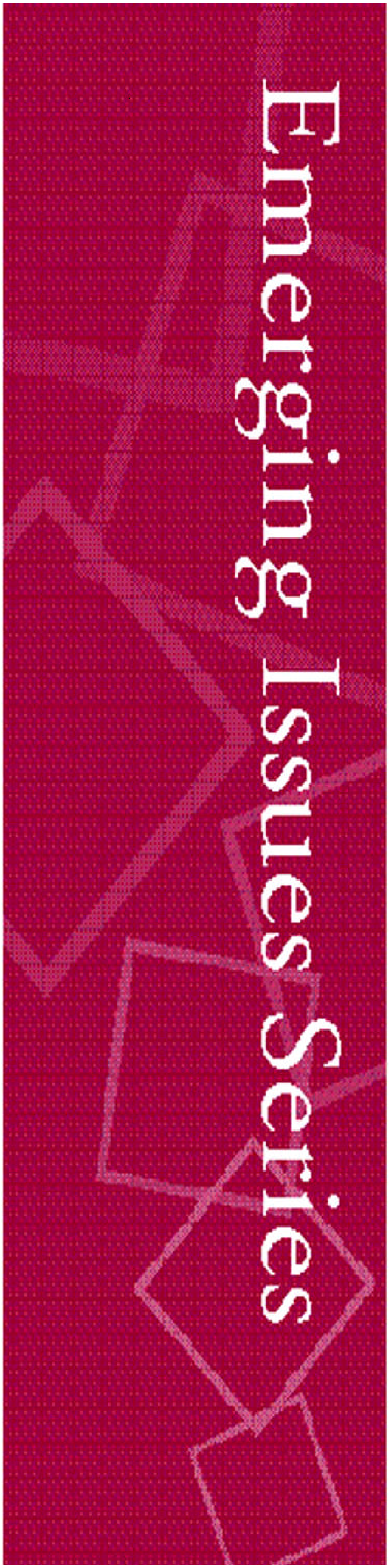



**Simple Forecasts of Bank Loan Quality in the Business Cycle**

Michele Gambera

Emerging Issues Series
Supervision and Regulation Department
Federal Reserve Bank of Chicago
April 2000 (S&R-2000-3)

# Simple Forecasts of Bank Loan Quality in the Business Cycle

**Michele Gambera**

**April 2000**

---

Address correspondence to Michele Gambera, Supervision and Regulation, Federal Reserve Bank of Chicago, 230 South LaSalle Street, Chicago, IL 60604-1413. Phone 312-322-5690, Fax 312-322-5894, E-mail: Michele.Gambera@Chi.Frb.Org. Requests for additional copies should be directed to the Public Information Center, Federal Reserve Bank of Chicago, P.O. Box 834, Chicago, Illinois 60690-0834, or telephone (312) 322-5111. The Federal Reserve Bank of Chicago's Web site can be accessed at: http://www.chicagofed.org/publications/workingpapers/emergingissues.cfm

# Simple Forecasts of Bank Loan Quality in the Business Cycle

**Michele Gambera***

April 2000

**Abstract**

Experience from models such as SEER suggests that bank financial conditions predict bank failures. However, it has been difficult to find a relationship between macroeconomic variables and bank failures. This paper shows ways in which simple time-series techniques can be used to forecast financial conditions of banks. The models include macroeconomic variables in order to consider systemic cyclical factors in forecasting. In addition, analysis of regression residuals is used to obtain relatively early warnings of unusual performance.

The empirical results suggest that a limited number of regional and national macroeconomic variables are often good predictors for problem-loan ratios, and that simple, bivariate VAR systems of one bank variable, one macroeconomic variable, and seasonal dummies can be quite effective. These variables include bankruptcy filings, farm income (particularly for states where farming has an important role), state annual product, housing permits, and unemployment. Analysis of the residuals is shown to be an interesting tool to detect unexpected changes in past-due loans. Impulse-response functions are a result of VAR estimation, which can be used for scenario analysis.

Studies of profitability and loan quality in banks usually focus on the financial factors of each bank, such as capitalization ratios or measures of efficiency. For example, Hiemstra *et al.* (1997) analyze how capital-to-asset ratios affect the probability of bank failure, while Swamy *et al.* (1998) consider each bank as a portfolio of assets and, in a model similar to the CAPM, study bank profits using 'portfolio' shares of each of the assets and their returns as regressors. The Federal Reserve has developed the Financial Institutions Monitoring Service (FIMS, often called SEER: see Cole *et al.*

*Policy and Emerging Issues Group, Federal Reserve Bank of Chicago, 230 S. LaSalle St., Chicago, IL 60604–1413. E-mail: `Michele.Gambera@chi.frb.org`. The views expressed are those of the author, and do not necessarily reflect official positions of the Federal Reserve Bank of Chicago or the Federal Reserve System.

(1995)), which consists of two models: one predicts the probability of bank failure in the following eight quarters, and the other forecasts examination ratings (CAMELS) for the financial institutions, which are to undergo exams in the following quarter. Those models are based on bank financial conditions from the latest data from the Report of Condition and Income (Call Report). Such models perform well, even without explicitly considering macroeconomic variables, because recent financial conditions are a good predictor of a bank's failure in the near future.

Given that financial conditions predict failures, it is interesting to predict financial conditions. This paper shows ways in which **simple** time-series techniques can be used to forecast financial conditions of banks, as well as to obtain relatively early warnings of unusual performance based on macroeconomic performance. The time-series approach is justified by the large dynamic (including cyclical and seasonal) components of the series of bank financial variables, such as non-performing loans. In turn, this is explained by the wide exposure of both banks and borrowers to volatile macroeconomic variables, e.g., interest rates, unemployment, and business failures. This study's departure from the traditional approach is therefore twofold: first, external (i.e., not bank-specific), cyclical macroeconomic factors are considered; second (and consequently), the approach emphasizes time series and not cross-sectional data.

Cross-sectional studies do not consider macroeconomic factors because (as can be expected) the current financial conditions of a single bank are the best predictors of its probability of failure in the near future. Time-series analysis is interesting in this framework because it focuses on the effects of business cycles on bank asset quality. This is a complement to the cross-sectional approach because cross sections take innovations in bank financial conditions as exogenous. Instead, this paper tries to identify factors that—based on the state of the economy—can determine and forecast a bank's financial ratios. Clearly, a panel (pooled time-series and cross-section) approach could exploit more of the available information, but it would increase the computational burden. The aim in this paper is to achieve the best forecasts under a constraint of using quite simple econometric tools.[1]

There is limited empirical evidence about the effects of macroeconomic factors on bank asset quality (see the recent analysis of rural banks using county- and state-level data by Meyer and Yeager (2000)). This is surprising because a large body of analysis focuses on the effects of banks on the macroeconomy, and so one would

[1]This author has been developing a model using panel approach to predict loan quality ratios of single banks jointly with Don Conner of the FR Board. A report should be ready in the second quarter of 2000.

expect the opposite relationship to have also been studied.[2] Therefore, this study has a practical use in the macroeconomic analysis of the dynamics of lending and asset quality; moreover, insights can be gained about future levels of problem loans and probabilities of failure—which are of direct interest to supervisors and market analysts.

Many models predicting bank failures do not consider the macroeconomic environment, thus probably underestimating cyclical factors, which are systemic. This paper focuses on the average behavior of groups of banks within the business cycle; the groups are formed considering the geographical location as well as the asset size of the individual banks. The tools used also permit analysis of the behavior of a single bank, and comparisons with the bank's peers.[3]
The models in this paper can be used:

1. To *forecast* future levels of past-due loans (predict asset quality).

2. To evaluate past *deviations* (spot potential problems using residual analysis).

3. To *stress-test* assets with a what-if analysis (assess risk sensitivity).

The **forecasts** in this paper suggest that a limited number of regional macroeconomic variables are often good predictors for problem-loan ratios. These variables include state bankruptcy filings, farming income (particularly for states where farming has an important role), state annual product, housing permits, and national unemployment.

Evaluation of past **deviations** analyzes the residuals (regression errors): when the actual data "surprise" the model, that is, when the residual is very large, a bank supervisor is warned that something unpredicted occurred. This yields an early-warning system.

**Stress-testing** exercises can be carried out in this paper using the models presented as well as impulse-response functions. Moreover, the estimated models can be easily translated into an electronic spreadsheet, which can be used by examiners in a

[2] A recent theoretical paper by Azariadis and Smith (1998) introduces a dynamic model including a business cycle and borrowers of different quality. The paper's results suggests that "credit crunches" may be an optimal response by banks during recessions when there is uncertainty about the quality of borrowers. Logically, an empirical investigation of the problem would require researchers to first assess the timing and the probable causes of variations in problem loans, which in turn are likely to induce a switch in bank lending patterns.

[3] In the model with Conner mentioned in footnote 1, banks are pooled using statistical techniques (cluster analysis) and not by grouping them by size and location of head office (which is done in this paper).

what-if scenario analysis. This is a potentially very interesting application, because bank supervisors are particularly interested in learning about negative changes in bank performance induced by cyclical downturns.

# 1 Models

Some relatively simple econometric techniques can be used to analyze bank data and obtain more precise information than one would get by simply charting the series.

A linear model (see box A) is a very easy forecasting tool. The dependent variable is the one that is to be predicted, and the regressors include past values of the dependent variable, as well as macroeconomic variables, and a deterministic part (e.g., a constant term).

Simplicity is the advantage of the linear approach: only OLS estimations (with possible corrections of the standard errors for serial correlation) are employed. The main disadvantage is that the method implies a computational burden: one has to look at each of the regressions (i.e., one for the one-step ahead, one for the two-step ahead) and exclude (one at a time) the regressor with the lowest t-statistic. If one has, as can be expected, twenty or thirty macroeconomic variables for each of the dependent variables, this can be exhausting.[4]

Vector-autoregressive models (VARs: see box B) are systems of linear regressions, and therefore quite easy to estimate. They have the advantage over the single-equation linear models to better consider the interactions between variables. VARs model a more complete dynamics.[5]

Ease of estimation (again, only OLS is required) and flexibility in representing a number of economic relationships make VARs particularly suitable for estimating and forecasting dynamic processes. However, including more than one macroeconomic variable implies introducing a large number of coefficients, which reduces the available degrees of freedom. This is why this paper only uses bivariate VARs—that is, one bank variable and one macroeconomic variable.

A shortcoming of most regression approaches is that different series are likely to

[4]However, this author found a simple stepwise way to automate the task. The Gauss code, utilizing Newey-West robust standard errors calculated with a publicly-available procedure written by Paul Söderlind of the Stockholm School of Economics, is available upon request.

[5]To avoid confusion, it is worth pointing out that VAR indicates vector autoregression, and *not* VaR, which means Value at Risk. VARs are common in macroeconomics and forecasting (see Hamilton (1994, chs. 10 and 11) or Enders (1995, ch. 5) for a detailed introduction). VaR models (see for example Crouhy, Galai, and Mark (2000)) are instead used to assess portfolio risk.

have different 'best' macroeconomic regressors. This may complicate comparisons of banks belonging to different geographical areas or peer groups: for example, one may be interested in knowing how an increase in corn prices may affect problem loans of banks in Iowa and Indiana, but perhaps corn prices are significant in predicting problem loans for banks in Iowa but not in Indiana, and therefore this inference cannot be obtained (more specifically, the statistic would suggest that the effect of such a price change is not significantly different from zero).

This paper shows how the models above do not just yield forecasts (see boxes C and D), but also other tools that can be used by bank supervisors, as well as by bankers, to evaluate bank performance. Regression residuals (see box E) and impulse-response functions (see box F) are among such tools. Moreover, the paper shows how forecast-evaluation criteria can be used to effectively choose the forecasting model most suitable to a user's needs.

## 2 Data

The bank data series analyzed in this paper are either state averages or Seventh District averages. The types of loans analyzed are agricultural (AG), commercial and industrial (C&I) and real estate (RE). The rest of this section gives some details on the sources and on how peer groups are formed. Data about the banks are from the Reports of Condition and Income in the National Information Center (NIC) data set.[6]

Two categories of problem loans are considered. The first category, called delinquencies (here abbreviated as DQ), consists of the total loans 30 to 89 days past due and still accruing divided by the total of outstanding loans of the considered type (i.e., AG, C&I, or RE). The second category is non-performing loans (NP), defined as the sum of the still accruing loans 90 or more days past due and all nonaccrual loans, divided by the total of outstanding loans of the considered type.

Peer groups are formed using commercial bank data from the Seventh Federal Reserve District. Banks are divided according to the amount of total assets: below \$300 millions ("small"), between \$300 millions and \$1 billion ("medium"), and above \$1 billion ("large"). Small member banks are divided by state, thus one obtains a total

[6] Aggregated NIC data for state averages can be obtained from the San Francisco Fed web site `http://xena.sf.frb.org/mdb/PANEL.AGGCALL/search.html`. These averages consider *all* banks within a state, regardless of asset size and FR district. Such data can be used with the same models, but are not used in this paper. Quarterly data about individual banks can be obtained from the Chicago Fed web site `http://www.frbchi.org/RCRI/rcri_database.html`.

of seven peer groups: one for large banks, one for medium banks, and five (one for each state) for small banks.

Table 1 summarizes the conventions used to label bank data. There are two types of problem loans, namely, DQ and NP, and three categories of loans, namely, AG, C&I, and RE, thus giving six series for each of the peer groups.

All data are quarterly, 1987Q1–1999Q4. It might be useful to have a deeper dataset, particularly to span multiple business cycles. Unfortunately, many of the series were redefined in the mid-1980s (see Kashyap and Stein (1997)), and at the same time numerous mergers changed the characteristics of individual institutions—which can cause undesirable discontinuities in the data. Ratios are used for all calculations in order to avoid adjusting for both inflation and small mergers (since both the numerator and the denominator change).

Forecasts are calculated, using competing methods, for each of three loan types, two performance measures, and seven peer groups, for a total of 32 variables.[7]

Macroeconomic data are from the Haver DLX data set.[8] Growth rates are used given the high probability of unit roots in macroeconomic series. Examples of such variables are the state unemployment rate, per-capita income, and bankruptcy filings, as well as nationwide measures such as the index of the National Association of Purchasing Managers, the unemployment rate in the US, and car sales in the US.

# 3 Results

## 3.1 District data: Large and medium banks

Sample empirical results for large and medium banks in the Seventh District are shown in this section. These institutions are considered at a district level and not divided by state, because they are very likely to have activities in broader geographical areas within the District and beyond.

Tables 2 and 3 list the significant regressors for the OLS models without (see box B, equation (1)) or with (as in equation (7)) seasonal dummy variables. Two regressions are calculated for each of the dependent variables. The first one, aimed at calculating the one-step-ahead forecast, has all regressors (except for seasonal dummy variables or

[7]Since AG DQ and NP are not available for small banks, the corresponding models were not estimated (thus the total is 32 and not 42 series as would be expected).

[8]Several of the macroeconomic series can be obtained from the St. Louis Fed web site, `http://www.stls.frb.org/fred/index.html`.

lagged dependent terms) lagged one period. The second regression, needed to calculate the two-step-ahead forecast, has all regressors lagged two periods. It can be seen that in most cases, both specifications of each equation have several significant regressors in common.

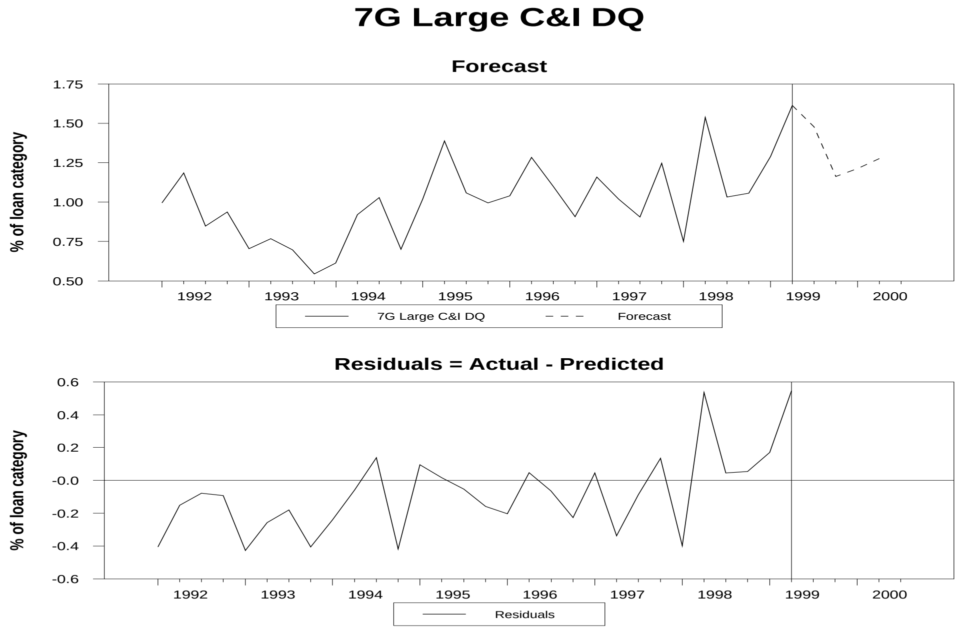
Tables 4 and 5 show some summary statistics from the VAR regressions for large and medium banks. The adjusted $R^2$ coefficients are generally high, as is usual in VAR systems. The lagged dependent variable terms are jointly significant at the 5 percent level in all equations but four (first and seventh row of table 4; first and seventh row of table 5). The macroeconomic variable is significant at the 5 percent or better confidence level in all cases. US housing permits, farming income, and unemployment appear to be quite good predictors of the problem-loan ratios for large and mid-sized commercial banks in the Seventh District.

The following figure presents an example of the results for a VAR with seasonal dummy variables using data of large- and medium-sized banks. The figure consists of two panels. The upper panel shows the variable since 1992 and its forecast up to the second quarter of the year 2000 (the forecast is depicted by the dashed line after the vertical bar). The lower panel shows the regression errors (residuals), defined as actual value less fitted value for the corresponding quarters. There are no residuals to show after the second quarter of 1999 because the model cannot predict its own future errors (remember that the data for the last two quarters of 1999 have been excluded from the sample for testing purposes). A positive residual indicates that the actual number is higher than the model predicts, and vice versa.

The upper panel in figure 1 shows that C&I DQ for large banks have been increasing since 1995. The forecast (dashed line) predicts that they will remain high but slowly decrease.

The lower panel of figure 1 shows immediately the usefulness of the residuals for bank surveillance (in an early warning framework). In the panel one easily notices that a series of positive and increasingly large residuals have occurred to C&I delinquencies for large banks since the second quarter of 1998. Therefore, C&I DQ have consistently topped the model's predictions. This is a source of concern for supervisors, because the size of the latest residual suggests that C&I DQ, at about 1.25% of all C&I loans, is about 50 basis points above the model's prediction—a sharp difference indeed. By relying on the size of residuals, supervisors can individuate groups of banks that have performed in a remarkably different way from what could be expected given the macroeconomic conditions (embedded in the model). Such groups may deserve

Figure 1: Commercial and Industrial Delinquencies for 7G Large Banks




sharper scrutiny to verify whether lending policies or bank strategies have changed, thus impairing safety and soundness of the system.

## 3.2 State data: Small banks

Banks whose total assets are less than \$300 millions typically have a limited number of branches, resulting in activities that are limited to a smaller geographic area. This is why small banks in this paper are grouped by state.

A peculiarity in the Call Report for banks with less than \$300 millions in total assets must be pointed out. For these banks, the C&I and RE NP and DQ series also include agricultural problem loans. This implies that the ratios, in absolute value, are not directly comparable with those of large and medium banks. At the same time, the presence of agricultural problem loans is likely to accentuate the seasonal pattern of the series.

Results of the OLS estimations for small-bank peer groups are not reported for space reasons. The interesting feature is that both national and state-level macroeconomic variables are significant. This confirms the results of Meyer and Yeager (2000), suggesting that state-level variables are good predictors of small bank performance.

Tables 6 and 7 summarize the results of the VARs for small banks. Again, the adjusted $R^2$ coefficients are quite high, and lags of the dependent variable are significant in all but one case (IN C&I DQ in table 7).

The macroeconomic variables are significant at the 5 percent level, with three exceptions in table 6 (first, tenth, and seventeenth rows), for which no significant regressor was found. In such cases, the macro regressor with the relatively highest adjusted $R^2$ was chosen.

Measures of income (including farming income) and unemployment appear to be good predictors of problem loans in all states. It is interesting to point out that bankruptcy filings often turn out to be significant regressors in these systems, even if they were not chosen at all times as best. One would normally think that past-due loans should predict bankruptcies, and not vice versa, because one first starts being late with payments, and then later may file for bankruptcy. One explanation for this is that people may choose to go bankrupt directly, avoiding pressures from lenders and collection agencies, and getting rid of debt in one, allegedly easy, shot.

Figures and results for individual states are not reported for brevity, and are available upon request (they are updated quarterly for internal use). As a whole, no common patterns emerge for the small banks in the five states of the Seventh District. This suggests that local economic factors tend to dominate national trends for the smaller institutions, and therefore that it is important to use state-level data and not only national data when forecasting the financial conditions of small banks.

# 4 Forecasts and simulations

## 4.1 Forecast evaluation

Different methods of forecast evaluation have been proposed. Some are surveyed by Diebold and Mariano (1995). Traditionally, a measure is used, which evaluates the distance between actual number and forecast, such as the mean square error (see Hamilton (1994, p. 73)). Another intuitive criterion is the confusion rate (see, for example, Swanson and White (1997)). The motivation of the latter criterion is as follows. One may care more to know if the variable goes up or down rather than knowing if it is going to change by a lot or by little. For example, one may be more interested to accurately predict whether a stock's price will increase or decrease, so to appropriately decide whether to buy or sell the stock. A confusion rate measures how often a predictor fails to forecast the correct direction of the movement of the relevant variable. For supervisory purposes, being able to predict that a key measure of bank performance will change direction can provide more important information than predicting the actual level of such variable.

For the purpose of forecast evaluation, each of the bank series is forecast using two different linear models (with or without seasonal dummies) and two VAR models (again, with or without seasonal dummies). The results for our sample are summarized in table 8.

Looking at the first two rows of table 8, one can see how both criteria seem to deteriorate with time, namely, the two-step-ahead forecast in our sample is significantly less accurate than the one-step-ahead. The use of seasonal dummies appears to improve forecasting accuracy.

The second two rows show that VAR models with seasonal dummies can yield a lower MSE and lower confusion rates in two-step-ahead forecast. Moreover, the VAR's predictive accuracy does not appear to deteriorate in the second quarter ahead—it actually improves. Looking at both criteria as a whole, it appears that a simple bivariate VAR with seasonal dummies provides a generally acceptable level of forecasting accuracy, with lower computational burden than a multivariate OLS model.

Some limited experiments run using more complicated models (for example, using nonlinear data transformations) yielded lower MSEs and, often, better confusion rates. In particular, Bayesian forecasting models appear to be more accurate (see for example Hamilton (1994, pp. 360–65)). The scope of this paper is, however, to see what can be obtained using the simplest models possible, and more complex frameworks are therefore beyond such scope.

## 4.2 Stress testing

A possible application of this framework is in stress testing. One could evaluate what-if scenarios such as the impact of a 100 basis points increase in US business bankruptcy filings on C&I NP for large banks in the Seventh District. This is a potentially interesting exercise, which however comes with two warnings: (i) the quality of the what-if exercise is as good as the economic relationship between the variables used (i.e., the macro variable must "cause" the bank variable), and (ii) because of their lack of theoretical structure, it is known that VAR models can predict only responses to small or medium shocks (thus they cannot reliably predict what might happen in a big crisis or in a catastrophic situation).

A very easy way to carry out a what-if analysis is by using the coefficients of the linear regressions: by just looking at the regression output one knows what the effect on the dependent variable will be, if one of the regressors increases or decreases by a certain amount: since the equations are linear, the coefficient corresponds to the first derivative with respect to that regressor. For example, if the coefficient of USR is 0.20, it means that an increase of unemployment by 30 basis points causes our forecast of the past-due loans variable in exam to increase by $30 \cdot 0.20 = 6$ basis points.

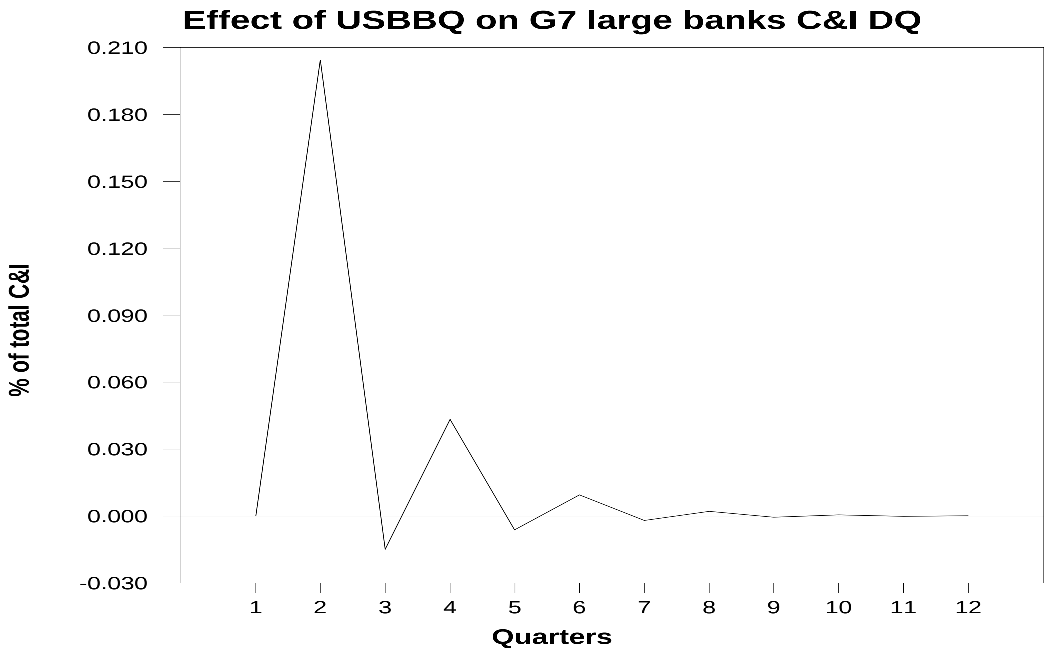
Because of the linear nature of the models used in this paper, it is not difficult to implement scenario analysis using a simple electronic spreadsheet for field use. Given the VARs estimated in this paper, any line in tables 4 to 7 yields a potentially valid what-if scenario. For example, from the first row of table 4 one sees that USEFQ (US farming income) can be used to stress-test AG DQ of the large banks in the Seventh District. It is important to point out that the tables only report the "best" significant variable: that is, there may be more than one significant variable, thus one could for example stress test 7G large bank AG DQ using either a system with USEFQ (farm income) or USABQ (all bankruptcy filings). When coding the system in a spreadsheet, one must remember that our VARs are systems of two equations, and therefore use recursion as explained in equations (8) and (9), as well as in the example of box D.

A slightly more complicated tool is the impulse-response function (IRF: see box F for more details).

Figure 2 shows the plot of an IRF. In this case, an increase by one standard deviation (883 basis points) in the growth rate of business bankruptcy filings in the US induces an expected increase in Seventh District large banks' C&I DQ by 20.5 basis points in the following quarter—an economically significant increase for a variable whose sample mean is 1.16%. The effect begins in the second quarter (because $USBBQ_{t-1}$ and not $USBBQ_t$ is a regressor in the equation where Seventh District large bank C&I DQ is the dependent variable), reaches a peak of plus 20.5 basis points in the second quarter, and fades out in an oscillating fashion, becoming slightly negative in the following quarter. The effect becomes practically negligible after the fourth quarter.

It is easy to understand how an IRF can be used in a what-if analysis. One just arbitrarily proposes a shock to the relevant macro variable, and looks at the IRF to see how the shock will affect the examined bank variable over time. Some software packages can also calculate confidence bands around the IRF, thus giving a better feel of the expected response of the bank variable to the hypothetical shock.

Figure 2: Sample Impulse-Response Function




# 5 Conclusion

This paper presents some results and a number of examples of how simple linear multivariate models and bivariate vector-autoregressive systems can be used to analyze and benchmark problem loans. The paper shows different uses of the models for forecasting, stress testing, and analysis of deviations. Examples of forecast evaluation criteria are presented.

A factor that emerges from examining the series is the strong seasonal component of the data. By simply looking at the data, it would be hard to understand whether an increase in one of the ratios is simply a seasonal event or is a more serious issue. This helps stress the importance of using statistical procedures, which can help supervisors discriminate between temporary and permanent and between seasonal regularity and actual shocks.

Residuals from VARs appear to yield interesting results, in particular when used for "control by exception", i.e., to identify potential trouble spots in the behavior of specific banks or groups of banks. Informal conversations with examiners have stressed that this is an important piece of information for supervisors in the field, who are interested in the financial trends of examined institution as well as in significant deviations from such trends.

Out-of-sample testing is used to evaluate the variables' forecasting power. The results suggest that VAR models are accurate in forecasting the direction of the variable (i.e., if it is expected to increase or decrease in the following quarters).

Scenario analysis, performed by simply plugging plausible arbitrary numbers into the equations or by using IRFs, can be easily performed and has several potential uses in bank surveillance.

The models presented in this paper can identify areas of risk in bank portfolios in four ways. First, they describe the links between macroeconomic dynamics and bank asset quality (variable significance). Second, they analyze past deviations in order to pinpoint possible trouble spots (residual analysis). Third, they yield reasonably accurate predictions of the future effects of the business cycle on asset quality. Fourth, they allow for scenario analysis (stress testing). Given their simplicity and their effectiveness, all four ways would be useful tools in the supervisory process.

# References


Azariadis, Costas, and Bruce Smith (1998), "Financial Intermediation and Regime Switching in Business Cycles." *American Economic Review,* 88(3), June, 516–36

Crouhy, Michel, Dan Galai, and Robert Mark (2000), "A Comparative Analysis of Current Credit Risk Models." *Journal of Banking and Finance,* 24(1–2), January, 59–117

Cole, Rebel A., Barbara G. Cornyn, and Jeffery W. Gunther (1995), "FIMS: A New Monitoring System for Banking Institutions." *Federal Reserve Bulletin,* 81(1), January, 1–15

Diebold, Francis X.; Mariano, Roberto S. (1995), "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics,* 13(3), July, 253–63

Enders, Walter (1995), *Applied Econometric Time Series,* New York, NY: Wiley

Hamilton, James D. (1994), *Time Series Analysis,* Princeton, NJ: Princeton University Press

Hiemstra, Stephen W., Stephen Kane, Thomas J. Lutton, and P.A.V.B. Swamy (1997), "A New Method of Forecasting Bank Failures and Insurance Fund Losses." Washington, DC: Office of the Comptroller of the Currency, working paper

Holtz-Eakin, Douglas, Whitney Newey, and Harvey S. Rosen (1988), "Estimating Vector Autoregressions with Panel Data." *Econometrica,* 56(6), November, 1371–95

Kashyap, Anil, and Jeremy C. Stein (1997), "Notes on Forming Consistent Time Series." Chicago, IL: Federal Reserve Bank of Chicago, working paper available online at `http://www.frbchi.org/rcri/consist.pdf`

Lütkepohl, Helmut (1993), *Introduction to Multiple Time Series Analysis,* 2nd ed., Berlin: Springer-Verlag

Meyer, Andrew P., and Timothy J. Yeager (2000), "Are Rural Community Banks Sensitive To Local Economic Activity?" St. Louis, MO: Federal Reserve Bank of St. Louis, working paper

Sims, Christopher A., James H. Stock, and Mark W. Watson (1990), "Inference in Linear Time Series Models with Some Unit Roots." *Econometrica,* 58(1), January, 113–44

Swamy, P.A.V.B., Thomas J. Lutton, and Philip F. Bartholomew (1998), "Whole Bank Risk: A Risk Measurement Technique." Washington, DC: Office of the Comptroller of the Currency, working paper

Swanson, Norman R., and Clive W.J. Granger (1997), "Impulse Response Functions Based on a Causal Approach to Residual Orthogonalization in Vector Autoregressions." *Journal of the American Statistical Association,* 92(437), March, 357–67

Swanson, Norman R., and Halbert White (1997), "Forecasting Economic Time Series Using Flexible versus Fixed Specification and Linear versus Nonlinear Econometric Models." *International Journal of Forecasting,* 13(4), December, 439–61

# A Box: Linear models

A simple way to forecast a variable is by using a linear model. One possible formulation is the following:

$$y_t = a_0 + a_1 y_{t-1} + \ldots + a_4 y_{t-4} + b_1 x_{1,t-1} + \ldots + b_n x_{n,t-1} + u_t, \tag{1}$$

which is the regression of "today's" $y_t$ on past values of $x$ and $y$. Note that $a$'s and $b$'s are coefficients (i.e., they are going to be estimated, but $x$ and $y$ are known). The variable $u$ represents innovations (a.k.a. shocks, residuals, or errors) affecting the dependent variable (e.g., an FOMC decision to increase interest rates).

Four lags of the dependent variable (i.e., $y_{t-1}$, $y_{t-2}$, $y_{t-3}$, and $y_{t-4}$) are included as regressors. Their function is to control for inertial and seasonal components in the variable that are not captured by the other regressors. This helps formulate a more parsimonious model and tends to limit serial correlation in the residuals.

# B Box: VAR models

VAR models look for predictors of relevant series in order to obtain forecasts. A bivariate VAR is a system of two linear equations, in which the regressors are lagged (past) values of the dependent variables (in our case, one of the variables is a bank financial variable, and the other is a macroeconomic variable). To obtain a **forecast** one only needs to substitute current values to the regressors, and the equation will yield the forecast. At the same time, small VARs using peer groups or state averages can be used to analyze deviations of a small group or of a single bank from a larger group—i.e., benchmarking.

Lag specifications (i.e., the number of past values of each variable included as regressors) are chosen using the Schwarz criterion (see Lütkepohl (1993, p. 132)).

Each of the bank variables is estimated in a number of bivariate systems, one for each of a series of macroeconomic variables. The output of each of the two-equation systems is then examined in order to find which of the macroeconomic variables is significant (using an $F$-test) in the equation where the bank variable is the dependent variable. The corresponding system is used to calculate forecasts and residuals. If more than one system has a macro variable significant at the 1% level, the equation with the highest adjusted $R^2$ is picked, and the corresponding system is used.

A simple, two-variable VAR consists of two equations in two variables (in the following example they are called $x$ and $y$) is shown in the following equations (2) and (3). The variables appear with their current values (namely, $x_t$ and $y_t$) as dependents (left side of equations). They also appear as regressors, with past (lagged) values (namely, $x_{t-1}$, $y_{t-1}$, and so on):

$$
\begin{aligned}
x_t &= f_0 + f_1 x_{t-1} + f_2 y_{t-1} + f_3 x_{t-2} + f_4 y_{t-2} + \ldots + \epsilon_t && (2) \\
y_t &= g_0 + g_1 x_{t-1} + g_2 y_{t-1} + g_3 x_{t-2} + g_4 y_{t-2} + \ldots + \nu_t. && (3)
\end{aligned}
$$

The two equations show that own past values of one variable have effects on the other one.

VAR models can have more than two variables. Higher-dimensional VARs, however, need more theoretical structure and are not suitable to the objectives of this paper (for issues such as orthogonalization, common in larger systems, see again Hamilton (1994) and Swanson and Granger (1997)).

# C Box: Forecasting with a linear model

If one substitutes more current values of the variables on the right-hand side, one obtains the one-step-ahead forecast, that is, the "fitted" value $\hat{y}_{t+1}$. This is shown in the following equation:

$$\hat{y}_{t+1} = a_0 + a_1 y_t + \ldots + a_4 y_{t-3} + b_1 x_{1,t} + \ldots + b_n x_{n,t}. \tag{4}$$

Note that the forecast equation does not include an error term because the model does not predict its own errors—that is, the model is expected to be correct on average.

One can also calculate a regression of $y_t$ on the same regressors as in equation (1), but with the right-hand side lagged (shifted back in time) once more, as in the following equation:

$$y_t = c_0 + c_1 y_{t-2} + \ldots + c_4 y_{t-5} + d_1 x_{1,t-2} + \ldots + d_n x_{n,t-2} + v_t, \tag{5}$$

which can be used to calculate $\hat{y}_{t+2}$ given the data available at time $t$:

$$\hat{y}_{t+2} = c_0 + c_1 y_t + \ldots + c_4 y_{t-3} + d_1 x_{1,t} + \ldots + d_n x_{n,t}. \tag{6}$$

An alternative formulation of the model in equation (1) is to only include one lag of the dependent variable, adding a set of seasonal dummy variables as in the following model:

$$y_t = a_{Q1}D_{1,t} + a_{Q2}D_{2,t} + a_{Q3}D_{3,t} + a_{Q4}D_{4,t} + a_1 y_{t-1} + b_1 x_{1,t-1} + \ldots + b_n x_{n,t-1} + u_t, \quad (7)$$

where the variables $D_1$ to $D_4$ are seasonal dummies: $D_1$ is equal to one if the observation refers to the first quarter of a year, and zero otherwise; $D_2$ is equal to one if the observation refers to the second quarter, and zero otherwise; and so on. If the seasonal component is rather stable in the series, a set of dummy variables can be quite useful.

# D Box: Forecasting with a VAR

If the estimated coefficients are substituted into one of the equations of a VAR system, one can obtain a (one-step-ahead) prediction of the dependent variable. For example, one could substitute the coefficient estimates from the system on the third row of table 4 into equations (2) and (3) and obtain:[9]

$$\widehat{CI}_{t+1} = 0.353 \cdot CI_t + 0.023 \cdot BKP_t + Seasonal_{1,t+1} \quad (8)$$

$$\widehat{BKP}_{t+1} = 2.450 \cdot CI_t - 0.425 \cdot BKP_t + Seasonal_{2,t+1} \quad (9)$$

where CI is the C&I DQ ratio for large banks in the Seventh District and BKP is the variation in business bankruptcy filings in the US (called USBBQ in the tables). Seasonal is a number that reflects periodic factors (e.g., there are few defaults of farmers in the Spring, because banks generally ask for their money back after the harvest has been sold). From the table one can see that both regressors are significant at the 1 percent level and the $\bar{R}^2$ coefficient is 0.28 in equation (8). As mentioned before, the analogous statistics relative to equation (9) are omitted for brevity.

Since all right-hand-side variables are known (their time is $t-1$), it is easy to do a simple one-step ahead forecast, that is, forecasting what C&I NPs can be expected to be next quarter given what is known about current C&I NPs and bankruptcy filings. From the regression one can also calculate confidence bands for the forecast (a statistic that can be loosely interpreted as a measure of reliability of the forecast). Given that

[9]The estimated coefficients are not reported in the tables for brevity. They are available from the author upon request.

BKP for the second quarter of 1999 was 13.1 and CI was 1.614%, and that the fourth-quarter value of Seasonal is $-0.606$, our forecast of CI for the fourth quarter of 1998 is $0.353 \cdot 1.614 + 0.023 \cdot 13.1 - 0.606 = 1.478\%$. Recall that the value for the second quarter was 1.614% of total C&I loans, thus a decrease in NPs is forecasted (and actually, a value of 1.202% is observed in the out-of-sample data). If one similarly forecasts BKP (using equation (9)), one obtains an estimate of BKP for the last quarter of 1998 ($-10.297$); this can be used *recursively* to forecast (two steps ahead) CI for the first quarter of 1999, obtaining a forecast of a further decrease 1.162% of total C&I loans of large banks in the Seventh District (out of sample, 1.105% is observed in the fourth quarter of 1999), and so on.

# E Box: Regression residuals

Particular attention is paid in this paper to the concept of regression residuals. A **residual** (also called regression error) is the deviation of the actual data from the values that the model would have "predicted" (also called "fitted"), given the behavior of the data during the sampling period.

An estimated residual $\hat{\epsilon}_t$ is defined as:

$$\hat{\epsilon}_t = y_t - \left(\hat{\alpha} + \hat{\beta} \cdot x_t\right) \equiv y_t - \hat{y}_t \tag{10}$$

where $y$ is the variable one wishes to analyze, $\hat{\alpha}$ and $\hat{\beta}$ are two estimated regression parameters, and $x$ is the regressor. The term between parentheses, that is rewritten as $\hat{y}_t$, is the linear projection of $y$ on $x$, also called the "predicted" value of $y$. The residual is therefore equal to the observed value $y$ less its predicted value $\hat{y}$. A nonzero residual suggests the presence of an unexpected shock to the relevant variable (that is, the actual value deviated from what the model predicted).

# F Box: Impulse Response Functions

Analytical solutions of systems such as that of equations (2) and (3) show that each of the variables is a weighted sum of current and past innovations to both variables, and the solution looks like the following equation (where $\alpha$ is some initial condition and $\beta$, and $\gamma$ are series of constant coefficients to be estimated.):

$$x_t = \alpha + \sum_{i=0}^{\infty} \beta_i \epsilon_{t-i} + \sum_{i=1}^{\infty} \gamma_i \nu_{t-i}. \tag{11}$$

It turns out that, if the coefficients of equations (2) and (3) are estimated, one can then mathematically derive the coefficients of equation (11) and of the corresponding solution for $y_t$. The solutions are called impulse response functions (IRF), and can be plotted as in figure 2. It can be noticed that the coefficients of more distant (i.e., older) innovations are smaller, because VAR models assume that innovations have effects that do not last forever (thus innovations fade over time).

IRF plots are useful for the what-if analysis of scenarios. See a practical application in section 4.2.

Table 1: Abbreviations used

| | |
|---|---|
| DQ | DELINQUENCIES: 30–89 days past due and still accruing / total |
| NP | NON-PERFORMING: (90 or more days past due still accruing + nonaccruals) / total |
| AG | loans financing agricultural production |
| C&I | commercial and industrial loans |
| RE | real estate loans (both residential and commercial included) |

NOTE: DQ is the ratio of loans of one category past due 30 to 89 days and still accruing to total loans of that category (for example, the ratio of real estate delinquent loans to total real estate loans). NP is the sum of the still accruing loans 90 or more days past due and all nonaccrual loans, divided by the total of outstanding loans of the considered category.

Table 2: OLS with seasonal dummies—Large and medium banks

| Dependent | Significant regressors |
|---|---|
| 7G Large AG DQ | USEFQ$_1$ USR$_1$ USTRU$_1$ USYPQ$_1$ |
| 7G Large AG DQ | USBBQ$_2$ |
| 7G Large AG NP | USABQ$_1$ USCSENT$_1$ USEFQ$_1$ USENFQ$_1$ USNAPM$_1$ USNBQ$_1$ USYPQ$_1$ |
| 7G Large AG NP | USABQ$_2$ USBBQ$_2$ USENFQ$_2$ |
| 7G Large C&I DQ | USABQ$_1$ USCSENT$_1$ USNAPM$_1$ USNBQ$_1$ USR$_1$ USTRU$_1$ USYPQ$_1$ |
| 7G Large C&I DQ | USABQ$_2$ USCAR$_2$ USNAPM$_2$ USNBQ$_2$ |
| 7G Large C&I NP | USBBQ$_1$ USCAR$_1$ USNBQ$_1$ |
| 7G Large C&I NP | USABQ$_2$ USNAPM$_2$ USNBQ$_2$ |
| 7G Large RE DQ | USABQ$_1$ USR$_1$ USTRU$_1$ |
| 7G Large RE DQ | USABQ$_2$ USNAPM$_2$ USNBQ$_2$ USR$_2$ USTRU$_2$ |
| 7G Large RE NP | USBBQ$_1$ USCAR$_1$ USCSENT$_1$ USENFQ$_1$ USR$_1$ USTRU$_1$ USYPQ$_1$ |
| 7G Large RE NP | USBBQ$_2$ USCAR$_2$ USCSENT$_2$ USENFQ$_2$ USTRU$_2$ USYPQ$_2$ |
| 7G Midsz AG DQ | USABQ$_1$ USBBQ$_1$ USCAR$_1$ USNAPM$_1$ USNBQ$_1$ USR$_1$ USTRU$_1$ USYPQ$_1$ |
| 7G Midsz AG DQ | USABQ$_2$ USENFQ$_2$ USNBQ$_2$ USYPQ$_2$ |
| 7G Midsz AG NP | USCSENT$_1$ USNAPM$_1$ USYPQ$_1$ |
| 7G Midsz AG NP | USABQ$_2$ USCAR$_2$ USNBQ$_2$ USTRU$_2$ |
| 7G Midsz C&I DQ | USABQ$_1$ USBBQ$_1$ USR$_1$ |
| 7G Midsz C&I DQ | USNBQ$_2$ USR$_2$ USTRU$_2$ |
| 7G Midsz C&I NP | USABQ$_1$ USBBQ$_1$ USEFQ$_1$ USNBQ$_1$ USR$_1$ USYPQ$_1$ |
| 7G Midsz C&I NP | USABQ$_2$ USBBQ$_2$ USCAR$_2$ USEFQ$_2$ USNAPM$_2$ USNBQ$_2$ USR$_2$ |
| 7G Midsz RE DQ | USABQ$_1$ USCSENT$_1$ USNAPM$_1$ USNBQ$_1$ USR$_1$ USTRU$_1$ |
| 7G Midsz RE DQ | USABQ$_2$ USBBQ$_2$ USNBQ$_2$ |
| 7G Midsz RE NP | USABQ$_1$ USNBQ$_1$ USR$_1$ USTRU$_1$ |
| 7G Midsz RE NP | USABQ$_2$ USBBQ$_2$ USNAPM$_2$ USNBQ$_2$ USR$_2$ USTRU$_2$ |

NOTE: Each of the equations includes four seasonal dummy variables and one lag of the dependent variable. For each of the dependent variables, the first regression is of the "current" dependent on all regressors lagged once (hence the 1 subscript), and the second regression has all regressors lagged two quarters. The abbreviations for the regressors are: USABQ all bankruptcy filings, USBBQ business bankruptcy filings, USCAR car sales, USCSENT University of Michigan consumer sentiment index, USEFQ farming income, USENFQ non-farm income, USNAPM National Association of Purchasing Managers index, USNBQ non-business bankruptcy filings, USR unemployment rate, USTRU housing permits, USYPQ personal income. All regressors reported (excluding seasonal dummies and lagged dependent) are at least significant at the 10% level, using Newey-West standard errors.

Table 3: OLS without seasonal dummies—Large and medium banks

| Dependent | Significant regressors |
|---|---|
| 7G Large AG DQ | USCAR$_1$ USEFQ$_1$ USNBQ$_1$ USYPQ$_1$ |
| 7G Large AG DQ | USBBQ$_2$ USCAR$_2$ USNBQ$_2$ |
| 7G Large AG NP | USENFQ$_1$ USNAPM$_1$ |
| 7G Large AG NP | USABQ$_2$ USBBQ$_2$ USNAPM$_2$ USR$_2$ |
| 7G Large C&I DQ | USABQ$_1$ USCAR$_1$ USCSENT$_1$ USNAPM$_1$ USTRU$_1$ USYPQ$_1$ |
| 7G Large C&I DQ | USR$_2$ |
| 7G Large C&I NP | USCAR$_1$ USCSENT$_1$ USNBQ$_1$ |
| 7G Large C&I NP | USABQ$_2$ USNAPM$_2$ USNBQ$_2$ USTRU$_2$ |
| 7G Large RE DQ | USABQ$_1$ USENFQ$_1$ USNBQ$_1$ USTRU$_1$ |
| 7G Large RE DQ | USABQ$_2$ USCAR$_2$ USNBQ$_2$ USTRU$_2$ USYPQ$_2$ |
| 7G Large RE NP | USBBQ$_1$ USCAR$_1$ USCSENT$_1$ USENFQ$_1$ USNBQ$_1$ USTRU$_1$ USYPQ$_1$ |
| 7G Large RE NP | USABQ$_2$ USBBQ$_2$ USCAR$_2$ USCSENT$_2$ USENFQ$_2$ USNAPM$_2$ USNBQ$_2$ USTRU$_2$ USYPQ$_2$ |
| 7G Midsz AG DQ | USBBQ$_1$ USCAR$_1$ USENFQ$_1$ USNAPM$_1$ |
| 7G Midsz AG DQ | USCAR$_2$ USENFQ$_2$ USTRU$_2$ USYPQ$_2$ |
| 7G Midsz AG NP | USABQ$_1$ USBBQ$_1$ USCSENT$_1$ USENFQ$_1$ USNAPM$_1$ USNBQ$_1$ USYPQ$_1$ |
| 7G Midsz AG NP | USCAR$_2$ |
| 7G Midsz C&I DQ | USCSENT$_1$ USTRU$_1$ |
| 7G Midsz C&I DQ | USABQ$_2$ USBBQ$_2$ USCAR$_2$ USEFQ$_2$ USNBQ$_2$ USR$_2$ USTRU$_2$ |
| 7G Midsz C&I NP | USABQ$_1$ USBBQ$_1$ USNAPM$_1$ USNBQ$_1$ USR$_1$ USTRU$_1$ USYPQ$_1$ |
| 7G Midsz C&I NP | USNAPM$_2$ USR$_2$ |
| 7G Midsz RE DQ | USABQ$_1$ USCAR$_1$ USCSENT$_1$ USNAPM$_1$ USNBQ$_1$ |
| 7G Midsz RE DQ | USEFQ$_2$ USENFQ$_2$ USR$_2$ USTRU$_2$ USYPQ$_2$ |
| 7G Midsz RE NP | USABQ$_1$ USEFQ$_1$ USENFQ$_1$ USNAPM$_1$ USNBQ$_1$ USR$_1$ USYPQ$_1$ |
| 7G Midsz RE NP | USBBQ$_2$ USEFQ$_2$ USENFQ$_2$ USNBQ$_2$ USYPQ$_2$ |

NOTE: Each of the equations includes one constant term and four lags of the dependent variable. For each of the dependent variables, the first regression is of the "current" dependent on all regressors lagged once (hence the 1 subscript, but the lags of the dependent variable are 1 to 4), and the second regression has all regressors lagged two quarters (while the lags of the dependent go from 2 to 5). The abbreviations for the regressors are: USABQ all bankruptcy filings, USBBQ business bankruptcy filings, USCAR car sales, USCSENT University of Michigan consumer sentiment index, USEFQ farming income, USENFQ non-farm income, USNAPM National Association of Purchasing Managers index, USNBQ non-business bankruptcy filings, USR unemployment rate, USTRU housing permits, USYPQ personal income. All regressors reported (excluding seasonal dummies and lagged dependent) are at least significant at the 10% level, using Newey-West standard errors.

Table 4: VAR with seasonal dummies—Large and medium banks

| Dependent variable | Lags | $\bar{R}^2$ | Lagged dep. F-test | Other variable | Other var. F-test |
|---|---|---|---|---|---|
| 7G Large AG DQ | 1 | 0.418 | 1.567 | USEFQ | 10.762*** |
| 7G Large AG NP | 8 | 0.761 | 11.037*** | USABQ | 4.313*** |
| 7G Large C&I DQ | 1 | 0.281 | 7.927*** | USBBQ | 13.233*** |
| 7G Large C&I NP | 1 | 0.814 | 200.682*** | USCSENT | 4.512** |
| 7G Large RE DQ | 2 | 0.716 | 25.488*** | USR | 12.592*** |
| 7G Large RE NP | 1 | 0.876 | 340.787*** | USCSENT | 5.993** |
| 7G Midsz AG DQ | 1 | 0.661 | 2.358 | USNAPM | 4.566** |
| 7G Midsz AG NP | 8 | 0.924 | 43.104*** | USTRU | 13.479*** |
| 7G Midsz C&I DQ | 2 | 0.725 | 45.105*** | USABQ | 5.938*** |
| 7G Midsz C&I NP | 1 | 0.742 | 119.540*** | USR | 10.815*** |
| 7G Midsz RE DQ | 4 | 0.775 | 20.288*** | USCAR | 5.516*** |
| 7G Midsz RE NP | 1 | 0.880 | 293.882*** | USR | 4.174** |

NOTE: The first column shows which of the banking variables is the dependent; the second column reports the number of lags in the model; the third column reports adjusted $R^2$ for that equation; the fourth column reports the F-test of the null hypothesis of all lags of the dependent variable being not significant; the fifth column lists which macroeconomic variable was picked according to statistical significance (the abbreviations are explained hereafter); the sixth column reports the F-test of the conventional null hypothesis on the lags of the macroeconomic variable. One asterisk (*) indicates that the lags of the variable are jointly significant at the 10 percent level, two asterisks (**) indicate significance at the 5 percent level, and three asterisks (***) indicate significance at the 1 percent level. All VARs include seasonal dummies. Lags (up to eight) are chosen according to the lowest Schwarz criterion of the system. All macroeconomic variables are in growth rates, all banking variables are in levels. List of the abbreviations used in the table: USCAR (Car sales), USR (Unemployment rate), USABQ (Total bankruptcy filings), USTRU (Permits, new, privately-owned buildings), ILTRU (Illinois building permits, new, privately-owned buildings), USNBQ (Non-business bankruptcy filings), USYPQ (), WITRU (Wisconsin building permits, new, privately-owned buildings). The second equation of each system, where the macroeconomic variable is the dependent, is not reported for simplicity and is available upon request. These regressions refer to the sample 1987Q1–1999Q2, in order to allow for out-of-sample validation.

Table 5: VAR with no seasonal dummies—Large and medium banks

| Dependent variable | Lags | $\bar{R}^2$ | Lagged dep. F-test | Other variable | Other var. F-test |
|---|---|---|---|---|---|
| 7G Large AG DQ | 8 | 0.493 | 1.652 | USCAR | 3.583*** |
| 7G Large AG NP | 8 | 0.681 | 4.732*** | USTRU | 2.819** |
| 7G Large C&I DQ | 3 | 0.285 | 3.831** | USBBQ | 5.086*** |
| 7G Large C&I NP | 1 | 0.819 | 207.595*** | USCSENT | 6.166** |
| 7G Large RE DQ | 7 | 0.794 | 3.937*** | USCSENT | 4.895*** |
| 7G Large RE NP | 1 | 0.878 | 345.561*** | USCSENT | 6.482** |
| 7G Midsz AG DQ | 8 | 0.758 | 1.396 | USTRU | 4.036*** |
| 7G Midsz AG NP | 8 | 0.834 | 18.652*** | USTRU | 5.996*** |
| 7G Midsz C&I DQ | 8 | 0.818 | 6.913*** | USTRU | 6.292*** |
| 7G Midsz C&I NP | 8 | 0.823 | 18.411*** | USTRU | 5.340*** |
| 7G Midsz RE DQ | 8 | 0.886 | 9.798*** | USTRU | 8.567*** |
| 7G Midsz RE NP | 7 | 0.915 | 43.934*** | USCAR | 3.589*** |

NOTE: The first column shows which of the banking variables is the dependent; the second column reports the number of lags in the model; the third column reports adjusted $R^2$ for that equation; the fourth column reports the F-test of the null hypothesis of all lags of the dependent variable being not significant; the fifth column lists which macroeconomic variable was picked according to statistical significance (the abbreviations are explained hereafter); the sixth column reports the F-test of the conventional null hypothesis on the lags of the macroeconomic variable. One asterisk (*) indicates that the lags of the variable are jointly significant at the 10 percent level, two asterisks (**) indicate significance at the 5 percent level, and three asterisks (***) indicate significance at the 1 percent level. All VARs include a constant term. Lags (up to eight) are chosen according to the lowest Schwarz criterion of the system. All macroeconomic variables are in growth rates, all banking variables are in levels. List of the abbreviations used in the table: USCAR (Car sales), USR (Unemployment rate), USABQ (Total bankruptcy filings), USTRU (Permits, new, privately-owned buildings), ILTRU (Illinois building permits, new, privately-owned buildings), USNBQ (Non-business bankruptcy filings), USYPQ (), WITRU (Wisconsin building permits, new, privately-owned buildings). The second equation of each system, where the macroeconomic variable is the dependent, is not reported for simplicity and is available upon request. These regressions refer to the sample 1987Q1–1999Q2, in order to allow for out-of-sample validation.

Table 6: VAR with seasonal dummies—Small banks

| Dependent variable | Lags | $\bar{R}^2$ | Lagged dep. F-test | Other variable | Other var. F-test |
|---|---|---|---|---|---|
| IA Small C&I DQ | 1 | 0.826 | 5.031** | USCAR | 1.050 |
| IA Small C&I NP | 1 | 0.966 | 1187.839*** | IAR | 4.778** |
| IA Small RE DQ | 1 | 0.848 | 108.314*** | IAEFQ | 6.810** |
| IA Small RE NP | 1 | 0.985 | 2713.797*** | IAR | 9.456*** |
| IL Small C&I DQ | 1 | 0.691 | 48.470*** | USNBQ | 11.932*** |
| IL Small C&I NP | 1 | 0.927 | 512.576*** | USR | 13.330*** |
| IL Small RE DQ | 1 | 0.642 | 6.167** | USR | 5.469** |
| IL Small RE NP | 1 | 0.888 | 360.566*** | ILR | 3.373* |
| IN Small C&I DQ | 1 | 0.538 | 24.853*** | INBBQ | 4.528** |
| IN Small C&I NP | 1 | 0.864 | 272.454*** | USCSENT | 2.637 |
| IN Small RE DQ | 2 | 0.824 | 71.912*** | USCAR | 5.668*** |
| IN Small RE NP | 1 | 0.956 | 1019.657*** | USR | 10.253*** |
| MI Small C&I DQ | 2 | 0.703 | 40.206*** | MIBBQ | 5.604*** |
| MI Small C&I NP | 1 | 0.895 | 381.302*** | USR | 5.791** |
| MI Small RE DQ | 1 | 0.941 | 701.898*** | USCSENT | 7.095** |
| MI Small RE NP | 1 | 0.944 | 798.578*** | USR | 4.691** |
| WI Small C&I DQ | 1 | 0.578 | 44.261*** | WIR | 1.665 |
| WI Small C&I NP | 4 | 0.929 | 139.442*** | USCAR | 2.789** |
| WI Small RE DQ | 2 | 0.814 | 49.852*** | USTRU | 5.987*** |
| WI Small RE NP | 1 | 0.973 | 1670.232*** | USENFQ | 16.884*** |

NOTE: The first column shows which of the banking variables is the dependent; the second column reports the number of lags in the model; the third column reports adjusted $R^2$ for that equation; the fourth column reports the F-test of the null hypothesis of all lags of the dependent variable being not significant; the fifth column lists which macroeconomic variable was picked according to statistical significance (the abbreviations are explained hereafter); the sixth column reports the F-test of the conventional null hypothesis on the lags of the macroeconomic variable. One asterisk (*) indicates that the lags of the variable are jointly significant at the 10 percent level, two asterisks (**) indicate significance at the 5 percent level, and three asterisks (***) indicate significance at the 1 percent level. All VARs include seasonal dummies. Lags (up to eight) are chosen according to the lowest Schwarz criterion of the system. All macroeconomic variables are in growth rates, all banking variables are in levels. List of the abbreviations used in the table: USCAR (Car sales), USR (Unemployment rate), USABQ (Total bankruptcy filings), USTRU (Permits, new, privately-owned buildings), ILTRU (Illinois building permits, new, privately-owned buildings), USNBQ (Non-business bankruptcy filings), USYPQ (), WITRU (Wisconsin building permits, new, privately-owned buildings). The second equation of each system, where the macroeconomic variable is the dependent, is not reported for simplicity and is available upon request. These regressions refer to the sample 1987Q1–1999Q2, in order to allow for out-of-sample validation.

Table 7: VAR with no seasonal dummies—Small banks

| Dependent variable | Lags | $\bar{R}^2$ | Lagged dep. F-test | Other variable | Other var. F-test |
|---|---|---|---|---|---|
| IA Small C&I DQ | 6 | 0.843 | 2.094* | USCAR | 4.889*** |
| IA Small C&I NP | 7 | 0.963 | 131.355*** | USR | 7.240*** |
| IA Small RE DQ | 8 | 0.878 | 19.197*** | USABQ | 6.745*** |
| IA Small RE NP | 3 | 0.969 | 468.888*** | USTRU | 9.500*** |
| IL Small C&I DQ | 8 | 0.852 | 6.068*** | USTRU | 11.203*** |
| IL Small C&I NP | 7 | 0.924 | 65.735*** | ILTRU | 4.086*** |
| IL Small RE DQ | 8 | 0.756 | 7.984*** | USNBQ | 5.436*** |
| IL Small RE NP | 4 | 0.859 | 39.529*** | USR | 4.045*** |
| IN Small C&I DQ | 4 | 0.594 | 1.356 | USR | 10.164*** |
| IN Small C&I NP | 4 | 0.851 | 49.485*** | USR | 6.016*** |
| IN Small RE DQ | 5 | 0.806 | 14.060*** | USR | 5.692*** |
| IN Small RE NP | 8 | 0.939 | 28.581*** | USTRU | 5.640*** |
| MI Small C&I DQ | 8 | 0.832 | 5.377*** | USTRU | 7.894*** |
| MI Small C&I NP | 4 | 0.895 | 81.476*** | USTRU | 3.363** |
| MI Small RE DQ | 4 | 0.918 | 82.529*** | USTRU | 5.862*** |
| MI Small RE NP | 1 | 0.943 | 762.367*** | USYPQ | 6.067** |
| WI Small C&I DQ | 6 | 0.677 | 4.175*** | USTRU | 7.117*** |
| WI Small C&I NP | 4 | 0.913 | 113.739*** | WITRU | 4.913*** |
| WI Small RE DQ | 4 | 0.788 | 16.008*** | USTRU | 8.385*** |
| WI Small RE NP | 3 | 0.950 | 289.663*** | WITRU | 4.770*** |

NOTE: The first column shows which of the banking variables is the dependent; the second column reports the number of lags in the model; the third column reports adjusted $R^2$ for that equation; the fourth column reports the F-test of the null hypothesis of all lags of the dependent variable being not significant; the fifth column lists which macroeconomic variable was picked according to statistical significance (the abbreviations are explained hereafter); the sixth column reports the F-test of the conventional null hypothesis on the lags of the macroeconomic variable. One asterisk (*) indicates that the lags of the variable are jointly significant at the 10 percent level, two asterisks (**) indicate significance at the 5 percent level, and three asterisks (***) indicate significance at the 1 percent level. All VARs include a constant term. Lags (up to eight) are chosen according to the lowest Schwarz criterion of the system. All macroeconomic variables are in growth rates, all banking variables are in levels. List of the abbreviations used in the table: USCAR (Car sales), USR (Unemployment rate), USABQ (Total bankruptcy filings), USTRU (Permits, new, privately-owned buildings), ILTRU (Illinois building permits, new, privately-owned buildings), USNBQ (Non-business bankruptcy filings), USYPQ (), WITRU (Wisconsin building permits, new, privately-owned buildings). The second equation of each system, where the macroeconomic variable is the dependent, is not reported for simplicity and is available upon request. These regressions refer to the sample 1987Q1–1999Q2, in order to allow for out-of-sample validation.

Table 8: Out-of-sample forecast evaluation

| Summary | MSE 1-step | MSE 2-step | CR 1-step | CR 2-step |
|---|---|---|---|---|
| OLS (4 lags of dependent) | 12.81% | 28.26% | 31.25% | 56.25% |
| OLS (seasonal dummies) | 11.18% | 18.49% | 37.50% | 40.63% |
| VAR (no seasonal dummies) | 10.35% | 7.66% | 50.00% | 37.50% |
| VAR (seasonal dummies) | 5.03% | 5.91% | 40.63% | 31.25% |

NOTE: MSE indicates mean squared error, CR the confusion rate. 1-step is the evaluation of the performance of the models in the one-step-ahead forecast (out of sample). 2-step is the evaluation of the forecast two quarters ahead (out of sample). OLS models include lags of the dependent variable as well as lagged macroeconomic variables. VAR bivariate models are described in the previous tables.

## Emerging Issues Series

A series of studies on emerging issues affecting the banking industry. Topics include bank supervisory and regulatory concerns, fair lending issues, potential risks to financial institutions and payment system risk issues. Requests for copies of papers can be directed to the Public Information Center, Federal Reserve Bank of Chicago, P.O. Box 834, Chicago, Illinois 60690-0834, or telephone (312) 322-5111.

These papers may also be obtained from the Internet at:
http://www.chicagofed.org/publications/workingpapers/emergingissues.cfm

## Emerging Issues Series