[[1]](#footnote-1)

Applying Multivariate Linear Regression to Major Financial Institutions Leading to the Financial Crisis of 2008

Javier Saldaña Jr.

# **INTRODUCTION**

The year 2008 will forever be remembered as the year of the most devastating financial crisis since the Great Depression of the 20th century. The United States’ Government Accountability Office (GOA) reports the US economy suffered an estimated $22 trillion in losses [1]. Major financial corporations were liquidated while others were forced to sale or merge in order to stay afloat. While Wall Street was suffering from the market panics, millions of Americans saw their sub-prime loans maturing and not being able to refinance a home that had lost a substantial amount of its value thanks to a slowing housing market. Although it didn’t take long for the economy to collapse, the crisis is the result of years of dangerous trade practices and poor oversight.

The US was rocked by the burst of the Dot Com bubble that burst in 2002, which saw the Dow Jones, NASDAQ, and S&P 500 wipe out 5 years’ worth of gains. However, a historically strong and reliable housing market was able to bring the US economy roaring back. Not only did the economy recover the value lost, but it was able to nearly double it within 5 years from the burst of the Dot Com bubble.

Liberal lending practices led to the creation of sub-prime loans, which would begin at a small fixed rate for a few years before their new rate would be initiated. These loans were referred to as sub-prime because they were generally made available for individuals with lower than standard credit strength. As a result, the financial institutions found themselves unable to get investors to take on this newly created investment opportunity for the risk of a default was too high. In order to hedge against the risk of default, financials institutions combined the toxic debt with superior debt, which would then bring the credit strength of the debt to a more attractive level for potential investors. The selling of this debt as collateral or investment opportunities is what began to set the stage for the 2008 financial crisis.

The housing market began to slow and housing prices were adjusted by the law of supply and demand in order to account for the flood of homes now available in the market. The early warning signs were dismissed by the US Federal Reserve and the International Monetary Fund for nobody believed a historically stable housing market would collapse. As the sub-prime loans began to mature, millions of Americans found themselves unable to make their mortgage payment, refinance/sell their home in a flooded market. Homes were foreclosed and any collateral made from the toxic debt was now worth pennies on the dollar as defaults rose. Eventually, the buyers of the sub-prime mortgage debt found their books filled with toxic debt they couldn’t get rid of on the secondary market.

The US Government began to see the severity of the issue and attempted to resurrect the housing market by slashing borrowing interest rates and opening bail out programs for companies filled with toxic debt in their books. This strategy, however, would not be enough when Fannie May and Freddie Mac found itself unable to liquidate and the US Government became the major shareholder in Fannie May and Freddie Mac. Within weeks, other major institutions found themselves in similar circumstances. It was at this point that the domino effect of the 2008 financial crisis resulted in a global recession and a US economy losing over 10 years’ worth of gains.

While the recession affected industries and countries all over the world, the greed of the US financial sector and lack of oversight are frequently viewed as the culprits primarily responsible for this crisis. We set out to study the strength of the correlation between the major financial institutions in the United States and the market indices (Dow Jones Industrial Average, S&P 500, and Nasdaq Composite Index). In addition, are there specific financial institutions that hold a higher coefficient than others with respect to the correlation to the market indices?

# **Descriptive Statistics**

The data used for this study was sourced from <https://finance.yahoo.com> and is available for the public to download. The data set includes market values from all entities listed in Table 1 (see appendix), which range from January 1, 2002 to December 31, 2009.

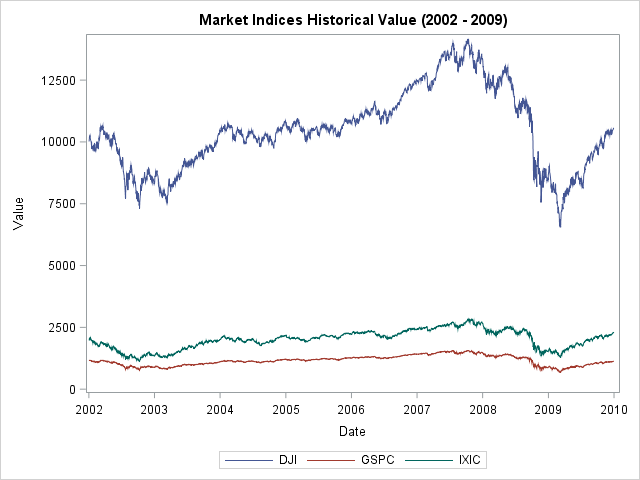
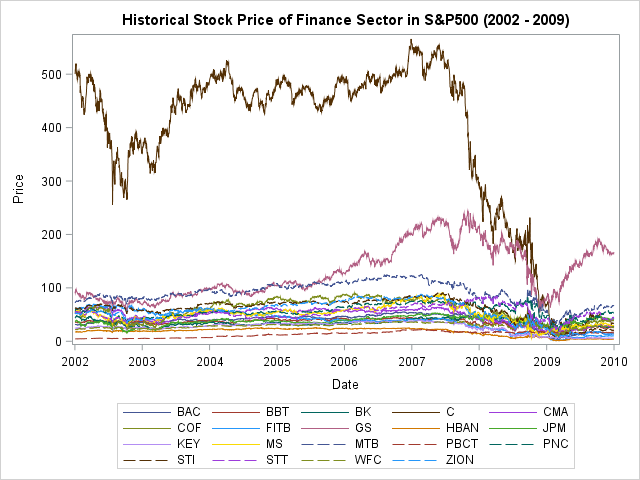
The indices are the three most popular stock market indices in the United States. Unlike standard stock data, they don’t carry a dollar value. Instead, their value is labeled as points, which are derived from proportions of the market (normally from companies listed in their index). The Dow Jones Industrial Average is calculated from the top 30 stocks in its index, which are all listed on the New York Stock Exchange. The Nasdaq Composite Index is based off the performance of 3,000 companies listed on its index. The S&P 500 comprises of the leading companies in the US market and includes over 80% of all available market capital. The S&P 500 is calculated based on its listed companies and the number of publicly available shares of their listed companies. Figure 1 plots the raw points of each index, which clearly shows an event occurred in 2008.

Figure 1 - Plot of raw index data values from 2002 to 2009

The other variables are major financial institutions all of which are either listed on the Dow Jones, S&P 500 and/or Nasdaq. Since the financial institutions are some of the largest institutions in the US, there is a high probability they were stakeholders in the 2008 financial crisis. The values for the stocks are their respected share price. Since the data is captured from 2002 through 2009, it is continual with the close share price marked on that given date. Unlike the indices, there is no specific calculation which determines the price of the share. The price reflects a true market value of the shares based on the supply and demand chain. While the indices depend on their listed companies to determine their points, the share prices are all independent of each other. However, there are extraneous variables which may affect them all since they all belong to the same sector. Figure 2 shows the plot of the raw share prices of the selected financial institutions. Just like with the indices plot, there is visible evidence of seasonality. In order to account for the seasonality, the data for all variables was transformed using a log transformation.

Figure 2 - Plot of major bank stock price from 2002 - 2009

# **Analysis**

With the data sorted and transformed accordingly, the following hypothesis may now be tested:

|  |  |  |
| --- | --- | --- |
| TABLE 2 | | |
| Index Model | Durbin-Watson without lag | Durbin-Watson with lag 1 |
| Dow Jones | 0.2794 | 2.1202 |
| S&P500 | 0.2674 | 2.1324 |
| Nasdaq | 0.1109 | 1.7916 |

Considering the companies selected are part of the same industry, there is a slight concern regarding the depth of independence between variables. Furthermore, it appears there is a trend in the price of the shares for the given companies. However, the residuals are normally distributed in the raw data. In order to address the trend violation, a log transformation helps de-trend the residuals slightly. A Durbin-Watson test confirms there is a positive autocorrelation in the three models. However, once we apply a lag of 1 to the model, we can see the Durbin-Watson statistic comes close to 2 and the autocorrelation is now addressed. Having addressed the auto correlation, the linear regression models may now be tested using a lag 1.

The following models were utilized to test the hypothesis above in order to determine if there is in fact a difference between the shares and whether they had any impact on the index. A linear regression was performed on the time series data with a lag of 1. The model outputs are listed in Table 3 and shows the coefficient of each variable in the model in addition to the p-value associated with the estimate to determine if it is statistically significant. The coefficient (listed as estimate in the table), shows the impact of the variable on the model. If the estimate is positive, then it would have a positive effect on the model whereas a negative coefficient would indicate it has a negative effect on the model. The p-value measures the significance of the coefficient at a 95% confidence level.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| TABLE 3 |  | |  | |  | |
|  | Dow Jones | | S&P 500 | | Nasdaq | |
| Variable | Estimate | p-value | Estimate | p-value | Estimate | p-value |
| Intercept | 6.8262 | <.0001 | 4.7373 | <.0001 | 4.7452 | <.0001 |
| Date | 0.0000506 | <.0001 | 0.0000245 | 0.0314 | 0.0000430 | 0.0004 |
| logBAC | 0.0617 | <.0001 | 0.0669 | <.0001 | -0.007296 | 0.5245 |
| logBBT | -0.0899 | <.0001 | -0.1046 | <.0001 | 0.0126 | 0.4583 |
| logBK | 0.0805 | <.0001 | 0.0713 | <.0001 | 0.0495 | <.0001 |
| logCMA | 0.0457 | 0.0021 | 0.0680 | <.0001 | 0.0132 | 0.3747 |
| logCOF | 0.0649 | <.0001 | 0.0865 | <.0001 | 0.0881 | <.0001 |
| logFITB | 0.0285 | 0.0010 | 0.0214 | 0.0092 | 0.0267 | 0.0008 |
| logGS | 0.0962 | <.0001 | 0.1292 | <.0001 | 0.1641 | <.0001 |
| logHBAN | -0.0601 | <.0001 | -0.0470 | <.0001 | -0.007955 | 0.3661 |
| logJPM | 0.0416 | 0.0033 | 0.009600 | 0.4779 | 0.1049 | <.0001 |
| logKEY | -0.0198 | 0.0638 | -0.0246 | 0.0157 | 0.007417 | 0.4733 |
| logMS | 0.0236 | 0.0320 | 0.0362 | 0.0005 | 0.0546 | <.0001 |
| logMTB | -0.0244 | 0.1764 | -0.006346 | 0.7114 | 0.0332 | 0.0466 |
| logPBCT | 0.0480 | 0.0004 | 0.0924 | <.0001 | 0.0797 | <.0001 |
| logPNC | 0.0307 | 0.0260 | 0.0511 | 0.0001 | 0.0135 | 0.3025 |
| logSTI | 0.1069 | <.0001 | 0.0676 | <.0001 | -0.0294 | 0.0296 |
| logSTT | 0.0666 | <.0001 | 0.0709 | <.0001 | 0.0520 | <.0001 |
| logWFC | -0.1786 | <.0001 | -0.1922 | <.0001 | -0.1193 | <.0001 |
| logZION | 0.0394 | 0.0010 | 0.0502 | <.0001 | -0.006994 | 0.5319 |

# **Interpretation and Conclusion**

Based on the statistical output of the models, it appears there is sufficient evidence to suggest at least one of the banks’ mean is different than the others (F statistic = 4644.39, p-value < 0.0001) and most of the banks had a statistically significant effect on the models (p-values < 0.05). In the Dow Jones model, the evidence suggest all banks with the exception of KEY and MTB were statistically significant at the 95% confidence level. The S&P 500 model finds that JPM and MTB were not statistically significant in the model at the 95% confidence level. In the Nasdaq model, BAC, BBT, CMA, HBAN, KEY, MTB, PNC, STI, and ZION were not statistically significant at the 95% confidence level.

Out of the banks that were statistically significant in the Dow Jones model, STI had the highest coefficient (0.1069) in the model, which means that for every 10% change in the mean of the Dow Jones, STI incurred a 28% change in the mean. For the S&P 500 model, Goldman Sachs has the highest coefficient (0.1292), which results in a 35% change in the mean for every 10% change in the mean of S&P 500. As for the Nasdaq model, GS once again produced the highest coefficient (0.1641), which equates to a 46% change in the mean for every 10% change in the Nasdaq.

While the evidence suggests STI and GS all had the strongest effect on their respective models, the study is based on an observational analysis which means we are unable to establish causal inferences may not be applied outside of the study. Furthermore, the companies selected for the study were not randomly selected which means the results may not be applied to other companies and/or sectors outside of the study. In other words, it is not safe to assume the banks are still leading the market with the same coefficients. The results also may not be applied to institutions outside of this study, which means that financial institutions outside of this scope may not be implicated in the study regardless of their involvement or lack thereof in the 2008 financial crisis.

# **Conclusion**

Based on the results, the evidence suggests that SunTrust Bank and Goldman & Sachs possessed the highest coefficient with respect to their models. This coincides with reports from the investigations into the 2008 financial crisis, which all confirm that Goldman Sachs and SunTrust both were involved in deceptive trade practices, which not only led to generous profits prior to the crash but also to severe losses. Goldman Sachs admitted to not only defrauding investors into purchasing toxic debt but also attempting to profit off the inevitable failure of that debt. The U.S. Government and Goldman Sachs reached an agreement in which Goldman Sachs would pay over $5 billion in restitution. SunTrust found itself in a similar situation when investigations into the practices of the company revealed that they were intentionally selling Financial Housing Authority backed-loans even though they didn’t meet the FHA requirements. Once the loans would default, they would claim the loss with the US Government. The investigation found that more than 50% of their FHA-backed loans did not meet the requirements. As a result, SunTrust was ordered to pay just shy of $1 billion to the US Government.

Further study into this topic would be recommended to investigate the volume being traded as well, instead of just stock prices. Volume being traded would give researchers the ability to determine whether the stock is selling profusely as a result of a crash or whether the price increase/decrease is related to extra variables not included in the study. Also, narrowing the scope of the study may help refine the model much more. The given data set included the tail of the Dot Com crash which may have had a slight influence. Yet, because the rise and crash were included in the model, it is difficult to determine whether the coefficient is attributed to the rise of the index or the crash specifically.

**Appendix**

|  |  |
| --- | --- |
| TABLE 1 | |
| Market Indices & Companies Selected | |
| Index/Company | Symbol |
| Dow Jones Industrial Average | DJI |
| S&P 500 | GSPC |
| NASDAQ | IXIC |
| Bank of America | BAC |
| BB&T | BBT |
| Bank of New York Mellon | BK |
| Citigroup | C |
| Comerica | CMA |
| Capital One Financial | COF |
| Fifth Third Bancorp | FITB |
| Goldman Sachs | GS |
| Hunting Bancshares | HBAN |
| JP Morgan Chase | JPM |
| KeyCorp | KEY |
| Morgan Stanley | MS |
| M&T Bank | MTB |
| People’s United | PBCT |
| PNC Financial | PNC |
| SunTrust Banks | STI |
| State Street | STT |
| US Bancorp | USBPO |
| Wells Fargo & Co. | WFC |
| Zions Bancorp | ZION |

Code

**proc** **sgplot** data = Index; \*plot raw data of indices;

title 'Market Indices Historical Value (2002 - 2009)';

series x = Date y = DJI;

series x = Date y = GSPC;

series x = Date y = IXIC;

xaxis LABEL = 'Date';

yaxis LABEL = 'Value';

**run**;

**proc** **sgplot** data = Index; \*plot raw data of bank stocks;

title 'Historical Stock Price of Finance Sector in S&P500 (2002 - 2009)';

series x = Date y = BAC;

series x = Date y = BBT;

series x = Date y = BK;

series x = Date y = C;

series x = Date y = CMA;

series x = Date y = COF;

series x = Date y = FITB;

series x = Date y = GS;

series x = Date y = HBAN;

series x = Date y = JPM;

series x = Date y = KEY;

series x = Date y = MS;

series x = Date y = MTB;

series x = Date y = PBCT;

series x = Date y = PNC;

series x = Date y = STI;

series x = Date y = STT;

series x = Date y = WFC;

series x = Date y = ZION;

xaxis LABEL = 'Date';

yaxis LABEL = 'Price';

**run**;

**data** Index1; set Index; \*log data to address seasonality;

logDJI = log(DJI);

logGSPC = log(GSPC);

logIXIC = log(IXIC);

logBAC = log(BAC);

logBBT = log(BBT);

logBK = log(BK);

logC = log(C);

logCMA = log(CMA);

logCOF = log(COF);

logFITB = log(FITB);

logGS = log(GS);

logHBAN = log(HBAN);

logJPM = log(JPM);

logKEY = log(KEY);

logMS = log(MS);

logMTB = log(MTB);

logPBCT = log(PBCT);

logPNC = log(PNC);

logSTI = log(STI);

logSTT = log(STT);

logUSBPO = log(USB);

logWFC = log(WFC);

logZION = log(ZION);

**run**;

**proc** **sgplot** data = Index1; \*plot log index data;

series x = Date y = logDJI;

series x = Date y = logGSPC;

series x = Date y = logIXIC;

title 'Log of Market Indices Historical Value (2002 - 2009)';

xaxis LABEL = 'Date';

yaxis LABEL = 'Value';

**run**;

**proc** **sgplot** data = Index1; \*plot log stock price data;

series x = Date y = logBAC;

series x = Date y = logBBT;

series x = Date y = logBK;

series x = Date y = logC;

series x = Date y = logCMA;

series x = Date y = logCOF;

series x = Date y = logFITB;

series x = Date y = logGS;

series x = Date y = logHBAN;

series x = Date y = logJPM;

series x = Date y = logKEY;

series x = Date y = logMS;

series x = Date y = logMTB;

series x = Date y = logPBCT;

series x = Date y = logPNC;

series x = Date y = logSTI;

series x = Date y = logSTT;

series x = Date y = logUSBPO;

series x = Date y = logWFC;

series x = Date y = logZION;

title 'Log of Stock Price Historical Value (2002 - 2009)';

xaxis LABEL = 'Date';

yaxis LABEL = 'Price Log';

**run**;

\*Check assumptions without transformation;

ods graphics on;

**proc** **glm** data = Index plots = all;

model DJI = Date BAC BBT BK CMA COF FITB GS HBAN JPM KEY MS MTB PBCT PNC STI STT WFC ZION;

output out = resids r = residsOLS;

**run**;

**proc** **sgplot** data = resids;

scatter x = Date y = residsOLS;

title 'Residuals of Dow Jones model';

yaxis LABEL = 'Residuals';

**run**;

**proc** **glm** data = Index plots = all;

model GSPC = Date BAC BBT BK CMA COF FITB GS HBAN JPM KEY MS MTB PBCT PNC STI STT WFC ZION;

output out = resids r = residsOLS;

**run**;

**proc** **sgplot** data = resids;

scatter x = Date y = residsOLS;

title 'Residuals of S&P 500';

yaxis LABEL = 'Residuals';

**run**;

**proc** **glm** data = Index plots = all;

model IXIC = Date BAC BBT BK CMA COF FITB GS HBAN JPM KEY MS MTB PBCT PNC STI STT WFC ZION;

output out = resids r = residsOLS;

**run**;

**proc** **sgplot** data = resids;

scatter x = Date y = residsOLS;

title 'Residuals of Nasdaq';

yaxis LABEL = 'Residuals';

**run**;

ods graphics off;

\*check for assumptions after transformations;

ods graphics on;

**proc** **glm** data = Index1 plots = all;

model logDJI = logBAC logBBT logBK logCMA logCOF logFITB logGS logHBAN logJPM logKEY logMS logMTB logPBCT logPNC logSTI logSTT logWFC logZION;

output out = resids r = residsOLS;

**run**;

**proc** **reg** data = Index1;

model logDJI = Date logBAC logBBT logBK logCMA logCOF logFITB logGS logHBAN logJPM logKEY logMS logMTB logPBCT logPNC logSTI logSTT logWFC logZION / vif;

**run**;

**proc** **sgplot** data = resids;

scatter x = Date y = residsOLS;

title 'Residuals of Dow Jones model';

yaxis LABEL = 'Residuals';

**run**;

**proc** **glm** data = Index1 plots = all;

model logGSPC = Date logBAC logBBT logBK logCMA logCOF logFITB logGS logHBAN logJPM logKEY logMS logMTB logPBCT logPNC logSTI logSTT logWFC logZION;

output out = resids r = residsOLS;

**run**;

**proc** **reg** data = Index1;

model logGSPC = Date logBAC logBBT logBK logCMA logCOF logFITB logGS logHBAN logJPM logKEY logMS logMTB logPBCT logPNC logSTI logSTT logWFC logZION / vif;

**run**;

**proc** **sgplot** data = resids;

scatter x = Date y = residsOLS;

title 'Residuals of S&P 500';

yaxis LABEL = 'Residuals';

**run**;

**proc** **glm** data = Index1 plots = all;

model logIXIC = Date logBAC logBBT logBK logCMA logCOF logFITB logGS logHBAN logJPM logKEY logMS logMTB logPBCT logPNC logSTI logSTT logWFC logZION;

output out = resids r = residsOLS;

**run**;

**proc** **sgplot** data = resids;

scatter x = Date y = residsOLS;

title 'Residuals of Nasdaq';

yaxis LABEL = 'Residuals';

**run**;

ods graphics off;

\*Test for autocorrelation;

\*Calculate Durbin-Watson for all 3 indexes;

**proc** **reg** data = Index1;

title 'Dow Jones Durbin-Watson Test';

model logDJI = Date logBAC logBBT logBK logCMA logCOF logFITB logGS logHBAN logJPM logKEY logMS logMTB logPBCT logPNC logSTI logSTT logWFC logZION / dw;

ods output dwstatistic = auto\_corr

(where=(label1="1st Order AutoCorrelation - AR1"));

**run**;

**proc** **reg** data = Index1;

title 'S&P 500 Durbin-Watson Test';

model logGSPC = Date logBAC logBBT logBK logCMA logCOF logFITB logGS logHBAN logJPM logKEY logMS logMTB logPBCT logPNC logSTI logSTT logWFC logZION / dw;

ods output dwstatistic = auto\_corr

(where=(label1="1st Order AutoCorrelation - AR1"));

**run**;

**proc** **reg** data = Index1;

title 'Nasdaq Durbin-Watson Test';

model logIXIC = Date logBAC logBBT logBK logCMA logCOF logFITB logGS logHBAN logJPM logKEY logMS logMTB logPBCT logPNC logSTI logSTT logWFC logZION / dw;

ods output dwstatistic = auto\_corr

(where=(label1="1st Order AutoCorrelation - AR1"));

**run**;

ods graphics on;

**proc** **autoreg** data = Index1 plots = all; \*Dow Jones model;

title 'Dow Jones model';

model logDJI = Date logBAC logBBT logBK logCMA logCOF logFITB logGS logHBAN logJPM logKEY logMS logMTB logPBCT logPNC logSTI logSTT logWFC logZION / dwprob nlag = **1**;

output out = p p = yhat pm = ytrend lcl = lcl ucl = ucl;

**run**;

**proc** **sgplot** data = p; \*Dow Jones model plot;

where Date >= **'01jan2008'd**;

title 'Dow Jones Market Value (2008 - 2009)';

band x = Date upper = ucl lower = lcl;

series x = Date y = logDJI / lineattrs = (color = black);

series x = Date y = yhat / lineattrs = (color = red) transparency = **0.5**;

series x = Date y = ytrend / lineattrs = (color = blue pattern = dash) transparency = **0.5**;

xaxis LABEL = 'Value';

**run**;

**proc** **autoreg** data = Index1 plots = all; \*S&P 500 model;

title 'S&P 500 model';

model logGSPC = Date logBAC logBBT logBK logCMA logCOF logFITB logGS logHBAN logJPM logKEY logMS logMTB logPBCT logPNC logSTI logSTT logWFC logZION / dwprob nlag = **1**;

output out = p p = yhat pm = ytrend lcl = lcl ucl = ucl;

**run**;

**proc** **sgplot** data = p; \*S&P 500 model plot;

title 'S&P 500 Market Value (2008 - 2009)';

where Date >= **'01jan2008'd**;

band x = Date upper = ucl lower = lcl;

series x = Date y = logGSPC / lineattrs = (color = black);

series x = Date y = yhat / lineattrs = (color = red) transparency = **0.5**;

series x = Date y = ytrend / lineattrs = (color = blue pattern = dash) transparency = **0.5**;

xaxis LABEL = 'Value';

**run**;

**proc** **autoreg** data = Index1 plots = all; \*NASDAQ model;

title 'NASDAQ model';

model logIXIC = Date logBAC logBBT logBK logCMA logCOF logFITB logGS logHBAN logJPM logKEY logMS logMTB logPBCT logPNC logSTI logSTT logWFC logZION / dwprob nlag = **1**;

output out = p p = yhat pm = ytrend lcl = lcl ucl = ucl;

**run**;

**proc** **sgplot** data = p; \*NASDAQ model plot;

where Date >= **'01jan2008'd**;

title 'NASDAQ Market Value (2008 - 2009)';

band x = Date upper = ucl lower = lcl;

series x = Date y = logIXIC / lineattrs = (color = black);

series x = Date y = yhat / lineattrs = (color = red) transparency = **0.5**;

series x = Date y = ytrend / lineattrs = (color = blue pattern = dash) transparency = **0.5**;

xaxis LABEL = 'Value';

**run**;

ods graphics off;

**References**

1. N. Clowers, “Financial Crisis Losses and Potential Impacts of the Dodd-Frank Act,” United States Government Accountability Office, Washington, DC, GAO-13-180, Jan. 16, 2013.

1. [↑](#footnote-ref-1)