in P&L volatility is illustrated for the August-October 1998 period when P&L volatility rose substantially. Table 6 compares model performances during this 3-month period. Even though the GARCH model VaRs are smaller over the full sample, the bank and GARCH VaRs are comparable during this period. For this 3-month period, the GARCH VaRs increased from 80 to 250 percent over their average VaRs during the 3 months prior to August 1998 for 4 of the 5 banks with violations. The bank VaRs in comparison were 20 percent lower to 30 percent higher than their respective averages over the preceding 3 months. As a result, the performances of the bank and GARCH VaRs are comparable in terms of average, aggregate and maximum violations.

While these results show that the GARCH VaR forecasts compare favorably with the banks' VaRs, the GARCH model is not unassailable. A plot of the GARCH violations in Figure 5, along with the results presented in Table 5, indicate that some clustering remains. Also, while the average violation rate for the GARCH VaRs is 1 percent, other statistics such as kurtosis indicate heavy tails in the GARCH P&L residuals. These results are due to the GARCH model's inability to adequately reflect the sharp increase in P&L volatility in the latter part of 1998.

Some further evidence of this is provided by the GARCH model parameter estimates for different sample periods. For banks 1 through 4, GARCH and ARCH parameters jump as the sample period is extended to include the period of heightened P&L volatility. The sum of the GARCH and ARCH parameters briefly reach one but subsequently decline below one as the

[6] Backtests were also carried out only for the out-of-sample forecasts, which account for about 75 percent of the full sample results. For the out-of-sample period, the average bank VaR was about the same as for the full sample, while the average of mean violations was somewhat lower. The average violation rate also was very close to .01. Average bank results for the backtests (coverage, dependence, etc) were very similar to those for the full sample period.

sample is further extended. Excluding 1998 from the sample period, the sum of the GARCH and ARCH parameters remain below one for all banks. These results are suggestive of an environment subject to regime shifts, which cannot be captured by the standard GARCH model (see Gray (1996)).[7]

III. Limitations of Structural Models

Our findings indicate that banks' 99th percentile VaR forecasts tend to be conservative and, for some banks, are highly inaccurate. In terms of forecast accuracy and the size of violations, the bank VaR forecasts could not out-perform forecasts based simply on an ARMA+GARCH model of the banks' daily P&L. These results are at least partly indicative of difficulties in building large-scale structural VaR models. However, we can identify some common modeling practices and regulatory constraints that lead to inaccurate forecasts.

The global trading portfolios of large trading banks contain tens of thousands of positions with several thousand market risk factors (interest rates, exchange rates, equity and commodity prices). Given the large number of positions and risk factors and the need to generate daily forecasts, it is impossible for the structural models to accurately measure the joint distribution of all material market risk factors, as well as the relationships between all risk factors and trading positions. To estimate the portfolio's risk structure, the banks' must make many approximations and parameters may be estimated only roughly. While the end result gives representation to a wide range of potential risks, the various compromises dampen any advantage in forecasting aggregate portfolio risk.

Bank portfolio models are limited in capturing time-varying volatility, which is particularly difficult within a structural model context. Formally accounting for, say, GARCH effects in the covariance matrix of a large set of market risks may be an insurmountable task (even for constant volatilities estimating stable covariance matrices can be difficult).[8] By reducing the risk factors to a univariate time series, the reduced-form model used here offers a

[7] We also estimated the VaR using an IGARCH model, where the sum of the ARCH and GARCH coefficients is constrained to equal one throughout the sample. Violations were again clustered in the August-October 1998 period.

[8] Further, regulatory modeling standards require that VaR estimates reflect market volatility over at least a one-year horizon, which precludes rapid adjustment to changes in current market volatility.

more tractable approach to estimating P&L mean and volatility dynamics. The reduced-form approach, however, does not account for changes in portfolio composition. A solution might be to estimate GARCH effects for historically simulated portfolio returns to current positions, rather than historically observed returns. The empirical properties and performance of this approach would require more study.[9]

In addition, there are several bank practices that tend to overstate VaRs. First, bank VaR models do not include forecasts of net fee income. Since net fee income is included in daily P&L, this will give the VaR forecasts a conservative bias *ceteris paribus*. The reduced-form VaR forecasts used here extrapolate P&L mean, as well as volatility, and thus implicitly include net fee income.[10]

For some banks, VaR models are not integrated across the entire trading portfolio. VaRs will be estimated for subgroups of positions, such as for foreign exchange positions and interest rate positions, and at a finer partition, and models may differ within groups. Subgroup VaRs are then aggregated under simplifying assumptions, such as perfect correlation or zero correlation between subgroups. Indeed regulatory requirements mandate that a simple sum of VaRs be used when aggregating separate VaRs across major risk categories. A simple sum does not allow for hedging or risk-diversification effects across the sub-portfolio returns. Of the bank models whose VaRs are among the most conservative, several make extensive use of the sub-portfolio addition procedure.

Finally, there are indisputable differences in banks' modeling experience and sophistication. Such differences have been documented during on-site model examinations by bank regulators and are apparent in our results. There is a clear division between the performance of the banks 1,2, and 6, where VaRs track near the lower range of P&L versus those of banks 3, 4 and 5 where this is not the case (see Figure 2). The former set of banks has more

---

[9] Barone-Adesi, Giannopoulos, and Vosper (1999) apply GARCH to historically simulated returns at the individual risk factor level under covariance parameter restrictions. Lopez and Walter (2000) report favorable results applying GARCH to portfolio returns as against applying GARCH at the risk factor level. Engle and Manganelli (1999) suggest reduced-form forecasting alternatives to GARCH. In particular, they advocate directly modeling the dynamics of the VaR rather than mean and variance dynamics. A reduced-form approach to VaR forecasting was originally suggested by Zangari (1997).

[10] Another omission is that that VaR forecasts for day *t*, based on end-of day *t-1* positions do not include intraday risk whose effects will be reflected in end-of-day P&L.

extensive experience and has the better VaR models. To date, the experience of examiners has been that building a satisfactory market risk model on a large scale is a long and arduous process.

IV. Conclusions

This study has presented the first direct evidence on the performance of Value-at-Risk models for large trading firms. The results show that the VaR forecasts for six large commercial banks have exceeded nominal coverage levels over the past two years and, for some banks, VaRs were substantially removed from the lower range of trading P&L. While such conservative estimates imply higher levels of capital coverage for trading risk, the reported VaRs are less useful as a measure of actual portfolio risk.

Despite the detailed information employed in the bank models, their VaR forecasts did not out-perform forecasts based simply on an ARMA plus GARCH model of the banks' P&L. Compared to these reduced-form forecasts, the bank VaRs did not adequately reflect changes in P&L volatility. These results may reflect substantial computational difficulties in constructing large-scale structural models of trading risks for large, complex portfolios.

Reduced-form or "time-series" models of portfolio P&L cannot account for positions' sensitivities to current risk factor shocks or changes in current positions. However, their parsimony and flexibility are convenient and accurate for modeling the mean and variance dynamics of P&L. While the forecasts used here did not account for current positions, the reduced-form approach is amenable to this if used in conjunction with historical simulation methods. In a larger sense, the P&L time series models are complementary to the large-scale models. The structural models are forward looking and they permit firms to examine the effects of individual positions on portfolio risk. Time series models may have advantages in forecasting and as a tool for identifying the shortcomings of the structural model.

To a certain extent, our study is limited by the fact that banks only forecast a single percentile of the portfolio distribution. Significantly more could be learned about the empirical performance of internal valuation models if density forecasts were recorded. Density forecast evaluation techniques described in Diebold, Gunther and Tay (1998) and Berkowitz (2000) provide researchers with substantially more information to assess the dimensions in which models need improvement and those in which models do well.

References

Barone-Adesi, G. K. Giannopoulos and L. Vosper (1999), "VaR without Correlations for Portfolios of Derivative Securities,"*Journal of Futures Markets*, 19, 583-602.

Basak, S. and Shapiro, A. (2000), "Value-at-Risk Based Risk Management: Optimal Policies and Asset Prices," forthcoming, *Review of Financial Studies*.

Basle Committee on Banking Supervision, *Amendment to the Capital Accord to Incorporate Market Risks*, January 1996.

Beder, T. S. (1995), "VAR: Seductive but Dangerous," *Financial Analysts Journal*, September, 12-24.

Berkowitz, J. (2000), "Testing the Accuracy of Density Forecasts," forthcoming, *Journal of Business and Economic Statistics*

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-327.

Christoffersen, P. (1998), "Evaluating Interval Forecasts," *International Economic Review*, 39, 841-862.

Christoffersen, P. and F. X. Diebold (2000), "How Relevant is Volatility Forecasting for Financial Risk Management," *Review of Economics and Statistics*, 82, 12-22.

Diebold, F. X., Gunther, T. A. and Tay, A. S. (1998), "Evaluating Density Forecasts," *International Economic Review*, 39, 863-883.

Dimson, E. and P. Marsh (1995), "Capital Requirements for Securities Firms," *Journal of Finance*, 50, 821-851.

Duffie, D. and J. Pan. (1997), "An Overview of Value at Risk", *Journal of Derivatives*, Spring 1997, 4, 7-49.

Engle, R. F. and S. Manganelli (1999), "CAViaR: Conditional Autoregressive Value-at-Risk by Regression Quantiles," manuscript, University of California, San Diego.

Figlewski, S. (1997), "Forecasting Volatility," *Financial Markets, Institutions, and Instruments 6*, Monograph series of the Stern School of Business, NYU.

Gray, S. (1996), "Modeling the conditional distribution of interest rates as a regime-switching process," *Journal of Financial Economics*, 42, 27-62.

Hendricks, D. (1996), "Evaluation of Value-at-Risk Models Using Historical Data," *Economic Policy Review*, Federal Reserve Bank of New York, April, 39-69.

Jorion, P. (2000), "How Informative are Value-at-Risk Disclosures," manuscript, University of California, Irvine.

Jorion, P. (2001). *Value-at-Risk: the New Benchmark for Controlling Market Risk.* Chicago: McGraw-Hill.

JP Morgan (1996). *RiskMetrics*, Technical Document. 4th Edition. New York.

Lopez, J. and Walter, C. (2000). "Evaluating Covariance Matrix Forecasts in a Value-at-Risk Framework," manuscript, Federal Reserve Bank of San Francisco.

Kupiec, P. (1995). "Techniques for Verifying the Accuracy of Risk Measurement Models," *Journal of Derivatives*, 3, 73-84.

Marshall, C. and M. Siegel (1997), "Value-at-Risk: Implementing a Risk Measurement Standard," *Journal of Derivatives*, 1, 91-111.

Pritsker, M. (1997), "Evaluating Value at Risk Methodologies: Accuracy versus Computational Time," *Journal of Financial Services Research*, 12, 201-242.

Zangari, P., 1997. "Streamlining the Market Risk Measurement Process," *RiskMetrics Monitor*, 1, 29-35.

Table 1. Bank P&L and VaR Summary Statistics

	Obs	Daily P&L					Daily VaR		
		mean	standard deviation	99^{th} percentile	excess kurtosis	skew	mean VaR	number violations	mean violation
Bank 1	569	.964	1.00	-1.78	11.63	-.993	-1.87	3	-2.12
Bank 2	581	.737	1.00	-2.26	4.53	.094	-1.74	6	-.741
Bank 3	585	.375	1.00	-2.73	23.87	-3.13	-4.41	3	-3.18
Bank 4	573	.595	1.00	-1.59	2.31	.860	-5.22	0	
Bank 5	[a]746	.253	1.00	-2.78	3.41	-.617	-5.62	1	-.775
Bank 6	586	.608	1.00	-.967	142.1	-8.25	-1.72	3	- 5.84

Notes: Daily profit and loss from trading activities reported by large commercial banks for January 1998 through March 2000. Each bank's data are divided by its average return to protect the confidentiality of individual institutions. [a]Data begins in May 1997. Mean violation refers to the loss in excess of the VaR.

Table 2. Bank P&L and VaR Summary Statistics
August to October 1998

		Daily P&L					Daily VaR		
	Obs	mean	standard deviation	minimum	excess kurtosis	skew	mean VaR	number violations	mean violation
Bank 1	63	.175	1.76	-7.01	4.58	-1.32	-2.32	3	-2.12
Bank 2	64	.076	1.89	-4.26	1.46	.787	-2.28	5	-.862
Bank 3	65	-.907	1.84	-8.68	7.65	-2.53	-4.63	3	-3.18
Bank 4	63	.0453	.773	-1.89	0.99	-.434	-4.66	0	
Bank 5	65	.0637	1.60	-5.51	1.99	-.837	-5.09	1	-.775
Bank 6	65	.171	2.28	-16.40	41.1	-5.92	-1.42	2	-7.99

Notes: Daily profit and loss data as reported by large commercial banks in the wake of the Russian default crisis, August 1998 to October 1998. For further details on the data, see Table 1.

Table 3. Correlations of Profit and Loss and VaR Across Individual Banks

A. P&L Correlation Coefficients

	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5	Bank 6
Bank 1	1.00					
Bank 2	.434	1.00				
	(10.1)					
Bank 3	.206	.102	1.00			
	(4.81)	(2.39)				
Bank 4	.164	.085	.358	1.00		
	(3.84)	(1.98)	(8.36)			
Bank 5	.053	.171	.117	.122	1.00	
	(1.29)	(3.99)	(2.73)	(2.84)		
Bank 6	.154	.165	.197	.108	.108	1.00
	(3.60)	(3.85)	(4.59)	(2.52)	(2.53)	

B. VaR Correlation Coefficients

	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5	Bank 6
Bank 1	1.00					
Bank 2	-.033	1.00				
	(-.777)					
Bank 3	.122	.207	1.00			
	(-2.84)	(3.02)				
Bank 4	.064	-.779	-.202	1.00		
	(1.48)	(-5.72)	(-4.72)			
Bank 5	-.184	.749	.072	-.742	1.00	
	(-4.30)	(-8.59)	(1.67)	(-17.4)		
Bank 6	-.404	-.229	-.220	.119	.131	1.00
	(-9.45)	(8.64)	(-5.15)	(2.78)	(3.06)	

Notes: Correlation coefficients for daily bank profit and loss and VaR calculated with a matched sample of 482 daily observations; t-statistics are shown in parentheses.

Table 4. Backtests of Bank Model VaRs

	Violation Rate	Coverage	Conditional Coverage	Independence	Serial Correlation
Bank 1	0.005	1.54 [.214]	1.57 [.455]	.0321 [.858]	-.00533 [.885]
Bank 2	0.010	.00693 [.934]	4.01 [.135]	4.00 [.046]	.158 [.016]
Bank 3	0.005	1.70 [.193]	8.81 [.012]	7.11 [.008]	.330 [.001]
Bank 4	0.000	NA	NA	NA	NA
Bank 5	0.001	8.922 [.003]	8.925 [.012]	.00271 [.959]	-.00137 [.995]
Bank 6	0.005	1.707 [.191]	1.738 [.419]	.0312 [.860]	-.00511 [.885]

Notes: Alternative backtests of bank VaRs. Data are daily and span from January 1998 to March 2000 (from May 1997 for bank 5). NA indicates the bank had no violations in the sample period. P-values are displayed in square brackets.

Table 5. Backtests of Time-Series Model
ARMA(1,1) + GARCH(1,1)

Summary Statistics

	Obs	Box-Ljung Stat	Mean VaR	Number Violations	Mean Violation
Bank 1	569	.284	-1.21	6	-.829
Bank 2	581	.106	-1.42	6	-.362
Bank 3	585	1.11	-1.41	13	-1.12
Bank 4	573	.356	-1.35	4	-.315
Bank 5	746	2.89	-2.10	12	-.772
Bank 6	586	.003	-2.40	2	-7.21

Backtests

	Violation Rate	Coverage	Conditional Coverage	Independence	Serial Correlation
Bank 1	0.011	0.018	3.976	3.958	0.158
		[.894]	[.137]	[.047]	[.016]
Bank 2	0.010	0.069	0.132	0.125	-0.010
		[.934]	[.936]	[.723]	[.436]
Bank 3	0.022	6.575	11.37	4.799	0.134
		[.010]	[.003]	[.029]	[.025]
Bank 4	0.007	0.584	0.640	0.056	-0.007
		[.445]	[.726]	[.812]	[.756]
Bank 5	0.016	2.369	4.156	1.787	0.068
		[.124]	[.125]	[.181]	[.050]
Bank 6	0.003	3.432	3.446	0.014	-0.003
		[.064]	[.179]	[.907]	[.964]

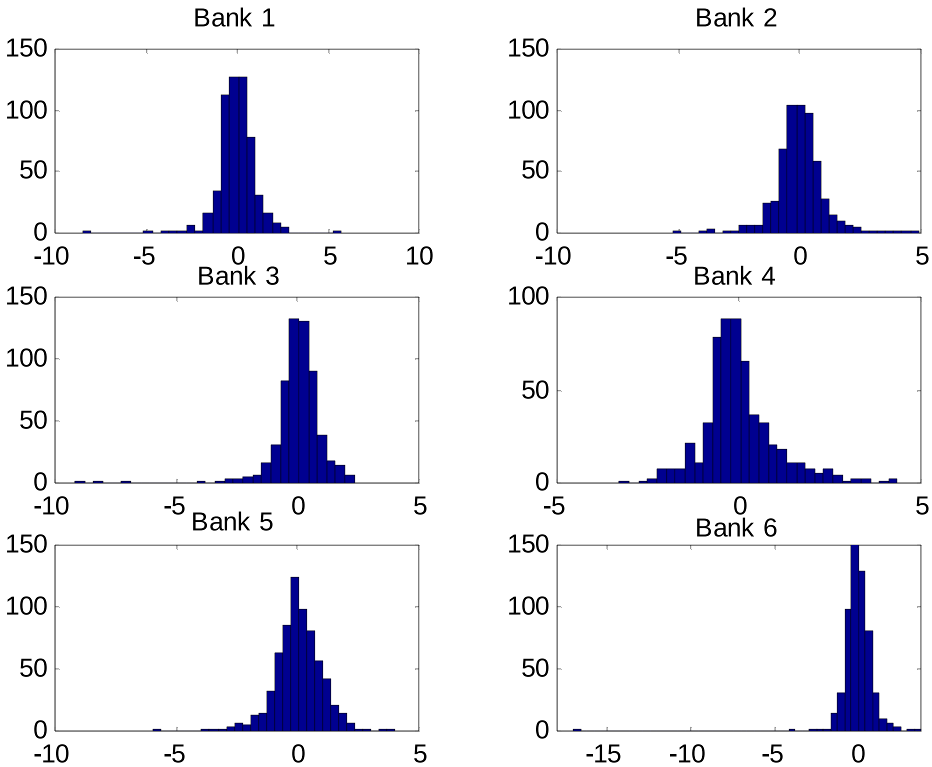
Notes: Alternative backtests of time series model VaRs. P-values are displayed in square brackets. Box-Ljung statistics are for first-order serial correlation in the squares of the standardized GARCH residuals. The 5% critical value is 3.84, the 10% value is 2.71.

Table 6. Bank and GARCH Model Comparisons
August to October 1998

		Bank VaRs				GARCH VaRs			
	obs	mean VaR	number viol	mean viol	max viol	mean VaR	number viol	mean viol	max viol
Bank 1	63	-2.32	3	-2.12	-4.70	-2.90	5	-.981	-4.33
Bank 2	64	-2.28	5	-.862	-2.46	-3.41	3	-.410	-.748
Bank 3	65	-4.62	3	-3.18	-4.13	-3.12	7	-1.45	-4.08
Bank 4	63	-4.66	0	NA	NA	-1.71	0	NA	NA
Bank 5	65	-5.08	1	-.775	-3.29	-2.97	4	-1.35	-2.87
Bank 6	65	-1.42	2	-7.99	-15.40	-2.93	1	-13.4	-13.4

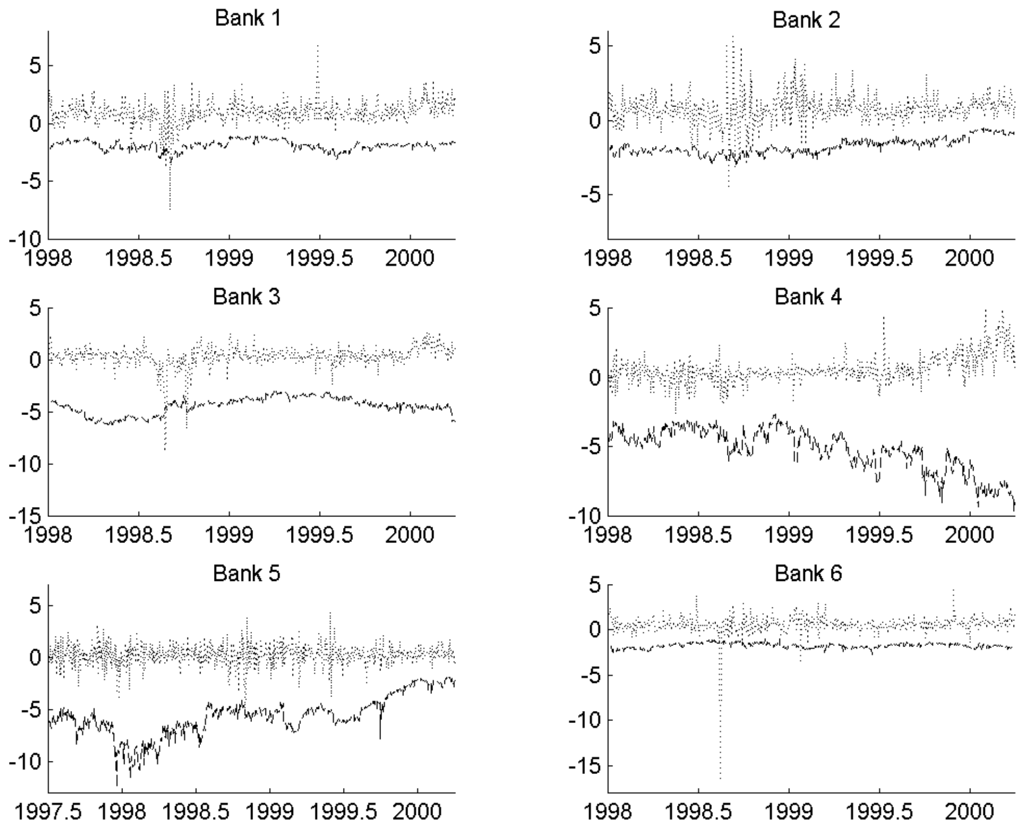
Notes: Table compares value-at-risk forecasts as reported by banks during the period of August 1998 to October 1998 to forecasts from a reduced form model. The GARCH VaR forecast is based on an ARMA(1,1) with GARCH(1,1) with conditionally Normal innovations. For further details see Table 1.

Figure 1. Bank Daily Profit/Loss Distributions



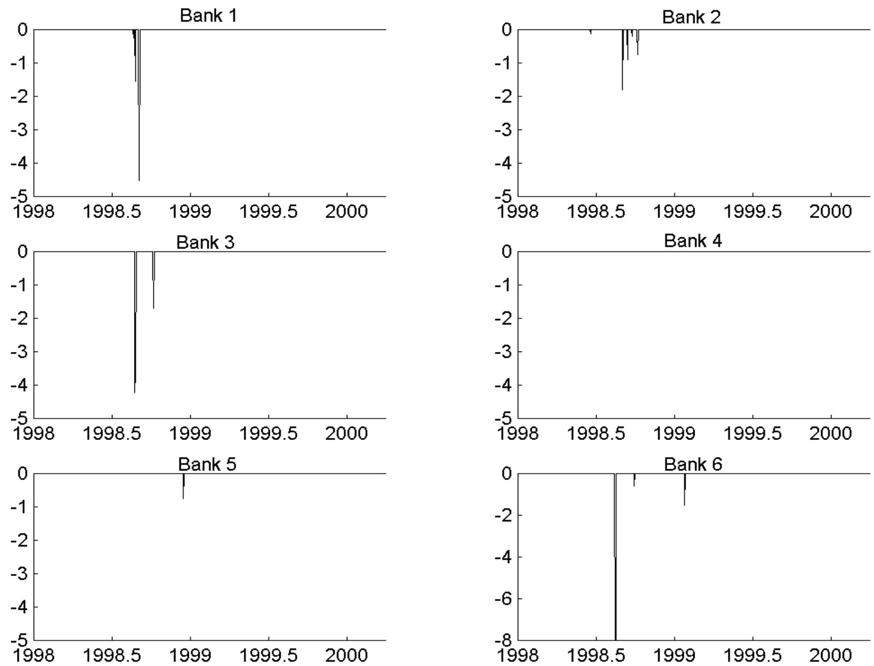
Notes:
Histograms of daily trading profit and loss reported for January 1998 through March 2000 (from May 1997 for bank 5). Data are de-meaned and expressed in standard deviations. See text for details.

Figure 2. Bank Daily VaR Models



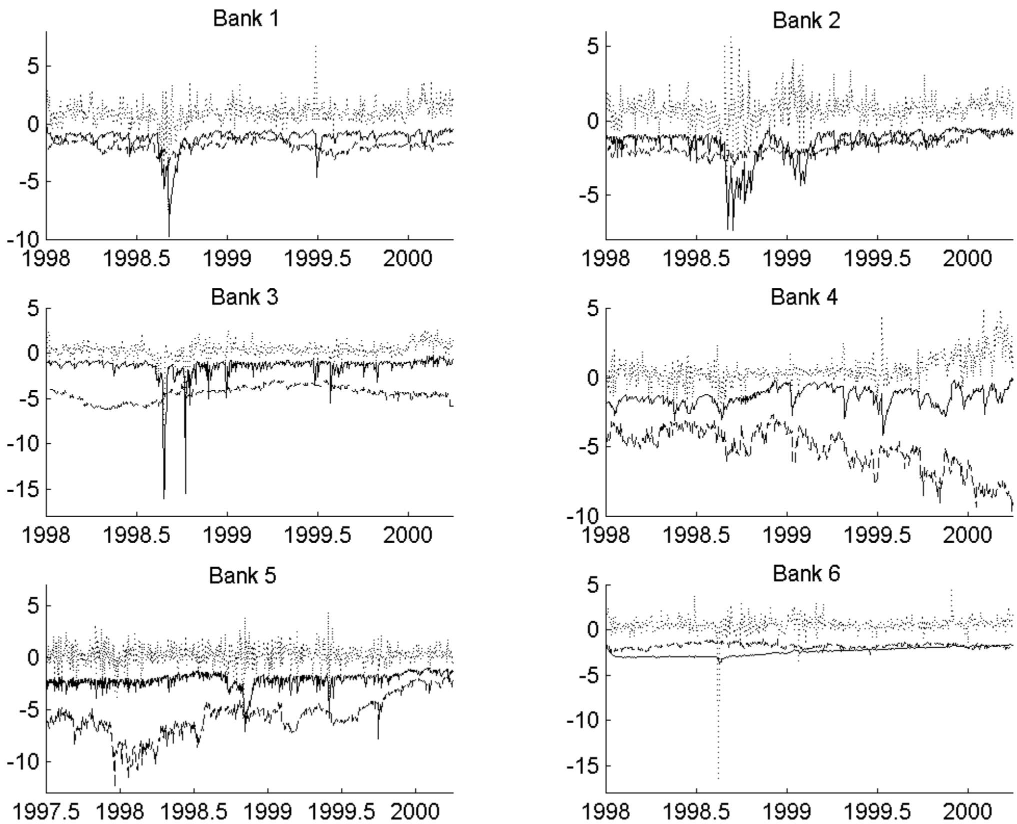
Notes: Time series of daily profit and loss (P&L) plotted with forecasts from bank internal VaR models (dashed line). The model is used to forecast the 1-day ahead 99th percentile of P&L.

Figure 3. Violations of Banks' 99% VaR



Notes: Plots show the daily P&L in excess of the bank's VaR for those days on which P&L drops below the VaR forecast for that day. Data are expressed in daily P&L standard deviations.

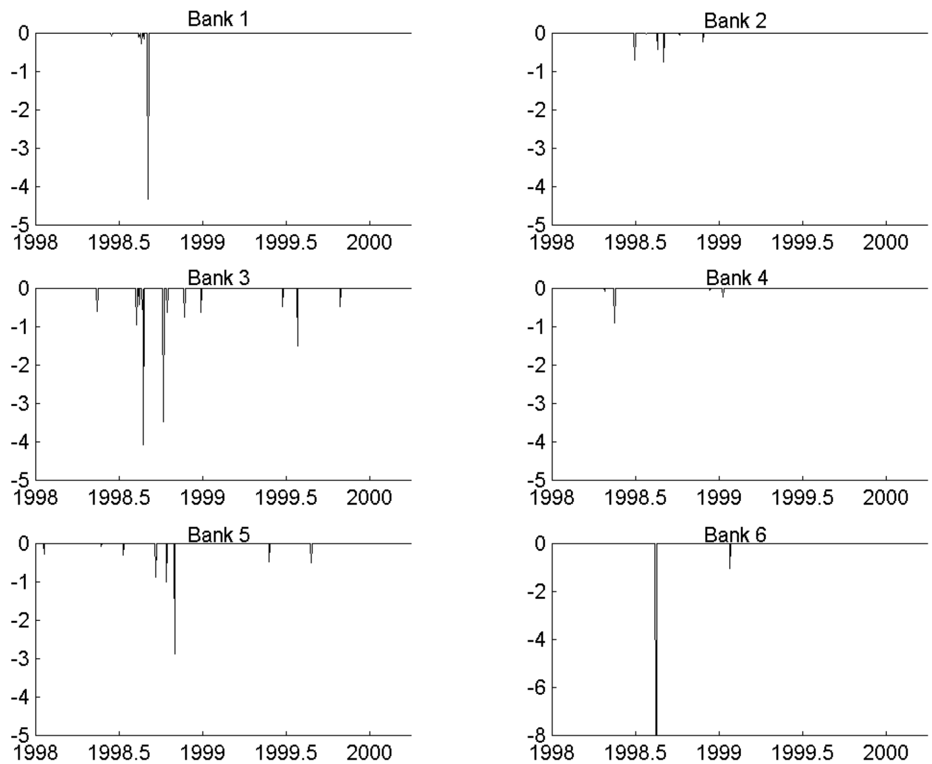
Figure 4. Daily Profit and Loss and VaR Forecasts



--- Bank VaR
— ARMA(1,1)+GARCH(1,1)

Notes: Time series of daily profit and loss (dotted lines) plotted with two model forecasts: an internal VaR model (dashed line) and a reduced form ARMA(1,1) with GARCH(1,1) Gaussian innovations in bank P&L (solid line). Both models are used to forecast the 1-day ahead 99th percentile of P&L.

Figure 5. Violations of the 99% GARCH-based VaR



Notes: Plots show the daily P&L in excess of the bank's VaR for those days on which P&L drops below the VaR forecast for that day. Data are expressed in daily P&L standard deviations.

Supplement to How Accurate are Value-at-Risk Models at Commercial Banks?

Jeremy Berkowitz
Graduate School of Management
University of California, Irvine
Irvine, CA 92697-3125

James O'Brien
Division of Research and Statistics
Federal Reserve Board
Washington, D.C. 20551

May 14, 2001

Berkowitz and O'Brien (April 2001) provide the first direct evidence on the performance of bank VaR models by analyzing the distribution of historical trading P&L and the performance of daily Value-at-Risk (VaR) estimates for the trading accounts of 6 large U.S. banks. All are large multinational institutions, meet the Basle "large trader" criterion, and include the largest US bank derivative dealers. The daily dollar trading return (P&L) and VaR data series are maintained by the banks to assess compliance with the Basle market risk capital requirements. Regulations stipulate that estimates are to be calculated for a 99 percent lower critical value of aggregate trading P&L with a one-day horizon. The forecasts provide a lower bound on aggregate trading P&L that should be breached 1 day in 100.

The VaR forecasts were evaluated first by testing the null hypothesis of a 99 percent coverage rate. Two important findings were that, unconditionally, the VaR estimates tend to be conservative relative to the 99th percentile of P&L and, for some banks, are highly inaccurate. Despite conservativeness, losses can substantially exceed the banks' VaRs and such events tend to be clustered. This suggests that the banks' models, besides a tendency toward conservatism, have difficulty forecasting changes in the volatility of P&L.

To further assess the performance of the banks' VaR models, VaR forecasts were compared with those from a standard GARCH model of the bank's P&L volatility, more precisely an ARMA(1,1)+GARCH(1,1) model of P&L. The GARCH model is a reduced form model of P&L and does not account for the composition or directly measure the risk of the current trading portfolio, as do the bank models. In principal, the banks' VaR models should deliver superior forecasts. Our results, however, indicated that the bank VaR models do not provide forecasts superior to a simple model of P&L volatility. The GARCH model of P&L generally provides for lower VaRs and is better at predicting changes in volatility. Because of the latter, the GARCH model permits comparable risk coverage with less regulatory capital. Implications for these results for VaR modeling and the potential usefulness of simple reduced-form risk forecasts were developed.

This supplement updates the results in Berkowitz and O'Brien (April 2001) with more data from subsequent reports by the banks. The original data for individual banks includes 1998 through the first quarter of 2000, about 600 daily observations on average. The new data extends the sample through the third quarter of 2000 for four banks and though the fourth quarter for two banks, with over 750 daily observations and a 25 percent increase in the average sample.

The results described below re-enforce those from the original data. Briefly, bank VaRs continue to be conservative in terms of a targeted 1 percent violation rate, and some models are hardly reflective of the lower tail of the observed distribution of P&L. VaR forecasts from a simple reduced-form model of bank P&L continue to provide a formidable benchmark for evaluating the bank VaR forecasts. The reduced-form VaRs on average achieve a 1 percent violation rate, although there is fair amount of variation across banks. The reduced-form VaRs, whilei decidedly less conservative than the bank VaRs, continue to at least match the performance of the latter in terms of the size of violations. Finally, as reported in Berkowitz and O'Brien (April 2001), cross-bank correlations in for 1-day P&L are positive and significant but low, although they are double in size over 5-day intervals.

I. Daily Trading P&L and VaR

As explained in more detail in Berkowitz and O'Brien (April 2001), the trading revenue for the 6 banks is based on close-of-day position values and, unless reported otherwise, represents the bank holding company's consolidated trading activities. Revenue includes gains and losses from daily marking to market of positions and fee income net of brokerage expenses related to the purchase and sale of trading instruments.

The daily VaR estimates are maintained by the banks for the purpose of forecast evaluation or "back-testing" and are required by regulation to be calculated with the same risk model used for internal measurement of trading risk. The VaRs are for a one-day ahead horizon and a 99 percent confidence level for losses, i.e., the 1% lower tail of the P&L distribution.

Daily P&L and VaR from the updated data are from January 1998 through September 2000 for 3 banks, from January 1998 through December 2000 for 2 banks, and from May 1997 through September 2000 for one bank. Summary statistics are reported in Table 1 for daily P&L and VaR. For these and other statistics reported below, each bank's daily P&L and VaR are divided by the bank's full-sample standard deviation of P&L to protect confidentiality. All banks report positive mean profits. Relative to the banks' P&L standard deviations, returns are somewhat higher than reported with the original data, reflecting the lessening influence of the high-volatility and low returns realized during the 1998 Asian financial and Russian debt crises (see below). Otherwise, results are similar. In particular, lower 99th percentile of the distributions of P&L the statistic forecasted by VaR, continue to be clustered at about 3 standard deviations below the mean. As a result, the excess kurtosis estimates (relative to the Normal distribution) displayed in column 5 are also large.

The last five columns of Table 1 show summary statistics for the banks' 99th percentile VaRs. Relative to the standard deviation of P&L, the bank VaRs continue to display a significant conservative bias. For 5 of 6 banks, the average VaR lies outside the lower 99th percentile P&L (column 7). The VaR for bank 2 is .86 of its 99th percentile P&L loss but, for the other 5 banks, VaRs range from 1.1 to 3.7 times their respective 99th percentile P&Ls. All violation rates are less than 1 percent and the mean violation rate is only .3 percent. Reflecting heavy tails and a left skew in P&L, the magnitudes of P&L losses *in excess of* the VaR can be large. This is indicated by the mean and maximum violations reported in the last two columns.

Histograms of daily P&L are presented in figure 1. At least 5 of the 6 banks exhibit extreme outliers, with a preponderance of the outliers in the left tail.

In Figure 2, the time series of each bank's P&L and corresponding one-day ahead 99th percentile VaR forecast (expressed in terms of the standard deviation of that bank's P&L) are displayed. The plots confirm the conservativeness of the VaR forecasts where violations of VaR are relatively few but large. The plots also show clear differences in VaR performances among banks. For banks 1, 2, and 6, VaRs are in the general vicinity of the lower range of their P&Ls, but for banks 3, and 4 this is not the case, while bank 5's VaR has improved.

More than three quarters of the violations of VaR seen in Figure 2 occurred as a consequence of high volatility in P&L between August and October 1998, in the wake of the Russian default and turmoil in Asian financial markets. Table 2 (column 1) shows that during this period, average returns are low, return standard deviations for most banks are exceptionally large, and the 99th percentiles are blown out. Figure 3 shows the timing and magnitudes of VaR violations over the full sample period.

These findings suggest that P&L may be correlated across banks. In the upper panel of Table 3, cross-correlations between banks' daily P&Ls are reported. Correlations are uniformly positive and with t-values mostly above 2. Nonetheless, the correlation coefficients for daily P&L are generally low, mostly below .2. Low correlations may reflect differences in portfolio compositions among banks. That is, even when market disruptions are widespread, shocks across different markets do not necessarily occur on the same calendar day. Additionally, trading firms have some discretion in the exact timing for reporting losses or gains in P&L, especially for inactively traded instruments. When P&L is aggregated over multi-day horizons, these idiosyncrasies may be less important. In fact, over 5-day holding periods, the P&L cross-correlations approximately double (not shown). If the August – October 1998 period of high P&L volatility is excluded, the 1-day correlations are moderately lower, by about 20 percent on average, while the 5-day correlations again are approximately twice the 1-day correlations.

The lower panel of Table 3 displays correlations for daily VaR across banks. The VaR correlations are as often negative as they are positive and no clear pattern of co-movement is evident. Results are qualitatively the same for different sample periods and time aggregations of the data. These findings are consistent with different patterns in the bank VaRs displayed in Figure 2 and contrast with consistently positive daily cross-correlations in P&L.

II. Testing Model Performance

In this section we study the forecast accuracy of the bank VaR estimates and their sensitivity to daily portfolio volatility. Denote the portfolio's P&L by r_t o that each day t the bank forecasts r_{t+1}. The VaR forecast is the quantity $\bar{r}$ such that $p(r_{t+1} < \bar{r}) = \alpha$ over the next trading day. Here $\alpha = .01$, so that the model predicts a lower bound on losses not to be exceeded with 99% confidence.

A. Forecast Evaluation

The traditional approach to validating such interval forecasts is to compare the targeted violation rate, α, to the observed violation rate. The first two columns (unconditional coverage) present likelihood ratio statistics and p-values for tests of bank VaR violation rates (shown in Table 1, column 9) against the null hypothesis of 1 percent. At a 10 percent significance test, the null of .01 would be rejected for 5 of 6 banks and for 3 banks at a 5 percent test level (column 2, including bank 4). While the power of these tests is not especially strong, there are stronger rejections of the null violation rate than with the smaller sample results in Berkowitz and O'Brien (April, 2001), Also, as noted above, the mean violation rate across banks is only .3 percent.

Christoffersen (1998) has developed tests for whether violations are independent, as would be expected if the VaR forecasts were optimal. Statistically, the variable defined as

$$I_t = 1 \quad \textit{if violation occurs}$$
$$= 0 \quad \textit{if no violation occurs}$$

should be an iid Bernoulli sequence with parameter α. Likelihood ratio tests of this null are easily constructed. These tests are referred to as conditional coverage, and LR statistics and p-values are reported in columns 3 and 4 (conditional coverage) of Table 4. At about a 5 percent significance level, the conditional null of .01 is rejected for all banks with a violation.

A useful feature of the likelihood framework is the following identity:

$$LR_{cc} = LR_{uc} + LR_{ind.}$$

That is, the conditional coverage test (LR_{cc}) can be decomposed into a test of the unconditional coverage (LR_{uc}), i.e., violation rate of α, plus a test that violations are independent (LR_{ind}). Columns 4 and 5 reports the results of LR tests for first-order serial dependence.[1] The p-values indicate two rejections at standard significance levels. While these results are limited to first-order serial dependence, as noted earlier, more than 3/4ths of the VaR violations occurred during a single three-month period.

B. Comparisons with a Benchmark Model

The clustering of violations suggests that the volatility of P&L may be time varying to a degree not captured by the models. To further pursue the potential for predictable volatility, we formulate an alternative VaR model determined from an ARMA(1,1) plus GARCH(1,1) model of portfolio returns. That is, we estimate the following reduced form model of r_t

$$(1) \quad r_t = \mu + \rho r_{t-1} + u_t + \lambda u_{t-1}$$

where u_t is an iid innovation with mean zero and variance σ_t. The volatility process σ_t is described by

[1] The tests are restricted to first-order dependence, rather than considering higher-order dependence as well, because of the small number of observations.

$$(2) \quad \sigma_t = \omega + \theta u_{t-1}^2 + \phi\sigma_{t-1}$$

where ω, θ and ϕ are parameters to be estimated. We apply the standard GARCH model where innovations are assumed to be conditionally Normal. Thus the 99% VaR forecast at time t is given by $\hat{r}_{t+1} - 2.33\hat{\sigma}_{t+1}$, where $\hat{r}_{t+1}$ is the predicted value of r_{t+1} from equation (1) and $\hat{\sigma}_{t+1}$ is the estimated volatility from equation (2).[2]

The ARMA and GARCH parameters are estimated each day with data available up to that point. To obtain stable estimates for the initial period, forecasts for days 1 through 165 are in-sample. Rolling out-of-sample forecasts begin after day 165, which is in the third week of August 1998 except for one bank (where it is May 1998). Out-of-sample estimates are updated daily. Given parameter estimates, we forecast the next day's 99% VaR assuming Normality of the GARCH innovations. The resultant forecasts, both within and out-of sample, are shown in Figure 4 by the solid line, along with the P&L and the internal model forecasts. One-day ahead reduced-form forecasts appear to track the lower tails of P&L remarkably well. Compared to the structural model, the time series model appears to do better at tracking the lower tail of the banks' P&L and adjusting to changes in volatility.

Summary statistics and backtests for the GARCH model VaRs are presented in the top panel of Table 5. The second column shows that the GARCH model successfully removes first-order persistence in banks' P&L volatility (as well as higher-order persistence). The average GARCH VaRs shown in column 3 are also considerably lower than average bank VaRs, except for bank 6, and the number of violations shown in column 4 average out to about 1 percent. The mean violation rate for the GARCH VaRs also is lower than that of the banks' VaRs. Thus, the results remain consistent with those in the shorter samples in Berkowitz and O'Brien (April 2001).

Because they are more conservative, bank VaRs might be expected to have higher mean violations but with lower maximum violations and lower aggregate violations (mean times

---

[2] Note that it is possible for the reduced form ARMA+GARCH VaR to be positive. This would occur if the conditional mean of the distribution is large so that $\hat{r}_{t+1} > 2.33\hat{\sigma}_{t+1}$.

frequency) than the less conservative GARCH VaRs, other things equal.[3] However, this is not the case. Aggregate violations (column 4 times column 5) and maximum violations for the GARCH VaRs are comparable to the bank model VaRs, even though the bank VaRs are more conservative. These results indicate a potentially important advantage for the reduced-form GARCH model. The magnitudes of the banks' VaR forecasts are used to determine regulatory capital requirements, and likely influence banks' internal capital allocations as well. The GARCH VaRs are able to deliver lower required capital levels without producing larger violations. As described below, this reflects the GARCH model VaRs greater responsiveness to changes in P&L volatility.

Formal back-tests of the GARCH models are presented in the bottom panel of Table 5.[4] As shown in the first two columns, at the 10-percent significance level, the null violation rate of .01 is rejected for 3 banks. The rate of rejection is lower than for the bank VaR models, with two rejections due to too many violations. For conditional coverage, the null of a .01 violation rate is rejected for four banks (columns 3 and 4) and for independence *per se* (columns 5 and 6) there is rejection for two banks. Overall, the individual bank backtests do not sharply differentiate between the bank and the GARCH VaRs in terms of forecast accuracy and the independence of violations. This would be consistent with the low power of these tests suggested in previous work (Kupiec (1995), Christoferson (1998), and Berkowitz (2000).

Despite the limitations of the backtests, the GARCH model's greater responsiveness to changes in P&L volatility is illustrated in results presented in Table 6. The top panel presents R-squares between the negative of the daily VaRs (VaRs are negative numbers) and absolute innovations in banks' daily P&L, a measure of daily volatility. The innovation is the difference between P&L and the 1-day ahead ARMA(1,1) forecast. For five of six banks the correlation is positive and for four banks, R-squares are significant at the .01 level. However, for the GARCH

---

[3] Under either a normal distribution or heavy-tailed distributions, such as the t distribution, the conditional expected value of lower tail returns is increasing in the lower critical tail value, while the unconditional, aggregate, and maximum expected values are inversely related to the lower critical tail value.

[4] Backtests were also carried out only for the out-of-sample forecasts, which account for about 75 percent of the full sample results. For the out-of-sample period, the average bank VaR was about the same as for the full sample, while the average of mean violations was somewhat lower. The average violation rate also was very close to .01. Average bank results for the backtests (coverage, dependence, etc) were very similar to those for the full sample period.

VaRs, all the correlations are positive and, as shown in the bottom panel of Table 6, the R-squares are more than twice as high. Five of six banks have significant R-squares and the sixth bank is significant at the .05 level. [5]

The greater responsiveness of the reduced-from VaRs is illustrated for the August-October 1998 period when P&L volatility rose substantially. Table 6 compares model performances during this 3-month period. Even though the GARCH model VaRs are smaller over the full sample, the bank and GARCH VaRs are comparable during this period. For this 3-month period, the GARCH VaRs increased from 80 to 250 percent over their average VaRs during the 3 months prior to August 1998 for 4 of the 5 banks with violations. The bank VaRs in comparison were 20 percent lower to 30 percent higher than their respective averages over the preceding 3 months. As a result, the performances of the bank and GARCH VaRs are comparable in terms of average, aggregate and maximum violations.

While these results show that the GARCH VaR forecasts compare favorably with the banks' VaRs, the GARCH model is not unassailable. A plot of the GARCH violations in Figure 5, along with the results presented in Table 6, indicate that some clustering remains. Also, while the average violation rate for the GARCH VaRs is 1 percent, other statistics such as kurtosis indicate heavy tails in the GARCH P&L residuals. These results are due to the GARCH model's inability to adequately reflect the sharp increase in P&L volatility in the latter part of 1998.

Some further evidence of this is provided by the GARCH model parameter estimates for different sample periods (not shown). For banks 1 through 4, GARCH and ARCH parameters jump as the sample period is extended to include the period of heightened P&L volatility. The sum of the GARCH and ARCH parameters briefly reach one but subsequently decline below one as the sample is further extended. Excluding 1998 from the sample period, the sum of the GARCH and ARCH parameters remain below one for all banks. These results are suggestive of

[5] The GARCH VaR includes the 1-day P&L forecast, as well as the 1-day volatility forecast. R-squares between the absolute P&L innovation and the GARCH volatility forecast alone are slightly higher than those using the GARCH VaR. It also is important to note that VaR is an estimate of the 99^{th} percentile of the distribution of P&L not of individual absolute shocks. The fact that a significant correlation is observed between changes in bank VaRs and subsequent P&L shocks suggests only that variation in VaR is consistent with changes in the distribution, i.e., volatility, of P&L. A zero or insignificant correlation, need not imply that the VaR is inaccurate in estimating the 99^{th} percentile P&L.

an environment subject to regime shifts, which cannot be captured by the standard GARCH model (see Gray (1996)).[6]

III. Conclusions

With more data on P&L and VaR for the banks studied in Berkowitz and O'Brien (April 2001), the results here offer further confirmation of the main conclusions reached with the smaller samples. For the majority of banks, variation in VaR is positively correlated with the absolute size of shocks to P&L. Nonetheless, bank VaR forecasts continue to be quite conservative relative to the 99th percentile coverage prescribed by the Market Risk Amendment rules. Also, the VaR forecasts for some banks are highly inaccurate.

VaRs based on the reduced-form or GARCH model of P&L and calibrated to a 99th percentile on average achieve a 1 percent violation rate, although with sizable variation in violation rates across banks. Nonetheless, the GARCH model VaRs are more responsive to changes in bank P&L volatility and, because of this, provide loss coverage comparable to the more conservative bank VaRs. While the GARCH P&L model used here is limited—for example changes in trading positions are not accounted for—the results continue to provide support for giving consideration to VaR forecasts based on volatility and trend estimates of an aggregated portfolio return.

[6] We also estimated VaR using an IGARCH model, where the sum of the ARCH and GARCH coefficients is constrained to equal one throughout the sample. Violations were again clustered in the August-October 1998 period.

References

Berkowitz, J. (2000), "Testing the Accuracy of Density Forecasts," forthcoming, *Journal of Business and Economic Statistics*

Berkowitz and O'Brien (2001), "How Accurate are Bank Value-at-Risk Models at Commercial Banks," April.

Christoffersen, P. (1998), "Evaluating Interval Forecasts," *International Economic Review*, 39, 841-862.

Gray, S. (1996), "Modeling the conditional distribution of interest rates as a regime-switching process," *Journal of Financial Economics*, 42, 27-62.

Kupiec, P. (1995). "Techniques for Verifying the Accuracy of Risk Measurement Models," *Journal of Derivatives*, 3, 73-84.

Table1. Bank P&L and VaR Summary Statistics

		Daily P&L						VaR Violations			
	observ	mean	standard deviation	99th percentile	excess kurtosis	skew	mean VaR	number	rate	mean	max
Bank 1	762	1.050	1	-1.732	10.744	-0.599	-1.938	3	0.004	-2.183	-4.673
Bank 2	711	0.801	1	-1.977	4.982	0.154	-1.706	6	0.008	-0.782	-1.902
Bank 3	712	0.491	1	-2.576	23.524	-2.947	-4.838	3	0.004	-3.281	-4.350
Bank 4	703	0.728	1	-1.543	1.657	0.659	-5.736	0	0.000	0.000	0.000
Bank 5	876	0.256	1	-2.920	4.738	-0.812	-5.110	3	0.003	-1.988	-3.991
Bank 6	780	0.698	1	-0.992	154.69	-8.161	-1.946	3	0.004	-6.390	-16.856

Notes: Daily P&L and VaR are reported from January 1998 through September 2000 for banks 2,3, and 4. and through December 2000 for banks 1 and 6. For bank 5, the data is from May 1997 through September 2000. Mean violation refers to the loss in excess of the VaR.

Table 2. Bank P&L and VaR Summary Statistics: August-October 1998

		Daily P&L			Daily VaR			
	observ	mean	standard deviation	99th percentile	mean VaR	number violations	mean violation	max violation
Bank 1	63	0.180	1.812	-7.231	-2.391	3	-2.183	-4.673
Bank 2	64	0.081	1.999	-4.502	-2.411	5	-0.910	-1.902
Bank 3	65	-0.936	1.897	-8.956	-4.772	3	-3.282	-4.350
Bank 4	63	-0.014	0.786	-1.898	-4.676	0	0	0
Bank 5	65	0.064	1.607	-5.538	-5.116	1	-0.780	-0.780
Bank 6	65	0.187	2.501	-15.567	-1.522	2	-8.748	-16.856

note:
Dollar values are divided by the banks' respective full sample standard deviations.

Table 3. Correlations of P&L and VaR Across Banks

A. P&L Correlation Coefficients

	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5	Bank 6
Bank 1	1.000	0.427	0.228	0.171	0.041	0.145
		(11.072)	(5.910)	(4.438)	(1.050)	(3.758)
Bank 2		1.000	0.119	0.081	0.139	0.156
			(3.079)	(2.102)	(3.615)	(4.057)
Bank 3			1.000	0.378	0.085	0.198
				(9.813)	(2.207)	(5.123)
Bank 4				1.000	0.094	0.105
					(2.435)	(2.722)
Bank 5					1.000	0.069
						(1.789)
Bank 6						1.000

B. VaR Correlation Coefficients

	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5	Bank 6
Bank 1	1.000	-0.033	0.109	0.064	-0.155	-0.328
		(-.852)	(-2.878)	(-1.667)	(-4.010)	(-8.503)
Bank 2		1.000	-0.144	-0.817	0.790	-0.298
			(-3.733)	(-21.195)	(20.531)	(-7.729)
Bank 3			1.000	0.261	-0.351	0.032
				(6.770)	(-9.106)	(.827)
Bank 4				1.000	-0.818	0.286
					(-21.215)	(-7.431)
Bank 5					1.000	-0.094
						(2.4461)
Bank 6						1.000

Notes: Correlations calculated with matched sample of daily P&L and VaR with 673 observations. t-statistics are shown in parentheses.

Table 4. Bank VaR Backtests

	unconditional coverage[1]		conditional coverage[2]		independence[3]	
	LR stat	p-value	LR stat	p-value	LR stat	p-value
bank 1	3.663	0.056	3.687	0.055	0.024	0.878
bank 2	0.182	0.670	4.568	0.033	4.387	0.036
bank 3	3.067	0.080	10.572	0.001	7.505	0.006
bank 4	NaN	NaN	NaN	NaN	NaN	NaN
bank 5	5.114	0.024	5.136	0.023	0.021	0.886
bank6	3.880	0.049	3.908	0.048	0.023	0.879

notes:

1. Unconditonal coverage: tests the null hypothesis that daily P&L will violate VaR at an expected rate of 1 percent. LR stat is the liklihood ratio value and p-value indicates the probability of the actual violation rate if the null of a 1 percent expected rate is true.
2. Independence: tests for first-order dependency in daily violations. p-value is the probability of the estimated LR value if violations are independent.
3. Conditional coverage: a joint test of a 1 percent expected violation rate and first-order independence.

Reference: Christofferson (1998)

Table 5. GARCH Model Results

5.1. GARCH Model 1-Day VaR Statistics[1]

				VaR Violations			
	observ	Box-Ljung[2]	mean VaR	number violation	rate	mean	max
bank 1	762	0.387	-1.157	6	0.008	-0.855	-4.467
bank 2	770	0.011	-1.356	6	0.008	-0.382	-0.784
bank 3	712	0.058	-1.344	13	0.018	-1.157	-4.211
bank 4	703	0.801	-1.189	5	0.007	-0.353	-1.110
bank 5	876	0.094	-2.082	14	0.016	-1.062	-3.646
bank 6	780	0.027	-2.396	2	0.003	-7.897	-14.654

5.2. GARCH VaR Backtests[3]

	unconditonal coverage		conditional coverage		independence	
	LR stat	p-value	LR stat	p-value	LR-stat	p-value
bank 1	0.037	0.543	4.892	0.027	4.521	0.034
bank 2	0.182	0.670	0.284	0.594	0.102	0.749
bank 3	3.959	0.047	9.469	0.002	5.510	0.019
bank 4	0.653	0.419	0.724	0.395	0.072	0.789
bank 5	2.692	0.100	4.222	0.040	1.530	0.216
bank 6	6.185	0.013	6.195	0.013	0.010	0.919

notes:

1. 1-day forecasts using ARMA(1,1)+GARCH(1,1).
2. Box-Ljung statistics are for first-order serial correlation in the squares of the standardized GARCH residuals.
3. For description of test statistics, see footnotes to Table 4.

Table 6. Bank VaR and GARCH VaR Adjustments to P&L Volatility

6.1. R-square: Daily VaR and Absolute P&L Innovations[1]

	bank 1	bank 2	bank 3	bank 4	bank 5	bank 6
Bank VaR	.023**	.079**	.000	.042**	.021**	(-).005
GARCH VaR	.118**	.167**	.118**	.035**	.054**	.009*

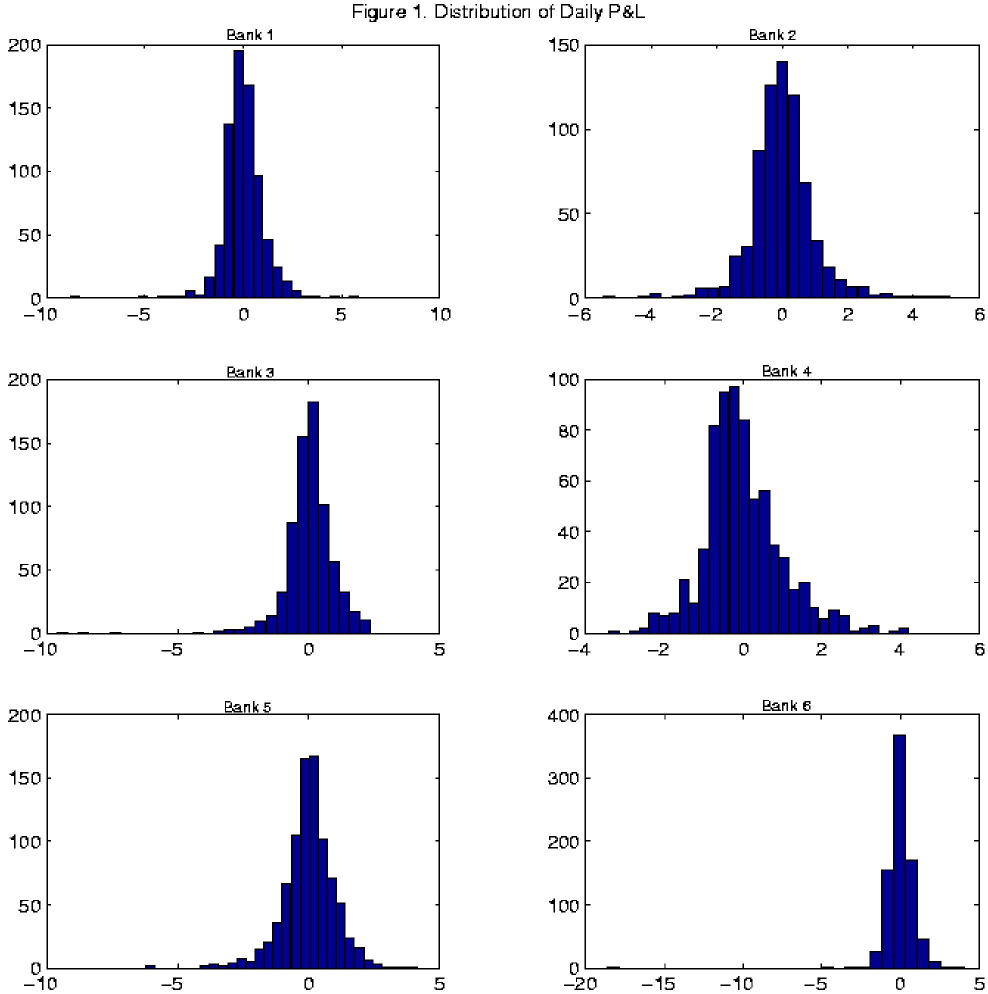
1. * (**) indicates significance at the .05 (.01) level

6.2. Bank and GARCH VaR: August-October 1998[2]

		VaR Violations			GARCH VaRs			
						VaR Violations		
	observ	number	mean	max	mean VaR	number	mean	max
Bank 1	63	3	-2.183	-4.673	-2.989	5	-1.011	-4.465
Bank 2	64	5	-0.910	-1.902	-3.605	3	-0.433	-0.790
Bank 3	65	3	-3.282	-4.350	-3.224	7	-1.495	-4.211
Bank 4	63	0	0	0	-1.697	0	0	0
Bank 5	65	1	-0.780	-0.780	-2.984	4	-1.356	-2.886
Bank 6	65	2	-8.748	-16.856	-3.206	1	-14.654	-14.654

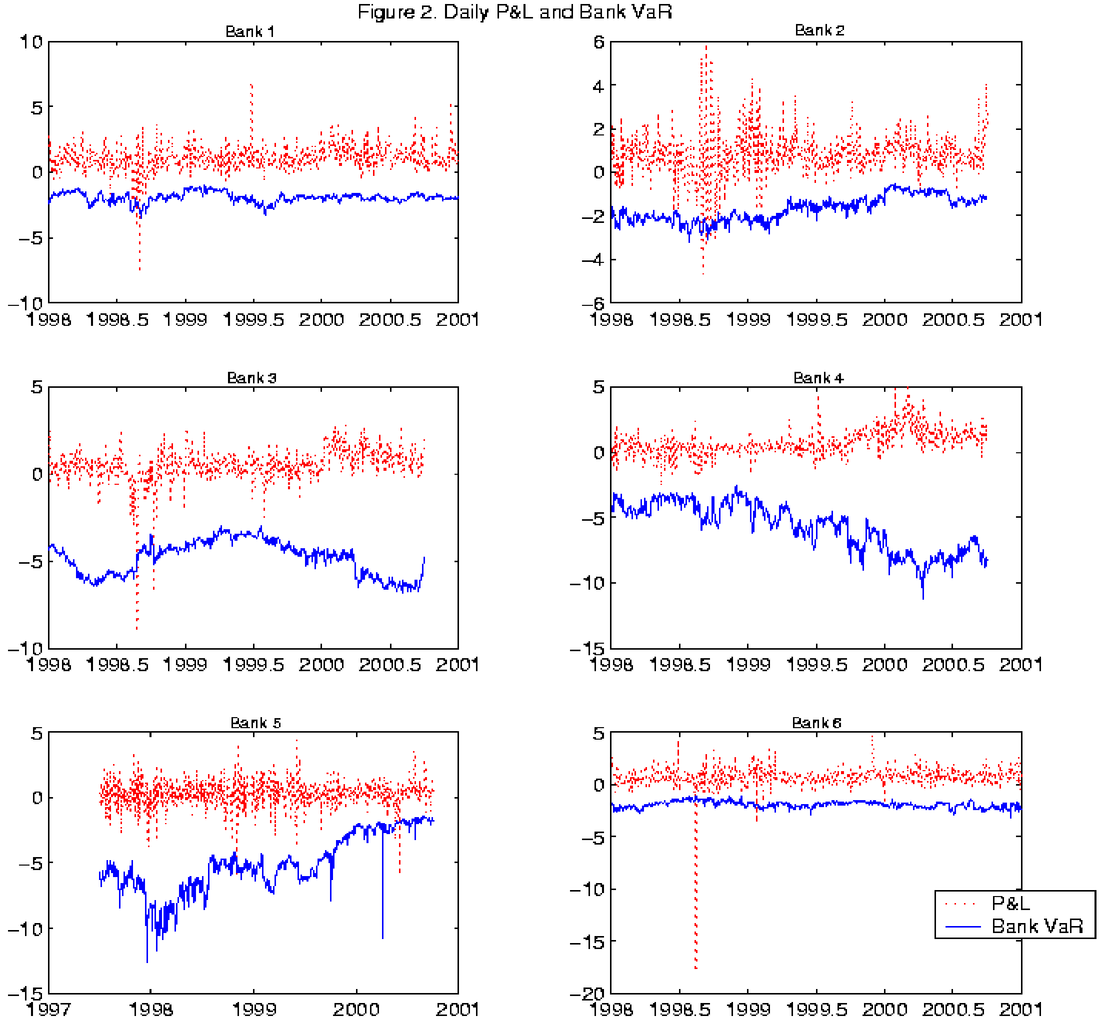
2. Dollar values are divided by full-sample bank P&L standard deviations.

Figure 1. Distribution of Daily P&L



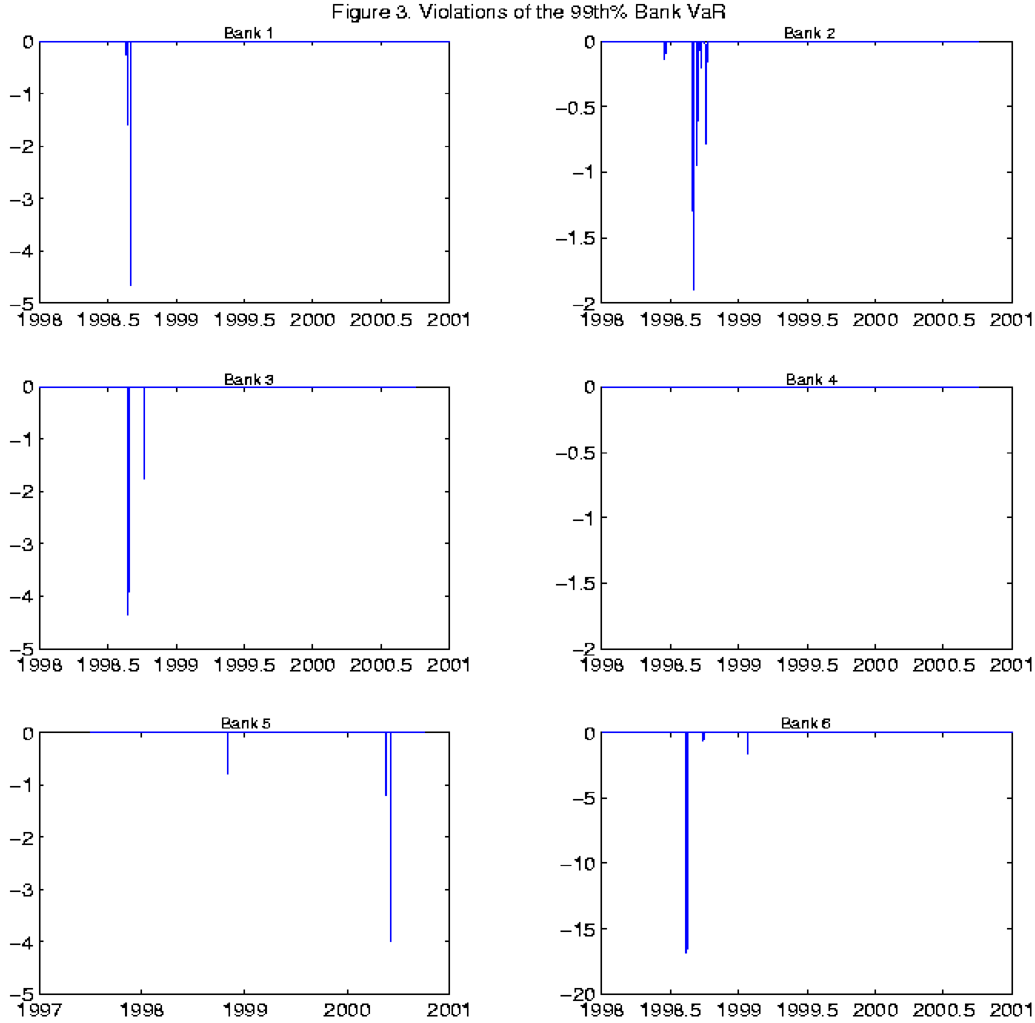
Notes: Histograms of daily profit and loss data reported by large commercial banks for January 1998 through October and December of 2000, with data starting in May7, 1997 for bank 5. Data are de-meaned and expressed in standard deviations. See text for details.

Figure 2. Daily P&L and Bank VaR



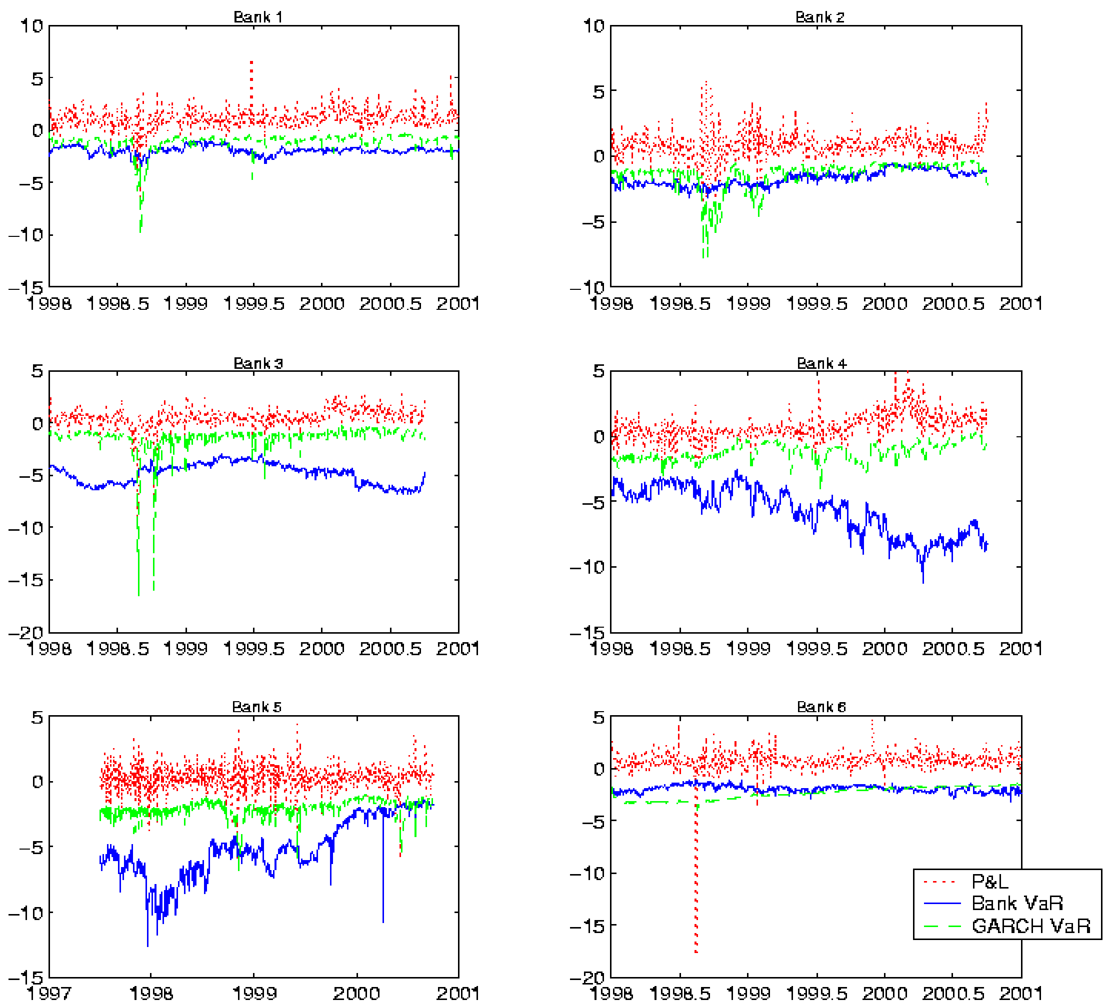
Notes: Time series of banks' daily trading profit and loss (dotted lines) plotted with forecasts from an internal VaR model (solid line) used to forecast 1-day ahead 99$^{\text{th}}$ percentile of P&L.

Figure 3. Violations of the 99th% Bank VaR



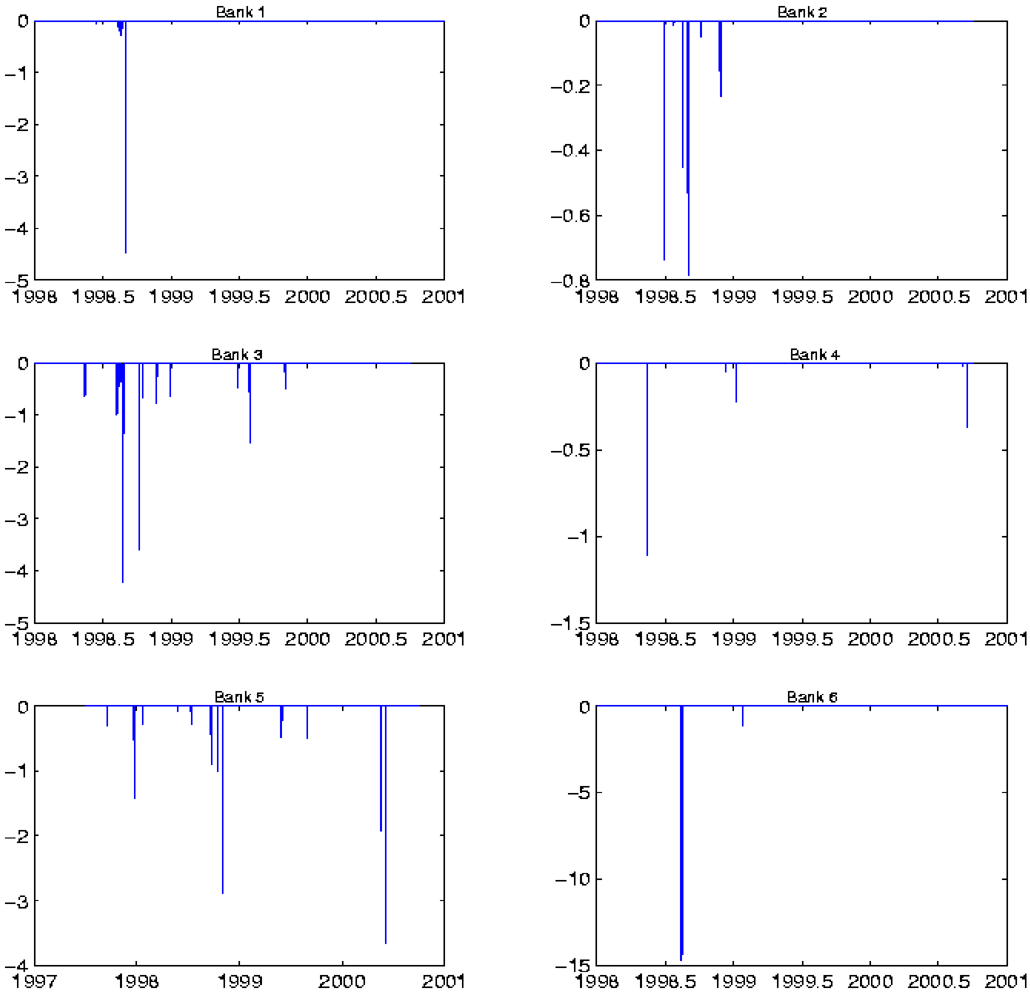
Notes: Plots show the daily P&L in excess of the bank's VaR for those days on which P&L drops below the VaR forecast for that day. Data are expressed in standard deviations.

Figure 4. Daily P&L, Bank VaR, and GARCH-based VaR



Notes: Time series of banks' daily trading profit and loss (dotted lines) plotted with two model forecasts of the 1-day ahead 99th percentile P&L. The two models are an internal VaR model (dashed line) and a P&L ARMA(1,1) with GARCH(1,1) model (solid line).

Figure 5. Violations of the 99th% GARCH VaR



Notes: Plots show the daily P&L in excess of the GARCH VaR for those days on which P&L drops below the GARCH VaR forecast for that day. Data are expressed in standard deviations.