Paper 3338-2015

Time Series Modeling and Forecasting—An Application to Bank's Stress-Testing

SAS® Global Forum 2015

Ania Supady-Chavan, KeyCorp

ABSTRACT

This paper provides with basic understanding of how large US bank holding companies (hereafter, BHC) can perform stress testing. It explains how banks in general can perform time series modeling of balance sheet to establish bank's position during simulated stress. It gives overview of basic process behind model building and validation for Comprehensive Capital Analysis and Review (hereafter, CCAR) purposes, which includes, but is not limited to, back testing, sensitivity analysis, scenario analysis, and model assumption testing. This paper examines the procedures that happen behind the scenes of model forecasting process as well as explores statistics which play crucial role in assessing model performance and forecasting.

INTRODUCTION

CCAR is Federal Reserve's annual assessment of the capital adequacy of large U.S bank holding companies. The fundamental purpose is to help ensure that those BHCs maintain sufficient capital and are able to withstand very stressful operating environments and maintain their financial stability. Bank stress-testing has been gaining attention since 2007. It consists of analyzing, by the use of quantitative models, the effect of macroeconomic shocks on bank's balance sheet. The stress testing consists also of scenario analysis, which aids in capital assessment during downturn of economy. "Maintaining ready access to capital is central to BHC's ability to absorb losses and continue to lend. Board of Governors of the Federal Reserve System explained in their review, that in November 2011, Federal Reserve adopted the capital plan rule. It stated that BHCs with consolidated assets of $50 billion or more were required to submit annual capital plans to the Federal Reserve for review. The submission must contain the following:

1. Detailed description of the BHCs internal process for assessing capital adequacy
2. The policies governing capital actions, such as common stock issuance, dividends and share purchases
3. All planed capital action over a nine-quarter planning horizon
4. The results of the stress tests conducted under number of scenarios that can assess the sources and uses of capital under baseline and stressed economic and financial scenarios". (Board of Governors)

The important note here is that each BHC must meet the regulatory capital requirements for each projected quarter of the planning horizon. A common equity tier 1 ratio is a measure of bank's solvency. It measures high-quality capital as a share of risk-weighted assets which is then used by regulators to grade bank's capital adequacy.

The stress testing begins with mapping out risk factors in a form of macroeconomic variables, which closely relate to balance sheet variables. The most important part of the exercise is allocation of those risk factors since they carry the information about the economic stress. There are typically three scenarios that BHCs utilize in the exercise: base, adverse and severely adverse. Each scenario is closer described in publications located in Federal Reserve.Gov.

There may be also internally calculated scenarios, unique to how BHC can determine on its own projected stress which will not be discussed here.

This paper focuses on several core concepts of effective stress testing, utilizing example of a simple model for transparency.

MODELING PROCESS

The industrywide data is used to illustrate stress testing techniques. The dependent variable chosen for this study is Commercial & Industrial loans (hereafter, C&I loans): Assets and Liabilities of FDIC-Insured Commercial Banks and Savings Institutions (Amounts in $ Millions). C&I loans represent an important line of business for banking industry and key source of funds for business sector. They usually charge flexible rates of interest that are tied to the bank prime rate or else to the LIBOR.

The graph below shows behavior of Commercial and Industrial loans industrywide over thirty year time span. In most recent years, there have been periods identified as recession. Following are the quarters which indicate recession: 2001.1, 2001.2, 2001.3, 2001.4, 2007.4, 2008.1, 2008.2, 2008.3, 2008.4, 2009.1, and 2009.2.

Figure 1: Commercial & industrial loans: Assets and Liabilities of FDIC-Insured Commercial Banks and Savings Institutions (Amounts in $ Millions).

This information introduces a challenge because most recent recession (1Q2008-2Q2009) had significant impact on the C&I loans. Certain liquidity events impacted the historical data used in the modeling. Research showed that during the early part of the recessionary period, there was increasing amount of concern around the safety and soundness of the entire banking industry. This liquidity crisis manifested itself at some banks in adverse deposit withdrawals (bank runs) as well as in significant funding of commercial loan commitments. Draws on commercial lines of credit provided borrowers with additional cash to protect themselves during the crisis. Looking at the history in the above chart it is clear that C&I loans experienced an unusual increase in balances during the early stages of the recession attributable to this liquidity event.

The first step in modeling process is a variable selection, which is crucial in establishing the stress testing events. The candidate macroeconomic variables are retrieved from source available to bank. It is a multi-stage process which comprises of a series of statistical tests and expert selection. Each variable is available along with the scenarios associated with the level of the stress mentioned previously. This paper will not discuss this topic due to its complexity. Final subset of macro variables must ensure that they are both statistically significant (i.e., non-zero) relationship with the dependent variable and economically relevant to the item being modeled. In order for the candidate model to be selected, both those conditions must be satisfied. Based on above criteria a candidate variable was selected to model C&I loans: Fixed Investment - Nonresidential - Total; (Bil. $; SAAR).

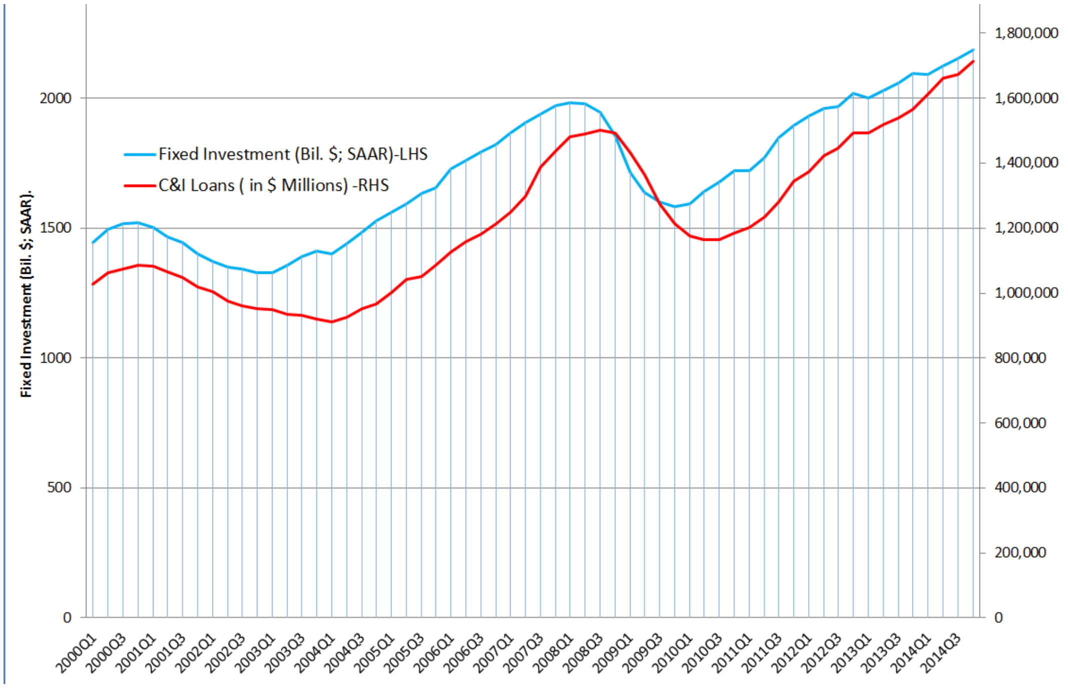


Figure 2 C&I loans and Fixed Investment

From the plot above we observe that the variables tend to track together well. Some lagging effect may be present and various lags of independent variable may be tested.

Fixed Investment represents investments in physical assets such as machinery, land, buildings, installations, vehicles, or technology. The correlation between C&I loans and fixed investments is quite logical as companies frequently require bank loans to fund their capital expansion. This logic applies to all levels of corporate borrowers, from small businesses to large corporate clients. Fixed Investments also reflect the level of confidence that business owners or managers have about the ability to increase earnings in the next few years. The reasoning is that they wouldn't be likely to tie up additional capital in fixed assets for several years or more, unless they thought it would be a commercially viable proposition in the longer term. This also adds to its explanatory power.

The next step is assessment of modeling methods available. Since balance sheet variables can be collected over time and expressed as time series, therefore time series modeling techniques are a natural choice for modeling objective. The goal is to arrive at a parsimonious model that makes the best use of data and provides practical and intuitive results.

This paper will discuss application of one method: Regression with ARMA errors model witch is flexible and long term focused.

$$Y_t = \beta X_t + \frac{\theta(B)}{\varphi(B)}\,\varepsilon_t$$

This method may be preferred by modelers who desire to maintain usual interpretation of regression coefficients and avoid models that create βs conditional on the values of previous values of response variable. Such issues have been noted in the time series literature (see e.g., Hyndman 2010). The stress testing exercise requires us to preserve the usual interpretation of a regression coefficient as a measure of sensitivity, meaning the effect of a unit change in a predictor variable on the response. Regression coefficients of other time series methods that include dependency of lagged response in the model do not possess this interpretation, because of dynamic dependence on the response. This can be observed in models when instead of regressing Y on X, one regresses diff(Y) on diff(X). The regression equation is now:

$$Y\hat{}_t - Y_{(t-1)} = \beta(X_{(t)} - X_{(t-1)})$$

This can be written as:

$$Y\hat{}_t = Y_{(t-1)} + \beta X_{(t)} - \beta X_{(t-1)}$$

This model brings prior values of both Y and X into the prediction. This is an important factor when model is not only generating fit but long term forecasts as well. Once the actual values of dependent series are no longer available, it is the prior forecast that generates one step ahead forecast. The above regression can become a dynamic regression when response variable is a function of exogenous inputs as well as lagged values of Y_t and lagged values of error terms.

$$\varphi(B)(Y_t - \sum_{i=1}^{q} \beta_i \varepsilon X_{t-i}) = \theta(B)\varepsilon_t$$

Modeling response or difference of the response using ARMA process tend to be short-term focus as the correlation between the actual responses and the one step ahead prediction tends to weaken with time. If the objective of the exercise is to find not only the best fitting model but mainly relevant macroeconomic drivers of the item being modeled, then regression with ARMA errors may be considered.

Dependent variable that is being modeled - C&I loans, can be expressed in terms of independent explanatory variables since it investigates the effect of changes in various factors on Y. This looks like simple linear regression, but applying it to time series often reveals autocorrelation in the error terms. In such cases, the ARMA model can be used to correct that. Modeling residuals in such way allows extracting information the errors contain that was not explained by independent variable.

The nature of the error terms requires using maximum likelihood estimation instead of least squares method. This method combines the input (explanatory) series with ARMA models for the errors. “The assumption of stationarity applies to the noise series. There is no requirement that the input series be stationary. If the input series are nonstationary, the response series will be nonstationary, even though the noise process may be stationary. If the nonstationary series are used for inputs, we fit the input variables first with no ARMA model for the errors and then examine the stationarity of the residuals before identifying the ARMA model for the noise part”. (SAS/STAT User’s Guide)

Same logic applies to the stationarity of the response series. If the ARMA model framework is not applied to the response series, the stationarity is irrelevant. It only applies if we rely on the past values of the response series to predict its future values. This method is widely exercised in time series modeling but is more complex and long term forecasts might not be reliable. We reduce the complexity of the model framework by depending on explanatory power of macro variables that build strong regression model which is later examined using sensitivity and scenario analysis in a stress testing approach.

There are several benefits of this approach. Regression with ARMA errors follows deterministic trend. Time series with a deterministic trend always revert to the trend in the long run (the effects of shocks are eventually eliminated), which can be a nice feature to have when estimating forecasted position for stress testing. Additionally, forecast intervals have constant width which can be a desired feature for sensitivity testing. It is easier to draw statistical inference from the model and give estimates proper interpretation. All the mentioned properties can be viewed as beneficial when aiming to arrive with the stress testing model for CCAR which should have straightforward interpretation of much simplified relationship to macroeconomic events and lastly, be predictive.

MODELING OUTCOME

Several attempts were made to fit a valid model using code below. The final candidate model does not contain any moving average terms.

```
ods graphics on;
PROC ARIMA DATA= SASGF2

PLOTS(UNPACK)=SERIES(ALL)
PLOTS= FORECAST(ALL);
IDENTIFY VAR=CL nlag=24  CROSSCORR=FI NOPRINT ;

estimate P = (1) Q = (0) INPUT = (2$FI)
method=ml maxiter=200   plot;
outlier maxnum=100 alpha=0.01 ;
RUN;
 FORECAST ID=time INTERVAL=quarter  LEAD=22
 OUT=PRED;

RUN;
```

The ARIMA Procedure

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	1699.5	121606.9	0.01	0.9888	0	CL	0
AR1,1	0.96680	0.04154	23.27	<.0001	1	CL	0
NUM1	754.52578	59.97279	12.58	<.0001	0	FI	2

Figure 3: The ARIMA procedure.

Based the output above, it is easy to write model specification.

$$\text{C\&I Loans}_t = 1699.50 + 754.52\ \text{FI1}_{t-2} + \varepsilon_t$$

$$\text{WHERE}$$

$$\varepsilon_t = \frac{1}{(1 + 0.9668\text{B}(1))}\delta_t$$

FI1_{t-2}: Fixed Investment - Nonresidential - Total; (Bil. $; SAAR).

Equation Interpretation:

A one billion dollars increase in current Nonresidential Fixed Investment will, on average, increase C&I Loans by $ $754.52\ \text{M}\ \$$ in two quarters from now, all else held constant.

MODEL VALIDATION AND TESTING

This section will discuss briefly several core concepts of effective model challenge including:

1. Model assumptions
2. Model fit
3. Backtesting
4. Sensitivity analysis
5. Scenario analysis

1. Model assumptions

There are several main model assumptions which include: stationarity of residuals; no serial correlation in residuals; statistical significance associated with exogenous variables which should not display feedback from dependent variable; no multicollinearity among independent series which should exhibit correct sign of coefficients supported by economic intuition. All of these assumptions must be met in order to be able to use regression analysis or maximum likelihood to accurately estimate coefficients. This way, we can make use of standard set of test statistics, such as the t- and F-statistics, to determine the statistical significance of the individual inputs used. Most modelers try to first establish significant terms in the model and residuals that are not autocorrelated by appropriately defining lag structure.

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	5.94	5	0.3121	0.258	0.065	0.053	-0.034	0.035	-0.139
12	7.53	11	0.7547	0.100	0.028	0.041	0.049	0.039	0.077
18	16.36	17	0.4982	-0.079	-0.128	-0.214	0.017	0.164	-0.111
24	17.76	23	0.7702	0.039	0.047	0.056	0.053	-0.066	0.023

Figure 4: The autocorrelation Check of Residuals.

The residuals diagnostics are shown above. The null hypothesis tested in each row is that no autocorrelation exists at any lag between, say, 1 and 6 (first line in the table). The p values in the table indicate that this null hypothesis is not rejected at all tested lags. Model has white noise residuals. That is, all of the systematic variation in the data is "captured" in the model specification, and all that is "left over" in the residuals is randomness. The results above indicate that the AR(1) model is adequately specified.

2. Model fit

The model fit was performed on 60 periods from 2000 Q1 to 2014 Q4. This means, we are using all the past values for C&I loans to develop model. The forecast was performed on 12 periods from 2014 Q4 to 2017 Q4 which is the last data point at the base scenario. The model was evaluated as valid and statistically significant. The model fit was derived from: 'abs (Modeled/Actual-1)' and equals 1.24%. We use this figure as an indication of the model accuracy. Below is the SAS output of model fit and forecast.

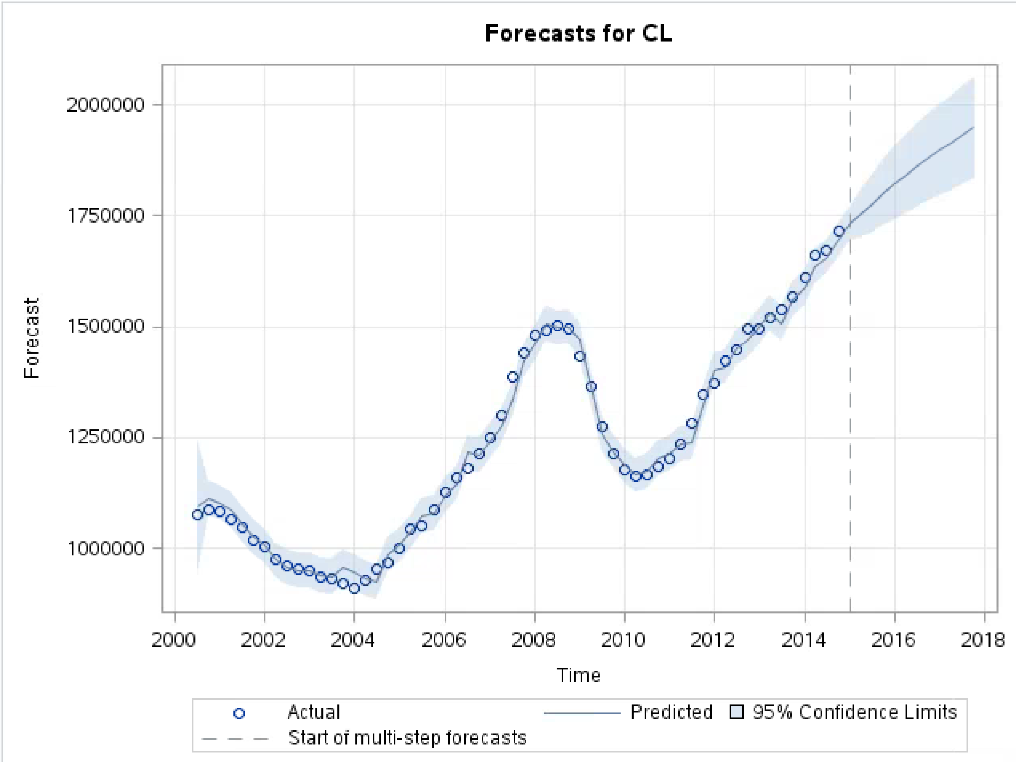


Figure 5: SAS output of model fit and forecast of C&I loans.

As a part of the model validation check, it is important to determine the source and power of the residuals. Error terms can play significant role in following the path of the response. Therefore, we check the explanatory power of macroeconomic variables by plotting regression without the error terms. The original estimates are used to do that, no the simple linear regression code.

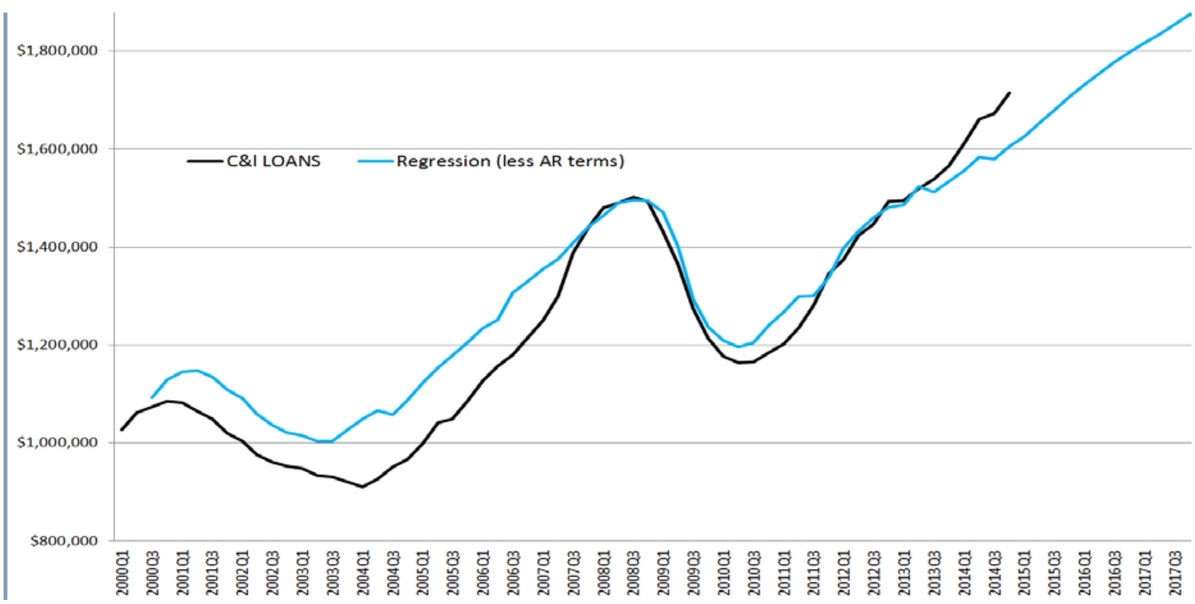


Figure 6: Actual vs. Regression without error terms

The process behind ARIMA syntax is quite simple for the first part of the equation: $1699.50 + 754.52\ \mathrm{FI1}_{t-2}$. This process generated the blue line in the chart above. This graph above means that, it is the explanatory variable that plays significant role in matching the path of the response, not the error term. The independent series does a good job in detecting the economic downturn and follows the general trend of the dependent series. The residuals are calculated by subtracting from the actuals the forecasted values generated by the process behind regression described above. Once there are no more actual values available, the next residual needs to be forecasted. The AR process determined by 'proc arima' is now applied as following:

$$\varphi_1 n_{t-1}\ ,\ \varphi_1 n_{t-2} \ldots \varphi_1 n_{t-n}$$

The last available residual (where actuals are still present), is multiplied by the AR1 term coefficient to calculate first residual of the forecasting step. This is a recursive process which ends with the last data point forecasted, in this case 2017.Q4. The summation of both equations generates the final model fit.

3. Backtesting

Backtesting will assess validity of model based on acceptable tolerance for back-testing results. Typically, the model's backtesting is designed in such way that at least last 9 quarters of the actuals are specified as holdout sample. It means we are treating the last 9 quarters of the response series as unobserved. This will allow evaluation of the forecasting performance of the candidate model. The reasoning behind this sampling design is that the model is used to perform nine quarter forecast for CCAR and backtesting must be consistent with model use.

The back testing was performed on 9 periods from 2012 Q4 to 2014 Q4 and resulted in average error of 3.3% calculated using 'abs (Modeled/Actual-1)' formula. Last four periods from 2014.Q1 to 2014.4 exhibit some growth with highest error recorded at 7%. The governance around reasonability of the model forecast error is put in place to control for model risk.

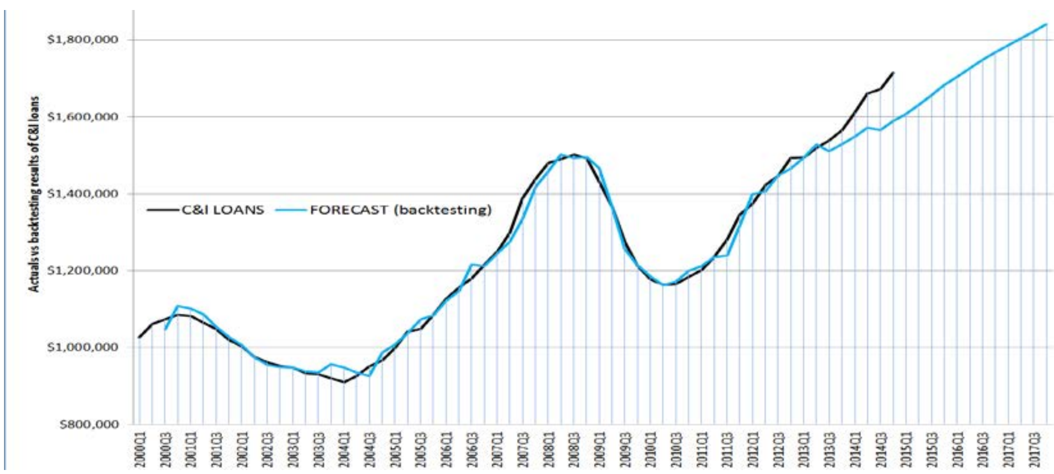


Figure 7: Backtesting error plot.

The training period contains 51 data points and is defined from 2000.Q1- 2012.Q3 where model was evaluated as valid and statistically significant. The outputs from ARIMA procedure are provided below.

The ARIMA Procedure

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	-44673.8	101512.2	-0.44	0.6599	0	CL	0
AR1,1	0.91386	0.05882	15.54	<.0001	1	CL	0
NUM1	755.89861	58.97236	12.82	<.0001	0	FI	2

Figure 8: The ARIMA procedure on backtesting period from 2000 Q1- 2012 Q3 forecasting 9 quarters out.

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.51	5	0.6226	0.235	0.022	-0.027	-0.021	-0.023	-0.086
12	4.46	11	0.9544	0.089	-0.038	0.036	-0.002	0.030	0.060
18	11.48	17	0.8307	-0.113	-0.100	-0.172	0.061	0.183	-0.058
24	16.36	23	0.8392	0.082	0.071	0.155	0.090	-0.023	-0.096

Figure 9: The Autocorrelation check of residuals.

Checking model assumptions should be also performed for each backtesting model. To test normality, the QQ-plot should not show systematic departure from the diagonal line and the histogram should take the bell shape. If the error distribution is significantly non-normal, the confidence interval may be too wide or too narrow, which later can result in unreliable sensitivity conclusions. Similar conclusions can be drawn from violations of homoscedasticity. If there is violation it makes it difficult to assess the true standard deviation of the forecast errors, usually resulting in confidence intervals that are too wide or too narrow.

To test the independence, the autocorrelation plot should not have large lags after lag 0. The residuals should not strongly depend on the past values or the lagged residuals. The Ljung-Box test can be used to conclude on that. In some cases, the autocorrelation is present but is mild, meaning the serial correlation is not strong. This can be examined by Durbin-Watson statistic. The DW statistic provides a test for significant residual autocorrelation at lag 1. The DW statistic is approximately equal to 2(1-a) where '*a'* is the lag-1 residual autocorrelation, so ideally it should be close to 2.0 or between 1.4 and 2.6 for a sample size of 50. In time series regression based models, the autocorrelation can be potentially serious. It mainly indicates that there is room for improvement to the model fit and other independent explanatory variables should be considered. The stationarity of the residuals should be checked by using code below:

```
ODS graphics on;
PROC ARIMA DATA= PRED;
IDENTIFY VAR=residual stationarity= (adf=(0 1 2 3 4 5 6));
run;
```

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-38.3198	<.0001	-5.54	<.0001		
	1	-39.6717	<.0001	-4.31	<.0001		
	2	-42.0149	<.0001	-3.64	0.0005		
	3	-44.5601	<.0001	-3.27	0.0016		
	4	-44.9698	<.0001	-2.82	0.0058		
	5	-95.3988	<.0001	-2.72	0.0077		
	6	-25.6822	<.0001	-1.96	0.0488		

Figure 10: Augmented Dickey-Fuller Unit Root Test on residuals.

The residuals are shown as stationary from the ADF unit root test.

In order to assess the long term forecasting performance of the model, series of in-sample tests can be performed. First, the model may be calibrated to full data set and next, same estimates can be used to forecast dependent variable on randomly selected windows of time of length 9 periods or more to stay consistent with the model use. The forecasting window can be incrementally moved in one direction, each time recording the forecast error. In order to assess model's stability over time, a training data set of say 50 data points and the out of sample holdout of 9 data points can be created. Next, the very last observation of the training set is added to the hold-out sample until we have reduced training set by at least half. This sort of sequential test will help decide whether the model is robust and stable in its performance metrics recorded such as: mean absolute error, mean error or backtesting error. This exercise will result in developing 20 plus models allowing recording the coefficient estimates each time. A common conclusion from such test is that models are sensitive to reduction of sample size and should be build utilizing sufficient historical data.

4. Sensitivity analysis

The purpose of this analysis is to determine how changes applied to values of an independent variable will impact dependent variable under a given set of assumptions. It is applied when attempting to determine the impact the actual outcome a particular variable will have if it differs from what was previously assumed. The most challenging part is determining the threshold of changes to the input variables to find out the first breaking point. In this model, that point is determined by observing when will the new forecast under new assumptions cross the original unadjusted confidence intervals in a forecasting period. The assumptions have to be made regarding variable used in the model: Fixed Investment and Corporate Profits.

In a regression model, it is fairly easy to derive value of X when the Y in a known value. Since the confidence intervals are provided - a value below next forecasted point of the lower bound and value above the next forecasted point of upper bound can be used to derive X. This would be the first breaking point of a model (with no intercept) when assumptions regarding independent value have changed. If current lower bound is 1,605,452 (M $) then its corresponding known value of FI is 2,155 (Bil. $). We can arrive at that by dividing value of lower bound by coefficient of independent variable. If we wanted to know the value of FI corresponding to 10% decrease of lower bound value, then we can derive it from: (1,605,452-(1,605,452*0.10))/755=1.915 which is a new value of FI. In addition to the regression, there is a model applied to the errors, which makes is more complex to find the first breaking point of the originally formulated model by the method described above. Modeler can try testing the sensitivity by applying percent change to the values of independent series during their last 9 quarters. In the plot below, we can view the result of the original model utilizing FI changed gradually up or down by 0.01, 0.02, 0.03…0.9.

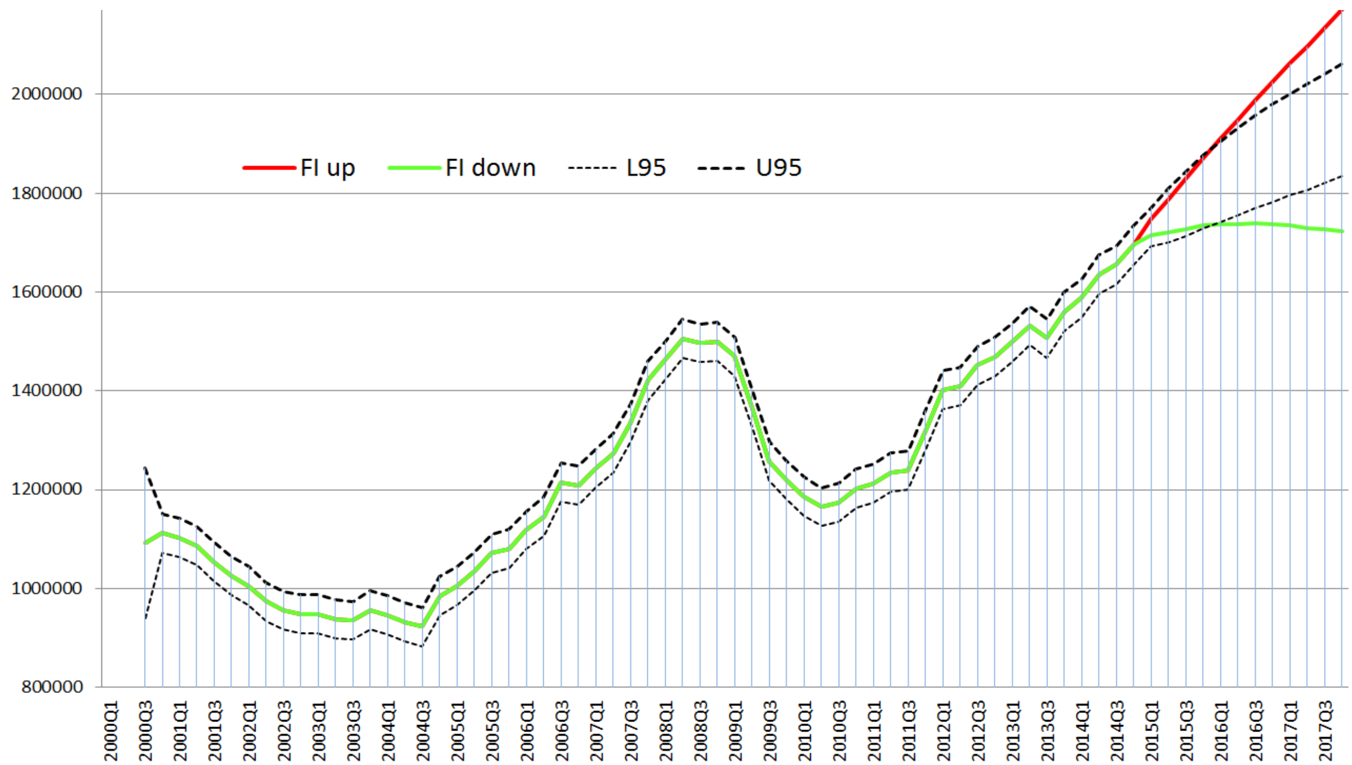
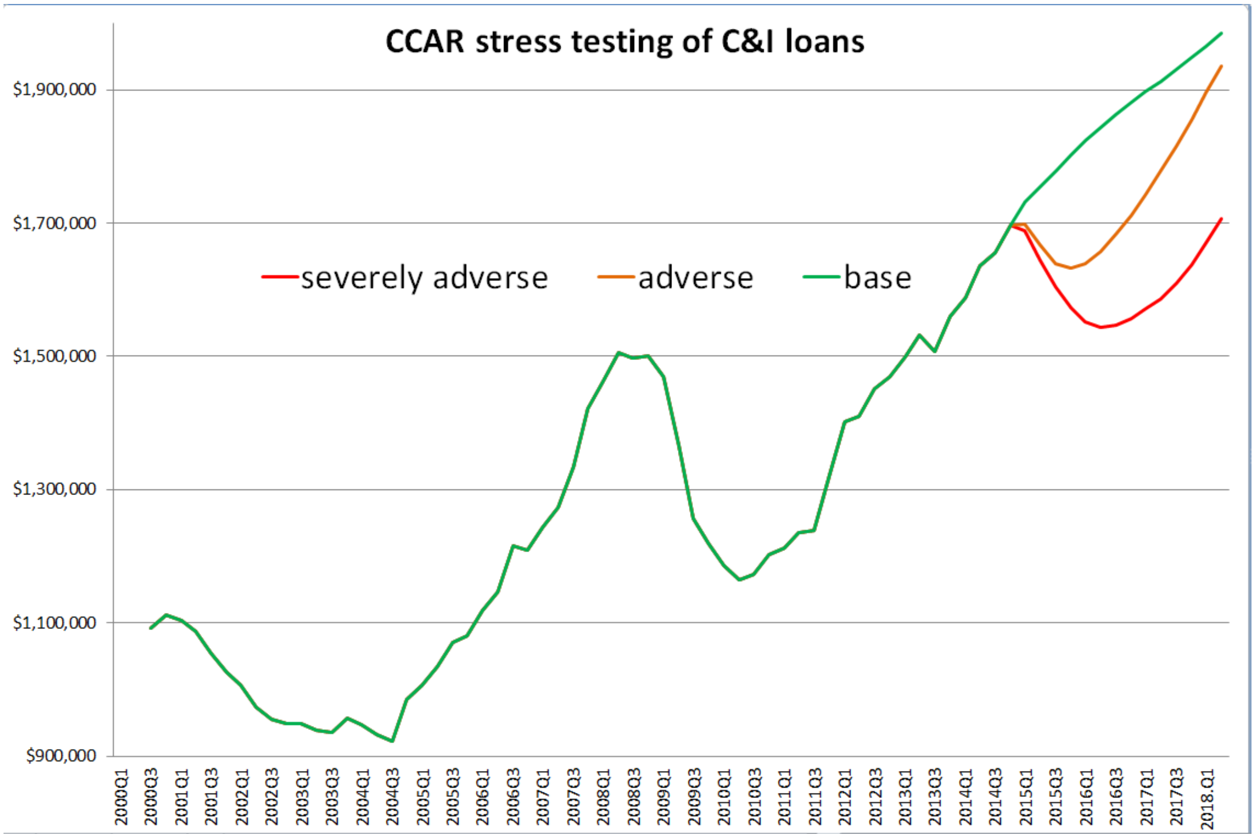


Figure 11: Breaking confidence intervals.

From the results above, we can conclude that the FI is fairly sensitive to changes to its original values. Model's first breaking point starts with just 5% change to FI. This part if usually discussed with SME to assess true risk behind model's output and management's ability to rely on it.

5. Scenario analysis

A central goal of the capital plan rule is to ensure that large BHCs have robust internal practices and policies to determine the appropriate amount and composition of their capital, given the BHC's risk exposures and corporate strategies and in line with supervisory expectations and regulatory standards. The supervisory scenarios in CCAR are also used in the Dodd-Frank Act stress tests. Under the Board's Dodd-Frank Act stress test rules, the Board is required to provide BHCs with a description of the supervisory macroeconomic scenarios no later than November 15 of each calendar year (Federal Reserve.Gov). BHCs prepare for this process by developing models and running scenarios provided by Federal Reserve, external vendor or otherwise simulated. The training of the candidate model was done utilizing base scenario, where no economic downturn is present. The following adverse and severely adverse scenarios aim at describing economy in distress and various stages of recovery. The main goal of this exercise is to arrive with most reliable and robust process of translating stress in economy into stress of the balance sheet of BHCs. The model should be sensitive enough so that clear separation of scenarios is present.



The following results allow assessing bank's position in commercial and industrial loans during stress. There is clear separation of each scenario and distance from base to adverse and base to severely adverse dictate varying recovery. The model is simple enough that it allows for traditional interpretation of parameter estimates while passing major model assumptions relevant to the goal of exercise. There is clear statistical relationship to dependent series which is also supported with SME opinion and research. The backtesting results and sensitivity analysis can point to risk factors of the model that are typical in most time series analysis. This is an industrywide model indicating how would C&I loans behave under stress scenarios, given Fixed Investment is their main driver.

CONCLUSION

This paper describes how BHCs can perform stress testing of their balance sheet variables using time series approach. The most intuitive and transparent method was selected: Regression with ARMA errors. Final candidate model used only autoregressive terms applied to residuals. The model performance was checked against model with no errors to establish predictive power of the explanatory variable. Model was back-tested and the error was evaluated. The sensitivity analysis helped establish threshold of minimum changes applied to FI in order to break model's original confidence intervals. Finally, scenario analysis was performed on the final model to establish banks position during stress for the C&I loans.

REFERENCES

Federal Deposit Insurance Corporation. “Quarterly Banking Profile.” Accessed March 7, 2015. https://www2.fdic.gov/qbp/timeseries/BalanceSheet.xls

Board of Governors of the Federal Reserve System http://www.federalreserve.gov/bankinforeg/ccar.htm

Hyndman, R. J. (2010, October). "The ARIMAX Model Muddle." *Rob J. Hyndman*. Retrieved February 7, 2013 from http://robjhyndman.com/researchtips/arimax/.

SAS Institute, 2008. “SAS/STAT 9.2 user's guide.”

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Ania Supady-Chavan
Sr. Quantitative Modeling Associate
KeyBank – Model Risk Control
ania_supady-chavan@keybank.com

DISCLAIMER

The opinions expressed in this document are solely the author’s and do not necessarily reflect those of KeyBank or any of its subsidiaries. Any graphs or tables shown are based on non-KeyCorp data and are for illustrative purposes only.