Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

An Evaluation of Bank VaR Measures for Market Risk During and Before the Financial Crisis

James O'Brien and Pawel J. Szerszen

2014-21

An Evaluation of Bank VaR Measures for Market Risk During and Before the Financial Crisis *

James O'Brien and Pawel J. Szerszen
Federal Reserve Board, Washington, DC, USA

This version: March 7, 2014

Abstract

We study the performance and behavior of Value at Risk (VaR) measures used by a number of large banks during and before the financial crisis. Alternative benchmark VaR measures, including GARCH-based measures, are also estimated directly from the banks' trading revenues and help to explain the bank VaR performance results. While highly conservative in the pre-crisis period, bank VaR exceedances were excessive and clustered in the crisis period. All benchmark VaRs were more accurate in the pre-crisis period with GARCH VaR measures the most accurate in the crisis period having lower exceedance rates with no exceedance clustering. Variance decompositions indicate a limited ability of the banks' VaR methodologies to adjust to the crisis-period market conditions. Despite their weaker performance, the bank VaRs exhibited greater predictive power for a measure of realized PnL volatility than benchmark VaR measures. Benchmark Expected Shortfall measures are also considered.

JEL classification: G01; G21; G28

Keywords: Market risk; VaR; Backtesting; Profit and loss; Financial crisis

*We thank Sean Campbell, Erik Heitfield and David Lynch for very helpful comments and suggestions. Excellent research assistance was provided by Daniel Kannell, Daniel Marts, and Nicole Abruzzo who also provided important added support. The analysis and conclusions set forth in this paper are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors.
Email: ⟨james.m.o'brien@frb.gov⟩, ⟨pawel.j.szerszen@frb.gov⟩.

1 Introduction

Value-at-Risk (VaR) is the potential loss on a financial position or portfolio with a designated horizon and probability of being exceeded. Its practical application began in the early 1990s when a number of large banks began employing VaR to measure market risk in their trading portfolios (see Mudge and Wee (1993)).

VaR has subsequently become the principal measure of market risk used by institutions for risk management, public reporting, and market risk regulatory capital requirements. Its widespread use owes to the attraction of a transparent measure of potential loss and banks' abilities to make it operational on a practical level. This includes the application of VaR to the entire portfolio, accounting for individual positions market exposures, and with regular updating to provide an ongoing measure of current risk.[1] During this time, however, VaR as a risk measure has been subject to criticism concerning both its estimation and usefulness.

Important estimation issues include determining the distribution of market price changes and variation in the distribution over time. For large banks, market factor exposures can exceed several thousand and cover a broad spectrum of market risks. Particularly because of its tractability for portfolios with high dimension market exposures, the most common measure used by banks has been a moving historical distribution of returns to current positions, "historical simulation" (HS). This methodology, however, has been strongly criticized.[2]

At the conceptual level, VaR's usefulness has also been argued to be limited, as it cannot indicate loss potential beyond the VaR quantile and can lack other desirable statistical properties. The principal alternative measure proposed is the expected loss conditioned on a VaR exceedance, Expected Shortfall (ES).[3]

Criticism of banks' VaR measures became vociferous during the financial crisis as the banks' risk measures appeared to give little forewarning of the loss potential and the high frequency and level of realized losses during the crisis period (see Nocera (2009)).[4] However, there has been little formal study of bank VaR crisis period performances.

In this paper, we study the behavior and performances of the VaR estimates of 5 large U.S. banks, the majority of which use standard HS VaR. The banks' daily VaRs and daily trading revenues are reported from the early 2000s through the financial crisis beginning in 2007 and extending through all major crisis events. The data are used internally and for regulatory capital purposes. The daily bank VaRs are estimated for a loss with a one percent chance of being exceeded, i.e., at a 1% quantile.

[1] For an extensive treatment on VaR measurement and practice, see Jorion (2007).

[2] For example, see Manganelli and Engle (2001), Pritsker (2006), and Andersen, Bollerslev, Christoffersen, and Diebold (2007)

[3] For a review and further study of VaR versus ES, see Yamai and Yoshiba (2005).

[4] Danielsson (2002) provides an early critique on VaR limits, especially in times of market instability. For VaR and ES performance during the financial crisis applied to market indices, see Kourouma, Dupre, Sanfilippo, and Taramasco (2011).

For the entire period of study, the VaR exceedance rates for all but one bank were not significantly different from the theoretical 1% quantile. However, a second important test is for independence or absence of clustering of VaR exceedances. For this test, for every bank, independence is strongly rejected. In fact, for all banks, almost all VaR exceedances occurred during the financial crisis.

In order to study the banks' VaR performances during the crisis period, the sample is split into pre-crisis and crisis periods, the latter beginning June 2007. The earlier period provides a benchmark, as well as allowing for comparison of results with prior literature. The banks' VaR performances are evaluated based on VaR accuracy (coverage) and independence among exceedances. Of additional interest is VaR correlation with a measure of 1-day ahead realized profit and loss (PnL) volatility.

In an early study Berkowitz and O'Brien (2002) found the daily VaRs of 6 large U. S. banks between 1998 and 2000 to be conservatively biased but with exceedance clustering during a sub-period of financial instability following the Russian debt default in August 1998. Further results suggested a simple GARCH-based VaR measure estimated from the banks' historical PnL was at least as accurate as the banks' internal VaRs. The broad results here are similar. Bank VaRs were overly conservative prior to the financial crisis but with excessive VaR exceedances and clustering during the crisis. A GARCH-based VaR estimated with bank historical PnL was more accurate for both periods and with no or limited evidence of exceedance clustering.

Perignon, Deng, and Wang (2008), Perignon and Smith (2010a) and Perignon and Smith (2010b) found bank VaRs to be conservatively biased with varying samples of banks and dates between 1999 and 2007(Q1).[5] Perignon and Smith (2010b) found no evidence of bank VaRs next-day PnL volatility forecast power. However, using U.S. bank VaR and PnL data reported quarterly for different periods between 1994 - 2002, Jorion (2002), Hirtle (2003), and Liu, Ryan, and Tan (2004) found significant correlations between the banks' VaRs and absolute next-quarter PnL. For a single bank between 2001 - 2004, Berkowitz, Christoffersen, and Pelletier (2011) reported significant correlations between business unit daily VaRs and next-day PnL volatility but also found evidence of business-line VaR inaccuracies. In this study, we also look at correlations between bank VaRs and a measure of next-day realized PnL volatility.

The analysis here is substantially expanded over earlier studies. This includes the use of benchmark VaRs estimated from the banks' historical PnL, specifically historical simulation (HS) VaR and VaR conditioned on a GARCH volatility forecast with normal and non-normal conditional return distribution variants. A substantial literature on VaR estimation applied

[5]Perignon, Deng, and Wang (2008) studied six major Canadian banks using daily data (1999 - 2005). Perignon and Smith (2010a) studied four U.S. banks using quarterly data (2001(Q4) - 2007(Q1)). Perignon and Smith (2010b) studied five major banks (Canadian, European, and one U.S) using daily data (2001 - 2004).

to returns associated with various market price series finds conditioning on a GARCH volatility forecast increases VaR accuracy, with further benefit from non-normal conditional return distributions.

Particular attention is also given to the timeliness of VaR adjustments to changing PnL volatility. Variance decomposition by frequency of bank and benchmark VaR measures' is especially helpful for this purpose.

As noted, bank VaRs were very conservative prior to the financial crisis, with few VaR exceedances. At least some of the conservativeness can be explained by reported PnL including net fee and related income, while VaR measures cover only position PnL. Despite their conservativeness, the pre-crisis bank VaRs showed significant power in forecasting next-day realized PnL volatility.

For the crisis-period, VaR exceedance rates for all banks were significantly above one percent and with significant clustering. However, there was a dichotomy in VaR performances, with three banks performing substantially better and improving over the period. These banks' VaRs also continued to exhibit significant PnL volatility forecast power. For the other two banks, VaRs increased very slowly over the entire period, despite continuing large PnL volatility and losses, and without significant PnL volatility forecast power.

The shift from bank VaR conservativeness in the pre-crisis period to a substantial understatement of risk during the crisis period is proximately explained by differences between the banks' PnL and VaR variation. The former is concentrated at higher frequencies, mostly within a one-quarter periodicity and increasing in the crisis-period. VaR variation, however, is concentrated at lower frequencies, mostly beyond one quarter periodicity for both periods. The continuance of low frequency variation accounts for the bank VaRs high exceedance rates and clustering in the early part of the crisis.

A gradual increase in banks' crisis-period VaRs accounts for high cross-bank VaR correlations, much higher than in the pre-crisis period. High cross-bank VaR correlation during a period of market instability, with perverse market feedback effects from the VaR constraint on trading, has been suggested as banks' similarly face the market turmoil.[6] However, the high correlation observed here reflects more the banks' common slow VaR adjustment and not immediate common VaR responses to market fluctuations.

For all benchmark VaR measures, pre-crisis period exceedance rates are mostly reasonably close to the 1% level, with little exceedance clustering. For the crisis period, the HS-benchmark VaRs experienced both excess exceedance rates and exceedance clustering. The GARCH-based benchmark VaRs experienced excess exceedance rates but fewer than those for the bank VaRs and benchmark HS VaRs and no exceedance clustering.

[6] See Basak and Shapiro (2001), Danielsson, Shin, and Zigrand (2004) and Danielsson, Shin, and Zigrand (2009). See also Adrian and Shin (2014) for bank VaR historical relation to market volatility and implications for economic stability.

The better accuracy of the benchmark VaRs in the pre-crisis period can be explained by smaller, on average, VaR levels much closer to the ex-post PnL 1% quantile. The crisis-period performance differences between the bank VaRs and benchmark VaRs, and between the GARCH-based VaRs and other VaR measures, including bank VaRs, can be explained by differences in VaR total variances and their frequency decompositions.

For the benchmark HS VaRs, both the pre-crisis and crisis period variation at higher frequencies is at a much lower level than those for the bank VaRs, even though the majority of banks use HS method to produce their VaRs. A possible explanation is that the bank VaRs' historical window is updated daily with current position exposures. The benchmark HS VaRs update daily only the newest and oldest PnL observations. This difference might also help explain the bank VaRs better PnL volatility forecast power.

The benchmark GARCH VaRs in the pre-crisis period had total variance similar to those for the bank VaRs but with moderately higher GARCH VaR variation at higher frequencies. However, for the crisis period, both the GARCH VaR total variance and its share at higher frequencies became much larger in response to the changing PnL volatility. This explains the GARCH VaR crisis-period better accuracy and exceedance independence. However, crisis-period GARCH volatility forecasts also over-reacted to large 1-day PnL shocks that adversely affected the GARCH VaRs' volatility forecast power relative to that for the better-performing banks' VaRs.

The results here indicate the limits of bank VaR measures in adapting to changing market conditions and particularly market instability. The benchmark GARCH VaR measure based on bank PnL has the modeling flexibility to account for variation in PnL volatility and may be improved upon. However, its use is limited in not being able to identify a bank's market risks at a disaggregated level. Additionally, the GARCH-based VaR measure still exhibited shortcomings in volatility forecasting during the financial crisis. These issues are considered further in the analysis below.

In a final exercise, ES (Expected Shortfall) measures were estimated with historical PnL using a GARCH-EVT model. The estimates were reasonably close to average losses given an exceedance during the pre-crisis period but lacked precision and understated mean losses conditioned on an exceedance in the crisis period. The latter results owe at least somewhat to GARCH VaR estimation limits during the crisis period, a time when ES is expected to be most informative. The results, however, would be consistent with concerns expressed by Yamai and Yoshiba (2005) about low ES precision for heavy-tailed distributions, particularly as ES requires a larger sample size than VaR to produce the same accuracy.

The rest of the paper is organized as follows. Section 2 provides a description of PnL and VaR data for the five banks studied here along with a brief description of bank VaR methodologies. Section 3 contains individual bank PnL and VaR analysis including an analysis of bank VaR temporal behavior. Benchmark VaR measures, including HS VaR, GARCH-normal and alternative non-normal GARCH VaRs are analyzed and presented in Section 4, along with implications of the results. Section 5 concludes.

2 Bank Data and VaR Estimation Practices

The bank data used here consists of individual bank historical daily trading revenues (PnL) and bank 1% VaRs provided to the Federal Reserve Board from five large U.S. banking institutions. The daily VaRs are used by banks in managing trading risk and for market risk regulatory capital purposes. Bank data starting dates differ somewhat but begin in the early 2000s for all but one bank and extend through all major market events. The starting date for the crisis period is June 2007, approximating the initial collapse of the mortgage securitization market and several mortgage financing firms. It also marks the beginning of an extensive period of much greater bank PnL volatility that persisted through the financial crisis. The pre-crisis period is based on the earliest dates for which bank data was available.[7]

Banks estimate 1-day-ahead VaR forecasts using daily market factor histories, with VaR updated daily and based on current (end-of-day) positions. Market exposures are estimated for market factors by geographical region, industry and factor type, where individual-name exposures may be represented by an aggregated market index. Market exposures included in bank VaRs can be several thousand or more. To estimate VaR, current exposures are applied to historical or analytically simulated market factor distributions based on historical windows ranging from 1 to 5 years and an average window of approximately 3 years. Updating frequency of market factor histories ranges from one day to one quarter, with an average lag of about one month and possibly some shortening of the lag during the crisis period. Also, bank VaRs typically measure risk only for mark-to-market position revaluations, while daily PnL used by banks and regulators also includes fees, commissions and interest income. On net, these additions to position PnL are typically positive, lowering reported VaR exceedance rates.[8]

VaR measures the minimum return on a portfolio with a designated confidence level. If the confidence level is $(1-q)$, the minimum return is the q-quantile value for the portfolio return distribution. Formally, let $VaR^q_{t,t+k}$ be the forecast on day t for a minimum k-period portfolio return $r_{t,t+k}$ on day $t+k, k \in \mathbb{N}$, satisfying the designated confidence level, $1-q$. If $F_{t,t+k}(\cdot)$ represents the *true* portfolio return distribution function for date $t+k$ conditional

[7]We do not report exact starting dates for the pre-crisis period and end dates for the crisis period because of confidentiality purposes.

[8]Under new Basel market risk rules effective in 2013, subjected banks must exclude such trading related income from trading revenues used for validating VaR measures.

on time t information set Ω_t, VaR is:

$$VaR^q_{t,t+k} \quad = \quad F^{-1}_{t,t+k}(q) \tag{1}$$

Since the value $F_{t,t+k}(q)$ of the *true* portfolio return distribution function is not known, it must be estimated, $\widehat{F}_{t,t+k}(q)$. For a given bank both bank VaRs and all benchmark VaRs differ in the prescription of $\widehat{F}_{t,t+k}(q)$ either by the set of available information at time t, i.e. banks know their positions and position sensitivities, or by the estimation method, e.g. parametric, nonparametric, etc. Bank VaRs and all benchmark VaRs considered here are measured at $q = 1\%$ with $k = 1$, i.e., a 1-day forecast horizon. In the sequel we simplify notation by denoting Ω_t- measurable variables with one index only, e.g. $r_{t-1,t} \equiv r_t$ and $VaR^q_{t,t+1} \equiv VaR^q_t$.

Banks calculate daily VaR using the history of market factors and factor exposures, i.e. positions. Formally, let x^b_t represent bank b's positions and f_t market factors at time t. Bank VaR measures are conditioned on the historical market factors and the bank's *current* market exposures, since both are available to banks. Letting L be the fixed length of historical data used to compute VaR, $\widehat{F}^{-1}_{t,t+k}(q) = \widehat{F}^{-1}_{t,t+k}(q, x^b_t, \{f_\tau\}_{\tau=t-L+1:t})$ is the bank's risk measure for time $t+1$ that depends on the *past* market factors and the bank's *current* positions at time t. In contrast, benchmark VaR measures used here are based only on *past* PnL realizations that, in turn, depend on respective *past* positions, since the history of positions is not directly observed.

Of the five banks studied here, three use the standard historical simulation (HS) method. Bank HS VaR is determined from position revaluations using historical market factors $\{f_\tau\}_{\tau=t-L+1:t}$ for a predefined historical window of size L and applied to *current* positions x^b_t producing a series of *pseudo*-returns $\{\hat{r}^b_\tau(x^b_t, f_\tau)\}_{\tau=t-L+1:t}$. Thus, bank HS VaR, $VaR^{b,q}_t$, is the q-quantile of $L-$ length history of *pseudo*-returns, $\hat{r}^b_{t-\tau_q+1}$, where $\tau_q \in \{1, \ldots, L\}$. Hence, $VaR^{b,q}_t \equiv \hat{r}^b_{t-\tau_q+1}(x^b_t, f_{t-\tau_q+1})$ depends on the historical market factors at time $t-\tau_q+1$, the bank's *current* positions at time t, and bank's market factor sensitivities. As noted, one bank applies some scaling to HS that makes VaR more sensitive to more recent market volatility. Another bank estimates an analytical factor distribution, using numerical simulation to obtain a portfolio distribution and 1% VaR.

3 Bank PnL and VaR Performance

3.1 Individual Bank Analysis

Descriptive statistics for pre-crisis and crisis period PnL and bank VaR for banks 1 through 4 are presented in Tables 1 and 3. These statistics are based on daily observations that start two years into the historical data, which is the starting point for the estimated benchmark VaRs to be considered below. Results for the complete pre-crisis sample are similar and presented in Tables 18 and 19 (with results for a more limited sample for bank 5). The historical data for each bank are standardized so that reported values do not represent actual dollar magnitudes.[9]

For bank PnL, several features in Table 1 are notable. One is the much higher pre-crisis period mean PnL, with comparatively low PnL standard deviations and loss rates compared to those for the crisis period. Also notable is the dramatically higher (absolute) 1% PnL quantile for the crisis period. Moreover, PnL was positively skewed in the pre-crisis and negatively skewed in the crisis period. For banks 2 and 4, the more extreme crisis-period statistics reflect a number of extreme 1-day losses.

For reference below, the variance decomposition by frequencies of daily PnL volatility is reported in Table 2. Frequency bands correspond to periodicities less than 5 days (weekly), less than 21 days (monthly), and less than 63 days (quarterly). The variance decomposition of |PnL|, a realized measure of PnL volatility, was performed using the Band-Pass Filter methodology of Baxter and King (1999) with $K = 30$ leads and lags.[10] The fraction of the total variance of |PnL| accounted for by each frequency range is also presented.

As shown in Table 2, the total variances of each bank's realized daily PnL volatility increased dramatically in the crisis period. Including both periods, 40 to 60 percent of the |PnL| volatility occurred within a 5-day periodicity, with approximately 60 percent reflecting extreme crisis period losses by banks 2 and 4. Most of the banks' daily PnL volatility is accounted for below the 63-day periodicity, 75 percent or more, and this fraction increased noticeably during the crisis period.

Pre-crisis and crisis period descriptive and test statistics for bank VaRs are provided in Table 3. The standardization used for the individual banks' PnL is also applied to the banks' historical VaR data so that the reported values do not represent dollar magnitudes. Consider first the pre-crisis period. For this period, exceedance rates are very low, close to zero for most banks. Coverage level p-values indicate a strong rejection of VaR unbiasedness. The conservativeness of the bank VaRs is apparent when compared to pre-crisis 1% PnL quantiles in Table 1, with banks' mean VaRs roughly twice those of the respective 1% PnL quantiles. The near absence of VaR exceedances also limits any testing of exceedance independence.[11] Following the Mincer and Zarnowitz (1969) method, we estimate the correlation between the 1-day VaR forecast and a range-based volatility measure |PnL| that is used to quantify bank VaR informativeness on near-term PnL volatility. The pre-crisis period correlations

[9]This first-stage standardization involves division by the standard deviation of banks' PnL calculated for the total sample including pre-crisis and crisis periods.

[10]We found that $K = 30$ leads and lags is sufficient to limit "leakage" and that the choice of higher feasible K does not have a significant effect on our results.

[11]As commented by Perignon, Deng, and Wang (2008), no VaR exceedances could support invalidity of independence.

in the last column are all positive, with three significant at 1-percent level and a fourth significant at 10-percent level.[12]

For the crisis period, there is a dichotomy in the performances of banks 1, 3, and 5 versus banks 2 and 4 shown in the bottom of Table 3.[13] First, consider banks 1, 3, and 5. For these banks, while VaRs averaged more than twice pre-crisis levels, exceedance rates were 2.7 to 4 percent, significantly in excess of the 1-percentile. The inadequacy of the VaR coverage is also evident in the three banks' mean VaRs ranging between 40 and 70 percent of their respective 1-percent PnL quantiles (Table 1). The under-estimation occurs despite the VaRs ignoring fee and related income, which provides a conservative bias. For independence, one test is limited to a first-order Markov process (Christoffersen (1998)). A second test is without this limitation. It is based on comparing observed time intervals (durations) between VaR exceedances and the expected duration conditioned on the VaR quantile (Christoffersen and Pelletier (2004)). More details on these tests are provided in the Technical Appendix.[14] For banks 1 and 5 both tests reject independence at standard test levels, with rejection notably weaker for bank 3. Despite excess exceedances, the crisis period VaRs for banks 1, 3, and 5 continue to be significantly correlated with next-day |PnL|.

For banks 2 and 4 exceedance rates were 10 and 5 percent respectively. The mean VaRs for these two respective banks did not exceed 30 percent of their respective 1-percent PnL quantiles. These banks also have much larger mean exceedances and the VaR - |PnL| bivariate correlations are not significant at standard test levels. The bank VaRs adjusted extremely slowly to the high crisis period PnL volatility and the banks also reported a number of very large losses.

To further interpret bank VaR behavior and performance, temporal behavior of the bank VaRs is considered. Specifically, VaR variance frequency decompositions are presented in Table 4. In general, better-performing VaRs would be expected to have better alignment with bank |PnL| variance decomposition among different periodicity ranges. However, in contrast to the bank |PnL| variance decompositions (Table 2), most of the banks' VaR variances are accounted for at periodicities exceeding 63-days with very little within 5 days. Further, while |PnL| variance contributions at periodicities within 63 days increased measurably during the crisis period, banks' VaR variance contributions remained mainly at lower frequencies.

Figures 1 and 2 show pre-crisis and crisis period scatter plots for bank daily VaR and

[12]The banks' full sample results shown in Tables 18 and 19 contain an extra two years and show similar correlations.

[13]We note that two banks in the first group and one bank in the second used standard HS VaR.

[14]See Kupiec (1995) for tests of the power of unconditional coverage. Berkowitz, Christoffersen, and Pelletier (2011) evaluate the size and power properties of the tests used here and other tests using desk-level bank PnL and VaR measures. Also, see Campbell (2006) for a review of VaR backtesting procedures.

next-day PnL combinations.[15] In the plots, the daily VaRs are aligned with next-day PnL. The 45-degree solid blue line demarcates VaR exceedances, where the "large dots" and ellipsoids below the line indicate respectively individual PnL exceedances or groups of PnL exceedances.[16] The scatter plot can indicate a relation between a bank's VaR and the level or dispersion of its PnL. Also, the distribution of PnL conditional on VaR violation can indicate if exceedances are related to the level of the bank's VaR.

For the pre-crisis period, Figure 1 shows the strong tendency for PnL to be positive for all banks, which can be partly explained by banks including net fee and interest and related income in the PnL measure. Except for bank 1, a second common feature is more widely scattered PnL at higher (in absolute) VaR levels.[17] This would be consistent with the correlations between banks' VaR and next-day |PnL| reported in Table 3 above.[18]

For the crisis period, Figure 2 shows that VaRs for the three better-performing banks 1, 3 and 5 cover a much wider range, though with comparatively few VaR and PnL pairs at the larger VaR values. As with the pre-crisis period, the dispersion of PnL tends to be greater at the higher VaR levels. However, most exceedances occur at the smaller VaR levels. This, coupled with excess exceedance rates, gives some indication of inadequate VaR adjustment to the crisis period conditions. For banks 2 and 4, VaRs remained in a narrow range with substantial excess exceedance rates and with some large losses. This behavior is closely tied to the banks' crisis-period VaR dynamics, which is described in more detail in Section 3.2.

3.2 Bank VaR Temporal Behavior

Historical PnL and VaR averaged across banks is first considered. For each period, both daily bank VaR and PnL are first standardized by the the individual bank's PnL mean and standard deviation for the respective period. This is to normalize the scale of the banks' trading activity so as to give VaR behavior for the different banks equal weight.[19]

In Figure 3, the average bank historical PnL and VaR are shown for pre-crisis and crisis periods. Such aggregates help to reveal the temporal behavior of PnL and VaRs while providing confidentiality of individual banks. For both periods, the dominance of high frequency variation in PnL realized volatility, |PnL|, and low frequency variation in VaR,

[15] Several outlier points are not shown in the figures to hide bank identities.

[16] For the crisis period we split PnL exceedances in several subgroups, with varying number of points, rather than show individual VaR violations to further conceal banks' identities. The plotted ellipsoids are the smallest containing all points within each subgroup and provide information about the constituents' dispersion.

[17] In the sequel we make references to absolute VaR levels.

[18] The VaR for bank 2 appears to have the same value on multiple days. This reflects a substantial period where VaR was in a very narrow range and small daily variation is not clearly visible in the Figure.

[19] This standardization ensures that the number and the timing of VaR exceedences is unchanged for each bank after standardization.

reported for individual banks in Tables 2 and 4, can clearly be seen in the aggregates.

For the pre-crisis period, average bank VaRs also show a gradual increase (in absolute value) in the second half of the period, following an increase in bank PnL volatility. This tendency of higher VaRs to be associated with greater PnL dispersion was seen in the scatter plots (Figure 1). Figure 3 further indicates that significant and positive correlations between VaR and PnL realized volatility presented in Table 3 reflect a gradual increase in pre-crisis PnL volatility accompanied by VaR increases. Of particular note is the highly conservative average bank VaR. While this reflects overly conservative individual bank VaRs documented above (Table 3), the averaging of bank VaRs somewhat exaggerates the conservativeness relative to that of the individual banks' VaR.[20]

For the crisis period, average bank PnL in Figure 3 continues to have a very large share of high frequency variation but with a much higher variance, which is also seen in Table 2 for individual banks. The high PnL volatility starts early and continues over the crisis period. Despite this, the average bank VaR increases at a slow rate and is highly persistent, until late 2008. This pattern reflects the five individual banks, although the VaR rate of increase was substantially higher for the three better-performing banks that resulted in a dichotomy for banks' VaR and |PnL| correlations. As seen in Table 3, for the two banks with large crisis-period loses banks' VaR and |PnL| correlations are not significant, while for the three better performing banks the correlations are significant and positive.

An additional feature of the crisis period is very high cross-bank VaR correlations shown in Table 5. Earlier studies have argued that, in periods of broad market instability, bank VaRs used for market risk management and regulatory capital may move strongly together, exaggerating market instability through a feedback effect of risk management on market conditions. These analyses assume bank VaRs are changing with *current* market conditions that are having a common effect on bank market returns.[21] However, the results here indicate that the high cross-bank VaR correlations reflect a common but very slow adjustment–a matter of months–to the more volatile market conditions and bank PnL. It is not clear that this pattern of bank VaR adjustment can produce the strong feedback effects on market conditions predicted in these analyses.[22]

Relating average bank PnL and VaR time series to bank VaR behavior and performance is limited without a temporal matching between an individual bank's VaR and its PnL. To satisfy limitations on the use of historical graphics of individual bank data, while providing more information on bank VaR temporal behavior and performance, an alternative bank

[20]Bank VaRs and benchmark VaRs are not additive, since bank PnLs do not satisfy comonotonicity property. As seen in Figure 3, benchmark GARCH VaRs that are studied in Section 4.2.1 and found to have a correct coverage, have a smaller average level than bank VaRs in the pre-crisis period.

[21]For example, see Basak and Shapiro (2001), Danielsson, Shin, and Zigrand (2004) and Danielsson, Shin, and Zigrand (2009).

[22]Adrian and Shin (2014) show that large financial institution VaR has a strong but lagged relation to implied market volatility, which they argue creates pro-cyclical bank risk/capital management.

data decomposition that matches bank's daily VaR with its PnL is used and presented in Figure 4 and Table 6.

On each day, the standardized banks' PnL and VaR are sorted into bin categories based on the relative size of the respective standardized banks' VaRs for that day.[23] The number of bins is set equal to the number of banks in each subperiod: 4 for the pre-crisis and 5 for the crisis period. A lower VaR conservativeness is associated with a higher bin number. On any day, the VaR and PnL for a particular size category will be for the same bank but the banks in a category will vary over time. There is significant mixing over time between the all four bank names during the pre-crisis period for all bins, and during the crisis period for the better performing banks 1, 3 and 5 in bins 1, 2 and 3. There is almost no mixing for bins 4 and 5 in the crisis period, with both bins comprised solely from the outlier banks 2 and 4. To further conceal identities, the sum of VaRs and PnLs for bins 4 and 5 are plotted rather than individual plots. Since both are outlier banks with similar behavior, this does not materially diminish the individual bank analysis.

Bank VaR and PnL order categories for the pre-crisis period are provided for all 4 bins, and for the crisis period for bins 1, 2, 3 and the aggregation of bins 4 and 5. This ordering of daily VaRs permits calculation of descriptive and test statistics at an individual bank level and with exceedances across bins aggregating to total bank exceedances. It allows for some further identification of the relation between bank VaR temporal behavior and performance.

Figure 4 shows the pre-crisis and crisis period order-category VaR and PnL temporal behavior and Table 6 provides ordered-category PnL and VaR statistics. For both pre-crisis and crisis periods, a similarity in VaR and PnL temporal behavior can be seen across order categories for the respective periods and that is also similar to that for the average bank VaRs in Figure 3. Thus, the temporal behavior described above for the average bank generally carries over to the individual banks. Further, because individual banks' VaR and PnL are paired, total VaR exceedances seen across order categories equal the bank VaRs total exceedances. In this respect, the order category VaRs are indicative of bank VaR performances over time.

Order category descriptive and performance statistics are presented in Table 6. A notable systematic difference between order categories within each period is that the higher the order category, the smaller is PnL variation, which holds for all bins with the exception of bin 'four + five' in the crisis period for which the PnL variation is the highest. Appropriately, with the exception of banks 2 and 4 in the crisis period, higher bank VaR levels (lower-numbered bins) are associated with higher PnL volatility.

[23]Our standardization procedure, similar to applied for bank aggregation, involves both demeaning and standardization by PnL first two sample moments. This allows bringing all banks data to the same level of magnitude and size.

In terms of performance statistics, the order category results are comparable to those reported for the individual banks in Table 3. Specifically, bank VaRs at every category level are overly conservative in the pre-crisis period and are not conservative enough in the crisis period. For the pre-crisis period, bank VaR correlation with next-day |PnL| tends to be more homogenous across bins than that for individual banks reported in Table 3. Moreover, bank VaR forecast power for next-day |PnL| increases with decrease in bank VaR conservativeness. Since VaR size and PnL volatility are positively related, this feature may reflect more difficulty in forecasting higher PnL volatility. Moreover, the homogeneity across bank VaR - next-day |PnL| correlations may owe to variation among individual banks within the order categories. For the crisis period, the bank VaR forecast power for next-day |PnL| behaves in a similar fashion as found for the pre-crisis period and increases with decrease in bank VaR conservativeness across all three bins that are composed of better performing bank VaRs but with more heterogeneity due to somewhat smaller variation among individual banks within the order categories. All correlations between bank VaR and next-day |PnL| are also found to be positive and significant for these bins.

For the aggregate bin 'four +five' with the outlier banks 2 and 4, the bank VaR correlation with the next-day realized PnL volatility is found to be statistically insignificant. Moreover, even though across all bins bank VaRs in the crisis period are not conservative enough, the extent of bias is the highest for the aggregate bin with the outlier banks. In sum, these two banks' response to the financial crisis was extremely weak in terms of both VaR level and a lack of timeliness in increasing VaR.

4 Benchmark VaR Measures

Alternative benchmark VaRs for each bank are estimated using the bank's respective historical daily PnL. Using bank daily PnL permits a more direct evaluation and interpretation of the banks' VaRs and their respective performances.

This use of historical PnL makes the benchmark VaRs dependent on historical bank positions and their market factor sensitivities. The bank VaRs have the important advantage of being conditioned on current positions. However, errors or other deficiencies in measuring current exposures will reduce the accuracy of the bank VaR estimations, whereas the benchmark VaRs will reflect realized but dated position market factor sensitivities. Bank VaR accuracy may also be reduced by failure to make timely updates in the historical market data.

In this study, daily 1% benchmark VaRs are estimated using a 2-year daily rolling history of the respective banks' daily PnL. The VaRs are based on 1-day out-of-sample forecasts.

4.1 Standard HS VaR

Standard HS is the most popular approach to bank VaR measurement and is used by three banks studied here. Benchmark HS VaR calculated at date t for bank b and for date $t+1$, $HSVaR_t^{b,q}$, is the q-quantile of $L-$ length history of daily returns $\{r_\tau^b\}_{\tau=t-L+1:t}$. Hence HS VaR is given by $HSVaR_t^{b,q} \equiv r^b_{t-\tau_q+1}(x^b_{t-\tau_q+1}, f_{t-\tau_q+1})$, where $\tau_q \in \{1, \ldots, L\}$, to explicitly denote PnL functional dependence on positions and market factors at time $t - \tau_q + 1$. Note that bank VaR methodologies differ from benchmark HS VaR in how the historical market factors are used in the VaR measurement, since the latter does not condition on bank's *current* positions.

Statistical results for HS VaRs for each bank are reported in Table 7. For the pre-crisis period, the HS VaR means are only a third to a half those for the bank VaRs. They are much closer to the 1% PnL quantiles, with an average VaR exceedance rate of .014. The null hypothesis of 1% coverage would be rejected only for bank 3. Independence and duration tests show little evidence of exceedance clustering. Correlations between the HS VaRs and next-day |PnL| are significant and positive for three banks.

For the crisis period, the HS VaR exceedance rates are much higher and the null hypothesis of 1% coverage is strongly rejected for all banks. The exceedance rates average about the same as the bank VaRs for banks 1, 3, and 5 (those with the better-performing VaRs) and are lower than those for banks 2 and 4. The maximum HS VaR exceedance levels are similar to those for the bank VaRs. As with the bank VaRs, HS VaRs also exhibit significant exceedance clustering at 5% test level. For banks 1, 3, and 5 the HS VaR - |PnL| correlations are considerably lower than those for bank VaRs and, at standard test levels, are significant only for one bank. For banks 2 and 4, however, the correlations are small and insignificant for bank VaRs and benchmark HS VaRs.

The HS VaR variance decomposition by frequency is reported in Table 8. As indicated, only a very small fraction of the HS VaR variation is attributed to periodicities within 63-days, which indicates the temporal rigidity of the benchmark HS VaRs. This greatly limits any timely adjustment of HS VaRs to changes in PnL (or market) volatility, especially in the crisis-period, and can explain the relatively weak predictive power for next-day |PnL| as shown in the last column in Table 7. The very low HS VaR variance contribution below one-quarter periodicity is consistent with the crisis-period exceedance clustering.

In Figures 5 and 6, pre-crisis and crisis period scatter plots of bank PnL and benchmark HS VaRs are presented. For both pre-crisis and crisis periods, the HS VaR scatter plots differ from the bank VaR scatter plots in two ways. One is that the smallest HS VaRs tend to start at noticeably lower levels than the bank VaRs. The second and more notable difference is that the HS VaR changes are relatively infrequent leading to clusters of PnL values associated with individual HS VaRs.

The first difference can be explained by reported bank PnL including fee, interest and related income from trading activity that is normally positive. This ancillary income is not included in the bank VaRs that measure only the potential loss on market positions. However, it is implicitly represented in the benchmark VaR measures, as these are estimated using reported bank PnL. Consequently, the benchmark VaR measures will tend to be smaller than the bank VaRs, with the latter having a conservative bias, ceteris paribus.

The second difference reflects temporal rigidity in the benchmark HS VaRs with a 2-year estimation window. The benchmark HS VaRs change at highly discrete intervals that leads to consecutive days of PnL aligning with the individual VaRs.

Figure 3 plots the time series PnL and benchmark HS VaR for the average bank for both the pre-crisis and crisis periods along with the bank VaR. These figures show the temporal rigidity of the HS VaR with weak responsiveness to changes in PnL volatility. For both periods, the average benchmark HS VaR and the average bank VaR show a similar trend pattern in each period. However, at higher frequencies, the average bank VaR shows considerably more variation than the average HS VaR.

Three of the five banks studied use the standard HS VaR methodology, including two of the crisis-period better-performing banks. If bank exposures were constant and measured accurately (i.e., consistent with historical PnL), the bank and benchmark HS VaRs should coincide for a given historical window. The much greater variation in the bank VaRs, particularly noticeable in the scatter plots, is thus consistent with the bank VaRs greater variability and better |PnL| forecast power, reflecting the banks conditioning on current position exposures.

Benchmark HS VaRs were also estimated with a 1-year historical window. The statistical results and HS VaR variances at different frequency ranges are reported in Tables 20 and 21, and PnL - VaR scatter plots in Figures 11 and 12. Most of the 1-year HS VaR variances at considered periodicity intervals (from $\leq$ 5-day to $\leq$ 63-day) are considerably higher than those for the 2-year HS VaRs. However, relative to those for the bank VaRs, they are all still extremely low. The pattern of the 1-year HS PnL - HS VaR scatter plots are similar to those with the 2-year window but with moderately larger ranges for most banks. The 1-year benchmark HS VaRs had modestly more accurate coverage in both periods than the 2-year HS VaR but no improvement in exceedance independence or correlation with realized PnL volatility.

4.2 VaR Measures with Time-varying Volatility

Alternatives to HS VaR considered here employ various degrees of parametric modeling to better account for time varying return volatility and alternative conditional distributions.

For the measures considered, the daily return r_t (PnL) can be expressed as:

$$\begin{aligned} r_t &= a_0 + a_1 r_{t-1} + \varepsilon_t && (2) \\ \sigma_t^2 &= b_0 + b_1 \sigma_{t-1}^2 + b_2 \varepsilon_{t-1}^2 && (3) \\ \varepsilon_t &\equiv \sigma_t z_t && (4) \end{aligned}$$

where z_t is a mean zero stochastic process with alternative specifications. Volatility in equation (3) is estimated as GARCH(1,1) specification and the expected return (PnL) in equation (2) as AR(1) process.[24] The VaR measures here are estimated with a 2-year rolling history.

Under the standard GARCH approach, the residual z_t distribution is normal. Results using the GARCH-normal process are first presented, followed by three alternative GARCH-based conditional return specifications. One is Filtered Historical Simulation (FHS), a method applicable to an HS VaR framework.[25] The second and third GARCH alternatives are parametric: GARCH-t and GARCH-EVT (Extreme Value Theory) based on the Generalized Pareto Distribution as in McNeil (1999) and McNeil and Frey (2000). For all models the PnL VaR is computed at .01 level.

4.2.1 GARCH-Normal VaR Measure

Results for the GARCH-normal VaRs are presented in Table 9. For the pre-crisis period, the average exceedance rate across the four banks is .009 and the null hypothesis of .01 cannot be rejected for any bank. The exceedance rates of GARCH-normal VaRs are closer to theoretical 1% level than the benchmark HS VaRs for all but bank 4. The GARCH-normal VaR independence and duration statistics indicate no significant clustering. However, correlations with next-day |PnL| are, in general, somewhat weaker than those for the banks' VaRs.

For the crisis-period and for the better-performing banks 1, 3, and 5, the GARCH-normal VaR exceedance rates are excessive but average slightly lower than those for the banks' VaRs. The size of the VaRs average about two-thirds of the banks' 1-percent PnL quantiles. Exceedance average and maximum sizes are similar to those for the bank VaRs. Notably, unlike the bank (and HS) VaRs, the GARCH-normal VaRs show little evidence of exceedance clustering. However, their correlations with next-day |PnL| are slightly lower than those for the banks' VaRs for all but one bank. For banks 2 and 4, the GARCH-normal VaRs exhibit coverage performance comparable to that for banks 1, 3, and 5 and

[24] ARMA(1,1)-GARCH(1,1) specification was also estimated with no significant differences in findings with AR(1)-GARCH(1,1) specification considered here. The AR model was free of convergence issues when applied to more general, fat tailed, residual specifications. This property is important because of the large number of rolling-window estimations performed for all sample days and for each bank.

[25] See Barone-Adesi, Giannopoulos, and Vosper (1999). Also, see Pritsker (2006).

much superior to the bank VaRs. However, the GARCH-normal VaR - |PnL| correlations for banks 2 and 4 are close to zero as are the bank VaR correlations.

The GARCH-normal VaRs' variances at different frequencies shown in Table 10 help in interpreting these results. In contrast to the bank VaRs, GARCH-normal VaR variation at higher frequencies shows large increases in the crisis period, except for bank 4, and also much larger increases in total variance. Moreover, the GARCH-normal VaR variation at all considered frequency ranges is much higher in the crisis period than that for the bank VaRs and closer to the values for the observed bank |PnL|. The ability of the GARCH-normal VaRs to adjust quickly to the changes in |PnL| can account for the smaller exceedance rates and the absence of clustering of VaR exceedances during the crisis period.

These comparisons between bank VaR and benchmark GARCH-normal VaR variance decompositions indicate that the GARCH-based specification allowed for the increased share of high frequency VaR variation and therefore it was much more responsive to the increase in the crisis-period PnL volatility. This explains the better coverage and independence test results. Surprisingly, GARCH-normal VaR - |PnL| correlations were lower for most analyzed banks than those for the bank VaRs. An explanation is provided in Table 11. The first two columns report correlations between the GARCH volatility forecast, $\hat{\sigma}_{t,t+1}$, and the realized volatility, $|PnL_{t+1}|$, and their respective p-values. These correlations are roughly in line with the VaR - |PnL| correlations in Table 9. The next two columns are the correlations between the estimated volatility, $\hat{\sigma}_{t+1}$, and realized volatility forecast error, $|PnL_{t+1}| - \hat{\sigma}_{t+1}$. The negative correlations, which are much larger in the crisis period, indicate overshooting of the GARCH volatility forecasts.

Figures 7 and 8 show pre-crisis and crisis period scatter plots for bank PnL and benchmark GARCH-normal VaRs. As with HS VaRs, the range of both the pre-crisis and crisis-period GARCH-normal VaRs begins at smaller levels than those for the bank VaRs. Except for this, the HS VaR and GARCH-normal scatter plots for the two periods are very different, with the latter having a much wider range and lacking the rigidity reflected in the HS VaR scatter plots. GARCH-normal VaR exceedances are also fewer and spread over a wider range of VaRs for most banks in the crisis period.

There are also noticeable differences between the bank VaRs and GARCH-normal VaRs. For both periods, the GARCH-normal VaRs display a wider range of variation. The GARCH-normal VaRs show more concentration at the low end of the range in the pre-crisis period, while bank VaR values show more concentration at the low end of the range in the crisis-period. This is consistent with bank VaRs being overly conservative in the pre-crisis period but not conservative enough in the crisis period. Moreover, GARCH-normal crisis period exceedances occur over a wider range and are not concentrated at low VaR levels. These features would suggest that the GARCH-normal VaRs show stronger covariation with bank PnL volatility than do bank VaRs. While this appears to be the case over much of the VaR ranges, GARCH-normal VaRs at the high-end of the range are not associated with abnormally high |PnL|.

Figure 3 shows the temporal behavior of the GARCH-normal VaR for the average bank. Although somewhat muted in reflecting VaR variation at an individual bank level, both pre-crisis and crisis period figures exhibit the more pronounced higher frequency variation reported in Table 10. In comparison to the bank VaR behavior, the GARCH-normal VaR displays a much greater shift of the share of the VaR variation at higher frequencies in the crisis period as compared to the pre-crisis period, as well as much higher overall variation. Such GARCH-normal VaR characteristics are in-line with the similar shift for the |PnL|. This again reflects the sensitivity of the GARCH volatility forecast. What tends to offset its forecasting power, however, is the over-reaction of the GARCH volatility forecasts on large PnL outliers visible in the Figure.

4.2.2 Alternative non-Normal GARCH VaR Measures

A substantial literature has argued and with some evidence shown that relaxing the conditional normality assumption can improve the GARCH-based VaR performances.[26] Three alternative GARCH-based VaRs with non-normal residual return specifications are considered: Filtered Historical Simulation (FHS), GARCH-t, and GARCH-EVT based on an extreme value theory approach to estimate the lower-tail of the return distribution. These specifications allow for fat tails in conditional return distributions.

Filtered Historical Simulation (FHS) is a method applicable to an HS VaR framework. FHS VaR uses the AR - GARCH normalized estimated conditional residuals, allowing for leptokurtosis and skewness by applying the HS nonparametric approach to estimate the quantiles. For FHS VaR, GARCH-based volatility forecast $\hat{\sigma}_{T+1}$ for an out-of-sample date $T+1$ is first computed based on the output from the AR-GARCH model. The estimated innovations from the 2-year estimation period, $\varepsilon_{T-i},\ i = 0, 1, \ldots, L$, are then weighted by the inverse of their respective in-sample GARCH volatility estimates and then scaled by the next-day volatility forecast described above. Such scaling normalizes the historical residuals to reflect the current-period volatility forecast. Finally, the desired quantiles of the next-day PnL are derived based on the in-sample distribution of scaled innovations, which allows for both leptokurtosis and skewness.

GARCH-t specification provides a heavy-tail parametric specification that replaces the normal distribution assumption underlying GARCH-normal. The shape of the distribution, particularly the heaviness of the tails, depends on the t-distribution's degrees of freedom (df) parameter. The GARCH-t VaR differs from the FHS VaR in two respects. First, it offers a

[26] VaR performance studies include McNeil and Frey (2000), Bao, Lee, and Saltoglu (2006), Kuester, Mittnik, and Paoella (2006), Ergen (2010), Kourouma, Dupre, Sanfilippo, and Taramasco (2011), and Adcock, Areal, and Oliveira (2012) among others.

fully parametric approach to model thickness of the conditional return distribution. Second, it is based on a joint estimation of the degrees of freedom parameter controlling thickness of tails with all other model parameters whereas FHS offers a two-step procedure conditional on the first-stage GARCH-normal estimation step. Thus, the estimated expected return and volatility parameters may naturally differ from those under the GARCH-normal and FHS VaRs. More specifically, the extra flexibility of controlling the thickness of conditional return distribution allows for some off-load of the impact of outliers from the volatility estimates. A caveat lies in the potential over-parametrization of the GARCH-t specification.

A fully parametric modeling of the lower tail is also applied using Extreme Value Theory (EVT), specifically the Generalized Pareto Distribution (GPD). For a broad family of continuous distributions, GPD is the limiting distribution of the tail and, since the threshold quantile is located further out in the tail, is used to estimate the 1% VaR.[27] McNeil and Frey (2000) provide a two-stage method to model return tails. First, a standard AR(1)-GARCH(1,1) model in equations (2) - (4) is fit to the PnL data by pseudo maximum likelihood (PML) under a normality assumption and without any assumptions on the true model standardized innovations z_t. The estimates of parameters, conditional mean and variance are obtained, which are used to derive model-implied scaled residuals z_t. In the second step EVT is used to model the tail behavior of standardized residuals by fitting the GPD to the implied scaled residuals z_t below the threshold of choice u. The set $\{z_t : z_t < u\}$ defines the tail of the distribution and can be well approximated by the GPD. All observations below the threshold are treated as iid GPD realizations. Since the GPD is a two parameter family of distributions, both parameters are jointly estimated using maximum likelihood method. The estimates of quantiles $VaR_t^{q,z}$ of interest for $q = .01$ can be easily obtained by utilizing analytical formulas in McNeil (1999). Analogous to the GARCH-normal and FHS VaRs, the GARCH-EVT quantile estimate is finally transformed to PnL quantiles given the following equation:

$$VaR_t^{q,r} = \mu_t + \sigma_t * VaR_t^{q,z} \tag{5}$$

The choice of threshold u plays a crucial role in the EVT method and is essentially a compromise between maximizing the number of observations in the tail and increase in the goodness of approximation of the distribution of scaled residuals by the GPD. In our application we use a constant proportion of 20% of observations in the tail. This provides the minimum number of observations in the tail that guarantees convergence of the maximum likelihood estimation procedure.

For these three alternative benchmark VaR measures, pre-crisis and crisis-period mean VaRs, exceedance rates with coverage and independence tests, and VaR correlations with

[27] See McNeil (1999). For EVT applications to VaR also see McNeil and Frey (2000), Longin (2000), Yamai and Yoshiba (2005).

next-day realized |PnL| are presented in Tables 12, 13, and 14 along with bank, HS, and GARCH-normal VaR statistics for comparison.

For the pre-crisis period, the FHS and GARCH-EVT VaRs have similar mean VaRs (Table 12, top panel) and exceedance rates (Table 13, top panel). Exceedance rates across banks average modestly above one percent and higher than the GARCH-normal VaR exceedance rates. Nevertheless the null hypothesis of 1% coverage cannot be rejected at 5% level. The GARCH-t VaRs are consistently larger, with lower exceedance rates with 1% coverage rejected for banks 3 and 4. Specifically the conservativeness of the GARCH-t VaRs reflect heavier-tailed conditional residuals, since mean estimated GARCH volatilities are found to be close to those for GARCH-normal specification. Exceedance independence is not rejected across all GARCH-based models, except for GARCH-t VaR for one bank.

For banks 1, 3, and 5 (the banks with better-performing bank VaRs and no extreme losses), the crisis-period GARCH-non-normal VaRs have means larger than the GARCH-normal VaRs and hence moderately lower exceedance rates that are closer to 1% level although not enough to support the null hypothesis of correct coverage. For FHS and GARCH-t VaRs, these somewhat improved results owe to conditional residual 1-percent quantiles being larger than the standard normal distribution 1-percent quantile value. For GARCH-EVT VaRs, conditional residual quantiles are small absolutely, but comparatively large negative covariances between GARCH-based volatility and conditional residual quantile provide, on net, some improvement over the standard GARCH-normal model.[28]

For banks 2 and 4, who experienced extreme crisis-period losses, GARCH-non-normal VaRs, might seem most appropriate. However, GARCH-normal VaRs are larger (Table 12, bottom panel) and have lower exceedance rates (Table 13, bottom panel) that are closer to 1% level than the GARCH non-normal alternatives. The smaller FHS and GARCH-EVT VaRs and higher exceedance rates reflect lower conditional residual 1% quantiles than for standard normal distribution. For GARCH-t, low GARCH volatility estimates combined with estimated conditional residual quantiles close to standard normal distribution produced relatively small VaRs and higher exceedance rates. This reflects the joint estimation of the GARCH-t volatility and degrees of freedom (df) parameters, whereas the FHS and GARCH-EVT VaRs are both conditioned on the GARCH-volatility forecasts. Specifically, extreme |PnL| realizations for banks 2 and 4 are "off-loaded" to a heavy tail df parameter, rather than producing a buildup of volatility forecasts following negative PnL outliers. Since volatility is persistent, the outliers substantially contribute to higher average risk measures in GARCH-normal based models but not so in GARCH-t model.[29]

VaR correlations with next day realized PnL volatility for non-normal VaR measures

[28] Following equation (5), the average VaR over the sample period for a particular bank can be expressed as $\overline{VaR^{q,r}} = \overline{\mu} + \overline{\sigma} \times \overline{VaR^{q,z}} + Cov(\sigma, VaR^{q,z})$.

[29] Ergen (2010) analyzes over-parameterization in joint estimation of GARCH-t parameters and considers a GARCH skewed-t VaR for stock indices of emerging market countries.

are presented in Table 14. For FHS and GARCH-EVT VaRs, correlations can differ from GARCH-normal VaR correlations only from differences in estimated conditional residual quantiles. For GARCH-t, correlations may differ due to differences in the estimated conditional residual quantiles, the GARCH volatility, and AR estimates. The Table shows however that the |PnL| correlations using the non-normal VaR measures are roughly of the same order of magnitude as those for the GARCH-normal specification.

Results for non-normal VaR variance decomposition are presented in Table 15 for the variance accounted for within a 63-day periodicity. The variance decomposition results for the FHS and GARCH-EVT VaRs closely agree with those for the GARCH-normal VaRs. GARCH-t VaR variance contributions within 63-day periodicity are generally lower in the crisis-period and similar to other GARCH-based models in the crisis period for better performing banks 1, 3 and 5. However, the GARCH-t VaR results are significantly higher for banks 2 and 4 with PnL outliers in the crisis period. This reflects the joint volatility and df parameter estimation muting the GARCH-volatility response coming after extreme losses by "off-loading" its response to the decrease in heavy tail df parameter.[30]

The general superiority of GARCH-based VaR measures over HS VaR is consistent with results in earlier studies. Variation in the GARCH-t VaR performance results is also consistent with previous studies for GARCH symmetric-t VaR. However, previous studies have generally found non-normal GARCH-based VaRs superior to GARCH-normal in coverage and independence, particularly allowing for skewness, and also have found EVT VaR measures tend to be superior (see footnote 19). However, these studies did not consider potential GARCH volatility overreaction to outlier PnL returns and effects on GARCH volatility forecast power.

4.2.3 Implications and Issues

The benchmark GARCH-based VaRs used here were more accurate than the banks' internal VaR measures in terms of forecasting VaR exceedances and useful for assessing bank VaR performances. A bank or regulator, with knowledge of the bank's current market positions, could potentially improve on the historical PnL GARCH-based VaR measure. The bank or regulator would revalue *current positions* using the historical market prices to obtain a pseudo-historical PnL, as done in standard historical simulation. A GARCH or possibly other volatility forecast could be estimated using the pseudo-historical PnL and then applied to the PnL VaR forecast. Being based on current exposures, this may serve as a better or complementary measure in evaluating the adequacy of bank VaR forecasts on a real time basis.[31]

[30] As with the 63-day comparisons, 5-day and 21-day periodicity comparisons show similarity across different GARCH VaR measures, with occasional large difference for GARCH-t VaR.

[31] Alexander and Sarabia (2011) also suggest a formal measure of the accuarcy of a bank's VaR where historical PnL is used to determine historical quantiles. These quantiles, along with the historical bank

This portfolio-level VaR measure will be limited in identifying risks at the sub-portfolio level and for forecasting conditions in the financial markets to which the banks have exposures. It thus will be of limited use for managing risks in the trading portfolio. For this purpose, measurement of more disaggregated postions risks and correlations is needed. Bank trading postion exposures, however, include a large number of positions in various equity, interest rate, foreign exchange, commodity, and credit markets. The large dimensionality of market exposures makes measurement of the market dynamics to which the banks' have exposures difficult. Various approaches to reduce the number of risk parameters that would need to be estimated have been suggested but with little application to date.[32]

Even with measurement of market dynamics, the results here indicate difficulties in estimating risk in a period of substantial market instability. During the crisis period, both bank and benchmark VaRs significantly underestimated the banks' 1-percent PnL quantiles. While the GARCH-based VaRs had better accuracy, they overreacted to large 1-day PnL realizations. The use of non-normal GARCH-based VaR measures provided only moderate improvement in loss coverage over the normal GARCH-based VaR.

A body of research points to the usefulness of recognizing several components contributing to return variance, a more persistent stochastic volatility with possible jumps in volatility, and less persistent jumps in returns.[33] Market instability is characterized by both significant variation in price volatility and price jumps. The accuracy of the GARCH or related time-varying volatility models may be improved in some manner by restricting the estimated volatility's response to large price changes.[34] Besides loosing some accuracy in the volatility forecast, this approach still does not account for the jump risk. Ultimately, an adequate measure of market risk needs to encompass both market instability features.[35]

In a limited attempt here, a jump-diffusive stochastic volatility model for bank PnL that combined a Poisson jump process with stochastic volatility was estimated. From this estimation, daily rolling benchmark VaRs were calculated for each bank. Crisis period VaR exceedances, however, were frequent and large as the VaR estimates responded slowly to the heightened PnL volatility and did not greatly raise the jump activity rate estimates

VaRs, are used to measure bank VaR accuracy (via bias and error dispersion), i.e. "VaR model-risk."

[32]References include Barone-Adesi, Giannopoulos, and Vosper (1999) for multivariate FHS; Andersen, Bollerslev, Christoffersen, and Diebold (2007) for dimension reduction in multi-variate GARCH methods; Aramonte, Rodriguez, and Wu (2013) for application of GARCH to a set of identified latent common factors underlying market factors; and Embrechts, Puccetti, and Rüschendorf (2013) for the use of copulas in establishing bounds on multi-factor dependence.

[33]See Andersen, Bollerslev, and Diebold (2007) and references therein.

[34]Muler and Yohai (2008) and Boudt, Danielsson, and Laurent (2011) suggest a limited approach that constrains the GARCH volatility estimate.

[35]For GARCH volatility models with jumps, see Maheu and McCurdy (2004) and Duan, Ritchken, and Sun (2006). For more general VaR measurement using stochastic volatility models with Levy jumps, see Szerszen (2009). Gibson (2001) provides an early attempt to incorporate jumps into a VaR measure.

during the period. These results owe to the long estimation period needed to precisely estimate jump activity rates and conditional jump size distribution due to an assumption of i.i.d. Poisson jumps, and suggest a need to recognize temporal dependence in jumps or their activity rate's co-dependence with volatility.[36]

4.3 Expected Shortfall

It is well known that VaR as a quantile measure is limited in measuring tail risk. Briefly considered here is Expected Shortfall (ES), which has been the principal risk measure recommended as an alternative or supplement to VaR for providing a measure of the size of potential loss. Under the GARCH assumption with conditional normality, ES is analytically determined conditioned on PnL being below the estimated .01 quantile. Using FHS, ES is the mean of the filtered historical PnL for values below the FHS VaR estimate. For GARCH-t and EVT measures, ES is the numerically or analytically determined expected value of the PnL conditioned on its being below the respective VaR estimate.

Average daily ES estimates for the pre-crisis and crisis periods for each bank using the different ES measures are reported in Table 16. For the pre-crisis period, the ES averages for the different measures, excluding GARCH-t, are similar, with GARCH-t consistently more conservative. For the crisis period, and for banks 1, 3 and 5, the ES averages for the GARCH-normal measure are smaller absolutely than those for the GARCH non-normal measures, which are generally about the same size.

ES size comparisons among the different VaR measures parallel the EVT VaR comparisons, as might be expected since the size of the loss given an exceedance depends on how exceedance, or VaR violation, is measured. Consistent with this and excluding banks 2 and 4, the more conservative crisis-period GARCH non-normal VaRs are accompanied by more conservative ES values than GARCH-normal ES estimates.

For banks 2 and 4, the crisis period GARCH-t ES estimates are smaller than those for the other ES measures. These results parallel the relatively smaller GARCH-t VaRs for these banks described above and with the same explanation. As described in more detail for the GARCH-t VaR results, they reflect the off-loading of extreme PnL realizations to a heavy tail df parameter, greatly reducing the levels of GARCH-estimated volatilities observed for the other GARCH-based VaRs. As with the GARCH-t estimates, this off-loading and the persistence of volatility are responsible for the two banks' relatively small average ES estimates.

In a critique of the ES measure, Yamai and Yoshiba (2005) point to a small ES estimation precision for heavy tailed distributions, which is when ES is needed the most. Further, ES requires a larger sample size than VaR to produce the same accuracy. Some evidence of

[36] The results for stochastic volatility model with jumps are not reported here but could be provided on request.

lack of ES estimation precision for bank PnL can be seen in Table 16. All studied methods produced much closer ES results in the pre-crisis period than in the crisis period. The differences in ES crisis-period estimates were greatest for banks 2 and 4, with these two banks having the most pronounced outliers, skewness, and kurtosis (Table 1). A viable explanation is that the sample could be not rich enough to provide precise ES estimates and depends heavily on model specification.[37]

To provide a closer look at ES performance, further results for the GARCH-EVT ES measure are presented in Table 17. For the pre-crisis period, the estimated unconditional mean ES values are about one-third larger than the EVT VaRs (Table 12). The unconditional and conditional mean ES values are roughly the same. While they are less than the actual mean ES values (realized losses) conditioned on an exceedance, they are in the same range. Rounding out this picture, the PnL exceedance rate for ES averaged .0058 across the four banks versus .012 for the average GARCH-EVT VaR exceedance rate, a moderately greater measure of potential loss.

Compared to the pre-crisis period, in the crisis period the unconditional mean estimated ES values are much larger and also larger than the estimated conditional and realized ES measures. However, the mean ES estimates conditioned on a VaR exceedance are much smaller than the actual ES values (actual losses) conditioned on a VaR exceedance. This difference in unconditional and conditional ES estimates reflects the GARCH-based ES estimates over-reaction to large changes changes in PnL, similar to that occurring with the GARCH-based VaR estimates. Even though unconditionally GARCH-EVT ES measure was overly conservative, it was not conservative enough when conditioned on VaR violation. As documented by the correlation with the next-day |PnL| volatility measure, the ES estimates had little forecast power in the crisis period, in contrast to the pre-crisis period. ES exceedance rates, mean exceedances, and maximum exceedances in Table 17 provide information on the ES loss "coverage".

5 Conclusions

This study examined the performance and behavior of VaR for a number of large banks during and before the financial crisis. For the pre-crisis period, bank VaRs were overly conservative with few VaR exceedances. This contrasted with mostly unbiased benchmark HS and GARCH VaRs using only historical bank PnL. For the crisis period, however, bank VaR exceedances were substantially excessive and exhibited clustering, with two of the five banks having significantly poorer performances. The benchmark GARCH VaRs had fewer crisis-period exceedances than the bank VaRs and did not exhibit exceedance clustering.

[37]See Kourouma, Dupre, Sanfilippo, and Taramasco (2011) for comparison of VaR and ES measures for stock and commodity indexes during the financial crisis.

This better performance during the financial crisis reflected the GARCH VaRs stronger and more immediate adjustment to the crisis-period volatility as reflected in GARCH-based volatility dependence on bank PnL regression residuals.

Covering a much longer and more recent history, the pre-crisis and crisis period bank VaR results here are similar to those for bank VaRs in the late 1990s - early 2000 period reported by Berkowitiz and O'Brien (2002). Specifically, now as then, bank VaRs are conservative when times are normal but understate risk in a period market instability, and also when compared to benchmark GARCH-based VaR measures.

Bank VaR methodologies are not designed to forecast near-term variation in market volatility. At least in part this may reflect difficulties in modeling market volatility with trading positions covering a wide variety of financial markets. Incorporating volatility modeling would seem necessary for improving bank VaR adjustment to changing market conditions. More active consideration could be given to current proposals for measuring market factor dynamics for high-dimensional portfolios, while using a PnL volatility-conditioned VaR forecast as a supplement to current bank VaR measures.

While application of GARCH or similar volatility forecasts might materially improve VaR performances, results here suggest remaining forecast limitations in a period of substantial market instability. A more complete measurement would require accounting for multiple variance components, e.g. jumps. Absent this, supplemental measures, such as stress exercises, appear to be logical.

Finally a general shortcoming of VaR is that it does not measure potential loss when VaR is exceeded. This is more likely to be important in a period of financial turmoil. Some limited results here in estimating an ES measure suggest that accurate estimation may also be significantly more difficult in a such a period.

References

Adcock, C., N. Areal, and B. Oliveira (2012, August). Value-at-risk forecasting ability of filtered historical simulation for non-normal GARCH returns. *SSRN*.

Adrian, T. and H. S. Shin (2014, February). Procyclical leverage and value at risk. *The Review of Financial Studies 27*(2), 373–403.

Alexander and Sarabia (2011). Value-at-risk model risk. *SSRN*.

Andersen, T., T. Bollerslev, P. Christoffersen, and F. Diebold (2007). Practical volatility and correlation modeling for financial market risk management. In *The Risks of Financial Institutions*, NBER Chapters, pp. 513–548. National Bureau of Economic Research, Inc.

Andersen, T., T. Bollerslev, and F. Diebold (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics 89*(4), 701–720.

Aramonte, S., M. Rodriguez, and J. Wu (2013). Dynamic factor value-at-risk for large, heteroskedastic portfolios. *Journal of Banking & Finance 37*(11), 4299–4309.

Bao, Y., T.-H. Lee, and B. Saltoglu (2006). Evaluating predictive performance of value-at-risk models in emerging markets: A reality check. *Journal of Forecasting 25*, 101–128.

Barone-Adesi, G., K. Giannopoulos, and L. Vosper (1999). VaR without correlations for portfolios of derivative securities. *Journal of Futures Markets 19*(5), 583–602.

Basak, S. and A. Shapiro (2001). Value-at-risk-based risk management: Optimal policies and asset prices. *The Review of Financial Studies 14*(2), 371–405.

Baxter, M. and R. G. King (1999). Measuring business cycles: Approximate band-pass filters for economic time series. *The Review of Economics and Statistics 81*(4), pp. 575–593.

Berkowitz, J., P. Christoffersen, and D. Pelletier (2011). Evaluating value-at-risk models with desk-level data. *Management Science* (57), 2213–2227.

Berkowitz, J. and J. O'Brien (2002). How accurate are value-at-risk models at commercial banks? *The Journal of Finance 57*(3), pp. 1093–1111.

Boudt, K., J. Danielsson, and S. Laurent (2011). Robust forecasting of dynamic conditional correlation GARCH models. *Working Paper*.

Campbell, S. D. (2006). A review of backtesting and backtesting procedures. *Journal of Risk 9*(2), 1–17.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review 39*(4), 841–862.

Christoffersen, P. and D. Pelletier (2004). Backtesting value-at-risk: A duration-based approach. *Journal of Financial Econometrics 2*(1), 84–108.

Danielsson, J. (2002). The emperor has no clothes: Limits to risk modelling. *Journal of Banking & Finance 26*(7), 1273–1296.

Danielsson, J., H. S. Shin, and J.-P. Zigrand (2004). The impact of risk regulation on price dynamics. *Journal of Banking & Finance 28*(5), 1069–1087.

Danielsson, J., H. S. Shin, and J.-P. Zigrand (2009). Risk appetite and endogenous risk.

Duan, J., P. Ritchken, and Z. Sun (2006). News arrival, jump dynamics, and volatility components for individual stock retruns. *Mathematical Finance 16*(1), 21–52.

Embrechts, P., G. Puccetti, and L. Rüschendorf (2013). Model uncertainty and VaR aggregation. *Journal of Banking & Finance 37*(8), 2750–2764.

Ergen, I. (2010, February). VaR prediction for emerging stock markets: The implications for fat tails and fair comparison of EVT methods and student-t distribution. *Federal Reserve Bank of Richmond Working Paper*.

Gibson, M. (2001). Incorporating event risk into value-at-risk. *Board of Governors of the Federal Reserve System Working Paper 2001-17*.

Hirtle, B. (2003, September). What market risk capital reporting tell us about bank risk. *FRBNY Policy Review*, 37–54.

Jorion, P. (2002). How informative are value-at-risk disclosures. *The Accounting Review 77*(4), 911–931.

Jorion, P. (2007). *Value at Risk: The New Benchmark for Managing Financial Risk* (3rd ed.). McGraw-Hill.

Kourouma, L., D. Dupre, G. Sanfilippo, and O. Taramasco (2011, April). Extreme value at risk and expected shortfall during financial crisis. *Working Paper*.

Kuester, K., S. Mittnik, and M. Paoella (2006). Value at risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics 4*(1), 53–89.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives 3*(2), 73–84.

Liu, C. C., S. Ryan, and H. Tan (2004). How banks' value-at-risk disclosures predict their total and priced risk: Effects of bank technical sophistication and learning over time: Accounting, disclosure, and the cost of capital. *Review of Accounting Studies 9*(2-3), 265–294.

Longin, F. (2000). From value at risk to stress testing: The extreme value approach. *Journal of Banking & Finance 24*, 1097–1130.

Maheu, J. and T. McCurdy (2004). News arrival, jump dynamics, and volatility components for individual stock retruns. *The Journal of Finance 59*(2), 795–793.

Manganelli, S. and R. Engle (2001, August). Value-at-risk models in finance. *European Central Bank Working Paper 41*.

McNeil, A. (1999). Extreme value theory for risk managers. In *Internal Modelling and CAD II*, pp. 93–113. RISK Books.

McNeil, A. and R. Frey (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance 7*(3-4), 271–300.

Mincer, J. and V. Zarnowitz (1969, March). The evaluation of economic forecasts. In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, NBER Chapters, pp. 1–46. National Bureau of Economic Research.

Mudge, D. and L. Wee (1993, December). Truer to type. *Risk*, 16–19.

Muler, N. and V. Yohai (2008). Robust estimates for GARCH models. *Journal of Statistical Planning and Inference 138*(10), 2918–2940.

Nocera, J. (2009, Jan). Risk management. *New York Times*.

Perignon, C., Z. Y. Deng, and Z. J. Wang (2008). Do banks overstate their value-at-risk? *Journal of Banking & Finance 32*, 783–794.

Perignon, C. and D. Smith (2010a). Diversification and value-at-risk. *Journal of Banking & Finance 34*(1), 55–66.

Perignon, C. and D. Smith (2010b). The level and quality of value-at-risk disclosure by commercial banks. *Journal of Banking & Finance 34*(2), 362–377.

Pritsker, M. (2006). The hidden dangers of historical simulation. *Journal of Banking & Finance 30*(2), 561–582.

Szerszen, P. (2009). Bayesian analysis of stochastic volatility models with levy jumps: Application to risk analysis. *Board of Governors of the Federal Reserve System Working Paper 2009-40.*

Yamai, Y. and T. Yoshiba (2005). Value-at-risk versus expected shortfall: A practical perspective. *Journal of Banking & Finance 29*(4), 997–1015.

Appendix: Backtesting VaR measures

In this Appendix, we briefly describe the battery of tests used throughout the paper that serve to statistically evaluate both unconditional and conditional performance of VaR measures. We base our exposition on the Christoffersen (1998) likelihood ratio framework and Christoffersen and Pelletier (2004) duration framework. Setting the stage, define the sequence of one-period ahead VaR measures $\{VaR_{t,t+1}\}, t = 1, \ldots, T-1$ coming from a benchmark or bank model. Since VaR forecasts can be seen as interval forecasts, we define a violation, or an exceedance, indicator variable I_t, by:

$$I_{t+1} = \begin{cases} 1 & \text{if } r_{t,t+1} < VaR^q_{t,t+1}, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The conditional coverage hypothesis is true if the following condition holds

$$E(I_{t+1}|\Omega_t) = q \tag{2}$$

for all t, which implies that the indicator sequence I_t is i.i.d. *Bernoulli*(q) distributed if $VaR_{t,t+1}$ is based on a true model.

The Christoffersen (1998) likelihood ratio test for unconditional coverage tests is $E(I_{t+1}) = q$, the weaker condition implied by (2). Under the null of correct coverage the likelihood is

$$L(q; I_1, I_2, \ldots, I_T) = (1-q)^{n_0} q^{n_1}.$$

The likelihood ratio test statistic is given by

$$LR_{uc} = -2log\left(L(q; I_1, I_2, \ldots, I_T)/L(\widehat{q}; I_1, I_2, \ldots, I_T)\right),$$

where $\widehat{q}$ is the relative number of VaR violations in the sample. The LR_{uc} statistic is asymptotically $\chi^2(1)$ distributed.

The Christoffersen (1998) test for independence of VaR violations assumes under the alternative that the hit process I_t follows a first-order Markov process. Under the null of independence, the probability of violation should be independent from the preceding realization of the hit sequence. More formally, consider a binary first-order Markov chain with transition matrix given by:

$$Q = \begin{pmatrix} 1-q_{01} & q_{01} \\ 1-q_{11} & q_{11} \end{pmatrix}$$

where $q_{ij} = Prob(I_{t+1} = j | I_t = i)$. The likelihood function for this chain is given by

$$L(Q; I_1, I_2, \ldots, I_T) = (1-q_{01})^{n_{00}} q_{01}^{n_{01}} (1-q_{11})^{n_{10}} q_{11}^{n_{11}}.$$

where n_{ij} is the number of hit realizations with the value i followed by the value j. Under the null, $q_{01} = q_{11} = q_1$ with the respective transition matrix Q_{ind}. Finally. after replacing probabilities in transition matrices Q and Q_{ind} by their sample counterparts and ending up with the respective matrices corresponding to the maximum likelihood estimates $\widehat{Q}$ and $\widehat{Q}_{ind}$, we have the following likelihood ratio statistic:

$$LR_{ind} = -2log\left(L(\widehat{Q}; I_1, I_2, \ldots, I_T)/L(\widehat{Q}_{ind}; I_1, I_2, \ldots, I_T)\right).$$

with LR_{ind} test statistic asymptotically $\chi^2(1)$ distributed.

As noted above, testing for independence with the application of Christoffersen (1998), LR_{ind} test statistic strongly depends on the first-order Markov assumption but offers an easily applicable method. However, the dependence structure of the hit sequence I_t can be of different nature. Christoffersen and Pelletier (2004) offers a viable approach to test for dependence without making strong assumptions. The test relies on the duration based test statistic. Lets denote by $D_i = t_i - t_{i-1}$ the time between two hits in the I_t sequence and by $N(T)$ the total number of hits in $\{I_t\}$ sequence. The first and the last duration times are special cases, since they can be either left-censored ($C_1 = 1$ or 0 otherwise) or right-censored ($C_{N(T)} = 1$ or 0 otherwise) respectively. Under the null of no dependence, the duration D_i should exhibit a flat hazard rate q, which is consistent with the exponential distribution $f_{exp}(D; q) = qexp(-qD)$. For the alternative we choose a simple Weibull distribution with $f_W(D; a, b) = a^b b D^{b-1} exp(-(aD)^b)$, which collapses to the exponential distribution if b=1. As noted by Christoffersen and Pelletier (2004), the maximum likelihood estimate for the unconstrained model with a Weibull alternative can be obtained by maximizing with respect only to one of the parameters b, since a solves the following first order condition

$$\widehat{a} = \left(\frac{N(T) - C_1 - C_{N(T)}}{\sum_i^{N(T)} D_i^b}\right)^{(1/b)}$$

Finally, the likelihood ratio statistic LR_{dur} is given by

$$LR_{ind} = -2log\left(L(D; \widehat{a}(1), 1)/L(D; \widehat{a}(\widehat{b}), \widehat{b})\right).$$

where $\widehat{b}$ is the ML estimate of the parameter b under a Weibull specification. The specific form of the likelihood function directly follows from the assumed independence of duration times D_i and is a simple product of Weibull likelihoods for each observation.[38]

[38] The first and the last duration times are special cases because of their possible left or right censoring. In case of censoring, the likelihood component of observing duration D is a function of cumulative distribution function.

Table 1: **Bank Daily PnL:** Bank daily PnL statistics are reported for pre-crisis and crisis periods for each bank. Reported statistics include sample dates, mean, standard deviation (std dev), historical 1% PnL quantile (.01 quant), daily loss frequency (PnL < 0), and PnL kurtosis and skewness. Normal distribution is rejected for all banks at 1% level using Jarque-Bera test. See Table 18 for full sample bank PnL statistics. For confidentiality purposes only calendar years of the first available data points are reported. The crisis period sub-sample starts on June 1st 2007 but no information on the end of the crisis period sample is provided for each bank.

			Pre-Crisis Period				
bank	dates	mean	st dev	.01 quant	PnL < 0	kurtosis	skewness
bank 1	2003 - May 2007	.5613	.5262	$-.795$	.101	5.790	.301
bank 2	2002 - May 2007	.1390	.1204	$-.141$	.076	5.802	.814
bank 3	2001 - May 2007	.4668	.5780	$-.704$	.170	6.312	.508
bank 4	2003 - May 2007	.2612	.2460	$-.149$	.087	38.966	3.271
			Crisis Period				
bank	start dates	mean	st dev	.01 quant	PnL < 0	kurtosis	skewness
bank 1	June 2007	.1977	1.7080	-5.355	.390	4.977	$-.707$
bank 2	June 2007	$-.1365$	1.9132	-8.291	.406	114.758	-9.391
bank 3	June 2007	.8071	1.7995	-3.805	.276	8.183	$-.649$
bank 4	June 2007	.0809	1.7506	-3.248	.301	155.602	-11.405
bank 5	June 2007	.2893	1.1129	-2.836	.333	10.957	.391

Table 2: **Bank |PnL| Variance by Frequency:** Bank pre-crisis and crisis period |PnL| variances (total variance) and |PnL| variance decompositions by frequencies are reported in the top and bottom panels. Frequency bands correspond to periodicities less than 5 (weekly), 21 (monthly), and 63 (quarterly) days. Cumulative fraction of |PnL| variance is reported at each frequency range. Variance decomposition of |PnL| uses the Band-Pass filter with K = 30 leads and lags.

	total variance	$\leq$5-day/total	$\leq$21-day/total	$\leq$63-day/total
		Pre-Crisis Period		
bank 1	.199	.5591	.8525	.9057
bank 2	.012	.5305	.7948	.8533
bank 3	.242	.4364	.7091	.7495
bank 4	.053	.4635	.7593	.8033
		Crisis Period		
bank 1	1.338	.5375	.8559	.9543
bank 2	3.253	.6040	.9251	.9800
bank 3	1.641	.4029	.7965	.9689
bank 4	2.677	.5752	.9109	.9655
bank 5	.647	.5566	.7808	.7972

Table 3: **Bank Daily VaR:** Pre-crisis period and crisis period bank VaR descriptive and performance statistics are reported for each bank. Descriptive statistics include daily bank VaR mean, standard deviation (std dev) and VaR exceedance rate (ex rate), as well as mean and maximum (max) exceedances (ex). Performance-related statistics include p-values for measures of 1% VaR accuracy (cover), exceedance independence (indep), duration between exceedances (dur) and VaR correlation with next-day |PnL| (with p-values in parentheses). The mean and max statistics are unsigned.

Pre-Crisis Period									
			ex	mean	max	———	p-values	———	cor$(-VaR_t,$
bank	mean	std dev	rate	ex	ex	cover	indep	dur	$\|PnL_{t+1}\|)$
bank 1	1.46	.35	.004	.27	.62	.014	.008	.140	.051 (.089)
bank 2	.32	.06	.000	–	–	–	–	–	.156 (.000)
bank 3	1.27	.46	.001	.99	1.25	.000	.941	–	.410 (.000)
bank 4	.52	.11	.000	–	–	–	–	–	.301 (.000)
Crisis Period									
			ex	mean	max	———	p-values	———	cor$(-VaR_t$
bank	mean	std dev	rate	ex	ex	cov	indep	dur	$\|PnL_{t+1}\|)$
bank 1	3.31	1.48	.040	1.51	3.71	.000	.000	.000	.132 (.008)
bank 2	.69	.21	.100	2.33	26.53	.000	.631	.699	.071 (.151)
bank 3	2.97	.56	.034	1.25	6.60	.000	.080	.132	.162 (.001)
bank 4	.97	.22	.051	2.88	25.06	.000	.000	.001	−.035 (.483)
bank 5	1.49	.74	.027	.76	4.02	.005	.002	.021	.389 (.000)

Table 4: **Bank Daily VaR Variance by Frequency:** Pre-crisis and crisis period daily bank VaR variances (total variance) and bank VaR variance decompositions by frequencies are reported in the top and bottom panels. Frequency bands correspond to periodicities less than 5 (weekly), 21 (monthly), and 63 (quarterly) days. Cumulative fraction of bank VaR variance is reported at each frequency range. Variance decomposition of bank VaR uses the Band-Pass filter with K = 30 leads and lags.

	total variance	≤5-day/total	≤21-day/total	≤63-day/total
Pre-Crisis Period				
bank 1	.122	.0820	.2058	.3606
bank 2	.004	.0436	.1101	.1756
bank 3	.211	.0086	.0311	.0559
bank 4	.013	.0159	.0457	.0928
Crisis Period				
bank 1	2.176	.0267	.0516	.1037
bank 2	.044	.0097	.0308	.1037
bank 3	.315	.0322	.1234	.2243
bank 4	.050	.0196	.0778	.1537
bank 5	.547	.0134	.0499	.1072

Table 5: **Cross-Bank PnL and VaR Correlations:** Cross-bank daily PnL correlations and VaR correlations respectively are reported for each bank for the pre-crisis and crisis periods. Cross-bank PnL correlations are reported in the upper half and cross-bank VaR correlations in the lower half of each panel. Pre-crisis cross-bank PnL p-values are all significant at the 1% level. Cross-bank VaR correlation p-values are significant at the 1% level, except that for Banks 1 and 3. Crisis period cross-bank PnL correlation p-values $\geq .15$ are significant at the 1% level. All crisis period cross-bank VaR p-values are significant at the 1% level.

$corr(VaR_i, VaR_j) \backslash corr(PnL_i, PnL_j)$

	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5
Pre-Crisis Period					
	Bank 1	Bank 2	Bank 3	Bank 4	
Bank 1		0.3430	0.2734	0.2005	
Bank 2	0.1672		0.3110	0.1470	
Bank 3	-0.0615	0.1136		0.2281	
Bank 4	-0.111	0.354	0.7050		
Crisis Period					
	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5
Bank 1		-0.0206	0.1875	0.2772	0.4450
Bank 2	0.6738		0.0849	0.0150	0.2500
Bank 3	0.7596	0.6907		0.1547	0.2773
Bank 4	0.7344	0.5802	0.8025		0.0681
Bank 5	0.8206	0.6145	0.7372	0.7469	

Table 6: **Bank VaR Order Categories:** On each day, the standardized banks' PnL and VaR are sorted into bin categories with the largest (most conservative) bank VaR in category one and smallest (least conservative) VaR in the highest numbered bin (four for pre-crisis and five for crisis period). The bins 4 and 5 for the crisis period are further aggregated into one bin called "bin 4 + 5" to ensure that each bin is based on data for more than one bank. For each period and order category, both PnL and VaR mean and standard deviations (st dev) are reported. Exceedance rates (ex rate) and correlations (with p-values in parentheses) for daily −VaR and PnL are also reported.

order category	PnL mean	PnL st dev	VaR mean	VaR std dev	ex rate	cor($-VaR_t$, $\lvert PnL_{t+1}\rvert$)
Pre Crisis Period						
one	0.0003	1.0809	-4.1547	0.5177	0.0023	0.1534 (0.0000)
two	0.0447	0.9942	-3.6875	0.3965	0.0011	0.1554 (0.0000)
three	-0.0289	0.9571	-3.2478	0.4195	0.0011	0.1730 (0.0000)
four	-0.0161	0.9612	-2.931	0.4529	0.0000	0.1812 (0.0000)
Crisis period: bins 1,2 and 3						
one	0.0220	1.0465	-2.3937	0.6767	0.0253	0.1563 (0.0018)
two	-0.0026	1.0209	-1.9123	0.5631	0.0405	0.2508 (0.0000)
three	-0.0194	0.9283	-1.5388	0.5746	0.0354	0.3704 (0.0000)
Crisis period: bin 4 + 5						
four+five	0.0000	1.4247	-0.8720	0.2067	0.0658	-0.005 (0.921)

Table 7: **HS VaR:** Pre-crisis period and crisis period HS VaR descriptive and performance statistics are reported for each bank. Descriptive statistics include HS VaR mean, standard deviation (std dev) and VaR exceedance rate (ex rate), as well as mean and maximum (max) exceedances (ex). Performance-related statistics include p-values for measures of 1% VaR accuracy (cover), exceedance independence (indep), duration between exceedances (dur) and VaR correlation with next-day |PnL| (with p-values in parentheses). The mean and max statistics are unsigned.

Pre-Crisis Period									
			ex	mean	max	p-values			cor$(-VaR_t,$
bank	mean	std dev	rate	ex	ex	cover	indep	dur	$\|PnL_{t+1}\|)$
bank 1	.84	.08	.007	.42	1.12	.335	.044	.886	−.093 (.002)
bank 2	.10	.06	.016	.07	.17	.065	.445	.317	.187 (.000)
bank 3	.54	.21	.020	.36	2.64	.001	.275	.850	.333 (.000)
bank 4	.14	.03	.012	.10	.36	.547	.604	.132	.121 (.000)
Crisis Period									
			ex	mean	max	p-values			cor$(-VaR_t$
bank	mean	std dev	rate	ex	ex	cov	indep	dur	$\|PnL_{t+1}\|)$
bank 1	3.84	1.23	.038	1.63	4.58	.000	.001	.001	.050 (.324)
bank 2	1.64	1.43	.068	2.54	26.66	.000	.040	.546	.053 (.288)
bank 3	3.14	.67	.027	1.70	7.31	.005	.028	.026	.071 (.147)
bank 4	1.60	.75	.041	3.22	24.24	.000	.028	.010	−.019 (.701)
bank 5	1.31	.48	.051	.75	3.89	.000	.018	.005	.275 (.000)

Table 8: **HS VaR Variance by Frequency:** Pre-crisis and crisis period HS VaR variances (total variance) and HS VaR variance decompositions by frequencies are reported in the top and bottom panels. Frequency bands correspond to periodicities less than 5 (weekly), 21 (monthly), and 63 (quarterly) days. Cumulative fraction of HS VaR variance is reported at each frequency range. Variance decomposition of HS VaR uses the Band-Pass filter with K = 30 leads and lags.

	total variance	≤5-day/total	≤21-day/total	≤63-day/total
Pre-Crisis Period				
bank 1	.006	.0035	.0147	.0486
bank 2	.004	.0005	.0050	.0242
bank 3	.043	.0007	.0051	.0253
bank 4	.001	.0012	.0076	.0282
Crisis Period				
bank 1	1.521	.0014	.0091	.0364
bank 2	2.054	.0010	.0069	.0275
bank 3	.451	.0015	.0220	.1213
bank 4	.567	.0008	.0060	.0277
bank 5	.234	.0012	.0083	.0373

Table 9: **GARCH-normal VaR:** Pre-crisis period and crisis period GARCH-normal VaR descriptive and performance statistics are reported for each bank. Descriptive statistics include GARCH-normal VaR mean, standard deviation (std dev) and VaR exceedance rate (ex rate), as well as mean and maximum (max) exceedances (ex). Performance-related statistics include p-values for measures of 1% VaR accuracy (cover), exceedance independence (indep), duration between exceedances (dur) and VaR correlation with next-day |PnL| (with p-values in parentheses). The mean and max statistics are unsigned.

Pre-Crisis Period									
			ex	mean	max	———	p-values	———	cor$(-VaR_t,$
bank	mean	std dev	rate	ex	ex	cover	indep	dur	$\lvert PnL_{t+1}\rvert)$
bank 1	.70	.21	.010	.41	.98	.993	.095	.840	−.004 (.882)
bank 2	.12	.08	.011	.05	.10	.836	.612	.623	.245 (.000)
bank 3	.72	.37	.009	.28	1.40	.706	.626	.382	.211 (.000)
bank 4	.23	.11	.007	.11	.27	.267	.778	.655	.111 (.001)
Crisis Period									
			ex	mean	max	———	p-values	———	cor$(-VaR_t$
bank	mean	std dev	rate	ex	ex	cov	indep	dur	$\lvert PnL_{t+1}\rvert)$
bank 1	2.93	1.75	.033	1.48	3.74	.000	.435	.650	.166 (.001)
bank 2	3.02	4.09	.034	4.41	26.59	.000	.081	.911	.005 (.925)
bank 3	2.86	1.56	.031	1.28	8.09	.000	.357	.600	.078 (.111)
bank 4	3.06	2.93	.024	4.79	23.59	.014	.018	.097	−.035 (.482)
bank 5	1.69	1.37	.031	.52	4.19	.000	.416	.704	.295 (.000)

Table 10: **GARCH-normal VaR Variance by Frequency:** Pre-crisis and crisis period GARCH-normal VaR variances (total variance) and GARCH-normal VaR variance decompositions by frequencies are reported in the top and bottom panels. Frequency bands correspond to periodicities less than 5 (weekly), 21 (monthly), and 63 (quarterly) days. Cumulative fraction of GARCH-normal VaR variance is reported at each frequency range. Variance decomposition of GARCH-normal VaR uses the Band-Pass filter with K = 30 leads and lags.

	total variance	≤5-day/total	≤21-day/total	≤63-day/total
Pre-Crisis Period				
bank 1	.045	.0602	.2803	.4879
bank 2	.007	.0137	.0385	.0864
bank 3	.138	.0484	.1025	.1680
bank 4	.012	.2140	.4621	.5784
Crisis Period				
bank 1	3.075	.1340	.4213	.8064
bank 2	16.765	.2769	.7633	.8965
bank 3	2.422	.0952	.2327	.5312
bank 4	8.607	.1252	.2391	.3220
bank 5	1.889	.0622	.1951	.3234

Table 11: **GARCH-normal Volatility Correlations:** Bivariate correlations and p-values between 1-day ahead GARCH-normal volatility forecasts ($\hat{\sigma}_{t+1}$) and next-day realized PnL volatility ($|PnL_{t+1}|$) are reported in the first two columns, with p-values. Correlations between GARCH-normal volatility forecasts and realized GARCH-normal volatility forecast error ($|PnL_{t+1}| - \hat{\sigma}_{t+1}$) are presented in the last two columns with p-values.

	$cor(\hat{\sigma}_{t,t+1}, \|PnL_{t+1}\|)$		$cor(\hat{\sigma}_{t+1}, \|PnL_{t+1}\| - \hat{\sigma}_{t+1})$	
	cor	p-value	cor	p-value
		Pre-Crisis Period		
bank 1	.035	.247	-.148	.000
bank 2	.254	.000	-.083	.005
bank 3	.318	.000	-.052	.048
bank 4	.272	.000	.045	.179
		Crisis Period		
bank 1	.185	.000	-.312	.000
bank 2	-.006	.898	-.753	.000
bank 3	.133	.007	-.326	.000
bank 4	-.037	.451	-.792	.000
bank 5	.307	.000	-.391	.000

Table 12: **VaR Measure Comparisons: Mean VaRs:** Pre-crisis and crisis period comparisons of mean VaRs among bank VaR and the different benchmark VaRs (HS, GARCH-normal, FHS, GARCH-t and GARCH-EVT) are shown for each bank. All table entries are unsigned.

Pre-Crisis Period						
	Bank VaR	HS	GARCH-N	FHS	GARCH-t	GARCH-EVT
bank 1	1.46	0.84	0.70	0.79	0.86	0.81
bank 2	0.32	0.10	0.12	0.11	0.15	0.11
bank 3	1.27	0.54	0.72	0.59	0.85	0.61
bank 4	0.52	0.14	0.23	0.16	0.32	0.20
Crisis Period						
	Bank VaR	HS	GARCH-N	FHS	GARCH-t	GARCH-EVT
bank 1	3.31	3.84	2.93	3.34	3.32	3.57
bank 2	0.69	1.64	3.02	1.99	1.70	2.31
bank 3	2.97	3.14	2.86	3.51	3.10	3.33
bank 4	0.97	1.60	3.06	2.28	1.40	1.97
bank 5	1.49	131	1.69	1.75	1.88	1.86

Table 13: **VaR Measure Comparisons: Exccedance Rates:** Pre-crisis and crisis period comparisons of exceedance rates among the bank VaR and different benchmark VaRs (HS, GARCH-normal, FHS, GARCH-t and GARCH-EVT) are shown for each bank.

Pre-Crisis Period						
	Bank VaR	HS	GARCH-N	FHS	GARCH-t	GARCH-EVT
bank 1	0.004^{b}	0.007^{b}	0.010	0.010	0.007^{b}	0.008
bank 2	0.000	0.016	0.011	0.012	0.008	0.013
bank 3	0.001^{a}	0.020^{a}	0.009	0.014	0.003^{a}	0.015
bank 4	0.000^{a}	0.012	0.007	0.013	0.003^{a}	0.010
Crisis Period						
	Bank VaR	HS	GARCH-N	FHS	GARCH-t	GARCH-EVT
bank 1	$0.040^{a,b}$	$0.038^{a,b}$	0.033^{a}	0.028^{a}	0.028^{a}	0.025^{a}
bank 2	0.100^{a}	$0.068^{a,b}$	0.034^{a}	0.049^{a}	0.071^{a}	0.036^{a}
bank 3	0.034^{a}	$0.027^{a,b}$	0.031^{a}	0.019	0.022^{a}	0.019
bank 4	$0.051^{a,b}$	$0.041^{a,b}$	$0.024^{a,b}$	0.036^{a}	0.036^{a}	$0.044^{a,b}$
bank 5	$0.027^{a,b}$	$0.051^{a,b}$	0.031^{a}	0.029^{a}	0.017	0.027^{a}

[a] VaR 1% coverage rejected at 5% level.

[b] Exceedance independence rejected at 5% level.

Table 14: **VaR Measure Comparisons: Correlation ($-VaR_t$, $|PnL_{t+1}|$):** Pre-crisis and crisis period comparisons of bivariate correlations between VaR measures and next-day realized PnL volatilty (|PnL|). The VaR-|PnL| correlations for bank VaR and the different benchmark VaRs (HS, GARCH-normal, FHS, GARCH-t and GARCH-EVT) are shown for each bank.

Pre-Crisis Period						
	Bank VaR	HS	GARCH-N	FHS	GARCH-t	GARCH-EVT
bank 1	.051	-.093	-.004	-.026	-.020	-.009
bank 2	.156*	.187*	.245*	.220*	.243*	.230*
bank 3	.410*	.333*	.211*	.220*	.231*	.219*
bank 4	.301*	,121*	.111*	-.034	.145*	.015
Crisis Period						
	Bank VaR	HS	GARCH-N	FHS	GARCH-t	GARCH-EVT
bank 1	.132*	.050	.166*	.169*	.172*	.170*
bank 2	.071	.053	.005	-.000	-.008	-.004
bank 3	.162*	.071	.078	.089	.073	.084
bank 4	-.035	-.019	-.035*	-.035	.011	-.024
bank 5	.389*	.275*	.295*	.291*	.321*	.302*

*Significance at 5% level.

Table 15: **VaR Measure Comparisons: VaR Variance Contributions (periodicity less than one quarter):** Pre-crisis and crisis period comparisons of VaR variance contributions for periodicity $\leq$ 63 days among the bank VaR and different benchmark VaRs (HS, GARCH-normal, FHS, GARCH-t and GARCH-EVT) are shown for each bank.

Pre-Crisis Period						
	Bank VaR	HS	GARCH-N	FHS	GARCH-t	GARCH-EVT
bank 1	0.361	0.049	0.488	0.389	0.272	0.464
bank 2	0.176	0.024	0.086	0.072	0.090	0.078
bank 3	0.056	0.025	0.168	0.190	0.132	0.174
bank 4	0.093	0.028	0.578	0.603	0.192	0.649
Crisis Period						
	Bank VaR	HS	GARCH-N	FHS	GARCH-t	GARCH-EVT
bank 1	0.104	0.036	0.806	0.777	0.774	0.792
bank 2	0.104	0.027	0.897	0.861	0.969	0.926
bank 3	0.224	0.121	0.531	0.501	0.484	0.484
bank 4	0.154	0.028	0.322	0.412	0.954	0.487
bank 5	0.107	0.037	0.323	0.328	0.273	0.280

Table 16: **Average Expected Shortfall measures:** Average Expected Shortfall (ES) is shown for different benchmark VaR methodologies for each bank for pre-crisis and crisis periods. Average ES is the mean estimated daily ES value for the particular category. All table entries are unsigned.

Pre-Crisis Period				
bank	GARCH-normal	FHS	GARCH-t	GARCH-EVT
bank 1	.88	1.17	1.28	1.19
bank 2	.16	.15	.22	.15
bank 3	.89	.79	1.17	.77
bank 4	.29	.27	.50	.27
Crisis Period				
bank	GARCH-normal	FHS	GARCH-t	GARCH-EVT
bank 1	3.42	5.23	4.56	4.88
bank 2	3.48	10.02	3.00	8.07
bank 3	3.39	4.41	4.08	4.35
bank 4	3.54	9.15	2.40	5.45
bank 5	2.00	2.54	2.47	2.47

Table 17: **GARCH-EVT ES:** GARCH-EVT ES estimates and supporting statistics are shown for each bank for pre-crisis and crisis-periods. Columns 1 and 2: Unconditional mean ES and standard deviation. Column 3: Mean ES conditional on VaR exceedance. Column 4: Mean VaR exceedance conditional on exceedance occurance. Columns 5, 6,and 7 and 8 respectively: exceedance rate, mean exceedance, maximium exceedance, correlation between ES and next-day realized PnL, given PnL less than estimated ES. P-values are in parentheses.

Pre-Crisis Period								
bank	est unc mean ES	est unc std ES	est cond mean ES	act cond mean ES	ex rate	mean ex	max ex	cor($\|PnL\|, ES$ $\|PnL < VaR)$
bank 1	1.186	.300	1.179	1.192	.004	.247	.607	.754 (.019)
bank 2	.153	.099	.132	.146	.007	.047	.085	.867 (.000)
bank 3	.766	.462	.806	.920	.008	.266	1.472	.902 (.000)
bank 4	.265	.106	.156	.185	.004	.134	.291	−.275 (.474)
Crisis Period								
bank	est unc mean ES	est unc std ES	est cond mean ES	act cond mean ES	ex rate	mean ex	max ex	cor($\|PnL\|, ES$ $\|PnL < VaR)$
bank 1	4.881	2.781	3.155	3.896	.018	1.285	3.080	.766 (.010)
bank 2	8.071	9.582	4.295	5.823	.017	7.695	26.488	−.046 (.870)
bank 3	4.353	2.221	3.361	4.198	.007	2.781	6.677	.435 (.281)
bank 4	5.446	4.671	2.106	3.852	.022	4.948	21.560	.233 (.352)
bank 5	2.474	2.138	1.398	1.684	.017	.660	3.955	.422 (.197)

Table 18: **Bank Daily PnL (all observations):** See description for Table 1.

Pre-Crisis Period (all observations)						
bank	dates	mean	st dev	PnL < 0	kurtosis	skewness
bank 1	2001 - May 2007	.5468	.5270	.109	5.582	.155
bank 2	2001 - May 2007	.1355	.1086	.062	6.455	.877
bank 3	1999 - May 2007	.4279	.5198	.147	7.570	.739
bank 4	2001 - May 2007	.2317	.2152	.073	44.883	3.440
bank 5	2005 - May 2007	.4181	.4130	.143	4.004	.301

Table 19: **Bank Daily VaR (all observations):** See description for Table 3.

Pre-Crisis Period (all observations)									
			ex	mean	max	———	p-values	———	cor($-VaR_t$,
bank	mean	std dev	rate	ex	ex	cover	indep	dur	$\lvert PnL_{t+1}\rvert$)
bank 1	1.51	.37	.003	.26	.62	.001	.009	.292	.057 (.022)
bank 2	.28	.07	.000	–	–	–	–	–	.167 (.000)
bank 3	1.09	.51	.001	.99	1.25	.000	.949	–	.435 (.000)
bank 4	.45	.14	.001	.04	.04	.000	.970	–	.341 (.000)
bank 5	.78	.15	.000	–	–	–	–	–	.016 (.684)

Table 20: **1-year HS VaR:** See description for Table 7.

Pre-Crisis Period									
			ex	mean	max	———	p-values	———	cor($-VaR_t$,
bank	mean	std dev	rate	ex	ex	cover	indep	dur	$\lvert PnL_{t+1}\rvert$)
bank 1	.78	.26	.011	.40	1.17	.699	.155	.807	−.063 (.021)
bank 2	.09	.06	.016	.06	.16	.042	.360	.492	.171 (.000)
bank 3	.60	.43	.017	.32	2.58	.007	.315	.620	.294 (.000)
bank 4	.14	.04	.011	.10	.36	.699	.588	.550	.065 (.026)
Crisis Period									
			ex	mean	max	———	p-values	———	cor($-VaR_t$
bank	mean	std dev	rate	ex	ex	cov	indep	dur	$\lvert PnL_{t+1}\rvert$)
bank 1	4.35	1.48	.038	1.28	4.30	.000	.014	.000	.003 (.958)
bank 2	6.89	6.25	.034	4.09	26.62	.000	.081	.041	.019 (.708)
bank 3	3.30	.75	.024	1.59	7.18	.014	.018	.051	.073 (.140)
bank 4	2.39	1.12	.041	3.03	23.39	.000	.000	.003	−.026 (.603)
bank 5	1.68	.69	.039	.70	3.28	.000	.002	.006	.243 (.000)

Table 21: **1-year HS VaR Variance by Frequency:** See description for Table 8.

	total variance	≤5-day/total	≤21-day/total	≤63-day/total
Pre-Crisis Period				
bank 1	.066	.0015	.0102	.0386
bank 2	.004	.0019	.0098	.0400
bank 3	.187	.0013	.0065	.0275
bank 4	.002	.0024	.0125	.0427
Crisis Period				
bank 1	2.196	.0029	.0159	.0488
bank 2	39.102	.0028	.0140	.0444
bank 3	.562	.0057	.0277	.1331
bank 4	1.262	.0037	.0194	.0579
bank 5	.478	.0031	.0172	.0601

Figure 1: **Bank PnL and bank VaR: Pre-crisis period.** Scatter plots for Bank PnL and bank VaR are presented for 4 banks in the pre-crisis period. The 45-degree line separates points with and without VaR violations. The large solid points denote PnL and VaR pairs with VaR violations.

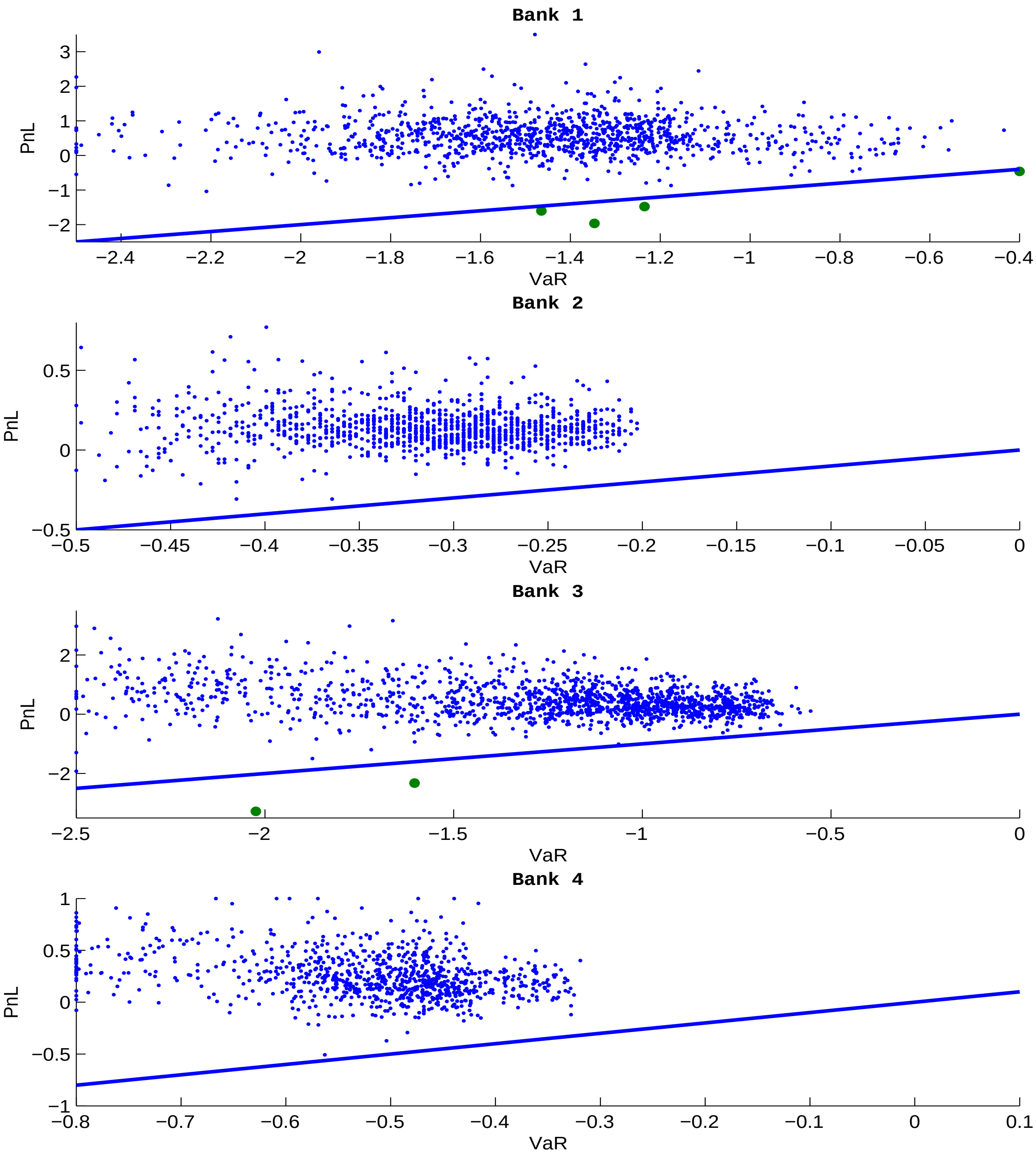


Figure 2: **Bank PnL and bank VaR: Crisis period.** Scatter plots for Bank PnL and bank VaR are presented for 5 banks in the crisis period. The 45-degree line separates points with and without VaR violations. The solid points and ellipsoids below the line denote respectively individual exceedances or groups of exceedances. The plotted ellipsoids are the smallest containing all points within each subgroup.

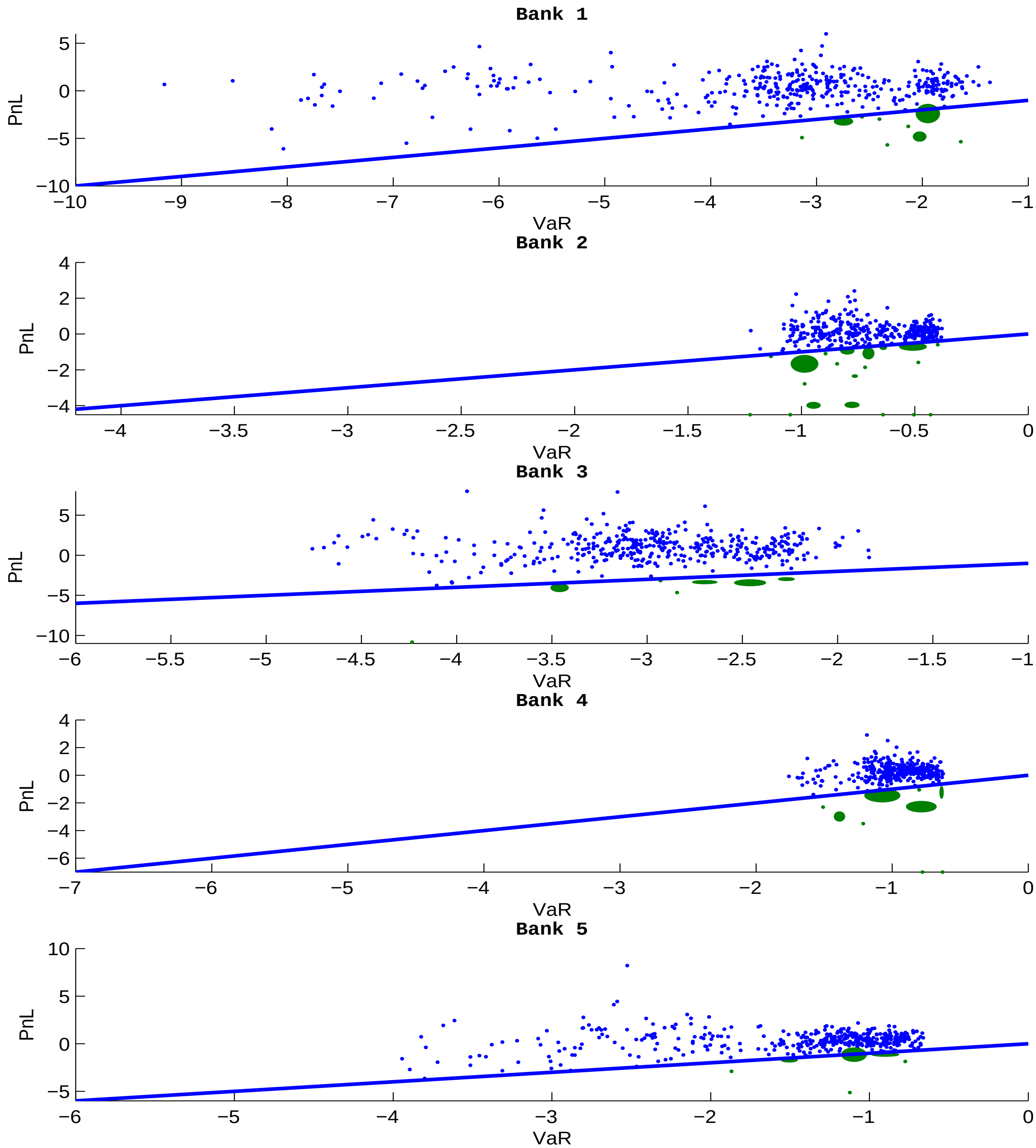
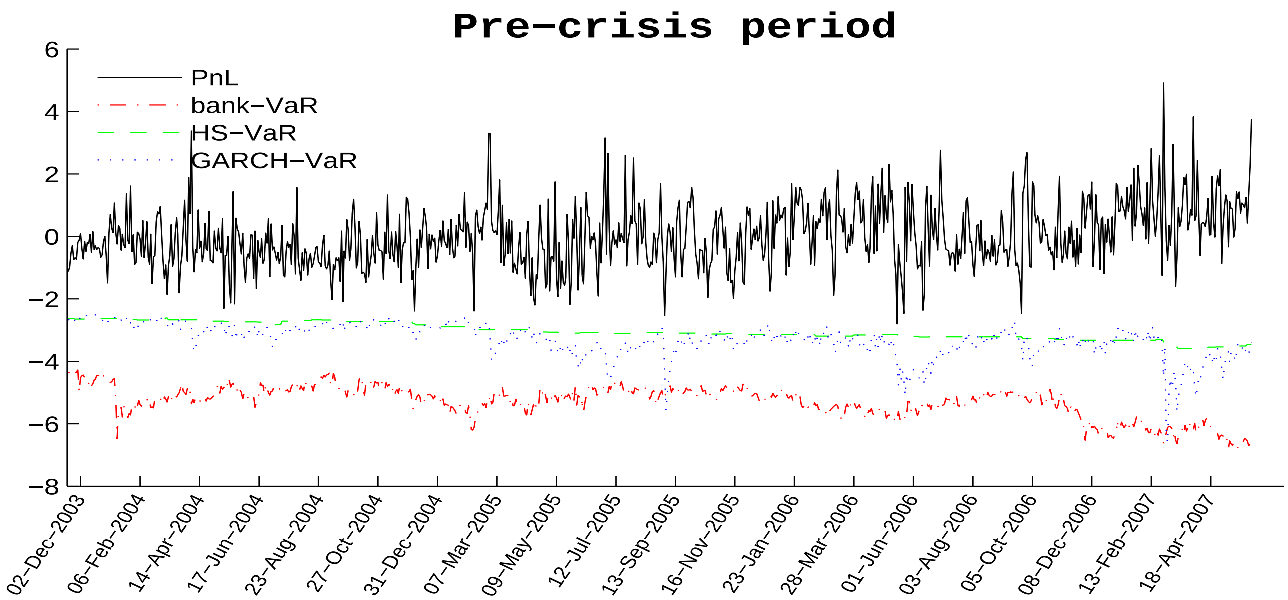


Figure 3: **Average Bank PnL, Bank VaR, HS VaR and GARCH-normal VaR.** We plot average bank PnL, bank VaR and benchmark VaRs for the pre-crisis period and the crisis-period. For each period, both daily VaRs and PnLs are first standardized by the the individual bank's PnL mean and standard deviation for the respective period and then aggregated. The solid line denotes average bank PnL, the dashdot line denotes bank VaR, the dashed line denotes benchmark HS VaR and the dotted line denotes benchmark GARCH-normal VaR. By construction, the plot covers the dates for which data for all banks is available from December 2003 to December 2008.



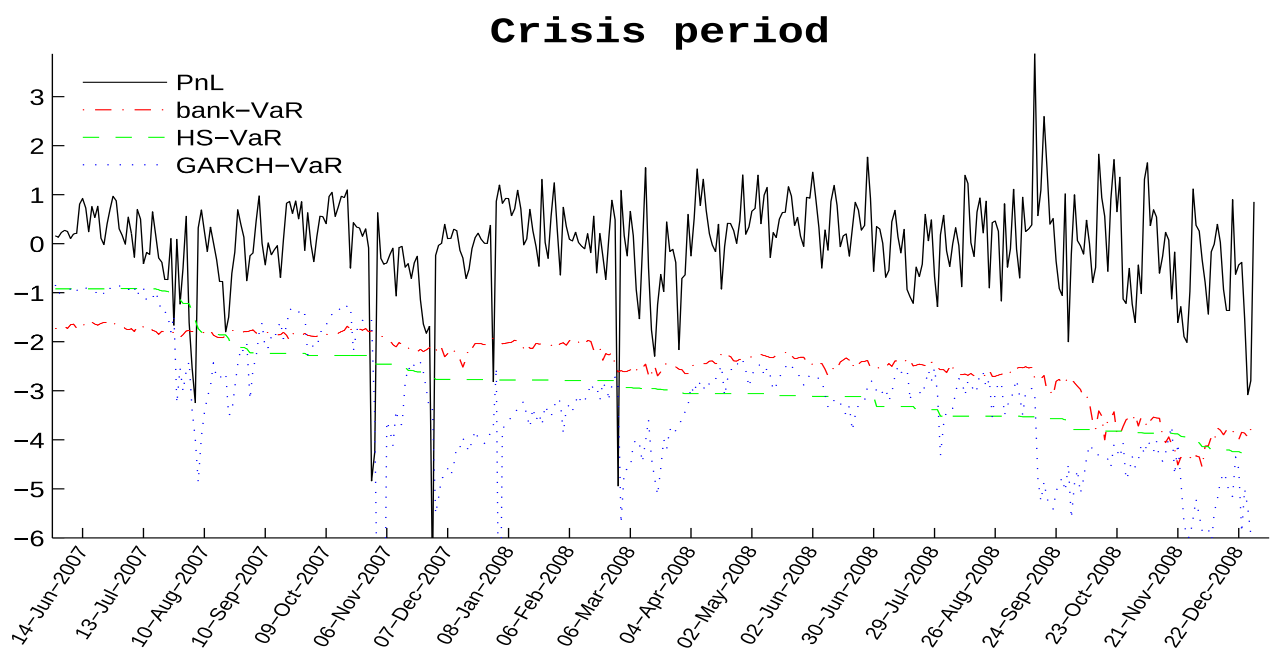


Figure 4: **Bank PnL and bank VaR across ordered categories.** The Figure shows the pre-crisis and crisis period bank order-category VaR and PnL. On each day, the standardized banks' PnL and VaR are sorted into bin categories based on the relative size of the respective standardized banks' VaRs for that day. The number of bins is set equal to the number of banks in each subperiod: 4 for the pre-crisis and 5 for the crisis period. A lower VaR conservativeness is associated with a higher bin number. The sum of VaRs and PnLs for bins 4 and 5, that are composed solely of two outlier banks, are plotted rather than individual plots. By construction, the plot covers the dates for which data for all banks is available from December 2003 to December 2008.

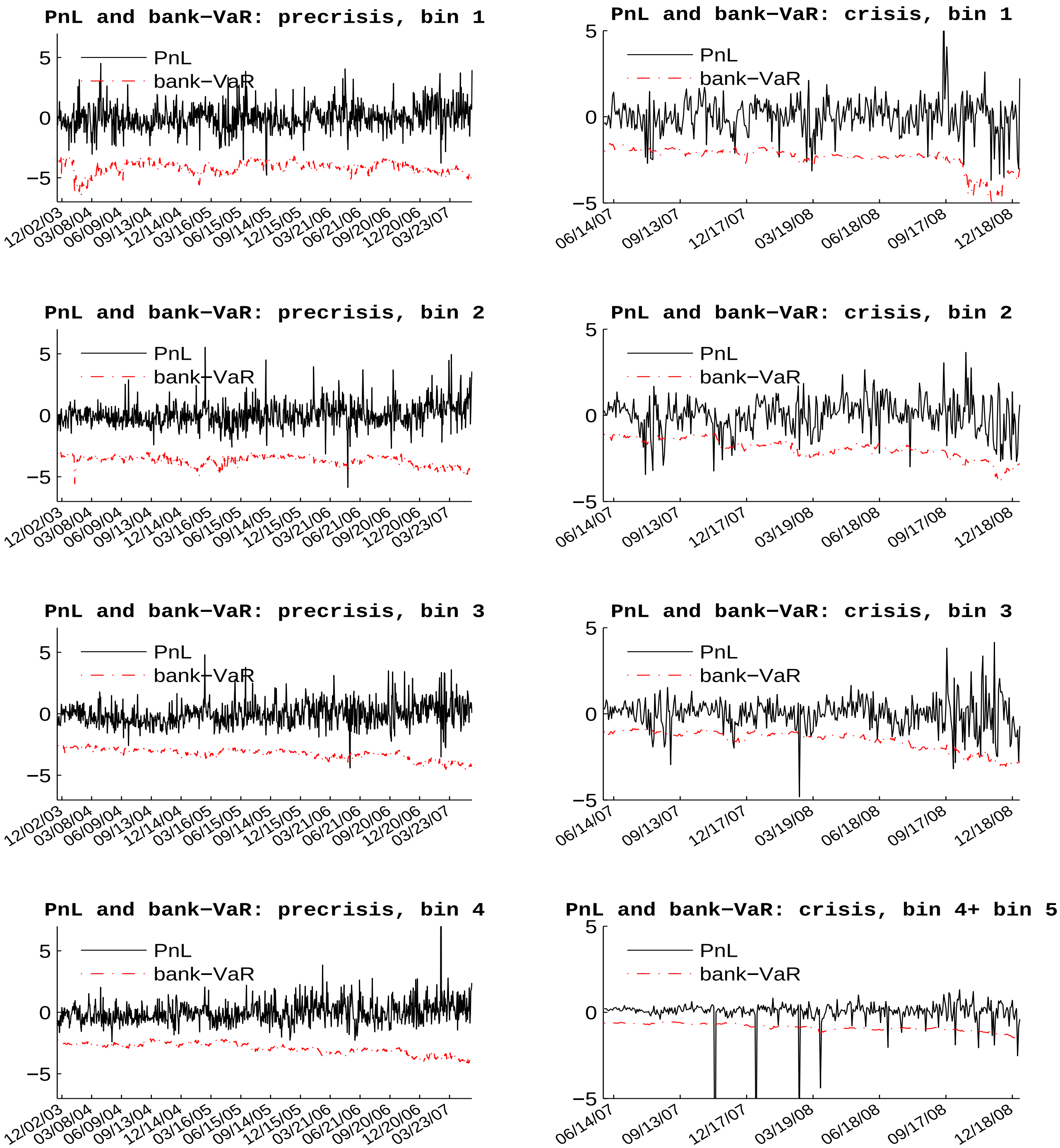


Figure 5: **Bank PnL and HS VaR: Pre-crisis period.** Scatter plots for Bank PnL and benchmark HS VaR (2-year window) are presented for 4 banks in the pre-crisis period. The 45-degree line separates points with and without VaR violations. The large solid points denote PnL and VaR pairs with VaR violations.

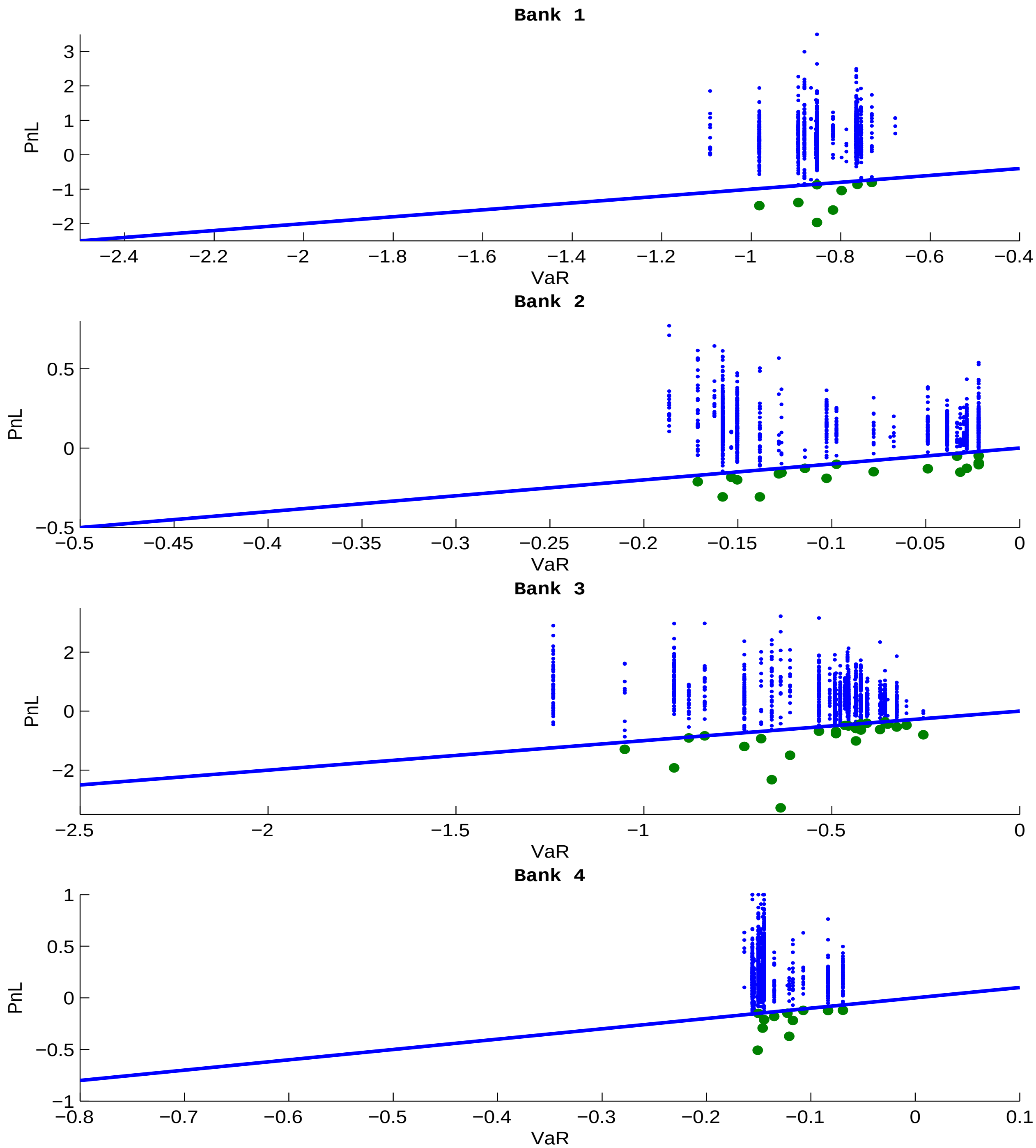


Figure 6: **Bank PnL and HS VaR: Crisis period.** Scatter plots for Bank PnL and benchmark HS VaR (2-year window) are presented for 5 banks in the crisis period. The 45-degree line separates points with and without VaR violations. The large solid points denote PnL and VaR pairs with VaR violations.

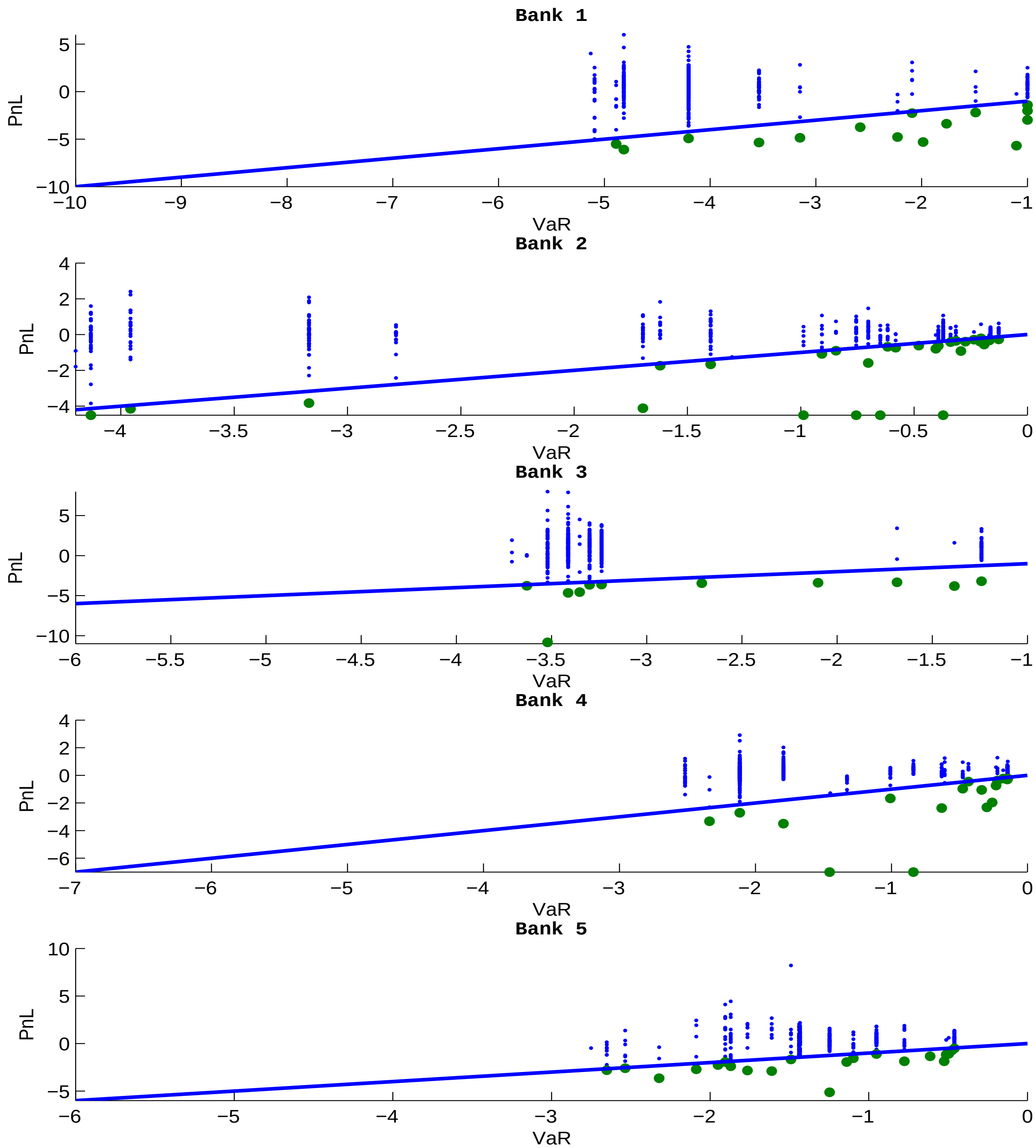


Figure 7: **Bank PnL and GARCH-normal VaR: Pre-crisis period.** Scatter plots for Bank PnL and benchmark GARCH-normal VaR (2-year window) are presented for 4 banks in the pre-crisis period. The 45-degree line separates points with and without VaR violations. The large solid points denote PnL and VaR pairs with VaR violations.

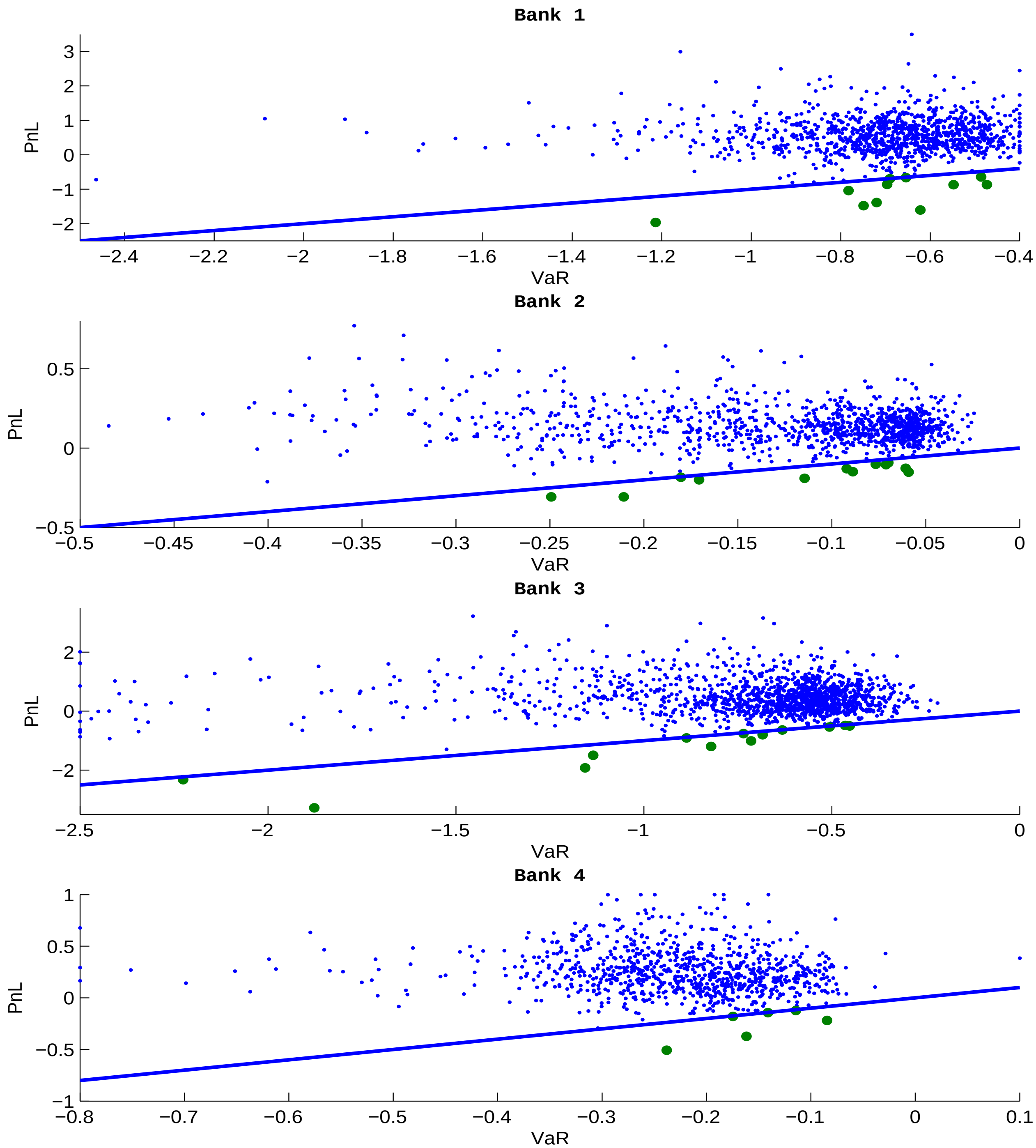


Figure 8: **Bank PnL and GARCH-normal VaR: Crisis period.** Scatter plots for Bank PnL and benchmark GARCH-normal VaR (2-year window) are presented for 5 banks in the crisis period. The 45-degree line separates points with and without VaR violations. The large solid points denote PnL and VaR pairs with VaR violations.

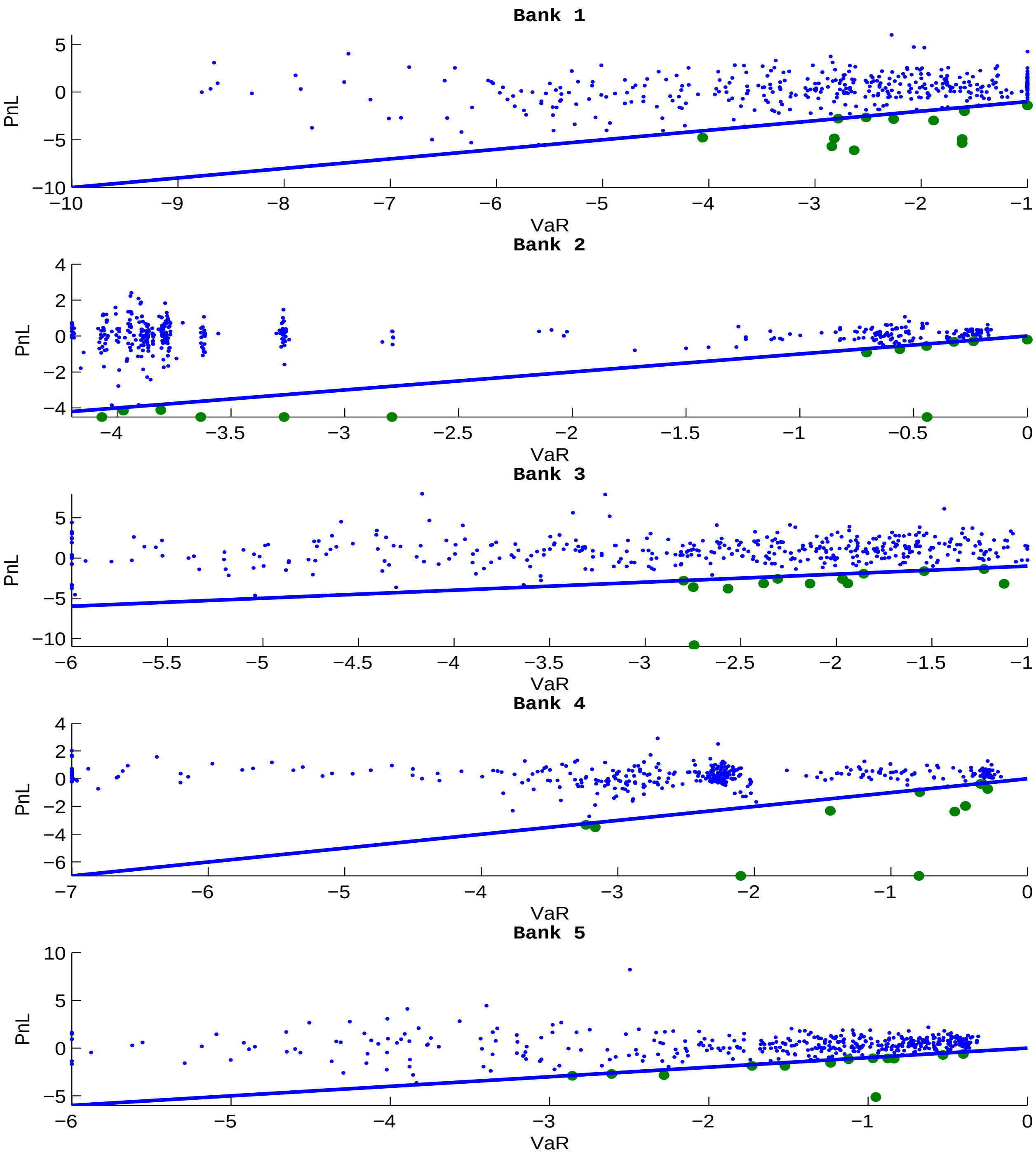


Figure 9: **Bank PnL and FHS VaR: Pre-crisis period.** Scatter plots for Bank PnL and benchmark FHS VaR (2-year window) are presented for 4 banks in the pre-crisis period. The 45-degree line separates points with and without VaR violations. The large solid points denote PnL and VaR pairs with VaR violations.

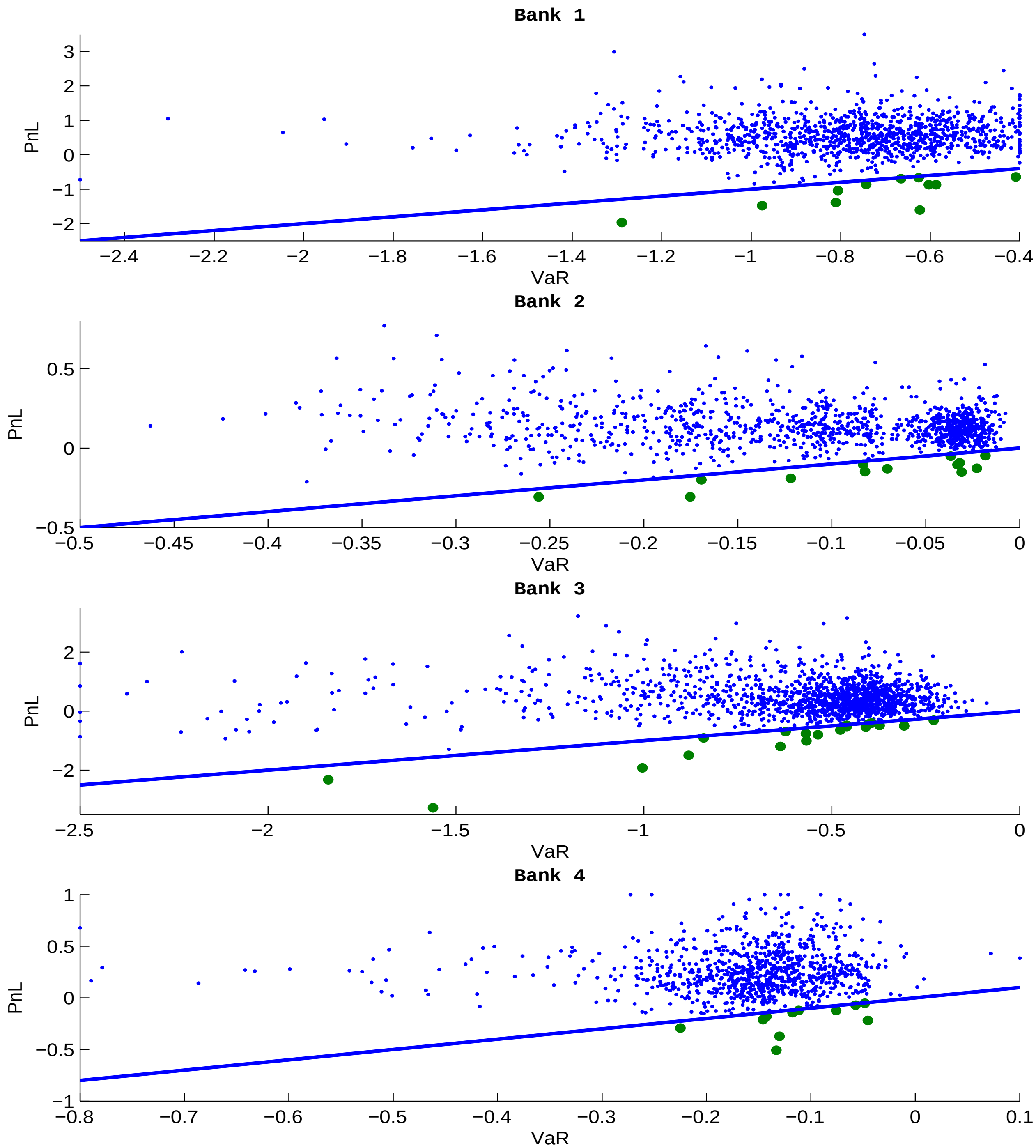


Figure 10: **Bank PnL and FHS VaR: Crisis period.** Scatter plots for Bank PnL and benchmark FHS VaR (2-year window) are presented for 5 banks in the crisis period. The 45-degree line separates points with and without VaR violations. The large solid points denote PnL and VaR pairs with VaR violations.

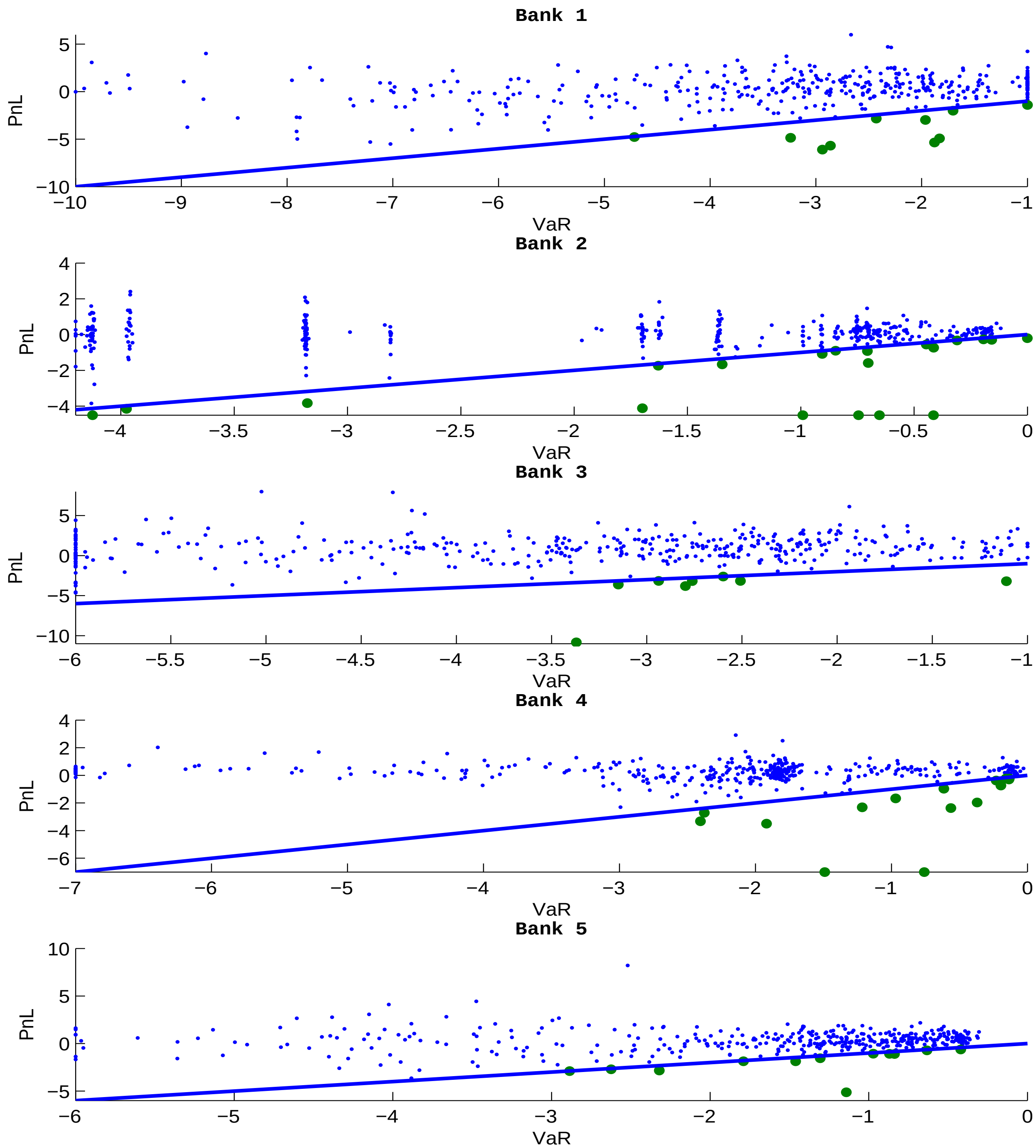


Figure 11: **Bank PnL and 1-year HS VaR: Pre-crisis period.** Scatter plots for Bank PnL and benchmark HS VaR (1-year window) are presented for 4 banks in the pre-crisis period. The 45-degree line separates points with and without VaR violations. The large solid points denote PnL and VaR pairs with VaR violations.

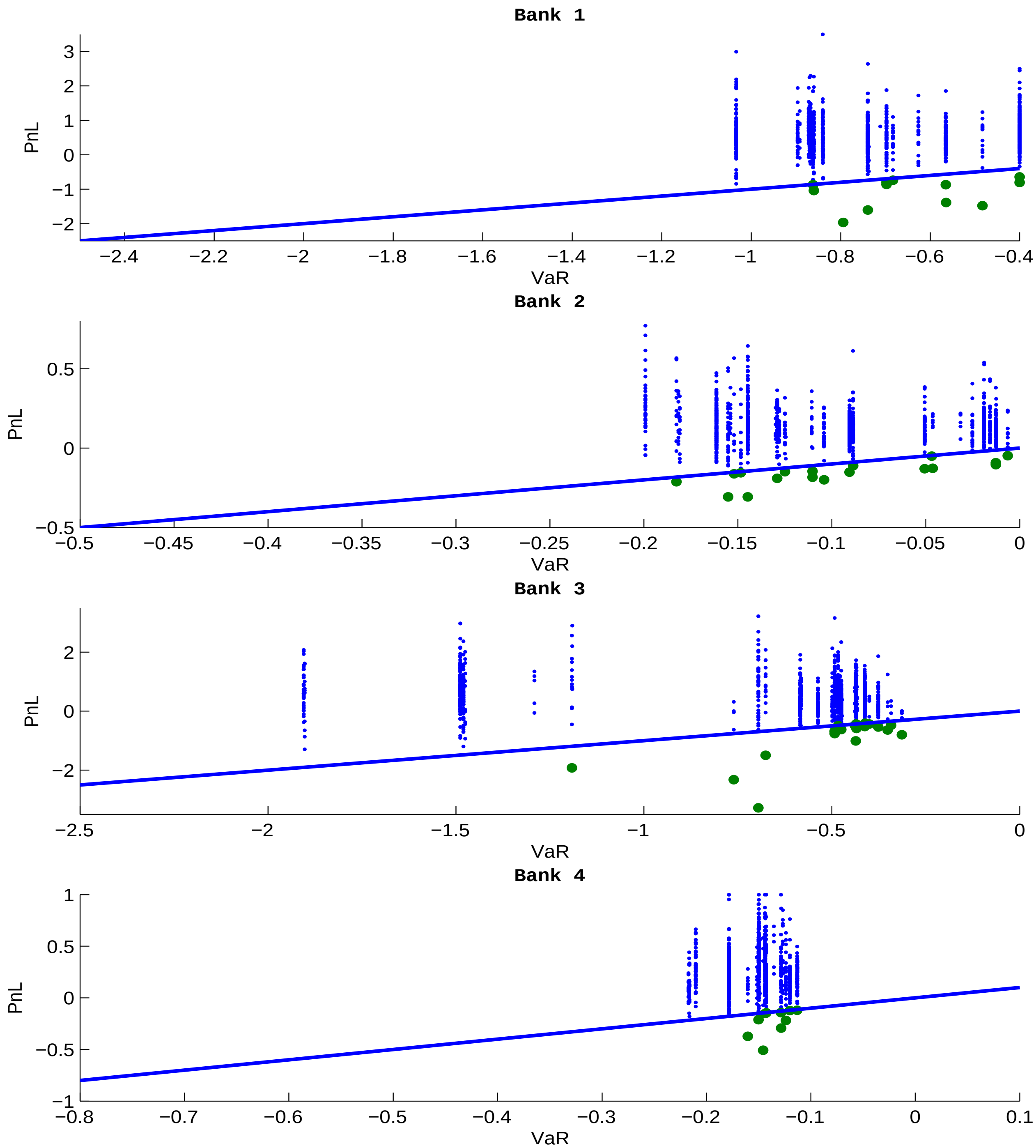


Figure 12: **Bank PnL and 1-year HS VaR: Crisis period.** Scatter plots for Bank PnL and benchmark HS VaR (1-year window) are presented for 5 banks in the crisis period. The 45-degree line separates points with and without VaR violations. The large solid points denote PnL and VaR pairs with VaR violations.