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

JEREMY BERKOWITZ and JAMES O'BRIEN*

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 (VaR) 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.

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 banks are the principal dealers. To manage market risks, major trading institutions have developed large scale risk measurement models. While approaches may differ, all such 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 nonfinancial firms as well.

The general acceptance and use of large scale 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., Hendricks (1996), Marshall and Siegel (1997) and 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 six large U.S. banks. All are large multinational institutions and meet the Basle "large trader" criterion—with trading activity equal to at least 10 percent of total assets or $1 billion. The banks include the largest U.S. 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 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, we identify modeling practices and regulatory constraints that may affect the precision, particularly the conservativeness, of the VaR forecasts.

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 are not better than simple models of 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.

---

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

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. Section IV 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 six 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, that is, the 1 percent 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 intraday position changes. The bank models also omit net fee income, although it is included in reported trading P&L.

Summary statistics are reported in Table I 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. Sizable differences in average P&L and standard deviation across banks (not shown) correspond to differences in the size of trading activity, although column two of the table also indicates significant disparity in mean P&L relative to its variation. In column four, we report the 99th percentile of losses of the P&L distributions, over the full sample. These losses, coming once in 100 days, are quite large

Table I

Bank P&L and VaR Summary Statistics

This table reports daily profit and loss data reported by large commercial banks for January 1998 through March 2000. Each bank's data are divided by its sample standard deviation to protect the confidentiality of individual institutions. Mean violation refers to the loss in excess of the VaR.

		Daily P&L					Daily VaR		
	Obs	Mean	Standard Deviation	99th Percentile	Excess Kurtosis	Skew	Mean VaR	Number Violations	Mean Violation
Bank 1	569	0.964	1.00	−1.78	11.63	−0.993	−1.87	3	−2.12
Bank 2	581	0.737	1.00	−2.26	4.53	0.094	−1.74	6	−0.741
Bank 3	585	0.375	1.00	−2.73	23.87	−3.13	−4.41	3	−3.18
Bank 4	573	0.595	1.00	−1.59	2.31	0.860	−5.22	0	
Bank 5	746[a]	0.253	1.00	−2.78	3.41	−0.617	−5.62	1	−0.775
Bank 6	586	0.608	1.00	−0.967	142.10	−8.25	−1.72	3	−5.84

[a]Data begin in May 1997.

and are clustered at about three standard deviations below the mean. As a result, the excess kurtosis estimates (relative to the Normal distribution) displayed in column five are also large.

The last three columns of Table I show summary statistics for the banks' 99th percentile VaRs. For five of six banks, the average VaR lies outside the lower 99th percentile P&L, with VaRs for four banks ranging from 1.6 to over three times their respective 99th percentile P&Ls. At the 99th percentile, P&L would be expected to violate VaR five times in 500 trading days. However, only one bank experienced more than three violations. In this sense, the VaR forecasts appear quite conservative, a finding that is given more attention in the analysis below.

A shortcoming of VaR as a risk management tool is that it conveys nothing about the size of violations when they do occur (e.g., Basak and Shapiro (2000) and Berkowitz (2001)). It is therefore of some interest to examine the empirical evidence on the magnitude of exceedances.

While violations of VaR in our dataset are infrequent, the magnitudes can be surprisingly large. For two banks, the mean violation is more than two standard deviations beyond the VaR. For one bank, it is more than five standard deviations beyond the mean VaR. To get a sense of the size of these violations, consider a parametric such as the Normal as a benchmark. Under a Normal distribution, the probability of a loss just one standard deviation beyond a 99 percent VaR is 0.04 percent. The probability of a loss two standard deviations beyond 99 percent is virtually zero. For a Student-*t* distribution with five degrees of freedom, which is quite fat-tailed, the probability of a one standard deviation exceedance is only 0.3 percent, and of a two standard deviation, 0.1 percent. In this sense, losses of the magnitude seen in our sample are quite far beyond the VaR.
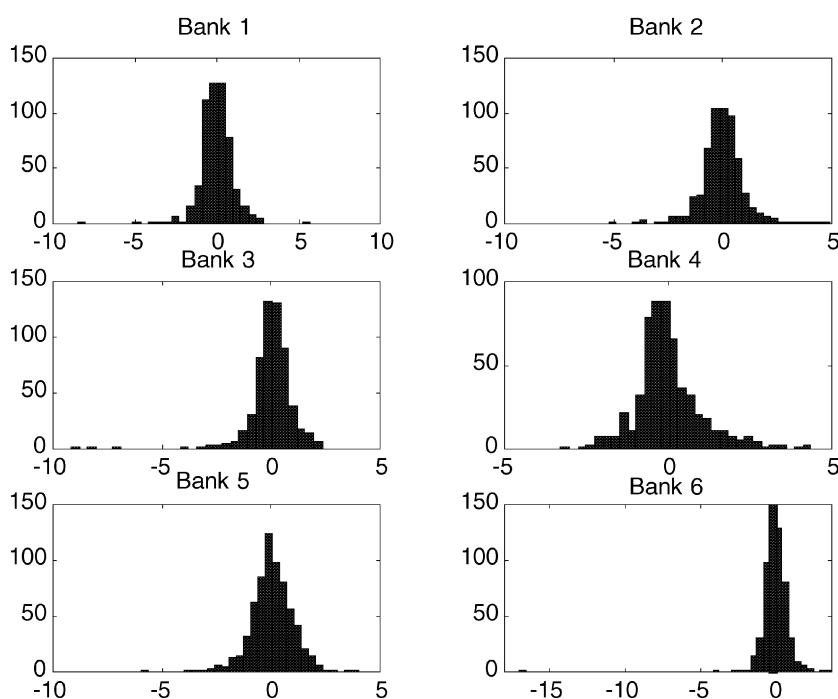
Figure 1. Bank daily profit/loss distributions. Histograms of daily profit and loss data reported by large commercial banks for January 1998 through March 2000. Data are demeaned and expressed in standard deviations. The scale of the subplots differs across banks and is indicated on the x-axis. See text for further details.

Histograms of P&L are presented in Figure 1 for the six banks. In all histograms, daily P&L are demeaned and divided by their standard deviation. At least five of the six banks exhibit extreme outliers, with a preponderance of the outliers in the left tail. Both the skewness estimates reported in Table I 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 99th 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 differences in VaR performances among banks. For banks one, two, and six, VaRs are in the general vicinity of the lower range of their P&Ls, but for banks three, four, and five, this is not the case. The VaRs for these three banks also appear to exhibit trends. In particular, bank four's VaR trends down while bank five's VaR trends up.

The large losses in Figure 2 occurred during the turbulent period in world debt markets between August and October 1998, marked by the devaluation of the Russian ruble, Russian debt default (August) and the near-collapse of
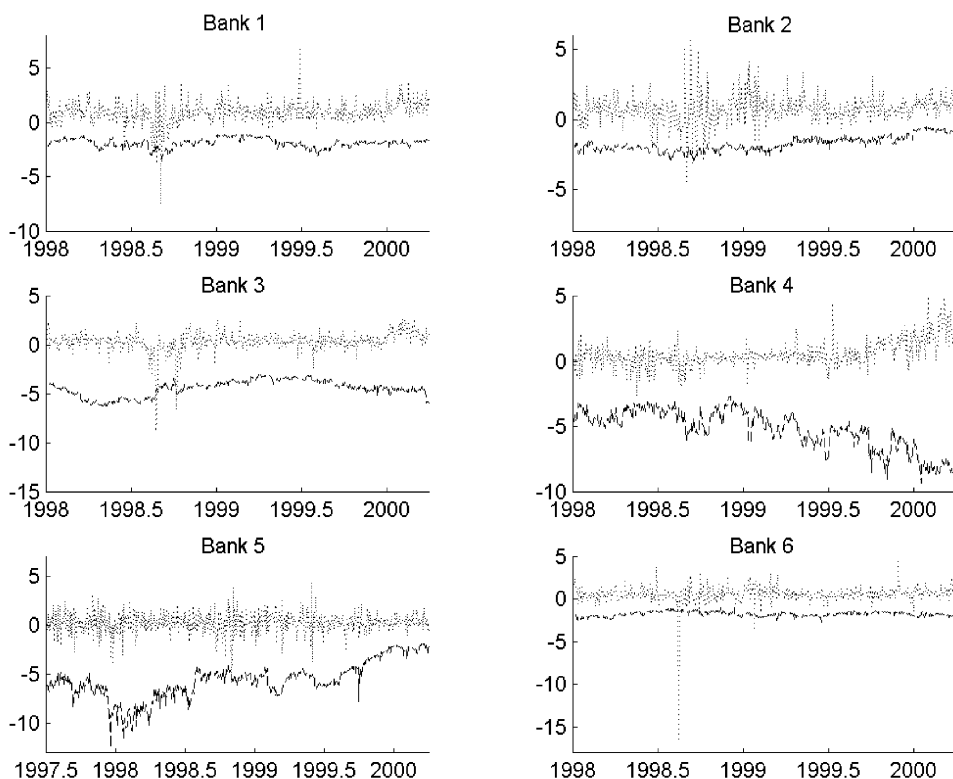
Figure 2. Bank daily VaR models. Time series of daily profit and loss data as reported by six commercial banks for January 1998 through March 2000 (dotted lines) plotted with forecasts from an internal VaR model (dashed line). The model is used to forecast the one-day-ahead 99th percentile of P&L.

a large U.S. hedge fund (September). Table II (column one) shows that during this period, average returns are lower, standard deviations of the P&L for most banks are exceptionally large. As shown in column three, 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.

Based on quarterly financial reports, the poor performance for most of the banks primarily reflects losses on interest rate positions, although some banks also reported losses in other trading activity (foreign exchange, equity, and commodities). While counterparty defaults on derivative contracts spiked up in this period, the dollar magnitudes still made only a small contribution to trading losses.

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 III, cross-correlations between banks' daily P&Ls are

Table II

**Bank P&L and VaR Summary Statistics,
August 1998 to October 1998**

This table reports 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 I.

		Daily P&L					Daily VaR		
	Obs	Mean	Standard Deviation	Minimum	Excess Kurtosis	Skew	Mean VaR	Number Violations	Mean Violation
Bank 1	63	0.175	1.76	−7.01	4.58	−1.32	−2.32	3	−2.12
Bank 2	64	0.076	1.89	−4.26	1.46	0.787	−2.28	5	−0.862
Bank 3	65	−0.907	1.84	−8.68	7.65	−2.53	−4.63	3	−3.18
Bank 4	63	0.0453	0.773	−1.89	0.99	−0.434	−4.66	0	
Bank 5	65	0.0637	1.60	−5.51	1.99	−0.837	−5.09	1	−0.775
Bank 6	65	0.171	2.28	−14.2	41.1	−5.92	−1.42	2	−7.99

reported. While uniformly positive, the correlation coefficients for daily P&L are generally low, mostly below 0.2. The daily correlations are low even for the subset of observations August to 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 multiday horizons, these idiosyncrasies may be less important. For example, over five-day holding periods, the P&L cross-correlations approximately double (not shown).

The lower panel of Table III displays correlations for daily VaR across banks. The VaR correlations are as often negative as they are positive, and no clear pattern of comovement 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 five-day average VaRs. 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. We 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 $\bar{r}_{t+1}$ such that $pr(r_{t+1} < \bar{r}_{t+1}) = \alpha$ over the next trading day. Here $\alpha = 0.01$, so that the model predicts a lower bound on losses not to be exceeded with 99 percent confidence.
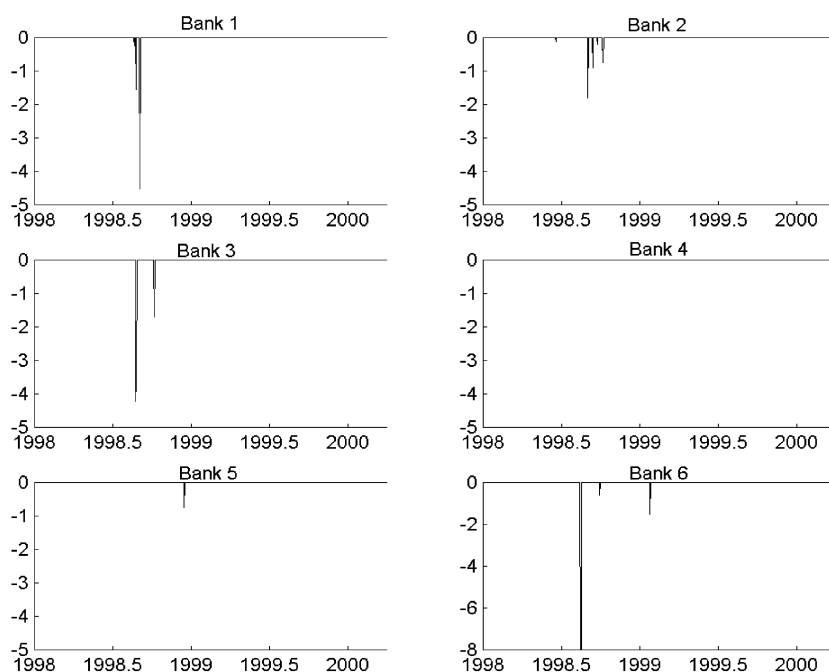
Figure 3. Violations of banks' 99 percent VaR. Plots show the daily P&L for those days on which P&L drops below the forecasted 99th percentile given the internal bank models. Data are expressed in standard deviations.

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 IV reports the actual rates at which violations occurred for the six banks. The average violation rate across banks is less than one-half of one percent. Column two reports likelihood ratio (LR) statistics for the null of a one percent violation rate. The p-values, shown in square brackets, are the probabilities of the likelihood ratio values exceeding the observed value under the one percent null.

These p-values indicate that one of the coverage rates is significantly different from one percent 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) and Berkowitz (2001)).

Table III

Correlations of Profit and Loss and VaR across Individual Banks

These panels report correlation coefficients for bank profit and loss and Value-at-Risk calculated with a matched sample of 482 daily observations; *t*-statistics are shown in parentheses.

	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5	Bank 6
Panel A: P&L Correlation Coefficients						
Bank 1	1.00					
Bank 2	0.434	1.00				
	(10.1)					
Bank 3	0.206	0.102	1.00			
	(4.81)	(2.39)				
Bank 4	0.164	0.085	0.358	1.00		
	(3.84)	(1.98)	(8.36)			
Bank 5	0.053	0.171	0.117	0.122	1.00	
	(1.29)	(3.99)	(2.73)	(2.84)		
Bank 6	0.154	0.165	0.197	0.108	0.108	1.00
	(3.60)	(3.85)	(4.59)	(2.52)	(2.53)	
Panel B: VaR Correlation Coefficients						
Bank 1	1.00					
Bank 2	−0.033	1.00				
	(−0.777)					
Bank 3	0.122	0.207	1.00			
	(−2.84)	(3.02)				
Bank 4	0.064	−0.779	−0.202	1.00		
	(1.48)	(−5.72)	(−4.72)			
Bank 5	−0.184	0.749	0.072	−0.742	1.00	
	(−4.30)	(−8.59)	(1.67)	(−17.4)		
Bank 6	−0.404	−0.229	−0.220	0.119	0.131	1.00
	(−9.45)	(8.64)	(−5.15)	(2.78)	(3.06)	

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

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

should be an i.i.d. 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 three of Table IV, 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 0.14.

<div align="center">

Table IV

Backtests of Bank VaR Models

</div>

This table shows alternative backtests of large-scale VaR models in operation at six commercial banks. Data are daily and span from January 1998 to March 2000. NA indicates the bank had no violations in the sample period. The *p*-values are displayed in square brackets. The symbols * and ** denote significance at the one and five percent levels, respectively.

	Violation Rate	Coverage	Conditional Coverage	Independence	Serial Correlation
Bank 1	0.005	1.54	1.57	0.0321	−0.00533
		[0.214]	[0.455]	[0.858]	[0.885]
Bank 2	0.010	0.00693	4.01	4.00*	0.158*
		[0.934]	[0.135]	[0.046]	[0.016]
Bank 3	0.005	1.70	8.81*	7.11**	0.330**
		[0.193]	[0.012]	[0.008]	[0.001]
Bank 4	0.000	NA	NA	NA	NA
Bank 5	0.001	8.92**	8.93*	0.00271	−0.00137
		[0.003]	[0.012]	[0.959]	[0.995]
Bank 6	0.005	1.71	1.74	0.0312	−0.00511
		[0.191]	[0.419]	[0.860]	[0.885]

NA: not applicable.

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}), that is, violation rate of α plus a test that violations are independent (LR_{ind}). Column four reports the results of LR tests for first-order serial dependence.[2] The *p*-values suggest that for two banks, given a violation on one day, there is a high probability of a violation the next day (higher than one percent). Similarly, the last column in Table IV 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.

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

---

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

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:

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

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

$$\sigma_t = \omega + \theta u_{t-1}^2 + \phi \sigma_{t-1}, \tag{2}$$

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 percent 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).

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. It cannot be used for sensitivity or scenario analysis. 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.

It is worth pointing out that by fitting the time series model to reported P&L, any systematic errors in the reported numbers are incorporated into the model. This would give the reduced-form model an advantage over the banks' models if the latter were not calibrated to reflect reported P&L. For example, if banks' smooth daily P&L, the reported numbers would have a tighter distribution than actual P&L. For present purposes, we simply accept the reported daily numbers.

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 percent 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.
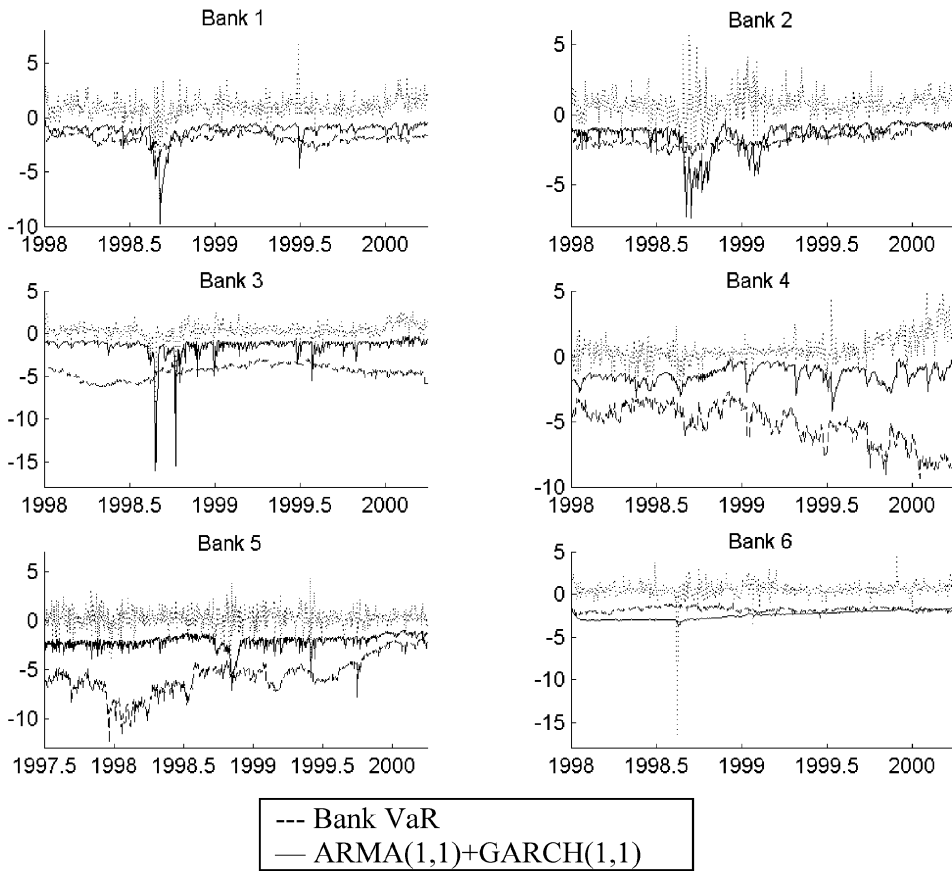
Figure 4. Daily profit and loss and a simple dynamic model. Time series of daily profit and loss data as reported by selected commercial banks for January 1998 through March 2000 (dotted lines) plotted with two model forecasts. The two models are an internal VaR model (dashed line) and a reduced form ARMA(1,1) with GARCH(1,1) Gaussian innovations. Both models are used to forecast the one-day-ahead 99th percentile of P&L.

Summary statistics and backtests for the GARCH model VaRs are presented in Table V. 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 time-series VaRs shown in column three are also lower than average bank VaRs, except for bank six, and the number of violations shown in column four averages out to about one percent. Thus, on average, the time-series VaRs achieve the targeted violation rate and a 99th percentile VaR coverage. The mean violation rate for the time-series VaRs also is lower than that of the banks' VaRs.

<div align="center">

Table V

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

</div>

This table displays alternative backtests of time-series value-at-risk forecasts. The *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 five percent critical value is 3.84, the 10 percent value is 2.71. The symbols * and ** denote significance at the one and five percent levels, respectively.

		Panel A: Summary Statistics			
	Obs	Box–Ljung Stat	Mean VaR	Number Violations	Mean Violation
Bank 1	569	0.284	−1.21	6	−0.829
Bank 2	581	0.106	−1.42	6	−0.362
Bank 3	585	1.11	−1.41	13	−1.12
Bank 4	573	0.356	−1.35	4	−0.315
Bank 5	746	2.89	−2.10	12	−0.772
Bank 6	586	0.003	−2.40	2	−7.21

		Panel B: Backtests			
	Violation Rate	Coverage	Conditional Coverage	Independence	Serial Correlation
Bank 1	0.011	0.018	3.97	3.96*	0.158*
		[0.894]	[0.137]	[0.047]	[0.016]
Bank 2	0.010	0.069	0.132	0.125	−0.010
		[0.934]	[0.936]	[0.723]	[0.436]
Bank 3	0.022	6.57**	11.4**	4.79*	0.134*
		[0.010]	[0.003]	[0.029]	[0.025]
Bank 4	0.007	0.584	0.640	0.056	−0.007
		[0.445]	[0.726]	[0.812]	[0.756]
Bank 5	0.016	2.37	4.16	1.787	0.068*
		[0.124]	[0.125]	[0.181]	[0.050]
Bank 6	0.003	3.43	3.45	0.014	−0.003
		[0.064]	[0.179]	[0.907]	[0.964]

While this last result would be expected simply because the bank VaRs are more conservative, more conservative VaRs also should produce smaller aggregate violations and maximum violations.[3] However, this is not the case. Aggregate violations (column four times column five) and maximum violations (see below) for the time-series 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

---

[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.

Bank and GARCH Model Comparisons, August 1998 to October 1998

The table compares Value-at-Risk forecasts as reported by large commercial 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 I.

		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	−0.981	−4.33
Bank 2	64	−2.28	5	−0.862	−2.46	−3.41	3	−0.410	−0.748
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	−0.775	−3.29	−2.97	4	−1.35	−2.87
Bank 6	65	−1.42	2	−7.99	−25.32	−2.93	1	−13.4	−13.4

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 time-series VaRs are able to deliver lower required capital levels without producing larger violations. As described below, this reflects the GARCH model VaR's greater responsiveness to changes in P&L volatility.

Formal backtests of the GARCH models are presented in the bottom panel of Table V.[4] The backtest results provide little basis to distinguish between the GARCH and bank VaR modeling approaches. In terms of coverage, one time-series VaR model is rejected at standard significance levels. Even though the time-series VaRs on average have a one percent violation rate and the bank models less than a one-half 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 time-series 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 to October 1998 period when P&L volatility rose substantially. Table VI compares model performances during this three-month period. Even though the GARCH model VaRs are smaller over the full sample, the bank and time-

---

[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 0.01. Average bank results for the backtests (coverage, dependence, etc.) were very similar to those for the full sample period.
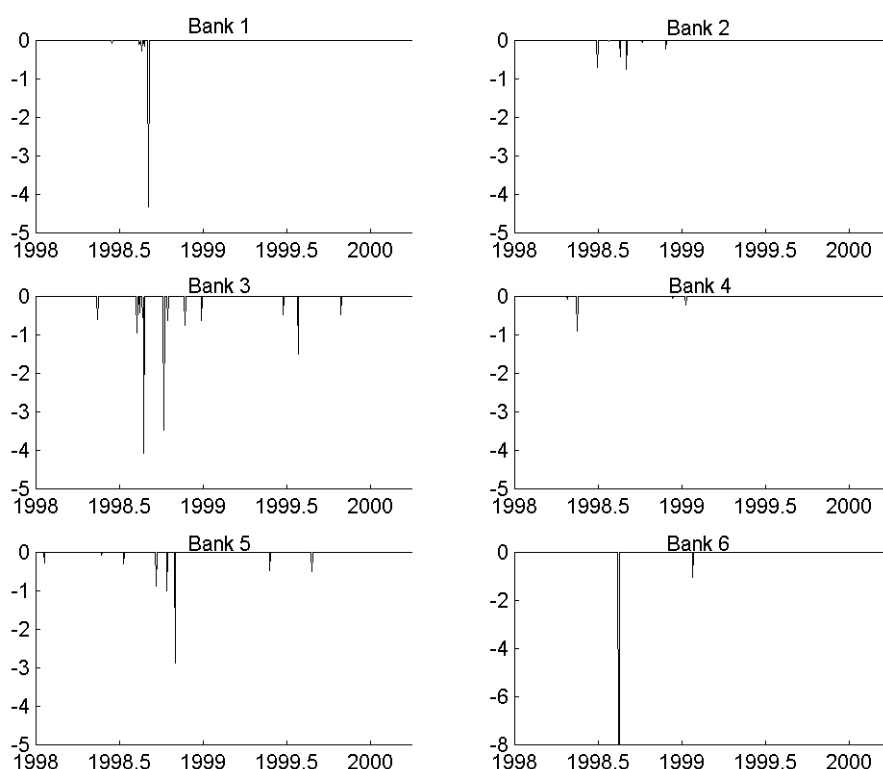
Figure 5. Violations of the 99 percent GARCH-based VaR. Plots show the daily P&L for those days on which P&L drops below the forecasted 99th percentile from the time-series models. Data are expressed in standard deviations.

series VaRs are comparable during this period. For this three-month period, the time-series VaRs increased from 80 to 250 percent over their average VaRs during the three months prior to August 1998 for four of the five banks with violations. The bank VaRs in comparison were 20 percent lower to 30 percent higher than their respective averages over the preceding three months. As a result, the performances of the bank and time-series VaRs are comparable in terms of average, aggregate, and maximum violations.

While these results show that the time-series 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 V, indicate that some clustering remains. Also, while the average violation rate for the GARCH VaRs is one 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 one through four, 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 an environment subject to regime shifts, which cannot be captured by the standard GARCH model (see Gray (1996)).[5]

III. Limitations of Bank VaR 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 outperform 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. We also 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 make many approximations, and parameters are often estimated only roughly. While this may appear to give representation to a wide range of potential risks, the various compromises tend to reduce any forecasting advantage.

The limitations of structural modeling extend to capturing time-varying volatility. Few if any of the structural-based models makes any systematic attempt to capture time variation in the variances and covariances of market risks. As for evaluating exposures to liquidity or other market crises, banks are mostly limited to performing stress exercises on their portfolios.[6] By reducing the risk factor to a univariate time series, our reduced-form model offers a more tractable approach to estimating P&L mean and volatility dynamics. While the reduced-form approach does not account for changes

---

[5] 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 to October 1998 period.

[6] We are aware of one bank that extends its VaR horizon for positions that are sizable relative to the market, thereby allowing for a possible slow liquidation. For more on this approach to modeling liquidity risk see Berkowitz (2000).

in portfolio composition, this limitation can be relaxed by estimating GARCH effects for historically simulated portfolio returns to current positions, rather than historically observed returns.[7]

Regarding the conservative bias in bank VaRs, several bank model features along with regulatory constraints may provide potential explanations. As indicated above, all banks exclude net fee income from VaR forecasts. When compared to actual P&L, this will give bank VaRs a conservative bias, since net fee income is believed to be a large part of average (positive) trading P&L. The actual bias cannot be determined, since banks do not separately report net fee income.

If we were to adjust reported P&L downward by, say, subtracting one half of the bank's average P&L from its reported P&L, we find that violation rates increase but that they are still conservative on average. Subtracting 100 percent of each bank's average P&L from reported P&L produces an average violation rate close to one percent.

In either case, these P&L adjustments would leave essentially unchanged the violation rates of the three banks with the most conservative VaRs. Moreover, this kind of level shift in the P&L has the effect of worsening the clustering phenomenon. The less conservative VaR results in more violations in the tumultuous August to October 1998 period. We would argue that a better approach would be to include forecasts of net fee income in the VaRs. The reduced-form VaR forecasts used here extrapolate P&L mean, as well as volatility, and thus implicitly include net fee income.[8]

A second practice that also may contribute to conservative bank VaRs is that VaRs may be estimated for subgroups of positions, such as for foreign exchange positions and interest rate positions. To obtain a VaR for the global portfolio, subgroup VaRs are simply summed. Since the subportfolio VaRs are each intended to be calibrated to a 99th percentile, the summation will overstate global 99th percentile VaR as it allows for no diversification or hedging affects among the subportfolios. Of the banks whose VaRs are among the most conservative, several make extensive use of the subportfolio addition procedure.

Certain regulatory standards may also contribute to the VaRs being both conservative and displaying limited response to changes in volatility. Regarding the latter, regulatory guidelines require that VaR estimates reflect market volatility over at least a one-year horizon, which precludes rapid

---

[7] 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).

[8] Another omission that could introduce systematic bias in the VaRs is 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.

adjustment to changes in current market volatility. Forecasts may be conservative in part because regulations require that banks whose global VaR is an aggregate of subportfolio VaRs must use the simple summation procedure. Further, the only formal regulatory test of bank VaRs is a one-sided "backtest"—a bank is deemed to have failed the backtest if there are more than four violations of VaR over the past 250 days. This can lead to a higher capital requirement and may provide an incentive for the bank to be conservative in its forecast.

To the extent that there are incentives for banks to be conservative, we expect them to be inversely related to management's confidence in its model. Indeed, we find that the VaRs of banks three, four, and five are much more conservative than those of banks one, two, and six. The former set of banks has less modeling experience and generally has less sophisticated models than the latter banks.

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 outperform forecasts based simply on an ARMA + 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. We also identify modeling practices and regulatory constraints that might harm VaR accuracy.

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 (2001) 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, Giovanni, Kostas Giannopoulos, and Les Vosper, 1999, VaR without correlations for portfolios of derivative securities, *Journal of Futures Markets* 19, 583–602.

Basak, Suleyman, and Alexander Shapiro, 2000, Value-at-Risk based risk management: Optimal policies and asset prices, *Review of Financial Studies* 14, 371–405.

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

Berkowitz, Jeremy, 2000, Breaking the silence, *Risk*, October, 105–108.

Berkowitz, Jeremy, 2001, Testing density forecasts with applications to risk management, *Journal of Business and Economic Statistics* 19, 465–474.

Christoffersen, Peter F., 1998, Evaluating interval forecasts, *International Economic Review* 39, 841–862.

Christoffersen, Peter F., and Francis X. Diebold, 2000, How relevant is volatility forecasting for financial risk management, *Review of Economics and Statistics* 82, 12–22.

Diebold, Francis X., Todd A. Gunther, and Anthony S. Tay, 1998, Evaluating density forecasts, *International Economic Review* 39, 863–883.

Engle, Robert F., and Simone Manganelli, 1999, CAViaR: Conditional autoregressive Value-at-Risk by regression quantiles, Manuscript, University of California, San Diego.

Gray, Stephen F., 1996, Modeling the conditional distribution of interest rates as a regime-switching process, *Journal of Financial Economics* 42, 27–62.

Hendricks, Darryl, 1996, Evaluation of Value-at-Risk models using historical data, *Economic Policy Review*, Federal Reserve Bank of New York, April, 39–69.

Jorion, Philippe, 2000, How informative are Value-at-Risk disclosures, Manuscript, University of California, Irvine.

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

Kupiec, Paul, 1995, Techniques for verifying the accuracy of risk measurement models, *Journal of Derivatives* 3, 73–84.

Lopez, Jose, and Christian Walter, 2000, Evaluating covariance matrix forecasts in a Value-at-Risk framework, Working Paper 00-21, Federal Reserve Bank of San Francisco.

Marshall, Chris, and Michael Siegel, 1997, Value-at-Risk: Implementing a risk measurement standard, *Journal of Derivatives* 4, 91–111.

Pritsker, Matthew, 1997, Evaluating Value-at-Risk methodologies: Accuracy versus computational time, *Journal of Financial Services Research* 12, 201–242.

Zangari, Peter, 1997, Streamlining the market risk measurement process, *RiskMetrics Monitor* 1, 29–35.