

# Factor-Based Prediction of Industry-Wide Bank Stress

*Sean Grover and Michael W. McCracken*

This article investigates the use of factor-based methods for predicting industry-wide bank stress. Specifically, using the variables detailed in the Federal Reserve Board of Governors' bank stress scenarios, the authors construct a small collection of distinct factors. We then investigate the predictive content of these factors for net charge-offs and net interest margins at the bank industry level. The authors find that the factors do have significant predictive content, both in and out of sample, for net interest margins but significantly less predictive content for net charge-offs. Overall, it seems reasonable to conclude that the variables used in the Fed's bank stress tests are useful for identifying stress at the industry-wide level. The final section offers a simple factor-based analysis of the counterfactual bank stress testing scenarios. (JEL C12, C32, C52, C53)

Federal Reserve Bank of St. Louis *Review*, Second Quarter 2014, 96(2), pp. 173-93.

---

**T**he Dodd-Frank Wall Street Reform and Consumer Protection Act requires a specified group of large U.S. bank holding companies to submit to an annual Comprehensive Capital Analysis and Review (CCAR) of their capital adequacy should a hypothetical “severely adverse” economic environment arise. Of course, such an analysis requires defining the phrase “severely adverse,” and one could imagine a variety of possible economic outcomes that would be considered bad for the banking industry. Moreover, what is stressful for one bank may not be stressful for another because of differences in their assets, liabilities, exposure to counterparty risk or particular markets, geography, and many other features associated with a specific bank.

The Dodd-Frank Act requires the Federal Reserve System to perform stress tests across a range of banks, so as a practical matter, it must choose a scenario that is likely to be severely adverse for all of the banks simultaneously. The current scenarios are provided in the “2014 Supervisory Scenarios for Annual Stress Tests Required under the Dodd-Frank Act Stress Testing Rules and the Capital Plan Rule” documentation.<sup>1</sup> The quarterly frequency series that characterize the “severely adverse” scenario include historical data back to 2001:Q1; but, more importantly, the series provide hypothetical outcomes moving forward from 2013:Q4 through

Sean Grover is a research associate and Michael W. McCracken is an assistant vice president and economist at the Federal Reserve Bank of St. Louis. The authors thank Bill Dupor and Dan Thornton for helpful comments.

© 2014, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

**Table 1****Data Used**

Variable name	Starting date	Source	Transformations
Charge-off rate on all loans, all commercial banks	1985:Q1	FRED	Percent
Net interest margin for all U.S. banks	1984:Q1	FRED	Percent
Real gross domestic product	1947:Q1	Haver Analytics	Annualized percent change
Gross domestic product	1947:Q1	Haver Analytics	Annualized percent change
Real disposable personal income	1959:Q1	Haver Analytics	Annualized percent change
Disposable personal income	1959:Q1	Haver Analytics	Annualized percent change
Unemployment rate, SA	1948:Q1	Haver Analytics	Percent
CPI, all items, SA	1977:Q1	Haver Analytics	Annualized percent change
3-Month Treasury yield, avg.	1981:Q3	Haver Analytics	Percent
5-Year Treasury yield, avg.	1953:Q2	Haver Analytics	Percent
10-Year Treasury yield, avg.	1953:Q2	Haver Analytics	Percent
Citigroup BBB-rated corporate bond yield index, EOP	1979:Q4	Haver Analytics	Percent
Conventional 30-year mortgage rate, avg.	1971:Q2	Haver Analytics	Percent
Bank prime loan rate, avg.	1948:Q1	Haver Analytics	Percent
Dow Jones U.S. total stock market index, avg.	1979:Q4	Haver Analytics	Annualized percent change
CoreLogic national house price index, NSA	1976:Q1	Haver Analytics	Annualized percent change
Commercial Real Estate Price Index, NSA	1945:Q4	Haver Analytics	Annualized percent change
Chicago Board Options Exchange market volatility index (VIX), avg.	1990:Q1	Haver Analytics	Levels

NOTE: Avg., average; CPI, consumer price index; EOP, end of period; NSA, non-seasonally adjusted; SA, seasonally adjusted.

2016:Q4 for a total of 13 quarters. There are 16 series  $X = (x_1, \dots, x_{16})'$  in the scenario. The real, nominal, monetary, and financial sides of the economy are all represented and include real gross domestic product (GDP) growth, consumer price index (CPI)-based inflation, various Treasury yields, and the Dow Jones Total Stock Market Index. Table 1 provides the complete list of variables. The paths of various international variables are also available in the documentation but are relevant for only a small subset of the banks undergoing stress testing. Hence, for brevity, they are omitted from the remainder of this discussion.

The path of the variables in this severely adverse scenario is safely described as indicative of those among the worst post-World War II recessionary conditions faced by the United States: The unemployment rate peaks at 11.25 percent in mid-2015, real GDP declines more than 4.5 percent in 2014, and housing prices decline roughly 25 percent. The other series are characterized by comparably dramatic shifts. Individually, each element of the scenario seems reasonable at an intuitive level. For example, recessions typically occur during periods when real GDP declines, and deeper recessions exhibit larger declines in real GDP. The given scenario is characterized by a large decline in real GDP; hence, the scenario could reasonably be described as a severe recession.

And yet, we are left with the deeper question of whether the severe recession is severely adverse for the banks. In this article, we attempt to address this question and provide indexes designed to measure and predict the scenario's degree of severity. The first step in our analysis requires choosing a measure of economic "badness" faced by a bank. We consider two measures: net charge-offs (NCOs) and net interest margins (NIMs). The NCO measure is the percent of debt that the bank believes it will be unlikely to collect. Banks write off poor-credit-quality loans as bad debts. Some of these debts are recovered, and the difference between the gross charge-offs and recoveries is NCOs. In a weakened economy, as loan defaults increase, NCOs would be expected to rise. The NIM measure is the difference in the rate of interest income paid to, say, depositors relative to the interest income earned from outstanding loans (in terms of assets). NIMs are a key driver of revenue in the traditional banking sector and, hence, movements in NIMs give an indication of the bank's overall health. Rather than attempt an analysis of individual banks, we instead use the charge-offs and interest margins aggregated across all commercial banks.

With these target variables ( $y$ ) in hand, our goal is to construct and compare a few macroeconomic factors (rather than financial or banking industry-specific factors) that can be used to track the strength of the banking industry. In particular, the goal is to obtain macroeconomic factors that are useful as predictors of the future state of the banking sector. We consider three distinct approaches to extracting factors ( $f$ ) from the predictors ( $X$ ) with an eye toward predicting the target variable ( $y$ ): principal components (PC), partial least squares (PLS), and the three-pass regression filter (3PRF) developed in Kelly and Pruitt (2011). We discuss each of these methods in turn in the following section.

We believe our results suggest that factor-based methods are useful as predictors of future bank stress. We reach these conclusions based on two major forms of evidence. First, based on pseudo out-of-sample forecasting exercises, it appears that factors extracted from these series in the CCAR scenarios can be fruitfully used to predict bank stress as measured using NIMs and, to a lesser extent, NCOs. In addition, when applied to the counterfactual scenarios, the factor models imply forecast paths that match our intuition on the degree of severity of the scenarios: The severely adverse scenarios tend to imply higher NCOs than do the adverse scenarios, which in turn tend to imply higher NCOs than do the baseline scenarios.

Not surprisingly given the increased attention to bank stress testing, a new literature strand focused on methodological best practices is quickly developing. Covas, Rump, and Zakrajsek (2013) use panel quantile methods to predict bank profitability and capital shortfalls. Acharya, Engle, and Pierret (2013) construct market-based metrics of bank-specific capital shortfalls. Our article is perhaps most closely related to those of Bolotnyy, Edge, and Guerrieri (2013) and Guerrieri and Welch (2012). Bolotnyy, Edge, and Guerrieri consider a variety of approaches to forecasting NIMs including factor-based methods, but they do so using level, slope, and curvature factors extracted from the yield curve rather than the macroeconomic aggregates specific to the stress testing scenarios. Guerrieri and Welch consider a variety of approaches to forecasting NIMs and NCOs using macroeconomic aggregates similar to those included in the scenarios, but they emphasize the role of model averaging across many bivariate vector autoregressions (VARs) rather than factor-based methods per se.

The next section offers a brief overview of the data used and estimation of the factors. We evaluate the predictive content of the indexes for NCOs and NIMs both in and out of sample. We then use the estimated factor weights along with the series provided in the hypothetical stress scenarios to construct implied counterfactual factors. The counterfactual factors provide a simple-to-interpret measure of the degree of stressfulness implied by the scenario.

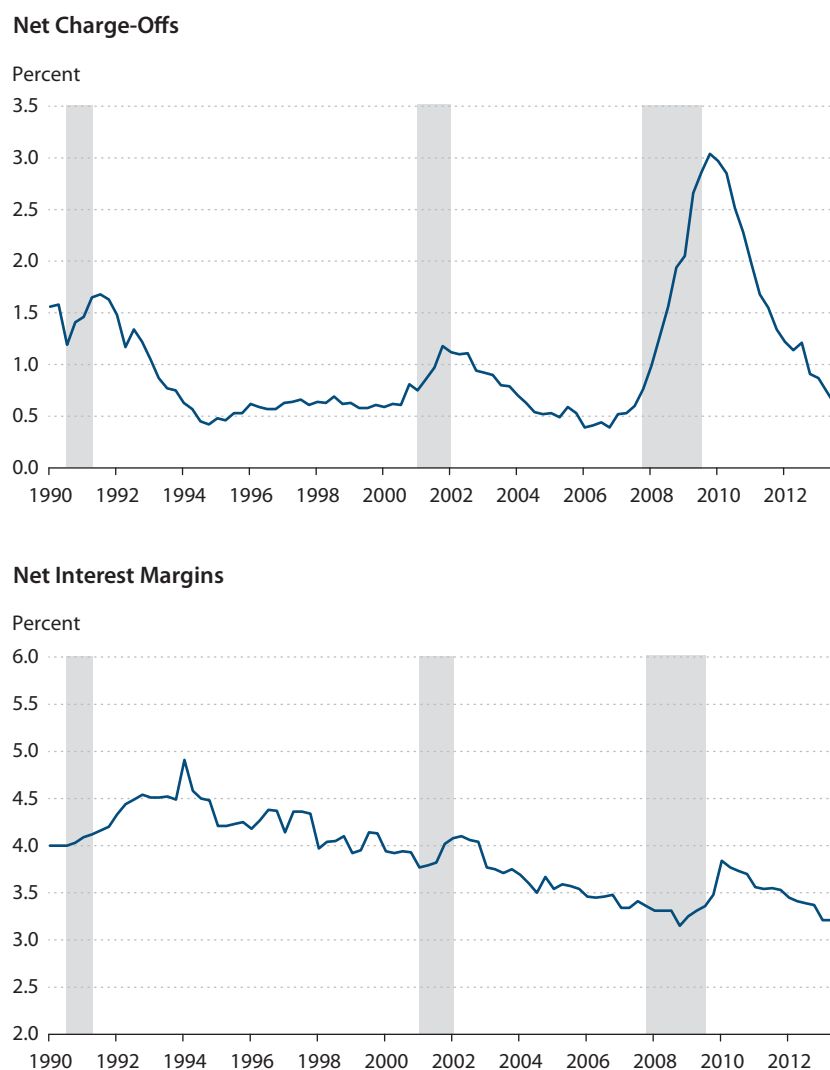
## DATA AND METHODS

As we just noted, our goal is to develop simple-to-interpret factors that can be used to predict industry-wide bank stress when measured through the lens of either NCOs or NIMs. In much of the literature on factor-based forecasting, these factors are often extracted based on scores, if not hundreds, of macroeconomic or financial series. For example, Stock and Watson (2002) extract factors using two datasets, one consisting of 149 series and another consisting of 215 series. Ludvigson and Ng (2009) use a related dataset consisting of 132 series. Giannone, Reichlin, and Small (2008) use roughly 200 series to extract factors and nowcast current-quarter real GDP growth. Smaller datasets have also been used. Ciccarelli and Mojon (2010) extract a common global inflation factor using inflation series from only 22 countries. Engel, Mark, and West (2012) extract a common U.S. exchange rate factor using bilateral exchange rates between the United States and 17 member countries of the Organisation for Economic Co-operation and Development.

In this article, we focus on factors extracted from a collection of the 16 quarterly frequency variables ( $X$ ) currently used by the Federal Reserve System to characterize the bank stress testing scenarios. We do so to facilitate a clean interpretation of our results as they pertain to bank stress testing conducted by the Federal Reserve System. See Table 1 for a complete list of these variables. The historical data for these series are from the Federal Reserve Economic Data (FRED) and Haver Analytics databases and consist of only the most recent vintages. The NCOs and NIMs data are available back to 1985:Q1 and 1984:Q1, respectively. Unfortunately, the Chicago Board Options Exchange Volatility Index (VIX) series used in the stress scenarios dates back to only 1990:Q1; hence, our analysis is based on observations from 1990:Q1 through 2013:Q3, providing a total of  $T = 95$  historical observations. We considered substituting a proxy for the VIX to obtain a longer time series. Instead, we chose to restrict ourselves to the time series directly associated with the CCAR scenarios since these are the ones banks must use as part of their stress testing efforts.

Figure 1 plots both NCOs and NIMs. NCOs rose during the 1991, 2001, and 2007-09 recessions, with dramatic increases during the most recent recession. NIMs reacted less dramatically during these recessions and, more than anything, seemed to trend downward for much of the sample before rising during the recent recovery. The plots suggest that both target variables are highly persistent. Since both Dickey-Fuller (1979) and Elliott, Rothenberg, and Stock (1996) unit root tests fail to reject the null of a unit root in both series, we model the target variable in differences rather than in levels. Specifically, when the forecast horizon  $h$  is 1 quarter ahead, the dependent variable is the first difference  $\Delta^1 = \Delta$ , but when the horizon is four, we treat the dependent variable as the 4-quarter-ahead difference  $\Delta^4$ . We should note,

**Figure 1**  
**History of Target Variables**



NOTE: The shaded bars indicate recessions as determined by the National Bureau of Economic Research.

however, that while we consider the  $h = 4$ -quarter-ahead horizon, the majority of our results emphasize predictability at the  $h = 1$ -quarter-ahead horizon. As to the predictors  $x_i$   $i = 1, \dots, 16$ , in most instances they are used in levels, but where necessary the data are transformed to induce stationarity. The relevant transformations are provided in Table 1.

We consider three distinct, but related, approaches to estimating the factors. By far, the PC approach is the most common in the literature. Stock and Watson (1999, 2002) show that PC-based factors ( $f^{PC}$ ) are useful for forecasting macroeconomic variables, including nominal

variables such as inflation, as well as real variables such as industrial production. See Stock and Watson (2011) for an overview of factor-based forecasting using PC and closely related methods. Theoretical results for factor-based forecasting can also be found in Bai and Ng (2006).

Although their use is standard, PC-based factors are not by construction useful for forecasting a given target variable  $\Delta^h y$ . To understand the issue, consider the simple  $h$ -step-ahead predictive model of the factors and target variables:

$$(1) \quad \Delta^h y_{t+h} = \beta_0 + \beta_1 f_t + \varepsilon_{t+h}$$

$$(2) \quad x_{i,t} = \alpha_{i0} + \alpha_{i1} f_t + v_{i,t} \quad i = 1, \dots, N,$$

where  $N$  denotes the number of  $x$  series used to estimate the factor.

In equation (1), the factor is modeled as potentially useful for predicting the target  $h$ -steps into the future. It will be useful for forecasting if  $\beta_1$  is nonzero and it is estimated sufficiently precise. Equation (2) simply states that each of the predictors  $x_i$  can be modeled as being determined by a single common factor  $f$ . When PC are used to estimate the factor, any link between equations (1) and (2) is completely ignored. Instead, the PC-based factors are constructed as follows: Let  $\tilde{X}_t$  denote the  $t \times N$  matrix  $(\tilde{x}_{t1}, \dots, \tilde{x}_{tN})'$  of  $N \times 1$  time- $t$  variance-standardized predictors  $\tilde{x}_{st}$   $s = 1, \dots, N$ . If we define  $\bar{x}_t$  as  $t^{-1} \sum_{s=1}^t \tilde{x}_{st}$ , the time- $s$  factor  $f_s^{PC}$  is estimated as  $\hat{f}_{s,t}^{PC} = \hat{\Lambda}'_t (\tilde{x}_{st} - \bar{x}_t)$ , where  $\hat{\Lambda}_t$  denotes the eigenvector associated with the largest eigenvalue of  $(\tilde{X}_t - \bar{X}_t)'(\tilde{X}_t - \bar{X}_t)$ . Clearly the target variable  $\Delta^h y$  plays no role whatsoever in construction of the factors, and specifically in the choice of  $\hat{\Lambda}_t$ , and hence there is no a priori reason to believe that  $\beta_1$  will be nonzero.

It might therefore be surprising to find that PC-based factors are often useful for forecasting. Note, however, that the  $x$  variables used to estimate  $f^{PC}$  are never chosen randomly from a collection of all possible time series available. For example, the Stock and Watson dataset (2002) was compiled by the authors based on their years of experience in forecasting macroeconomic variables. As such, there undoubtedly was a bias toward including those variables in the dataset that had already shown some usefulness for forecasting. PC-based forecasting is therefore a reasonable choice provided the collection of predictors  $x$  is selected in a reasonable fashion.

This is a well-known issue in the factor literature; hence, methods—most notably PLS—have been developed to integrate knowledge of the target variable  $y$  into the estimation of  $\Lambda$ . One recently developed generalization of PLS is delineated in Kelly and Pruitt (2011). Their 3PRF estimates the factor weights  $\Lambda$ , taking advantage of the covariation between the predictors  $x$  and the target variable  $\Delta^h y$ , but does so while allowing for ancillary information  $z$  to be introduced into the estimation of the factors.<sup>2</sup> Effectively, this ancillary information introduces an additional equation of the form

$$(3) \quad z_t = \gamma_0 + \gamma_1 f_t + v_t$$

into the system above. Without going into the derivation here, if we define the vector  $Z_t = (z_1, \dots, z_t)'$ , we obtain a closed-form solution for the factor loadings that takes the form

$$(4) \quad \hat{\Lambda}'_t = \hat{S}_{ZZ,t} (\hat{W}'_{XZ,t} \hat{S}_{XZ,t})^{-1} \hat{W}'_{XZ,t},$$

where  $J_t = I_t - \iota_t \iota_t'$  for  $I_t$ , the  $t$ -dimensional identity matrix  $\iota_t$  is the  $t$ -vector of ones,  $\hat{W}_{XZ,t} = J_N \tilde{X}'_t J_t Z_t$ ,  $\hat{S}_{XZ,t} = \tilde{X}'_t J_t Z_t$ , and  $\hat{S}_{ZZ,t} = Z_t' J_t Z_t$ . At first glance, it may appear that in this closed-form solution the estimated loadings do not account for the target variable; but as Kelly and Pruitt (2011) note, one can always choose to set  $z$  equal to the target variable, in which case the factor weights obviously do depend on the target variable. In fact, loosely speaking, if we set  $z$  equal to  $\Delta^h y$ , we obtain PLS as a special case. Kelly and Pruitt (2011) argue that there are applications in which using proxies instead of the target variable, perhaps driven by economic reasons, can lead to more-accurate predictions of the target variable. Drawing from this intuition and the results in Guerrieri and Welch (2012), in which the term spread is found to be a useful predictor of NCOs and NIMs, we use the first difference in the spread between a 10-year U.S. bond and a 3-month T-bill as our proxy when applying the 3PRF.

For each target variable we then have three sets of factors,  $\hat{f}_{s,t}^{PC}$ ,  $\hat{f}_{s,t}^{3PRF}$ , and  $\hat{f}_{s,t}^{PLS}$   $s = 1, \dots, t$ , that we can use to conduct pseudo out-of-sample forecasting exercises. Note that the PC and 3PRF factors are invariant to the target variable (and horizon) and, hence, we really have only four distinct factors. In the forecasting experiments we reestimate the factor loadings at each forecast origin. For example, if we have a full set of observations  $s = 1, \dots, T$  and forecast  $h$ -steps ahead starting with an initial forecast origin  $t = R$ , we obtain  $P = T + h - R$  forecasts, which in turn can be used to construct forecast errors  $\hat{\epsilon}_{t+h}$  and corresponding mean squared errors (MSEs) as  $MSEs = P^{-1} \sum_{t=R}^{T-h} \hat{\epsilon}_{t+h}^2$ .<sup>3</sup>

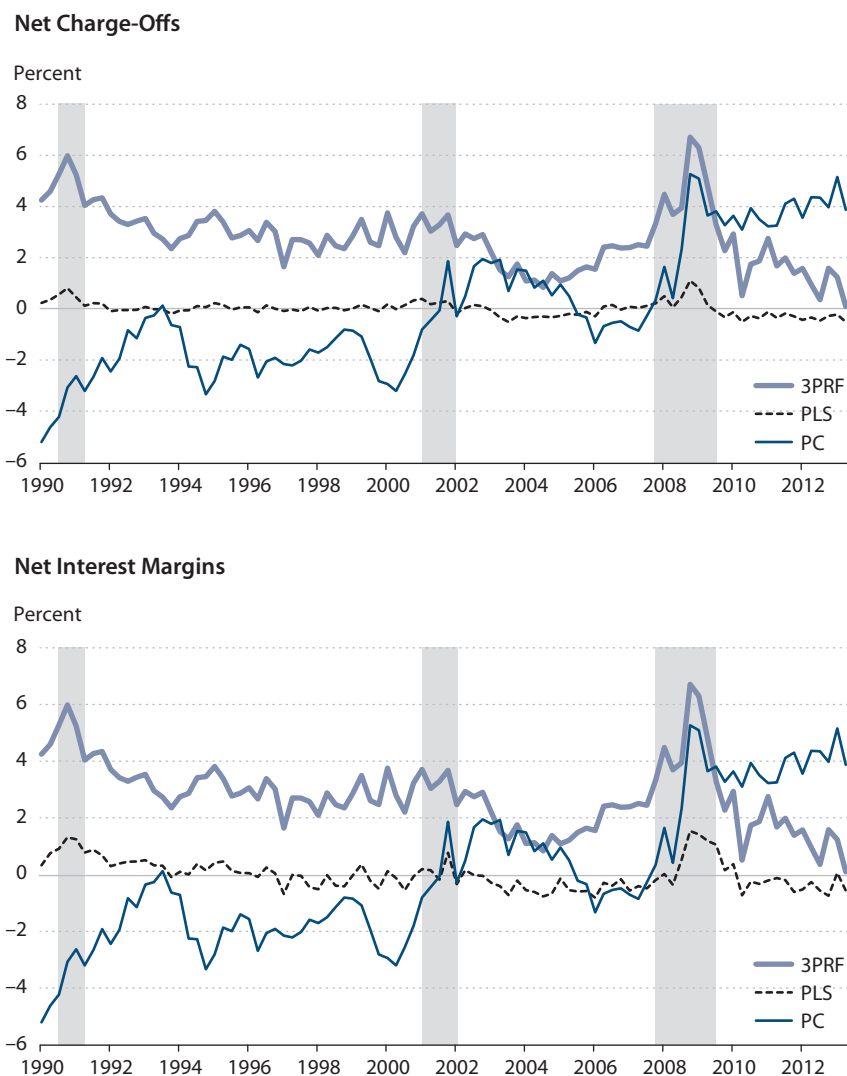
## RESULTS

In the following, we provide a description of the factors and their usefulness for forecasting bank stress at the industry level.

### The Factors

Figure 2 plots the three factors associated with the 1-step-ahead models estimated using the full sample  $s = 1, \dots, T$ . The top panel plots the factors when NCOs are the target variable, while the lower panel plots the factors when NIMs are the target variable. For completeness we include the PC and 3PRF factors in both panels despite the fact they are identical in each panel. In both panels, the PLS and 3PRF factors appear stationary. Both sets of factors have peaks during both the 1991 and 2007-09 recessions, but there do not appear to be any such peaks associated with the 2001 recession. In contrast, the PC-based factors are quite interesting. Whereas the PLS and 3PRF factors appear stationary, the PC factor appears to be trending upward. In fact, one could argue that the PC factor rises sharply at the onset of each recession and then flattens but never reverts to its pre-recession level. Since stationarity of the factors is a primary assumption in much of the literature on PC-based forecasting, one might suspect

**Figure 2**  
**Estimated Historical Factors**



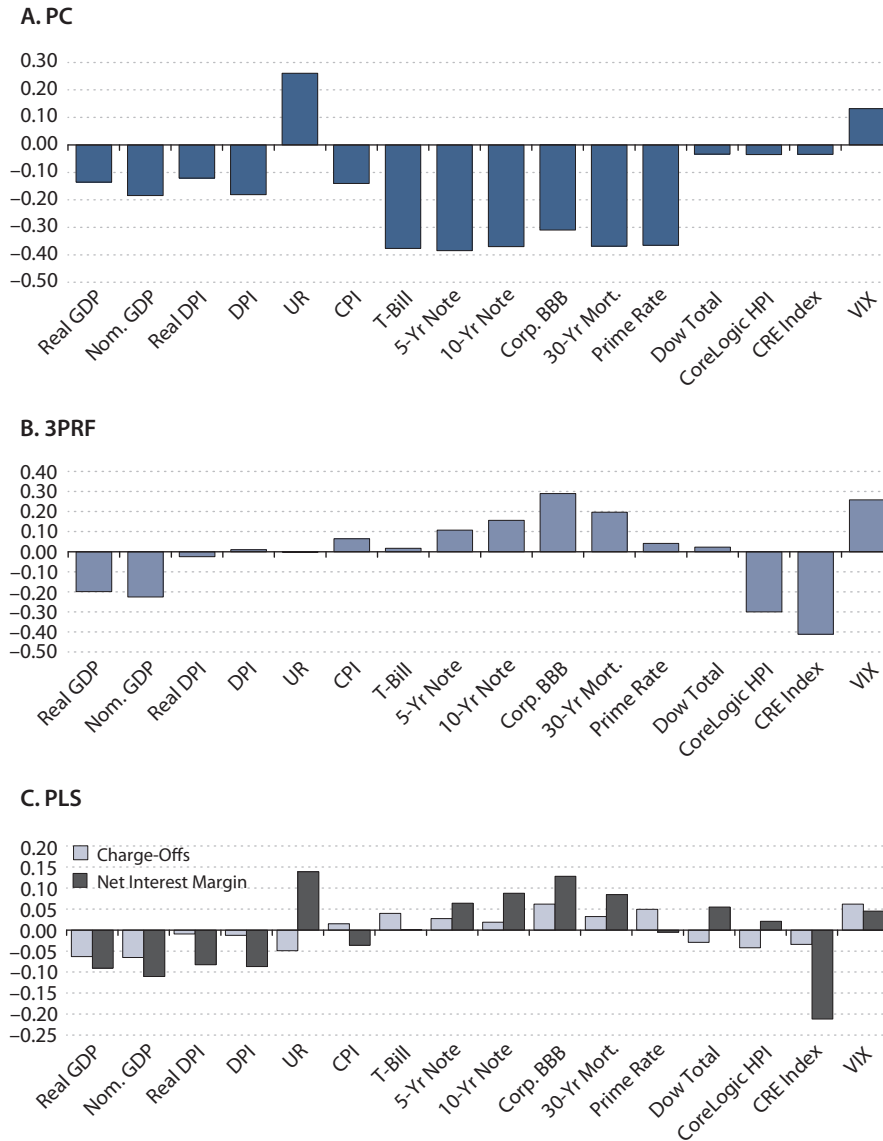
NOTE: The shaded bars indicate recessions as determined by the National Bureau of Economic Research.

that our PC-based factors will not prove useful for prediction. We return to this issue in the following section.

Figure 3 shows the factor weights  $\hat{\Lambda}$  associated with the 1-step-ahead models estimated over the entire sample. There are three panels, one each for the three approaches to estimating the factors. Panel A depicts the PC weights. Since these weights are constructed on data that have been standardized, the magnitudes of the weights are comparable across variables. In addition, since the variables have all been demeaned, negative values of the weights indi-



**Figure 3**  
**Factor Weights**



NOTE: BBB, Citigroup BBB-rated corporate bond; CRE, commercial real estate; DPI, disposable personal income; HPI, house price index; Mort., mortgage; Nom., nominal; UR, unemployment rate.

cate that if a variable is above its historical mean, all else equal, the factor will be smaller. With these caveats in mind, in most cases the weights are sizable, with values greater than 0.1 in absolute value. The exceptions are all financial indicators and include the Dow, the CoreLogic house price index, and the Commercial Real Estate price index. The largest weights are on the interest rate variables and are negative, sharing the same sign as that associated with real and nominal GDP as might be expected. In addition, the weights given to the unemployment rate and the VIX are positive, so they have the opposite sign of both nominal and real GDP as also might be expected.

Panel B of Figure 3 depicts the weights associated with the 3PRF-constructed factor. The weights are distinct from those associated with the PC-constructed factor, even though both sets of weights are fundamentally based on the same predictors. In particular, the weights in the financial and monetary variables are different. PC assign large negative weights on interest rates, while the 3PRF places positive weights on them. In addition, PC assign almost no weight on the Dow or either of the two price indexes, while the 3PRF assigns large negative weights on both the residential and commercial real estate price indexes. Both the PC and 3PRF assign significant, and comparably signed, weights on real and nominal GDP, as well as the VIX.

Panel C of Figure 3 has two sets of weights associated with PLS, one for each target variable. What is perhaps more interesting is how the magnitudes and signs of the weights change when the target variables are changed. As an example, note that the weight on the unemployment rate is large and positive when NIMs are the target but is modest and negative when NCOs are the target. And while PLS places a negative weight on the commercial real estate price index for both target variables, the weight is an order of magnitude larger for NIMs.

In some ways, however, the weights are similar across each of the three panels. All four sets of weights are negative for real and nominal GDP and all four are positive for the VIX. Perhaps the sharpest distinction among the three panels is the weight that PC place on the interest rate variables. Both PLS and the 3PRF place small or modestly positive weights on the interest variables, while PC place large negative weights on them.

### ***Predictive Content of the Factors***

We consider the predictive content of these factors based on both in-sample and out-of-sample evidence. Panel A of Table 2 shows the intercepts, slope coefficients, and  $R$ -squared values associated with the predictive regressions in equation (1) for each factor, target variable, and horizon estimated over the full sample. At each horizon and for each target variable the  $R$ -squared value associated with PC-based factors is smallest among the other two competitors. This is perhaps not surprising given the apparent nonstationarity of the PC-estimated factors as seen in Figure 2. When the 3PRF- and PLS-based factors are used, the  $R$ -squared values are typically neither large nor nontrivial. At the 1-quarter-ahead horizon, the  $R$ -squared values are roughly 0.20 and 0.30 for NCOs using 3PRF- and PLS-based factors, respectively. In all instances, the PLS-based factors are the largest among the three estimation methods. Even so, the PLS factors are constructed to maximize the correlation between the factor and the target variable; hence, it is not surprising that the  $R$ -squared values are larger for PLS. At the 1-quarter-ahead horizon, the  $R$ -squared values are much higher, particularly for PLS. Even so, the same

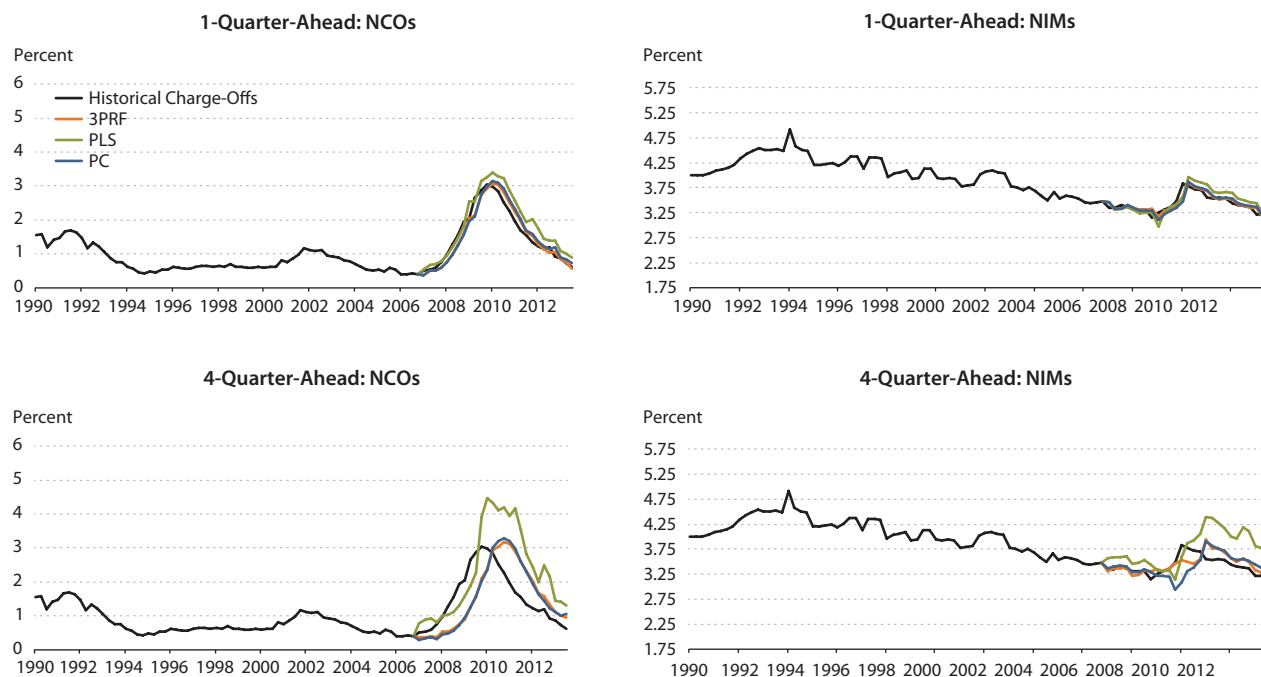
**Table 2**  
**In-Sample and Out-of-Sample Evidence on Predictability**

	1-Quarter-ahead horizon						4-Quarter-ahead horizon					
	NCO			NIM			NCO			NIM		
	3PRF	PLS	PC	3PRF	PLS	PC	3PRF	PLS	PC	3PRF	PLS	PC
<b>Panel A</b>												
Intercept	-0.165	-0.009	-0.009	-0.047	-0.008	-0.008	-0.442	-0.022	-0.022	-0.343	-0.032	-0.032
Slope	0.006	0.286	-0.004	0.014	0.049	0	0.070	0.430	-0.031	0.052	0.325	-0.001
R <sup>2</sup>	0.192	0.287	0.003	0.020	0.049	0	0.101	0.431	0.027	0.277	0.333	0
<b>Panel B</b>												
RMSE, naive	0.190			0.135			0.819			0.325		
Ratio	1.096	1.988	1.277	0.723	1.027	0.778	0.966	1.420	0.994	0.490	1.430	0.782
<b>Panel C</b>												
PLS (MSE- <i>t</i> )	-3.120*			-2.139*			-1.407			-2.278*		
PC (MSE- <i>t</i> )	-3.365*	2.830*		-1.863*	2.025*		-1.366	1.368		-1.305	1.385	
Naive (MSE- <i>F</i> )	-4.54	-20.17	-10.45	24.65*	-1.39	17.66*	1.96	-13.62	0.32	85.37*	-13.79	17.20*

NOTE: Panel A reports OLS regression output associated with equation (1) in the text. Panel B reports RMSE for the naive, no-change model, as well as the ratio of the RMSEs from the naive and factor-based model. Numbers below 1 indicate nominally superior predictive content for the factor model. Panel C reports MSE-*t* statistics for pairwise comparisons of the factor models and MSE-*F* statistics for comparisons between the naive and factor-based models. \* Indicates significance at the 5 percent level.

Figure 4

## Factor-Based Forecasts of Target Variables



caveat applies and, hence, it is not clear how useful the estimated factors will be for actual predictions based solely on the in-sample  $R$ -squared values.

We perform a pseudo out-of-sample forecasting exercise to get a better grasp of the predictive content of these factors. Specifically, because we are most interested in predictive content during periods of bank stress, we perform a pseudo out-of-sample forecasting exercise that starts in 2006:Q4 and ends in 2013:Q3 so the forecast sample includes the Great Recession and the subsequent recovery. Figure 4 plots the path of the forecasts for each target variable, horizon, and factor estimation method. It appears that at the 1-quarter-ahead horizon the factor-based forecasts typically track the arc of the realized values of both NCOs and NIMs. Perhaps the biggest exceptions are those associated with PLS, which tend to be much higher for each target variable throughout the entire forecast period. At the 4-quarter-ahead horizon, the predictions generally track the arc of the realized values but do so with a substantial lag. As was the case for the 1-step-ahead forecasts, those associated with PLS tend to be much higher than either the PC- or 3PRF-based forecasts.

Visually, the factor-based predictions seem somewhat reasonable, but it is not clear which of the three factor-based methods performs the best. Moreover, it is even less clear whether any of the models performs better than a naive, “no-change” benchmark would. Therefore, in Panel B of Table 2 we report root mean squared errors (RMSEs) associated with forecasts made

by the various factor-based models, as well as those associated with the naive benchmark. The first row reports the nominal RMSEs, while the second row reports ratios relative to the benchmark.

Particularly for NCOs, the naive benchmark tends to be a good predictor at both horizons and for both target variables. In all instances, the RMSE ratios are greater than or nearly indistinguishable from 1 and, hence, the naive model outpredicts the factor models. Predictions improve when NIMs are the target variable. At both horizons, 3PRF- and PC-based predictions are better than the naive benchmark. In some cases, the improvements are as large as 50 percent.

To better determine the magnitudes of the improvements and which factor estimation method performs best, we conduct tests of equal MSEs across each of the possible pairwise comparisons, holding the horizon and target variable constant. Specifically, we first construct the statistic

$$(5) \quad MSE-t = \frac{P^{-1/2} \sum_{t=R}^{T-h} \hat{d}_{ij,t+h}}{\hat{\sigma}_{i,j}}$$

for each of the pairwise comparisons across factor-type  $i \neq j \in \{3PRF, PLS, PC\}$ , where  $\hat{d}_{ij,t+h} = \hat{\varepsilon}_{i,t+h}^2 - \hat{\varepsilon}_{j,t+h}^2$ . At the 1-quarter-ahead horizon the denominator is constructed using  $\hat{\sigma}_{i,j} = \left( P^{-1} \sum_{t=R}^{T-h} (\hat{d}_{ij,t+h} - \bar{d}_{i,j})^2 \right)^{1/2}$ , while at the 4-quarter-ahead horizon we use a Newey-West (1987) autocorrelation-consistent estimate of the standard error allowing for three lags. We use these statistics to test the null of equal population-level forecast accuracy  $H_0 : E(\varepsilon_{i,t+h}^2 - \varepsilon_{j,t+h}^2) = 0$   $t = R, \dots, T - h$ . In each instance, we consider the two-sided alternative  $H_A : E(\varepsilon_{i,t+h}^2 - \varepsilon_{j,t+h}^2) \neq 0$ .

For the comparisons between the naive, no-change model and the factor-based models, we use the MSE- $F$  statistic developed in Clark and McCracken (2001, 2005),

$$(6) \quad MSE-F = \frac{\sum_{t=R}^{T-h} \hat{d}_{ij,t+h}}{P^{-1} \sum_{t=R}^{T-h} \hat{\varepsilon}_{j,t+h}^2},$$

for  $i = Naive$  and  $j \in \{3PRF, PLS, PC\}$ . This statistic is explicitly designed to capture the fact that the two models are nested under the null of equal population-level predictive ability. For the comparisons between the naive and the factor-based models, we consider the one-sided alternative  $H_A : E(\varepsilon_{naive,t+h}^2 - \varepsilon_{j,t+h}^2) > 0$  and hence reject when the MSE- $F$  statistic is sufficiently large.

For each of the pairwise factor-based comparisons, we treat the MSE- $t$  statistic as asymptotically standard normal and hence conduct inference using the relevant critical values. As a technical matter, the appropriateness of standard normal critical values has not been proven in the literature. Neither the results in Diebold and Mariano (1995) nor those in West (1996) are directly applicable to situations where generated regressors are used for prediction. Even so, we follow the literature and use standard normal critical values when the comparison is between two non-nested models estimated by OLS and evaluated under quadratic loss.

Unfortunately, this approach to inference does not extend to the MSE- $F$  statistic used for the nested model comparisons. Instead, we consider a fixed regressor wild bootstrap as described in Clark and McCracken (2012). The 1-step-ahead bootstrap algorithm proceeds as follows:

- (i) Generate artificial samples of  $\Delta y_t = y_t - y_{t-1}$  using the wild-form  $\Delta y_t^* = \varepsilon_t^* = \varepsilon_t \eta_t = (y_t - y_{t-1}) \eta_t$  for a simulated i.i.d. sequence of standard normal deviates  $\eta_t$ ,  $t = 1, \dots, T$ .
- (ii) Conduct the pseudo out-of-sample forecasting exercise precisely as was done in Table 2 but use  $\varepsilon_t^*$  as the dependent variable rather than  $y_t - y_{t-1} = \varepsilon_t$ . Construct the associated MSE- $F$  statistic.
- (iii) Repeat steps (i) and (ii)  $J = 499$  times.
- (iv) Estimate the  $(100 - \alpha)$  percentile of the empirical distribution of the bootstrapped MSE- $F$  statistics and use it to conduct inference.

For the 4-quarter-ahead case, the bootstrap differs only in how the dependent variable is simulated. Note that if  $y_t = y_{t-1} + \varepsilon_t$ , then  $\Delta^4 y = y_t - y_{t-4} = (y_t - y_{t-1}) + (y_{t-1} - y_{t-2}) + (y_{t-2} - y_{t-3}) + (y_{t-3} - y_{t-4}) = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \varepsilon_{t-3}$ . If we define the dependent variable as  $\Delta^4 y_t^* = \varepsilon_t^* + \varepsilon_{t-1}^* + \varepsilon_{t-2}^* + \varepsilon_{t-3}^* = \varepsilon_t \eta_t + \varepsilon_{t-1} \eta_{t-1} + \varepsilon_{t-2} \eta_{t-2} + \varepsilon_{t-3} \eta_{t-3}$  in step (i) of the algorithm, the remainder of the bootstrap proceeds as stated. Note, however, as was the case above, this bootstrap was developed to work in environments where the predictors were observed and not generated. Even so, we apply the bootstrap as designed in order to at least attempt to use critical values designed to accommodate a comparison between nested models.

The lower panel of Table 2 lists the results. Tests significant at the 5 percent level are indicated by an asterisk. First, consider the nested comparisons between the naive benchmark with each of the factor-based models. When NCOs are the target variable, we never reject the null of equal accuracy in favor of the alternative, suggesting that the factor-based models improve forecast accuracy. In contrast, when NIMs are the target variable, we find significant evidence that PC- and 3PRF-based forecasts improve on the naive model at both the 1- and 4-quarter-ahead horizons.

Now consider the pairwise comparisons across factor-based models. Here there is abundant evidence suggesting that the PLS-based approach to forecasting provides significantly less-accurate forecasts than those from PC- and 3PRF-based approaches. At each horizon for NIMs and at the 1-quarter-ahead horizon for NCOs, each of the MSE- $t$  statistics is significantly different from zero for any comparison with PLS. Somewhat surprisingly, at the 4-quarter-ahead horizon and for NCOs, we fail to reject the null of equal accuracy in each comparison with PLS despite the fact that the PLS-based forecasts have much higher nominal MSEs.

Overall, a few conclusions on the predictive content of the factors seem reasonable. First, and perhaps most obvious, the PLS-based factors do not predict well at any horizon and for either target variable. In each instance, at least one of the PC- or 3PRF-based models forecasts much more accurately in an MSE sense. This is despite the fact that in sample, the PLS approach had the highest  $R$ -squared value, even though, as already stated, this is by construction and not particularly informative. Second, at the 1-quarter horizon, the 3PRF-based factors perform

better than each of the competing factor-based forecast models. This is particularly true for NIMs. Even so, for NCOs this is small consolation given that the naive model does even better. Finally, at the longer horizons, the 3PRF-based factors are nominally the best but even so, the MSE- $t$  statistics indicate that they are not significantly smaller than those associated with the PC-based factors for either target variable.

## IMPLIED STRESS IN THE 2014 CCAR SCENARIOS

In the previous section, we established that the factors had varying degrees of predictive content for their corresponding target variable. In this section, we assess the degree to which the factors indicate stress in the context of the 2014 CCAR scenarios. That is, we attempt to measure the degree of stress faced by the banking industry in each of the baseline, adverse, and severely adverse scenarios when viewed through the lens of the implied counterfactual factors. We then conduct conditional forecasting exercises akin to those required by the Dodd-Frank Act. That is, we construct forecasts of NCOs and NIMs conditional on the counterfactual factors implied by the various scenarios and types of factor estimation methods.

### *The Counterfactual Factors*

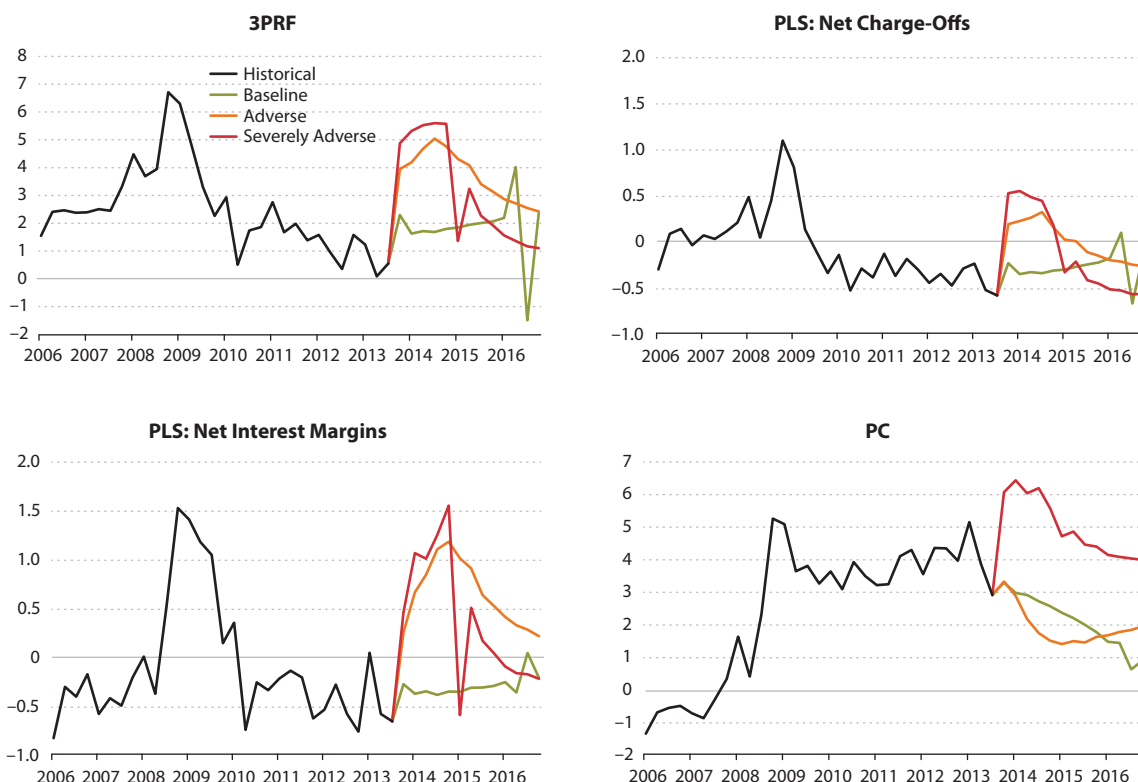
Figure 5 plots the factors associated with the 1-quarter-ahead horizon when the counterfactual scenarios are used to estimate the factors. To do so, we first estimate the four sets of factor weights through 2013:Q3, yielding  $\hat{\Lambda}_{2013:Q3}$ . Where necessary, we then demean and standardize the hypothetical variables in each of the baseline, adverse, and severely adverse scenarios using the historical sample means and standard deviations. Together these allow us to estimate the counterfactual factors  $\hat{f}_{t,2013:Q3}$  for all  $t = 2013:Q4, \dots, 2016:Q3$ .

There are four panels in Figure 5, with one each for the 3PRF- and PC-based factors. There are two panels for PLS, one for each target variable. The black lines in the panels show the factors estimated over the historical sample and end in 2013:Q3. The remaining three lines correspond to a distinct scenario starting in 2013:Q4 and ending in 2016:Q4.

The three panels associated with the PLS and 3PRF-based factors show comparable results for the degree of severity of the scenarios. In each case, the hypothetical severely adverse factors reach the highest levels and remain elevated for at least the first full year of the scenario. Somewhat surprisingly, in each of these three cases the factors associated with the adverse scenario tend to be only modestly lower than those associated with the severely adverse scenario over the first year. In fact, the factors associated with the adverse scenario remain elevated throughout much of the horizon even after those associated with the severely adverse scenario have declined to levels more closely associated with the baseline levels, which appear to remain near the historical mean of (the first difference of) NCOs and NIMs. Finally, it is instructive to note that the levels of the factors in the severely adverse scenario reach maximum values near those observed during the Great Recession.

The final panel in Figure 5 shows the results for the PC-based factors. As before, the PC-based factors associated with the severely adverse scenario reach the highest levels of all three

**Figure 5**  
**Counterfactual Factors Based on Scenarios**



scenarios. In contrast to the other three scenarios, here the hypothetical factors in the severely adverse scenario are always the largest of the three. Moreover, the levels are the same, if not a bit higher, than those observed during the Great Recession. Somewhat oddly, the factor associated with the adverse scenario is not particularly adverse and, in fact, often takes values lower than the factor associated with the baseline scenario.

**Conditional Forecasting**

In general, the paths of the counterfactual factors seem reasonable. The severely adverse factors are generally—albeit not always—higher than those associated with the adverse scenario, which in turn, are higher than those associated with the baseline scenario. Even so, this visual evaluation of the counterfactual factors does not by itself imply anything about the magnitude of stress faced by the banks over the scenario horizon. To get a better grasp of such stress, we conduct a conditional forecasting exercise akin to that required by the Dodd-Frank Act. Specifically, starting in 2013:Q3 we construct a path of 1-step-ahead through 4-step-ahead forecasts of both NCOs and NIMs, conditional on the counterfactual factors for each scenario.



In light of the apparent benefit of using the 3PRF-based factors, we focus exclusively on conditional forecasts using these factors.

Methodologically, our forecasts are constructed using the minimum MSE approach to conditional forecasting delineated in Doan, Litterman, and Sims (1984) designed for use with VARs. To implement this approach we first link the 1-step-ahead predictive equation from equation (1) with an AR(1) for the factors to form a restricted VAR of the form

$$(7) \quad \begin{pmatrix} \Delta y_t \\ f_t \end{pmatrix} = \begin{pmatrix} \alpha_y \\ \alpha_f \end{pmatrix} + \begin{pmatrix} 0 & \beta_y \\ 0 & \beta_f \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ f_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{ft} \end{pmatrix}.$$

We then estimate this system by OLS using the historical sample ranging from 1990:Q1 through 2013:Q3. The coefficients in the first equation line up with those already presented in Table 2. The coefficients in the second equation are not reported here but imply a root of the lag polynomial of roughly 0.80.<sup>4</sup> Using this system of equations, along with an assumed future path of the factors for all  $t = 2013:Q4, \dots, 2016:Q3$ , we construct forecasts across the entire scenario horizon. We do so separately for each target variable and each scenario. For the sake of comparison, we also include the path of the unconditional forecast associated with equation (7) using an iterated multistep approach to forecasting.

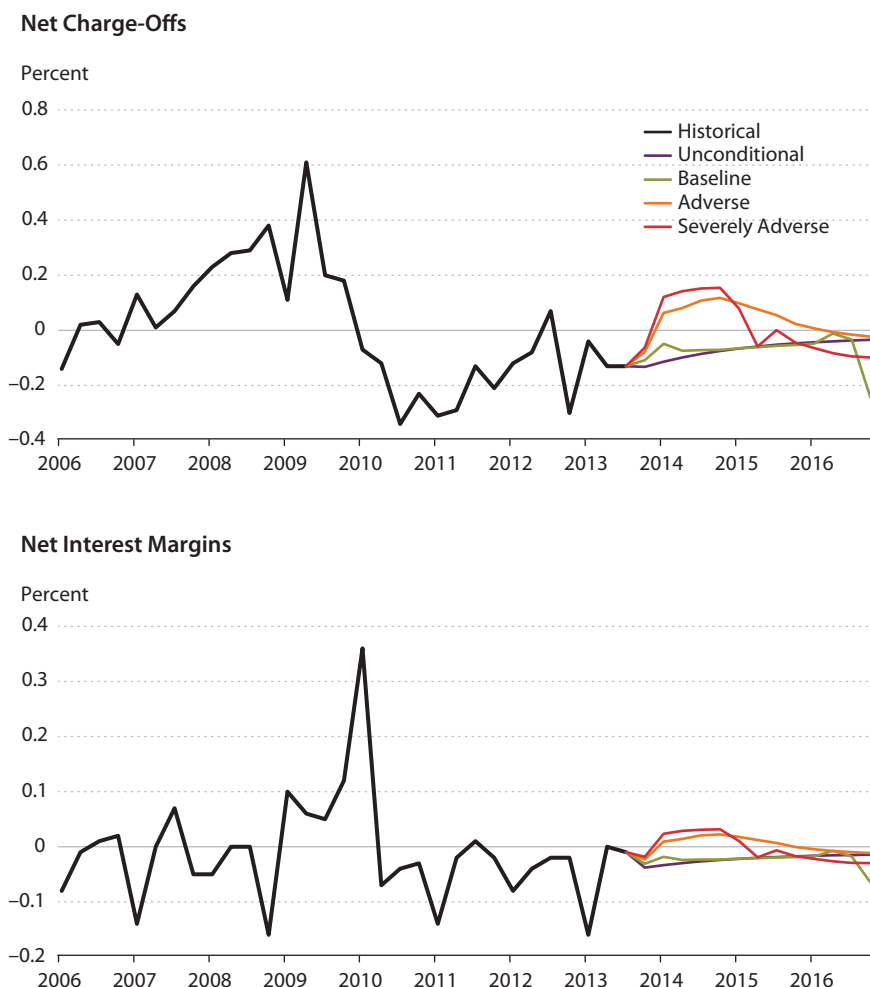
Figure 6 shows the conditional forecasts. The upper panel is associated with NCOs, while the latter is associated with NIMs. In each panel, we plot the historical path of the change in the target variable through 2013:Q3, at which point we plot four distinct forecast paths: one associated with the unconditional forecast and one associated with each of the three scenarios. The two plots are similar in many ways. The unconditional forecasts converge smoothly to the unconditional mean over the forecast horizon. As expected, the baseline forecasts differ little from those associated with the unconditional forecast.

In contrast, the adverse and severely adverse forecasts tend to be significantly higher than the unconditional forecast at least through mid-2015 for NCOs and early 2016 for NIMs. Through the end of 2014, the severely adverse predictions are higher than any of the other scenarios at any point over the forecast horizon. However, in early 2015 the forecasts associated with the severely adverse scenario decline sharply and thereafter track the baseline forecasts. The same path does not apply to the adverse scenario forecasts. They remain relatively higher until the end of 2015 before converging to those associated with the baseline forecasts.

The paths of the forecasts seem reasonable in the sense that the forecast for the severely adverse scenario is the most severe, followed by the adverse and then the baseline forecasts. Even so, the magnitudes of the forecasts, especially those associated with the severely adverse scenario, do not match the levels of either NCOs or NIMs observed during the Great Recession. On the one hand, this might be viewed as a failure of the model. But upon reflection, it is not so surprising that the forecasts do not achieve the levels observed during and immediately following the financial crisis. These forecasts are constructed based solely on a limited number of macroeconomic aggregates and do not include many of the series known to have been problematic during the crisis. These include, but are certainly not limited to, measures of financial market liquidity and counterparty risk, as well as measures related to monetary and fiscal

**Figure 6**

**Forecasts of Target Variables Conditional on Future Path of Factors**



policy. Variables such as money market rates and the TED spread (U.S. T-bill–eurodollar spread) would almost certainly accommodate these bank stress concerns but are not used here because they are not present in the CCAR scenarios. In short, while the conditional forecasts based on macroeconomic aggregates provide some information on the expected path of future NCOs and NIMs, they by no means paint a complete picture.

## CONCLUSION

In this article, we investigate the usefulness of factor-based methods for assessing industry-wide bank stress. We consider standard PC-based approaches to constructing factors, as well

as methods such as PLS and the 3PRF, that design the weights to relate the predictive content of the factor to a given target variable. As is common in the stress testing literature, we evaluate industry-wide bank stress based on the level of NCOs and NIMs.

We provide two types of evidence related to the usefulness of these indexes. Based on a pseudo out-of-sample forecasting exercise, we show that our 3PRF-based factor appears to typically provide the best or nearly the best forecasts for NCOs and NIMs. This finding holds at both the 1-quarter- and 4-quarter-ahead horizons. Admittedly, the best factor model performs worse than the naive, no-change model for NCOs, but for NIMs the factor models do appear to have marginal predictive content beyond that contained in the no-change model.

We then study how these factor-based methods might help policymakers evaluate the degree of stress present in the stress testing scenarios. To do so, we use the factor weights estimated using historical data, along with the hypothetical scenarios as presented in the official stress testing documentation, to construct hypothetical stress factors. At the 1-quarter-ahead horizon, these factors generate very intuitive graphical representations of the level of stress faced by the banking industry as measured through the lens of NCOs and NIMs. We conclude with a conditional forecasting exercise designed to provide some indication of the future path of NCOs and NIMs should one of the counterfactual scenarios occur. ■

## NOTES

- <sup>1</sup> The scenarios are available on the Federal Reserve Board of Governors website (<http://www.federalreserve.gov/bankinforeg/bcreg20131101a1.pdf>).
- <sup>2</sup> We focus exclusively on the case of one factor throughout and hence consider only one proxy variable throughout. The theory developed in Kelly and Pruitt (2011) permits more than one proxy variable and thus more than one factor.
- <sup>3</sup> For brevity, we let  $P$  denote the number of forecasts regardless of the horizon  $h$ .
- <sup>4</sup> We also considered lags of  $\Delta y_t$  in the second equation. In all instances, they were insignificantly different from zero at conventional levels.

## REFERENCES

- Acharya, Viral V.; Engle, Robert and Pierret, Dianne. "Testing Macroprudential Stress Tests: The Risk of Regulatory Risk Weights." NBER Working Paper No. 18968, National Bureau of Economic Research, April 2013; [http://www.nber.org/papers/w18968.pdf?new\\_window=1](http://www.nber.org/papers/w18968.pdf?new_window=1).
- Bai, Jushan and Ng, Serena. "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions." *Econometrica*, July 2006, 74(4), pp. 1133-50; <http://www.jstor.org/stable/3805918>.
- Bolotnyy, Valentin; Edge, Rochelle M. and Guerrieri, Luca. "Stressing Bank Profitability for Interest Rate Risk." Unpublished manuscript, 2013, Board of Governors of the Federal Reserve System.
- Ciccarelli, Matteo and Mojon, Benoit. "Global Inflation." *Review of Economics and Statistics*, August 2010, 92(3), pp. 524-35; [http://www.mitpressjournals.org/doi/pdf/10.1162/REST\\_a\\_00008](http://www.mitpressjournals.org/doi/pdf/10.1162/REST_a_00008).
- Clark, Todd E. and McCracken, Michael W. "Tests of Equal Forecast Accuracy and Encompassing for Nested Models." *Journal of Econometrics*, November 2001, 105(1), pp. 85-110; <http://www.sciencedirect.com/science/article/pii/S0304407601000719?via=ihub>.
- Clark, Todd E. and McCracken, Michael W. "Evaluating Direct Multistep Forecasts." *Econometric Reviews*, 2005, 24(4), pp. 369-404.
- Clark, Todd E. and McCracken, Michael W. "Reality Checks and Comparisons of Nested Predictive Models." *Journal of Business and Economic Statistics*, January 2012, 30(1), pp. 53-66.
- Covas, Francisco B.; Rump, Ben and Zakrajsek, Egon. "Stress-Testing U.S. Bank Holding Companies: A Dynamic Panel Quantile Regression Approach." Finance and Economic Discussion 2013-155, Board of Governors of the Federal Reserve System, September 2013; <http://www.federalreserve.gov/pubs/feds/2013/201355/201355pap.pdf>.
- Dickey, David A. and Fuller, Wayne A. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association*, 1979, 74(366), pp. 427-31; <http://www.jstor.org/stable/2286348>.
- Diebold, Francis X. and Mariano, Roberto S. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics*, July 1995, 13(3), pp. 253-63; <http://www.tandfonline.com/doi/abs/10.1080/07350015.1995.10524599#.UzRfSF8o6Uk>.
- Doan, Thomas; Litterman, Robert and Sims, Christopher. "Forecasting and Conditional Projection Using Realistic Prior Distributions." *Econometric Reviews*, 1984, 3(1), pp. 1-100; <http://www.tandfonline.com/doi/abs/10.1080/07474938408800053#.UzRdBF8o6Uk>.
- Elliott, Graham; Rothenberg, Thomas J. and Stock, James H. "Efficient Tests for an Autoregressive Root." *Econometrica*, July 1996, 64(4), pp. 813-36; <http://www.jstor.org/stable/2171846>.
- Engel, Charles; Mark, Nelson C. and West, Kenneth D. "Factor Model Forecasts of Exchange Rates." Working Paper No. 012, University of Notre Dame Department of Economics, January 2012; [http://www3.nd.edu/~tjohns20/RePEc/deendus/wpaper/012\\_rates.pdf](http://www3.nd.edu/~tjohns20/RePEc/deendus/wpaper/012_rates.pdf).

- Giannone, Domenico; Reichlin, Lucrezia and Small, David. "Nowcasting: The Real-Time Informational Content of Macroeconomic Data." *Journal of Monetary Economics*, May 2008, 55(4), pp. 665-76; <http://www.sciencedirect.com/science/article/pii/S0304393208000652#>.
- Guerrieri, Luca and Welch, Michelle. "Can Macro Variables Used in Stress Testing Forecast the Performance of Banks?" Finance and Economics Discussion Series 2012-49, Board of Governors of the Federal Reserve System, July 2012; <http://www.federalreserve.gov/pubs/feds/2012/201249/201249pap.pdf>.
- Kelly, Bryan T. and Pruitt, Seth. "The Three-Pass Regression Filter: A New Approach to Forecasting with Many Predictors." Research Paper No. 11-19, Fama-Miller Working Paper Series, University of Chicago Booth School of Business, January 2011; [http://faculty.chicagobooth.edu/bryan.kelly/research/pdf/Forecasting\\_theory.pdf](http://faculty.chicagobooth.edu/bryan.kelly/research/pdf/Forecasting_theory.pdf).
- Ludvigson, Sydney C. and Ng, Serena. "Macro Factors in Bond Risk Premia." *Review of Financial Studies*, December 2009, 22(12), pp. 5027-67.
- Newey, Whitney K. and West, Kenneth D. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, May 1987, 55(3), pp. 703-08; <http://www.jstor.org/stable/1913610>.
- Stock, James H. and Watson, Mark W. "Forecasting Inflation." *Journal of Monetary Economics*, October 1999, 44(2), pp. 293-335; <http://www.sciencedirect.com/science/article/pii/S0304393299000276?via=ihub>.
- Stock, James H. and Watson, Mark W. "Macroeconomic Forecasting Using Diffusion Indexes." *Journal of Business and Economic Statistics*, April 2002, 20(2), pp. 147-62; <http://dx.doi.org/10.1198/073500102317351921>.
- Stock, James H. and Watson, Mark W. "Dynamic Factor Models," in Michael P. Clements and David F. Hendry, eds., *Oxford Handbook of Economic Forecasting*. New York: Oxford University Press, 2011, pp. 35-59.
- West, Kenneth D. "Asymptotic Inference About Predictive Ability." *Econometrica*, September 1996, 64(5), pp. 1067-84; <http://www.jstor.org/stable/2171956>.

