Revenue Forecasts, Capital Adequacy, and the Uncertainty of Stress Test Results*

Valentin Bolotnyy
Harvard University

Rochelle M. Edge
Federal Reserve Board

Luca Guerrieri
Federal Reserve Board

May 6, 2015

Abstract

An array of time series models found in many applications to be near the frontier of forecast performance generate forecasts for large-bank net interest margins (NIMs) that capture very little of the out-of-sample variation in NIMs. Accordingly, these models, even when conditioned on the stressful interest rate risk scenarios in the Federal Reserve's bank stress tests, imply NIM forecasts that are statistically indistinguishable from those obtained under the baseline scenarios. However, this large degree of forecast uncertainty is not reflected in the current quantitative capital assessment framework.

Key Words: Net Interest Margins, Interest Rate Risk, Forecasting, Stress Tests.

JEL Classification: C53, E47, G21

* This paper was previously circulated under the title of "Stressing Bank Profitability for Interest Rate Risk." We thank Michel van der Wel and Christina Wang for comments on earlier drafts and Jon Faust, Divya Kirti, Andreas Lehnert, Barbara Rossi, and Emre Yoldas for helpful conversations. We are especially grateful to Jeremy Rudd for performing an independent check of the paper's aggregate forecast results. Bolotnyy acknowledges support from the NSF and the Paul & Daisy Soros Fellowship for New Americans. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.
Contact Information: vbolotnyy@fas.harvard.edu; rochelle.m.edge@frb.gov; luca.guerrieri@frb.gov (corresponding author).

1 Introduction

This paper considers the ability of an array of time series models to forecast aggregate and individual large-bank net interest margins (NIMs), conditional on interest rates and other variables, and ultimately finds that these models forecast poorly. The models that we consider are those that in other applications have been found to be near the frontier of forecast performance and are similar to the types models that are used to project revenues in the Federal Reserve's supervisory bank stress tests (see page 75 of FRB 2014b and pages 61 and 62 of FRB 2015). Among the models that are specified in levels—like most of the models that have been used in the macro-banking literature to study NIMs—only a few deliver conditional forecasts that can improve on a random walk out-of-sample and, even in such cases, the improvement is marginal. Analogous models specified in first differences do deliver conditional forecasts that improve more notably on a random walk—and in some cases also improve on forecasts for specific banks made by bank earnings analysts—but, in an absolute sense, these models' forecasts remain poor. That is, they imply forecast errors that are large and only marginally smaller than the variability of NIMs themselves, thereby indicating that these models predict very little of the variation in NIMs over time.

Our analysis is motivated by macroeconomic stress tests and the projected paths of (regulatory) bank capital that they imply—so-called, *pro forma* capital ratios. Stress tests and *pro forma* capital ratios play a prominent role in the U.S. post-crisis capital regulatory regime and are the reason we emphasize *conditional* forecasts. In particular, in the current U.S. regulatory regime, *pro forma* capital ratios represent an additional criterion for capital-adequacy standards since not only must banks' *current* bank capital ratios exceed regulatory (*e.g.*, Basel III) minima but their *pro forma* capital ratios over the next 9-quarters under baseline and stressed conditions must do so, as well.[1]

The poor forecast performance of NIM models has serious implications for regulators' use of stress tests and *pro forma* capital ratios in providing forward-looking assessments of bank

[1]Stress tests and *pro forma* capital ratios are also prominent in the post-crisis regulatory regimes of several other countries, such as the United Kingdom and the European Union (see, BOE 2014 and EBA 2014). Additionally, they have also been used in past banking crises to restore confidence in the banking sector, such as in the Nordic banking crises of the early 1990s, the U.S. financial crisis in 2008-09, and the sovereign-debt crisis in the euro area around the turn of the decade.

capital adequacy. In particular, we find that when we combine even the best of our NIM models with the stressful interest rate risk scenarios that were used by the Federal Reserve in the 2013 stress tests, the paths of NIMs that are implied by even very dire interest rate scenarios are statistically indistinguishable from those implied by the baseline scenario. We then calculate what the paths of NIMs associated with the baseline and stress scenarios imply for Pre-provision Net Revenue (PPNR) and then (based on the projected losses, revenues, and incomes that were published for the 2014 stress tests) calculate the implications for bank capital. We can also calculate what the forecast errors that surround the projected path of NIMs imply for the forecast errors surrounding the projected path of bank capital. Here we find that, mirroring the results for NIMs, the *differences* between the paths of bank capital implied by the different paths of NIMs are small when compared to the *uncertainty* surrounding each of these scenarios' paths of capital.

The organization of our paper is as follows. Section 2 provides a brief primer on the Federal Reserve's stress tests and motivates our focus on NIMs. Section 3 reviews the literature on NIMs with an eye to cataloging variables that may be useful for forecasting NIMs. Section 4 describes the data and models that we use to forecast aggregate NIMs and then presents our ultimately poor forecasting results. Section 5 follows up on the paper's main results with some robustness analysis. In particular, it considers the use of a longer sample and the use of other variables—not included in the stress-test scenarios that are published by the Federal Reserve—that might be useful for forecasting NIMs. The section also applies the same forecast analysis performed on aggregate NIMs to bank-specific NIMs for which we find that our model-based forecasts are competitive with analysts forecasts, especially for universal banks. Ultimately, however, we still find that these additional specifications do not alter our robust finding that NIM forecast errors are large and that NIM models predict very little of the variation in NIMs over time, conditional on macro scenarios.

Section 6 then investigates the implications of poor NIM forecasts for the use of stress tests and *pro forma* capital ratios in the post-crisis capital regulatory regime. Finally, section 7 concludes and suggests an alternative approach to stressing bank revenue generation that is less reliant on macro scenarios to assess future bank capital adequacy.

2 Primer on the Federal Reserve's stress tests and our focus on NIMs

Macroeconomic stress tests and the projected paths of (regulatory) bank capital that they imply—so-called, *pro forma* capital ratios—play a prominent role in the U.S. post-crisis capital regulatory regime.[2] They have been codified in two key rules—specifically, the 2012 rules implementing the stress testing requirements of the Dodd-Frank Act (DFA) and in the Federal Reserve Board's 2011 Capital Planning Rule (referred to as CCAR, an abbreviation for Comprehensive Capital Analysis and Review)—both of which apply to all U.S. bank holding companies (BHCs) with total consolidated assets of $50 billion or more, currently corresponding to the largest 31 BHCs.[3] These rules represent a significant change in the U.S. post-crisis capital regulatory regime, since they imply that banks no longer meet regulatory capital-adequacy standards if only their *current* bank capital ratios exceed regulatory (*e.g.*, Basel III) minima. Rather, these rules—and, in particular, CCAR—imply that to meet regulatory capital-adequacy standards, banks' baseline and stressed *pro forma* capital ratios over the next 9-quarters (as projected by supervisory authorities and given banks' planned capital distributions) must also exceed regulatory minima.

This major change in regulatory capital-adequacy standards reflects one of the key lessons that emerged in the 2008 financial crisis (and, indeed, has surfaced repeatedly in past banking crises), which is that creditor and counterparty confidence do not depend on a bank's *current* capital ratios but rather depend on *future* capital ratios, particularly under prolonged severe stress. Indeed, in the fall of 2008, many of the banks that found themselves facing creditors unwilling to rollover their debt and trading counterparties ceasing to want to contract with them were, in current terms, well capitalized. However, creditors' and trading counterparties' doubts as to these banks' ability to remain adequately capitalized in the future—especially,

[2]Our description of the Federal Reserve's stress testing program in this section is selective and focuses mainly on the details of these requirements that were relevant to guiding our analysis. We direct the reader to Hirtle and Lehnert (2014) for a broader description of the program.

[3]The number of BHCs that have participated in CCAR has been growing from the original 18 in CCAR 2012 and 2013, to 30 in CCAR 2014, and to 31 in CCAR 2015. See the Federal Reserve Board's (FRB's) 2012 rules implementing Section 165(i)(1) and (i)(2) of the DFA, FRS (2012), and Section 225.8 of the Board's Regulation Y (the 2011 Capital Planning Rule), FRS (2011).

in the event of economic and financial conditions deteriorating rapidly— led to credit drying up for these banks and in some cases to their demise.[4]

Building off of this key lesson, one of the main goals of the Federal Reserve's stress tests is to ensure that banks hold enough capital so as to remain adequately capitalized, even following a period of rapid deterioration in economic and financial conditions. Indeed, the CCAR stress tests tie a BHC's ability to make capital distributions to its capital ratios under expected and stressed economic and financial conditions. Specifically, in order to make its planned capital distributions, a BHC in the CCAR stress tests must demonstrate that—given its planned capital distributions—it can for the next 9-quarters continue to meet all minimum capital regulatory standards, under both the expected and stressed conditions described in the stress tests' scenarios.[5] The DFA stress tests are closely related to the CCAR stress tests but have a few key differences, one of which is that the results of the DFA stress tests are not as directly tied to banks' permitted capital distributions.

In performing both the CCAR and DFA stress tests, supervisory authorities must generate 9-quarter projections of *pro forma* bank capital ratios, given the macroeconomic and financial conditions specified in the stress test scenarios and assumed capital distribution plans. This exercise requires projections of all components of after-tax net income and other comprehensive income (OCI), since it is projections of these variables that are then cumulated—together with bank capital distributions—to yield the projected path of regulatory capital. Equation 1 shows more precisely all of the components of after-tax net income that cumulate to changes in regulatory capital (where these identities are taken from page 12 of FRB 2014b and page 9 of FRB 2015). The first line of equation 1 on the right hand side is PPNR, while the first three lines are pre-tax net income, the first four lines are post-tax net income, and the first five lines are the "accounting" concept of capital.

[4]See item 1.d on page 71441 of FRS (2013)—the Federal Reserve Board's 2013 policy statement on scenario design—for an articulation on this view.

[5]The CCAR stress test scenarios start in the fourth quarter of the year preceding the "reference" year of the stress test and end in the fourth quarter of the year following the reference year, for a total of 9-quarters. So, for example, the scenarios for the CCAR 2013 stress tests extend from 2012q4 through 2014q4.

$$
\begin{aligned}
&\Delta \text{ Regulatory capital} \\
&= \underbrace{\text{Net interest income (NII)} + \text{Non-interest income} - \text{Non-interest expense}}_{\text{Pre-provision net revenue (PPNR)}} \\
&+\text{Other revenues} - \text{Provisions for loan losses} - \text{Realized losses on securities} \\
&-\text{Realized losses on fair-value loans} - \text{Trading and counterparty losses} \\
&-\text{Taxes} + \text{Extraordinary items net of taxes} \\
&-\text{Net capital distributions to common and preferred shareholders} \\
&-\text{Deductions \& additions to regulatory capital } (\textit{e.g.}, \text{ OCI})
\end{aligned} \tag{1}
$$

Equation 1 allows us to clarify another key difference between the CCAR and DFA stress tests, which is that in the CCAR stress tests, net capital distributions (given in the second to last line in equation 1) are set equal to banks' *own* planned capital distributions to shareholders, whereas in the DFA stress tests, an "assumed" capital plan—to allow comparability—is assumed for all banks.

While all of the terms listed in equation 1 form part of the projections of future regulatory capital, we focus on the revenue-generation component and, in particular, on net interest margins (NIMs), defined as the ratio between net interest income (NII) and interest earning assets. This decision reflects a number of considerations. First, as evidenced from several past banking crises—including the first waves of bank failures in the Great Depression, the U.K. secondary banking crisis of the 1970s, the U.S. Savings and Loans crisis of the 1980s, and the Nordic banking crises of the 1990s—bank revenue losses (captured in the first line in equation 1) have the potential to be just as an important a source of stress for banks as losses stemming from loan defaults (captured by provisions in the second line). Second, also related to revenue generation, is that even in situations where bank losses resulting from loan defaults are the main source of stress to bank capital adequacy, bank revenue developments are still material for determining the resulting path of *pro forma* capital ratios.[6] Third, to the extent that revenue generation is a relevant determinant of bank capital, the stress tests are the only place in the capital regulatory regime where bank revenues materially enter into

[6] See Governor Tarullo's April 2012 speech "Developing Tools for Dynamic Capital Supervision" for an articulation of this view. Later in this section, we also provide some calculations that highlight the materiality of bank revenue levels for capital-ratio outcomes.

the assessment of bank capital adequacy. Other regimes for assessing bank capital adequacy, such as the Basel regime, are instead focused on the losses that banks face from their asset portfolios (that is, items captured in the second and third lines of equation 1) and not on the losses that they face from revenue generation.[7]

We would also note that the analysis that we undertake in this paper is more salient to the revenue-generation component of changes in bank capital. The NIM models that we consider are time series models and it is time series models that are for the most part used in the DFA/CCAR stress tests to project revenues. This is in contrast to models used in the DFA/CCAR stress tests to project loan, securities, and trading book losses, which are instead all estimated on very granular—that is, loan- and securities-level—data and use probability of default (PD) and loss given default (LGD) modeling frameworks.[8]

Within revenues, we focus on NIMs for two main reasons. First, for most banks, interest income is the main component of revenues. Interest income accounts for about two-thirds of income for the largest BHCs, while interest expenses typically (that is, outside of the zero lower bound environment) account for about 40 percent of bank expenses (excluding loss provisions). Indeed, in the historical examples discussed above, bank revenue losses stemmed from unfavorable interest rate configurations that depressed NIM outcomes.[9] Second, NIMs reflect the return from maturity transformation—that is, borrowing short and lending long—which is one of the key services provided by banks. Accordingly, it is essential for stress tests to be able to model this revenue process well.

[7]The Basel regime does include operational risk (which is considered to be a risk to revenues) in its evaluation of bank capital adequacy. However, operational risk is a very different form of risk to revenues to the the one that we consider in this paper—specifically interest rate risk. In addition, operational risk—while having been an important source of stress for specific institutions—has never lead to widespread risk in the banking sector as indeed interest rate risk has on several occasions in the past.

[8]See pages 65 to 71 of FRB (2014b) for a description of the models used to forecast losses in the stress tests and page 75 for a description of the revenue models. Note that in the 2015 DFA/CCAR stress tests one component of NIMs—specifically interest expenses for subordinated debt—began to be forecast using granular information—specifically, security-level information on BHCs' subordinated debt. See pages 61 to 62 of FRB (2015) for a discussion of this change.

[9]The exception is the first wave of bank failures in the Great Depression, which stemmed from elevated liability costs that in turn were brought about by fierce competition between banks for deposits.

3 Related literature on NIMs and variable selection

The literature on modeling NIMs has two strands. The first strand emphasizes the link between risk-free interest rates of various maturities and NIMs and lies more in the macro-banking tradition. The second strand focuses on the optimal margin set by banks and lies more in the micro-banking tradition. Our paper is more closely related to the macro-banking approach and, accordingly, we review the first strand in more detail.

Covas, Rump, and Zakrajsek (2014) and Hirtle, Kovner, and Vickery (2013) are two recent papers in the macro-banking tradition that, like our paper, have been motivated by the recent importance of stress testing. These papers both include models of NIMs, but since they present much larger aggregative models of banks' financial statements, their focus on NIMs is considerably less in-depth than in our paper. These papers also place relatively less emphasis on evaluating the conditional out-of-sample forecast performance, which we emphasize given the ultimate use of our models for generating *pro forma* capital ratios under stressed scenarios.[10] An attractive feature of these models is that, in addition to encompassing the components of banks' financial statements, they also have an aggregation framework to link changes in NIMs with changes in capital ratios.

Covas et al. (2014) and Hirtle et al. (2013) both link the *level* of NIMs to the *level* of the slope of the Treasury yield curve and the *level* of a short-term Treasury rate. These two variables capture bank earnings from maturity transformation services and deposit transactions services. The slope of the yield curve reflects the return to banks from maturity transformation and is thus expected to enter NIM models with a positive sign. The short-term market interest rate in NIM models reflects the fact that bank deposit rates, while typically lower than market rates (given the transactions services they provide), are constrained by the zero lower bound. As such, the short rate places an upper limit on what banks can earn from the provision of their transactions services, which means that the short rate is expected to enter the model, also, with a positive sign.

Other papers in the macro-banking strand of the NIM-modeling literature are English

[10] Covas, Rump, and Zakrajsek (2014) do present some forecast evaluation results in order to demonstrate the benefits of focusing on quantile projections. However, they only report results for net charge offs (NCOs) and PPNR and their focus is on density forecasts generated by their quantile regression model.

(2002), English, van den Heuvel, and Zakrajsek (2012), and Alessandri and Nelson (2015) that in contrast to the papers just discussed do not consider NIMs in the context of a broader stress testing model. All of these models, nonetheless, also link the *level* of NIMs to the *level* of the slope of the Treasury yield curve and the *level* of a short-term Treasury rate. With the exception of English (2002), all of the macro-banking NIM papers discussed thus far use firm-level time-series data and estimate panel regression models. Of these papers, only one, Covas et al. (2014), does not use least squares regression; rather, it uses quantile regression methods to capture the possibility of differential effects of key macro variables on banking variables during episodes of stress.

Although English (2002), English et al. (2012), and Alessandri and Nelson (2015) focus solely on NIMs (like our paper), they do not consider forecast performance. English (2002) studies NIMs, as well as its two components of interest income and interest expenses, for a number of advanced economies with the view to understanding how the aggregate banking sectors in these countries have managed the interest rate risk that they face on their earnings. English et al. (2012) consider the relationship between NIMs and the slope of the yield curve and short-term rates for a panel of banks, allowing for specific coefficients estimated for each bank to vary as linear functions of bank balance sheet variables that one *a priori* might think would be relevant to net interest income (such as the maturity gap between bank assets and liabilities, the deposit share of liabilities, and the loan share of assets). Alessandri and Nelson (2015), who use data for the United Kingdom, also consider the relationship between NIMs and the slope of the yield curve and short-term rates for a panel of banks, although here their ultimate interest is the possibility of hedging between banks' income generation from their lending and deposit-taking activities and their trading book activities.

The documentation that is available on the models used to project the components of PPNR in the DFA/CCAR stress tests (see page 75 of FRB 2014b and see pages 61 and 62 of FRB 2015) suggests that the way in which NIMs are modeled in the stress tests shares many features with the macro-banking NIM-modeling literature. For example, although the documentation does not state the precise variables that enter the equations related to NIMs, variables like Treasury yields are included in the list of macro variables used to model components of PPNR. One notable feature of the equations used to model NIMs in the stress tests is that they are quite disaggregated. For example, similar to English (2002), interest income and interest expenses are modeled separately. However, going beyond the disaggregation in English (2002), interest income is modeled in eight components while interest expenses are modeled in five components. Although the documentation states that statistical predictive power and economic interpretation inform the inclusion of macroeconomic variables in the models, the documentation does not mention out-of-sample forecast performance specifically.

Exceptions to the papers discussed thus far—that consider models of NIMs but do not look into forecast performance—are Guerrieri and Welch (2012) and Grover and McCracken (2014). These papers examine forecast performance, although they address a very different question to the one that we consider. In particular, these papers both examine whether any of the information contained in the sixteen U.S. variables that comprise the Federal Reserve Board's DFA/CCAR stress test scenarios have predictive content for key bank stress-test variables, like NIMs and net charge offs (NCOs). This interesting question for scenario design is clearly different to the one that we consider, which is the ability of time-series model to forecast NIMs in a way that would yield credible projections of *pro forma* capital ratios. Nonetheless, these papers suggest that it may be worth considering other stress-test scenario variables in some of our NIM-model specifications.

A separate strand of the NIM modeling literature comes from the micro-banking tradition, which focuses on the determinants of the loan rates and deposit rates that banks set as implied by the maximization of their profits. Papers in this tradition—such as Ho and Saunders (1981), Angbazo (1994), and Saunders and Schumacher (2000)—emphasize a different set of variables for modeling NIMs. These include the degree of competition facing banks in deposit and loan markets, banks' risk aversion to finding themselves with deposit supply without corresponding loan demand or loan demand without corresponding deposit supply, and the volatility of interest rates, which raises the likelihood that banks will face losses in one of the aforementioned situations. Empirical papers in this literature typically only include competition and interest-rate volatility in their models of NIMs, which are the variables that we also include when considering alternate model specifications.[11]

[11] While most papers modeling NIMs fall clearly in either the macro- or micro-banking literature, Flannery (1981) features elements of both literatures. In particular, he emphasizes the level and volatility of short-term Treasury rates, although he does not consider the other variables considered in the two literatures.

4 Forecasting aggregate NIMs: 2000q1 to 2008q3

We assess our forecasts of aggregate NIMs over 2000q1 to 2008q3 (and, in robustness analysis, also over 2000q1 to 2013q3). We use the assessment that ends in 2008q3 as our baseline (and preferred) sample due complications with the last half decade of data. The most notable of these complications is the zero interest rate environment, which poses significant challenges for modeling interest rates. There were also several structural changes with implications for NIMs over this period, including the Federal Reserve starting to pay interest on excess reserves in the fourth quarter of 2008, Financial Accounting Statements (FAS) 166/167 coming into effect at the beginning of 2010, and the full repeal of Reg. Q in the DFA of 2010. We attempt to smooth over these changes when we perform our robustness analysis in section 5.

4.1 Data for forecasting NIMs

4.1.1 Aggregate NIMs

For our forecast analysis of aggregate NIMs we use data from the quarterly Consolidated Reports of Condition and Income (Call Report) that every national, state member, and insured nonmember bank is required to file for the last day of each quarter by the Federal Financial Institutions Examination Council (FFIEC). Our definition of NIMs follows closely the aggregate series from the Call Report form. For our definition of net interest income, we use Item 3 in Schedule RI of the FFIEC 031 reporting form. For the definition of interest earning assets at quarter end, we use the Schedule RC and the totals from Schedules RC-A (excluding Item RC 1.a, noninterest-bearing balances and currency and coin), RC-B, RC-C, and RC-D. Our results are little changed by defining NIMs using an average of interest earning assets over the reporting quarter, instead of at the end of the quarter.

The Call Report data used in this analysis are adjusted for bank mergers and acquisitions, using structure data from the National Information Clearinghouse (NIC) on mergers and acquisitions. Foreign entities are excluded and domestic subsidiaries are aggregated up to the parent, bank holding company (BHC), level.[12] To get an aggregate banking sector

[12]NIM data are adjusted for mergers between commercial banks by comparing balance sheet values of interest income, interest expenses, and interest earning assets at the end of the quarter with those at the

measure of NIMs, we aggregate NIMs for the top 25 BHCs, as ranked by total interest earning assets, where this ranking is assessed quarterly.

The top panel of figure 1 shows the historical time series of NIMs while the middle panel shows the two series—Interest Income divided by Interest Earning Assets and Interest Expenses divided by Interest Earning Assets—that when differenced imply NIMs. NIMs show a large spike in 1988q4. We start out baseline sample in 1989q1 to avoid this spike, which stems from late payments from Brazil during the Latin American debt crisis.

4.1.2 Treasury yields

The yields data that we use in our analysis are derived using a smoothing technique from Gurkaynak, Sack, and Wright (2007), based on Nelson and Siegel (1987) and Svensson (1994), which allows for daily measures of an off-the-run Treasury yield curve. We use quarterly yields for twelve maturities in our models: 3-month, 6-month, 9-month, 1-year, 2-year, 3-year, 5-year, 7-year, 10-year, 15-year, 20-year, and 30-year.[13] The yields, which are plotted in the lower panel of figure 1, are quarterly averages of daily yields.

4.2 Models for forecasting NIMs based on Treasury yields

We generate forecasts of NIMs from an array of standard methods. We use quarterly frequency data and consider a forecast horizon up to ten quarters long, which is the horizon used in the DFA/CCAR stress tests. Our benchmark set of models include only Treasury yields—either untransformed or as part of yield-curve factors—and we forecast NIMs *conditional* on actual Treasury yields or actual, calculated yield-curve factors. All of our models, with the exception of our random walk benchmark model—which we call the no-change forecast—condition on some variables. We denote the variables upon which we condition our forecast by I and denote our s-step ahead NIM forecast conditional on I by $NIM_{t+s/I}$. Our benchmark is the no-change forecast (model 1).

Our presumption is that our NIM forecasting models will benefit from including a large

beginning of the quarter, accounting for amounts acquired or lost during the period because of mergers; see the appendix in English and Nelson (1998).

[13]Daily yields are published with a two-day lag under the base mnemonic SVENY at www.federalreserve.gov/econresdata/researchdata/feds200628_1.html.

number of yields and so we consider a broader set of yields than those used in the macro-banking NIM literature, which considers only 3-month and 10-year yields. Our motivation for exploring other maturities is that in practice it is more than just these two rates that are relevant for banks' net interest income. While it is the case that 10-year yields are the driving rate for 30-year fixed residential mortgages, other loans categories are influenced by yields of other maturities. For example, 6-month to 1-year yields are the more relevant driving rates for adjustable rate residential mortgages, 3-month yields for home equity lines of credit (HELOCs), 3- to 5-year yields for commerical real estate loans, and 3-month yields for commercial and industrial (C&I) loans. In addition, interest income from securities are influenced by yields across all maturities.

Our forecast models therefore include either a large number of yields or include yield-curve factors, specifically the level and slope, and sometimes also curvature, derived from a large number of yields. The first of our two specifications in which yields enter our models untransformed includes a large number of yields. The other specification includes a smaller number of yields. We consider the following models:

1. No-change forecast. This model, which forecasts NIMs as a random walk without a drift at all horizons s, is given simply by: $NIM_{t+s/t} = NIM_t$.

2. Many yields with forecast combination. This model consists of twelve bi-variate models of NIMs and yields that each generate a forecast that is then combined with uniform weights to form an overall forecast.

3. Fewer yields. This model includes only two or three yields; specifically, the 3-month and 10-year yield and sometimes also the 2-year yield.

We calculate the yield-curve factors for our NIM models in several different ways.

4. Observed factors. We calculate these factors as simple additive functions of a small number of yields of different maturities. In particular, the level factor L_t^o is set equal to the average of the 3-month, 2-year and 10-year yields, the slope factor S_t^o is the difference between the 10-year and the 3-month yields, and the curvature factor C_t^o is given by 2 times the 2-year yield minus the sum of the 3-month and 10-year yields. These factors follow the "observed factor" definitions used by Diebold and Li (2006).

5. Dynamic factors. We calculate these factors using a Nelson and Siegel (1987) model and following Diebold, Rudebusch, and Aruba (2007). The dynamic factor model that we consider takes the familiar state-space form:

$$\begin{pmatrix} L^d_{t+1} - \mu_L \\ S^d_{t+1} - \mu_S \\ C^d_{t+1} - \mu_C \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} L^d_t - \mu_L \\ S^d_t - \mu_S \\ C^d_t - \mu_C \end{pmatrix} + \begin{pmatrix} \eta_{Lt} \\ \eta_{St} \\ \eta_{Ct} \end{pmatrix}$$

$$Y^\tau_t = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau}}{\lambda\tau} & \frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \end{pmatrix} \begin{pmatrix} L^d_t - \mu_L \\ S^d_t - \mu_S \\ C^d_t - \mu_C \end{pmatrix} + e^\tau_t,$$

where dynamic factors are the smoothed estimates of L^d_t, S^d_t, and C^d_t from this model, when estimated and conditioned on information on our twelve (subsection 4.1.2) Treasury yields, $Y(\tau)_t$ up until period T.

6. Principal components factors. We calculate these factors from principal component analysis (PCA) of our twelve yields. The first, second, and third factors are interpretable as yield curve factors, L^p_t, S^p_t, and C^p_t.

7. Single partial least squares factors. Partial least squares (PLS) is similar to PCA, but whereas PCA selects factors that explain the largest fraction of the variance of yields, PLS selects factors to be those that explain the largest fraction of the covariance between yields and the dependent variable (in our case NIMs). PLS factors, for which the first three we denote F^1_t, F^2_t, and F^3_t, are less closely related to yield-curve factors. (In contrast to our other factor estimation methods, PLS factors will change as a result of changes to the specification of the NIM equation.)

For all of our models—with the exception of the many yields with forecast combination model—we are faced with a large number of decisions when it comes to specifying our NIM equation. Such decisions include:

i. Whether we should specify our models in levels, like the macro-banking NIM literature, or in first differences, given potentially nonstationary data;

ii. Whether we should generate forecasts that are more than one-quarter ahead with iterative forecasting models or with direct forecasting models;

iii. Whether our models should also condition on contemporaneous yields or yield-curve factors, which is a fairer evaluation of conditional forecast performance, or only condition on lags, since this helps to avoid endogeneity issues;

iv. Whether we should consider only the yield-curve level and slope factors in our models, like the macro-banking NIM literature, or also consider the yield-curve curvature factor, as emphasized in the yield-curve modeling literature;

v. Whether we should estimate single-equation AR models with yields or yield-curve factors or multi-equation VAR models so that we can generate "true conditional projections" as described by Doan, Litterman, and Sims (1984).[14]

   a. If we estimate single-equation AR models, whether we should add all yields or yield-curve factors to the model together or estimate several bi-variate models and from these form a combined forecast; and,

   b. If we estimate multi-equation VAR models, whether we should add all yields or yield-curve factors to the VAR model together or estimate several bi-variate VAR models and from these form a combined forecast.

The above list presents us with a large number of different ways to specify our models. Indeed, the above list leaves us with 48 alternatives for models 3 through 6, an almost as large a number of alternatives for model 7, and a smaller but still notable number of alternatives for model 1. The above list also does not address the issue of lag length, to which we return below.[15]

Although we generated our NIM forecasts for all the alternatives listed above, we editorialize our results by presenting the alternative that generates the better forecast performance and we follow this approach in choosing between the options given in item iv above and some of the options given in item v above. Related to item iv, for almost all of models 3 through 7, models with only two factors or yields (level and slope or 3-month and 10-year

[14]Note that if our NIM models are specified as direct forecasting models it is not possible to then also specify a VAR model from which to generate true conditional projections. Also, it is not clear how we would specify a VAR counterpart to our PLS model.

[15]Note that the above list also presumes that we want to forecast NIMs in one piece and not—as in English (2002) and FRB (2014b)—as subcomponents that we then aggregate up. Clearly, this additional way to forecast NIMs would add even more alternative specifications to the list.

yields) exhibit better forecast performance than models that also include the third factor or yield (curvature or the 2-year yield). For this reason, we only report results from two factor or two yield models. Similarly, related to item v, for almost all of these same models, models in which the two factors or yields enter together exhibit as good or better forecast performance than two bi-variate models from which forecasts are then combined. For this reason, we only report results that add our two factors or yields to the model together. Still related to item v, we find that it is somewhat mixed as to whether single-equation multivariate AR models or multivariate VAR models display better forecast performance. For this reason, we report both sets of results.

Similarly, for item i we present results from both our models in levels and in first differences, despite the fact that our models specified in first differences exhibit notably better forecast performance than our models specified in levels. We include the levels models for the reason that levels models are so prominent in the NIM macro-banking literature. Clearly, in specifying our models in levels we must be careful about the possibility of unit roots in our NIM, yield, and yield-curve factor series and thereby spurious regression results.[16] We address the issue of unit roots in our levels models by always including at least one lag of NIMs and one additional lag of yields (or the yield-curve factors), which is one of the standard approaches for addressing the possibility of spurious regressions. Another standard approach to dealing with the possibility of spurious regressions is to estimate the model in first differences, which is our other specification.

With regard to items ii and iii on our list of alternatives, our preference is to use iterative forecasts and to condition on contemporaneous yields or yield-curve factors. This preference stems from the fact that our iterative models in general have better forecast performance than our direct models and they also seem better suited than direct forecasts to thinking about conditional forecasts. Our preference toward conditioning on contemporaneous yields or yield-curve factors is based on the fact that giving our models contemporaneous information advantages them the most and seems to be the fairer evaluation. That being said, forecast performance does not appear to be much altered by including or not including contemporaneous yields, perhaps reflecting the fact that many bank assets, such as securities,

[16]We do test for the presence of unit roots in our NIMs and yields series, but the tests are inconclusive.

do not reprice, while those with floating rates reprice with some lag. For the most part, we primarily report iterative forecast models and models that condition on contemporaneous yields or yield-curve factors.

There is, however, one practical disadvantage with choosing interative over direct forecasts, which is that a lot of tests and methodologies that have been developed for rigorously assessing forecasts are only applicable for direct forecasting methods. Indeed, one methodology that we find very appealing—specifically, the decomposition developed by Rossi and Sekhposyan (2011) that formally examines *why* one model forecasts better than another—can only be implemented with direct forecasting models and only when current yields or factors are excluded from the model. Since we do want to investigate this question, we also generate some forecast results in which we choose options that will allow us to apply the Rossi and Sekhposyan (2011) decomposition.

To recap our preceding discussion: We consider levels and changes specifications of our NIM models. We primarily report results for iterative forecast models and we primarily report results for models in which yields and yield-curve factors are allowed to enter contemporaneously. We include only two yields or the first two yield-curve factors in our models and we add these to our models together (where this does not apply to our many yields with forecast combination model). We consider single-equation AR and multi-equation VAR variants of our models (where again this does not apply to our many yields with forecast combination model).

On the issue of lag length, we always opt for the short lags since we find that longer lags result in worse forecast performance. This means that our levels models include only one lag of the level of NIMs and only the current observation and one lag of the level of yields or yield-curve factors. Including at least one lag of NIMs and one additional lag of yields or the yield-curve factors in our levels models addresses the possibility of spurious regressions. Our changes models include only the contemporaneous change in yields or yield-curve factors.

4.3 NIM forecasting results over the 2000q1 to 2008q3 evaluation period

To compare the performance of the aggregate NIM models, we consider the root mean squared (forecast) errors (RMSEs) of our models' out-of-sample forecasts and we present results that use 10-year rolling windows for estimation. (We also generated the same results for recursive or expanding windows with analogous results.) The first 10-year rolling window spans 1989q4 to 1999q4 and the assessment window spans 2000q1 to 2008q3. We focus on forecasts that go out 10 steps ahead for the reason that the DFA/CCAR stress tests extend out 9 quarters. In constructing the out-of-sample forecasts, we condition on the observation of the right-hand-side variables in each model, with the exception of lagged NIMs. That is, we condition on the observed value of Treasury yields or factors based on the Treasury yield curve. If lagged NIMs are required, as they are for forecasts two or more steps ahead in our iterative forecast models, we use the NIM forecasts for preceding periods. We use the Diebold-Mariano-West (DMW) test to gauge the quantitative importance of differences in forecast performance across models.

We also generate in-sample forecasts to help understand relative out-of-sample forecast performance and, in doing so, we generate RSMEs for our in-sample forecast errors is the same way that Rossi and Sekhposyan (2011) do to perform their decomposition. That is, for each rolling window model for which we generate an out-of-sample forecast of some horizon (and an associated out-of-sample forecast error), we also generate an in-sample forecast for that same horizon and an associated in-sample forecast error that is right at the end of the sample. More specifically, letting $NIM^f_{m,t+s|t,a_n,z_n}$ denote the s-step ahead forecast of NIMs, jumping off from quarter t and thereby forecasting NIMs in quarter $t+s$ and using model m estimated over the sample quarter a_n to quarter z_n. For our out-of-sample forecasts $t = z_n$ in our above notation. While for our in-sample forecasts $t + s = z_n$.

Simple ocular comparisons of relative in-sample forecast performance across models with relative out-of-sample forecast performance allows us to get some sense of *why* one model may forecast better or worse out-of-sample than another. As noted by Rossi and Sekhposyan (2011), there are two main reasons why one model may forecast better or worse out-of-sample than another. One reason is that the model that forecasts better out-of-sample captures the data over history better and thereby has better in-sample predictive content. The other is that the model that forecasts better out-of-sample is subject to less overfitting in-sample. In general, if a model forecasts worse than another out-of-sample but better in-sample, the likely reason is in-sample overfitting. In contrast, if a model forecasts worse than another both out-of-sample and in-sample, the likely reason is that the model explains the data more poorly. Clearly, this analysis is much less precise than the Rossi and Sekhposyan (2011) decomposition, which we report in appendix A for direct forecasting models.

Figure 2 shows root mean squared errors (RMSEs) for in-sample forecasts (upper panel) and out-of-sample forecasts (lower panel) for our iterative forecasting models specified in levels. Table 1 reports the numbers underlying these results. As is evident from the in-sample results, all of the iterative levels models have lower RMSEs across all forecast horizons than Model 1—the no-change forecast. Note also that the standard deviation of NIMs over the forecast evaluation period is 0.22. This means that at most horizons and for most models the RMSEs for NIMs are smaller than the standard deviation of NIMs, which suggests that the in-sample forecasts of NIMs are helpful in predicting future values of NIMs.

For the most part, the encouraging in-sample results do not translate to good out-of-sample results, with most models forecasting worse than Model 1. Of the nine models that we consider, two models—specifically, Models 2 and 7—forecast worse out-of-sample over all horizons and four models—specifically, Models 3a, 4a, 5, and 6a—forecast better at short horizons and worse at long horizons. Three models—specifically, Models 3b, 4b, and 6b—do have better out-of-sample forecast performance than Model 1 at all horizons; in all cases these are models that are based around VAR specifications that use the Kalman-filter approach of Clarida and Coyle (1984) to generate true conditional projections. The fact that most of our models forecast better than Model 1 in-sample, but either worse or only slightly better out-of-sample suggests that even our best levels models suffer from in-sample overfitting. All of our models seem to be subject to broadly the same degree of overfitting, since relative performance across all models (except the no-change forecast) is quite similar, in-sample and out-of-sample.

Figure 3 and table 2 consider iterative forecasts again, but in this case for models specified in first differences. Figure 3 shows root mean squared errors (RMSEs) for in-sample forecasts (upper panel) and out-of-sample forecasts (lower panel) for the iterative forecasting models specified in changes and table 2 reports the numbers underlying these results. The results from the changes models do differ somewhat from those that we obtained from the models specified in levels. Although the in-sample finding that all models forecast better than the no-change forecast at all horizons carries through with the changes models, the improvement is smaller. Out-of-sample, however, all models continue to forecast better than the no-change model, which is in clear contrast with the findings for the levels models. The finding that the changes models forecast better than the levels models seems in line with the result of Campbell and Perron (1991) that when a unit root is suspected, using a first-difference specification typically leads to better forecast performance.

The models that have the best forecast performance also differ between levels and changes specifications. That is, while among the levels models, it was the VAR-based models with true conditional forecasts—specifically, Models 3b, 4b, and 6b— that yielded the best performing forecasts, among our changes models, it is the multivariate regression models—specifically, Models 3a, 4a, 5, and 6a—that yield the best performing forecasts. Models 2 and 7 generally tend to be the poorest-performing forecasting models, both for the levels and changes specifications.

Notably, however, the results of figures 2 and 3 all show that out-of-sample the forecasts errors from even our best performing models are large relative to the variation in NIMs, which implicitly means that our NIM forecasts explain very little of the variation in NIMs. For our levels models, in particular, the RMSEs of our forecasts at the 6-quarter or more horizon are at least as large as the standard deviation of NIMs over the forecast evaluation period. For our changes models, the RMSEs for some models are a little lower than the standard deviation of NIMs over the forecast evaluation period, but they are nonetheless still quite large.

Even before considering the broader implications of our results, the findings presented in this section offer some practical recommendations for modeling NIMs for stress testing purposes. In particular, the results indicate that although the forecast performance of the models specified in first differences are poor in an absolute sense, they nonetheless deliver notably better forecasts than similar models specified in levels and, as such, strongly suggest that at a minimum NIM models used for stress testing should be specified in first differences (as in Grover and McCracken 2014) rather than in levels, as they have been more-traditionally specified in the banking literature and a few recent stress-testing papers.

5 Summary of robustness analysis

In this section we summarize the robustness analysis that we perform surrounding the results presented in section 4. There are three parts to this analysis: First, an examination of our results over a longer assessment period, specifically 2000q1 to 2013q3; second, the inclusion in our models of some additional variables, such as those that have been suggested by the micro-banking literature and those that seem intuitively reasonable; and, third, an examination of bank-specific NIM forecasts for as many as possible of the 31 BHCs that participate in the DFA/CCAR stress tests. Our discussion of the robustness analysis here is brief, with details of the data, model-adjustment, and results provided in appendices B, C, and D.

For the results presented in section 4, our forecast evaluation period ends in 2008q3 due the complications posed by using the last half-decade of data. Our first robustness analysis—detailed in appendix B—therefore re-examines the results of section 4 for a longer evaluation period, attempting where possible to control for these complications. In practice, this meant adjusting the NIM series that we used to perform our regressions, which we discuss in detail in appendix B.1.

Results for the longer evaluation period are similar to those obtained in our benchmark analysis. That is, forecast performance deteriorates going between the in-sample and out-of-sample results for both our levels and changes models and out-of-sample forecast performance is better for our changes models relative to our levels models (though the difference in forecast performance over the longer sample is much less pronounced). In addition, the forecast errors from even our best performing models are large relative to the variation in NIMs, indicating that over the extended evaluation period NIM forecasts continue to explain little of the variation in NIMs.

The paper's main results focus on models based only on Treasury yields, which are the variables emphasized by the macro-banking NIM literature. The poor forecast performance of our models based only on Treasury yields raises the natural question of whether other variables might be able to improve our forecasts of NIMs and we consider this question in depth in appendix C. The first two variables that we add to our forecasting models in our auxiliary-variables analysis are those emphasized by the micro-banking literature; in particular, the degree of competition facing banks in deposit and loan markets and the volatility of interest rates. We also add the Gilchrist and Zakrajsek (2012) or GZ credit-spread measure, the Freddie Mac median age of a refinanced loan, and the full set of U.S. variables included in the Federal Reserve's DFA/CCAR stress test scenarios.

Our analysis finds that there is some scope to lower RMSEs by including some additional variables—such as measures of competition between banks, the GZ credit-spread measure, and measures of refinancing activity—in our forecasting models. Consequently, our analysis suggests that it might be reasonable to add some of these series to the set of variables published in the Federal Reserve's DFA/CCAR stress test scenarios. Nonetheless, these robustness results continue to show that the forecast errors of even our best performing models are large relative to the variation in NIMs.

Finally, we consider the NIM forecasts of individual BHCs and, in particular, for a group that is as close as possible to those that participate in the Federal Reserve's DFA/CCAR stress tests. In addition to being worthwhile for robustness analysis, considering BHC-specific NIM forecasts has a clear practical motivation, given that ultimately what supervisors require in evaluating BHCs' capital plans is projections of individual BHCs' components of net income, including net interest income. Moreover, BHC-specific NIM forecasts can also be compared against earnings analysts' forecasts of NIMs, which provide a challenging point of comparison since earnings analysts have a great deal of information available to them and also devote a lot of resources to forecasting bank financial-statement variables.

While our analysis on BHC-specific NIM forecasts offers favorable findings for our model-based forecasts in a relative sense—that is, our model-based forecasts are competitive with analysts' forecasts, especially for universal banks—in an absolute sense, our model-based forecasts remain poor. That is, the RMSEs of our model-based forecasts remain large relative to the standard deviation of individual-BHC NIMs, thus indicating that our models ultimately predict very little of the variation of individual-BHC NIMs over time.

6 Implications of our results

We return to the aggregate NIM models of section 4 and consider what their large forecast errors imply for the use of stress tests and *pro forma* capital ratios in providing forward-looking assessment of bank capital adequacy. We first consider what the large forecast errors of our aggregate NIM models imply for the uncertaintly surrounding the NIM projections that would be generated in a stress test. We then ask what the uncertainty surrounding stress test NIM projections would imply for the uncertainty surrounding stress test capital ratio projections—that is, *pro forma* capital ratio projections.

6.1 Implications for stress test NIM projections

To consider the implications of large forecast errors for NIM stress test projections we examine what our best-performing forecast models would imply for the paths of NIMs under the baseline, adverse, and severely adverse scenarios of the 2013 DFA/CCAR stress tests (see, FRB 2012). These scenarios are shown in figure 4. As can be seen from the left side of the figure, the severely adverse scenario—which also features a severe global recession—implies a downward shift in the profile of the yield curve (which at its largest is about 175 b.p. relative to baseline) accompanied by a modest flattening. Such an interest rate scenario is often referred to as a "bull flattener" scenario. As can be seen from the right side of the figure, the adverse scenario—which was motivated by a sudden jump in inflation that accompanies a moderate recession—implies an upward shift of the yield curve (which at its maximum is on the order of about 250 b.p. relative to baseline) as well as a flattening in the slope of the yield curve (which at its maximum implies a yield curve that is about 100 b.p. flatter than baseline). Such an interest rate scenario is often referred to as a "bear flattener" scenario. In terms of changes in the yield curve, these two scenarios capture vastly different interest rate configurations. However, since both involve a flattening of the yield curve, they are both potentially stressful to banks' net interest income.[17]

In considering the effects of the scenarios on NIMs we focus on the dynamic factor model

[17] The bull flattener adverse scenario and the bear flattener severely adverse scenario from the 2013 DFA/CCAR stress tests scenarios are very similar to their counterpart scenarios from the most-recent 2015 DFA/CCAR stress test (see FRB 2014a).

(Model 5) in our changes specification, since out of all of our yield-variables only specifications (that is, the models of subsection 4.3) this model resulted in the lowest RMSEs. Although the 2013 published DFA/CCAR stress test scenarios contained only two Treasury yields—the 3-month and the 10-year yields—whereas our forecast models require many more, we can obtain the paths of all the other yields that we need using our DFM framework and from these calculate the paths of the level and slope of the yield curve consistent with the DFA/CCAR scenarios, which are the variables that we need for our NIM model. The paths of the level and slope of the yield curve consistent with the DFA/CCAR baseline, adverse, and severely adverse scenarios are given in the upper four panels of figure 5. As is evident from these panels, the paths of the level and slope of the yield curve derived from the DFM framework are in line with the scenarios described above.

The lower two panels of figure 5 report the paths of NIMs under the three 2013 DFA/CCAR stress test scenarios given our dynamic factor model specified in changes (Model 5). The paths of NIMs in the baseline and severely adverse scenarios are shown in the lower left panel and the paths of NIMs in the baseline and adverse scenarios are shown in the lower right panel. The paths of the point forecasts appear sensible. The severely adverse scenario is unambiguously unfavorable to NIMs, since the yield curve is flatter (albeit only by a small amount), which reduces returns from maturity transformation, and the yield curve is lower, which reduces returns from the provision of transaction services. This outcome is evident from the lower left panel of figure 5.

The adverse scenario has more ambiguous implications for NIMs, since the yield curve is flatter, which reduces returns from maturity transformation, but is also higher, which boosts returns from the provision of transaction services, since now banks have greater scope to pay deposit rates below that of short-term market rates. The lower right panel of figure 5 does seem to suggest forces pulling in different directions. In particular, NIMs in the adverse scenario are marginally higher than in the baseline over the first half-year of the scenario and lower over subsequent periods.

The lower two panels of figure 5 also consider the uncertainty of the forecasts. In addition to showing point estimates of the paths of NIMs, the figure also shows bands surrounding the central forecast, which are constructed by adding and subtracting to the point forecast the

RMSEs for the dynamic factor model. For both scenarios the one-RMSE uncertainty bands dwarf by five to ten times (depending on the scenario) the variation implied by the scenarios for NIMs. In particular, the path of NIMs in the severely adverse scenario is (on average over the nine-quarters of the stress test's horizon) 0.06 percentage point (p.p.) lower than in the baseline. Similarly, in the adverse scenario, the path of NIMs is (on average) 0.03 p.p. lower. By comparison, the size of a one-RMSE NIM-forecast uncertainty band is 0.30 p.p. As such, the two severe scenarios are statistically indistinguishable from the baseline as well as from each other.[18]

6.2 Implications for stress test capital ratio projections

Since the key variable in the DFA/CCAR stress tests is bank capital—a variable not part of our model—in this subsection we roughly gauge what our NIM forecast errors would imply for uncertainty surrounding the projected path of capital in the stress tests. We base our calculations off of information contained in the stress test disclosure documents (see page 27 of FRB 2014b), which means that they are inherently rough, since the published stress test results do not provide the detail that is needed to make more precise calculations. Indeed, the stress test disclosure documents do not report NIMs. Rather, they report a series called PPNR, which is true PPNR plus income from mortgage servicing rights, losses from operational-risk events, mortgage put-back losses, and other real estate owned losses, where NII is a component of true PPNR.

Based on recent historical data, we estimate that a one RMSE-NIM forecast uncertainty band of 0.30 p.p. would imply a 0.15 p.p. higher or lower level of NIMs, a 4.5 percent higher or lower level of NII, and a 6.7 percent higher or lower path of PPNR (also assuming no change in non-interest income and non-interest expenses). As reported in equation 2, PPNR through the nine quarters of the stress test and in the severely adverse (SA) scenario was $316 billion. This means that PPNR being 6.7 percent higher or lower would amount to PPNR being $22 billion higher or lower over the nine quarters of the scenario.

[18]The results of figure 5 have also been generated from other variants of our models, such as those that also result in relatively low RMSEs when we specified our models in first differences. The result that the paths of NIMs implied by our stress test scenarios are statistically indistinguishable from those implied by the baseline continues to hold.

$$
\begin{aligned}
&\Delta \text{ Regulatory capital} \\
&\quad = \underbrace{\text{Net interest income (NII)} + \text{Non-interest income} - \text{Non-interest expense}}_{\text{PPNR } = \text{ \$316B in severely adverse (SA) scenario}} \\
&\quad \underbrace{+\text{Other revenues}}_{= \text{ \$0B in SA scenario}} \underbrace{-\text{Provisions for loan losses}}_{= \text{ }-\text{\$399B in SA scenario}} \underbrace{-\text{Realized losses on securities}}_{= \text{ }-\text{\$7B in SA scenario}} \\
&\quad \underbrace{-\text{Realized losses on fair-value loans} - \text{Trading and counterparty losses}}_{= \text{ \$29.3B in SA scenario}} \\
&\quad -\text{Taxes} + \text{Extraordinary items net of taxes} \\
&\quad -\text{Net capital distributions to common and preferred shareholders} \\
&\quad \underbrace{-\text{Deductions \& additions to regulatory capital } (\textit{e.g.}, \text{ OCI})}_{= \text{ }-\text{\$12B in SA scenario}} \qquad (2)
\end{aligned}
$$

Holding constant all of the losses, tax payments, and distributions projected for the severely adverse scenario of the stress tests (reported where available in equation 2), this would mean a \$22 billion greater or smaller reduction in capital. Since tier-one common capital in 2013q3 (the starting point of the stress tests) was about \$930 billion or 11.5 percent (as a ratio of risk weighted assets) for banks participating in the stress tests, a \$22 billion greater or smaller reduction in capital would mean a 0.25 p.p. greater or smaller reduction in the resulting path of the capital ratio.

We would argue that 0.25 p.p. greater or smaller level of the capital ratio at the end of the stress test scenario is consequential magnitude. One way to put this into context is to note that 0.5 p.p.—the size of the overall band—is the size of the increments by which the Basel III higher minimum capital requirements were phased in (see, BCBS 2011). Likewise, it is also the size of the increments by which capital surcharges increase as banks get larger in the Basel III framework for calculating the requirements of globally systemically important banks (see, BCBS 2013).

Note also that 0.25 p.p. is the uncertainty around bank capital implied by a conservative one standard deviation forecast error. A two standard deviation forecast error would imply a 0.5 p.p. higher or lower path of capital or a 1.0 p.p. error band. For comparison, we would also note that the interest-rate configurations in the severely adverse and adverse scenarios would—according to our best forecasting model—only lower bank capital ratios on the order of 0.1 p.p. or 0.05 p.p.

7 Conclusion

We consider the ability of an array of time series models to forecast aggregate and individual large-bank net interest margins (NIMs) conditional on interest rates. Among the models that we consider, some of which are similar to the types of models that have been used in the macro-banking literature to study NIMs, only a few deliver forecasts that can improve on a random walk out of sample. Even in such cases, the improvement is marginal. Analogous models that are specified in first differences do deliver forecasts that improve more notably on a random walk—and in some cases also improve on the forecasts for specific banks made by bank earnings analysts—but in an absolute sense, these model forecasts remain quite poor. In particular, these models yield forecast errors that are large and on the same order of magnitude as the variability of NIMs themselves. We also experiment with other specifications, such as models that include variables that have been emphasized in other literatures. Although a few variables—such as the share of assets in the shadow banking sector, the GZ credit spread, and the Freddie Mac median age of a refinanced loan—do result in improved conditional forecast performance, on balance we continue to find large forecast errors.

Our robust finding, of large forecast errors that are on the same order of magnitude as the variability of NIMs themselves, has significant implications for the use of stress tests and *pro forma* capital ratios by regulators in assessing bank capital adequacy. With our best performing NIM models, we show that even stress-test scenarios that have vastly different implications for interest rates have implications for NIMs that are dwarfed by the uncertainty surrounding the different scenarios' point-estimate projections. Moreover, the uncertainty surrounding NIM point-estimate projections translates into a consequential degree of uncertainty surrounding *pro forma* bank capital projections. However, this large degree of forecast uncertainty is not reflected in the current quantitative capital assessment framework.

While our paper has focused on U.S. stress tests and U.S. capital regulation, it is important to note that stress tests and *pro forma* capital ratios are also prominent post-crisis tools for assessing capital adequacy in several other countries and regions. The Bank of England, for example, in the release of its inaugural stress test results, underscored the use of stress test projections for assessing bank capital adequacy in a forward-looking manner (see pages 7 and 8 of BOE 2014). Likewise, the European Banking Authority, in its 2014 stress-test disclosure documents, emphasized the use of stress test results to inform the supervisory responses of relevant national authorities that in many cases it anticipated would involve capital actions (see page 36 of EBA 2014). Although the empirical analysis that we undertook in this paper was confined to the United States, other countries may well be facing the same issues with their capital-ratio projections, and thus the same risk of having their stress tests generate assessments of future bank capital adequacy that lack credibility.[19]

Given the challenges in forecasting NIMs that we have highlighted in this paper and the significant implications of these challenges for how regulators use stress tests and *pro forma* capital ratios, we believe that other options for projecting NIMs should be explored. These options include (i) projecting NIMs using more granular information on the income and expense paths of individual assets and liabilities and (ii) adding adverse shocks to the paths of revenue components projected in stress tests according to their historical volatilities.

Projecting NIMs using more granular information would likely be feasible for interest income, given that stress test variables like charge-offs, loan-loss provisions, and realized gains and losses on securities are already projected at the individual-asset level. Projecting interest expenses may be more challenging, however, since for most types of liabilities similarly granular information is not currently part of the stress-test data collection undertaken by supervisory authorities.[20] Moreover, the evolution of balance sheets would still be important for NIM outcomes, which means that forecasts would likely still be needed.

Adding adverse shocks to the paths of revenue components projected in stress tests according to their historical volatilities would represent a somewhat conceptual change to the process of stress testing, since it would imply that stress tests were no longer purely scenario-based. That being said, imposing shocks on categories of bank incomes and expenses is somewhat akin to the Basel value-at-risk approach to assessing capital adequacy (although Basel only considers credit losses). To the extent, however, that stress tests are the only place in the capital regulatory regime where the potential for bank revenue losses enters into

[19] Stress tests have been used at the height of crises—such as the early 1990s Nordic banking crisis, the U.S. 2008-09 financial crisis, and the euro-area sovereign-debt crisis—to restore confidence in the banking sector. The presence of large forecast errors around bank-capital projections would likely undermine the effectiveness of stress tests in this context as well.

[20] As noted earlier, in the 2015 DFA/CCAR stress tests, one liability expense component of NIMs, namely interest expenses for subordinated debt, began to be forecast using granular information.

the assessment of bank capital adequacy, incrementing macro scenarios with direct shocks to bank revenues may be necessary to enhance stress-test credibility.

References

Alessandri, P. and B. Nelson (2015). Simple Banking: Profitability and the Yield Curve. *Journal of Money, Credit, and Banking 47*(1), 143–175.

Angbazo, L. (1994). Commercial Bank Net Interest Margins, Default Risk, Interest Rate Risk, and Off-balance Sheet Banking. *Journal of Banking and Finance 21*(6), 55–87.

BCBS (2011). Basel III: A global regulatory framework for more resilient banks and banking systems. *BIS Publication*.

BCBS (2013). Globally systemically important banks: Updated assessment methodology and the higher loss absorbency requirement. *BIS Publication*.

BOE (2014). Stress Testing the UK Banking System: 2014 Results. *Bank of England Publication*.

Campbell, J. and P. Perron (1991). Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots. *NBER Macro Annual*, 141–220.

Clarida, R. and D. Coyle (1984). Conditional Projection by Means of Kalman Filtering. Technical Working Paper 36, NBER.

Covas, F., B. Rump, and E. Zakrajsek (2014). Stress-testing U.S. Bank Holding Companies: A Dynamic Panel Quantile Regressions Approach. *International Journal of Forecasting 30*(3), 691–713.

Diebold, F. X. and C. Li (2006). Forecasting the Term Structure of Government Bond Yields. *Journal of Econometrics 130*(2), 337–364.

Diebold, F. X., G. D. Rudebusch, and S. B. Aruba (2007). The Macroeconomy and the Yield Curve: A Dynamic Latent Factor. *Journal of Econometrics 131*(2), 309–338.

Doan, T., R. B. Litterman, and C. A. Sims (1984). Forecasting and Conditional Projection Using Realistic Prior Distributions. *Econometric Reviews 3*(1), 1–100.

EBA (2014). Results of the 2014 EU-wide Stress Tests. *EBA Publication*.

El-Ghazaly, H. and Y. Gopalan (2010). A Jump in Consumer Loans. Economic Synopses 2010-18, Federal Reserve Bank of St. Louis.

English, W. B. (2002). Interest Rate Risk and Bank Net Interest Margins. BIS Quarterly Review 2002-26, Bank for International Settlements.

English, W. B. and W. R. Nelson (1998). Bank Risk Rating of Business Loans. Finance and Economics Discussion Series 1998-51, Board of Governors of the Federal Reserve System.

English, W. B., S. J. van den Heuvel, and E. Zakrajsek (2012). Interest Rate Risk and Bank Equity Valuations. Finance and Economics Discussion Series 2012-26, Board of Governors of the Federal Reserve System.

Flannery, M. (1981). Market Interest Rates and Commercial Bank Profitability: An Empirical Investigation. *Journal of Finance 36*(5), 1085–1101.

FRB (2012). 2013 Supervisory Scenarios for Annual Stress Tests Required under the Dodd-Frank Act Stress Testing Rules and the Capital Plan Rule. Report, Board of Governors of the Federal Reserve System.

FRB (2013). 2014 Supervisory Scenarios for Annual Stress Tests Required under the Dodd-Frank Act Stress Testing Rules and the Capital Plan Rule. Report, Board of Governors of the Federal Reserve System.

FRB (2014a). 2015 Supervisory Scenarios for Annual Stress Tests Required under the Dodd-Frank Act Stress Testing Rules and the Capital Plan Rule. Report, Board of Governors of the Federal Reserve System.

FRB (2014b). Dodd-Frank Act Stress Test 2014: Supervisory Stress Test Methodology and Results. *Report to Congress*.

FRB (2015). Dodd-Frank Act Stress Test 2015: Supervisory Stress Test Methodology and Results. *Report to Congress*.

FRS (2011). Capital Plan Final Rule. *Federal Register 76*(231), 74631–48.

FRS (2012). Supervisory and Company-run Stress Test Requirements for Covered Companies Final Rule. *Federal Register 77*(198), 62378–96.

FRS (2013). Policy Statement on the Scenario Design Framework for Stress Testing. *Federal Register 78*(230), 71435–48.

Gilchrist, S. and E. Zakrajsek (2012). Credit Spreads and Business Cycle Fluctuations. *American Economic Review 102*(4), 1692–1702.

Grover, S. and M. McCracken (2014). Factor-based Prediction of Industry-wide Bank Stress. *Federal Reserve Bank of St. Louis Review*, 173–93.

Guerrieri, L. and M. Welch (2012). 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.

Gurkaynak, R., B. Sack, and J. Wright (2007). The U.S. Treasury Yield Curve: 1961 to the Present. *Journal of Monetary Economics 54*(8), 2291–2304.

Hirtle, B., A. Kovner, and J. Vickery (2013). The Capital and Loss Assessment under Stress Scenarios (CLASS) Model. FRBNY Staff Report 663, Federal Reserve Bank of New York.

Hirtle, B. and A. Lehnert (2014). Supervisory Stress Tests. FRBNY Staff Report 696, Federal Reserve Bank of New York.

Ho, T. S. Y. and A. Saunders (1981). The Determinants of Bank Interest Margins: Theory and Empirical Evidence. *Journal of Financial and Quantitative Analysis 16*(4), 581–600.

Nelson, C. and A. Siegel (1987). Parsimonious Modeling of Yield Curves. *Journal of Business 60*(3), 473–489.

Rossi, B. and T. Sekhposyan (2011). Understanding Models' Forecasting Performance. *Journal of Econometrics 164*(1), 158–172.

Saunders, A. and L. Schumacher (2000). The Determinants of Bank Interest Rate Margins: An International Study. *Journal of International Money and Finance 19*(6), 813–832.

Svensson, L. E. (1994). Estimating and Interpreting Forward Rates: Sweden 1992-4. Working Paper 4871, NBER.

Table 1: RMSEs of Iterative Forecasts, Levels on Levels, and 2000q1 to 2008q3 Evaluation Window

	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Step 10
In-Sample RMSEs										
1. No-Change Forecast	0.08	0.12	0.14	0.17	0.20	0.22	0.23	0.24	0.25	0.27
2. Yields, F. Combination	0.07	0.09	0.11	0.12	0.13	0.14	0.14	0.14	0.13	0.14
3a. 3M and 10Y, Multiv. Reg.	0.06	0.07	0.07	0.07	0.08	0.08	0.08	0.08	0.08	0.09
3b. 3M and 10Y, VAR	0.06	0.06	0.06	0.06	0.06	0.07	0.07	0.06	0.06	0.06
4a. Observed Factors, Multiv. Reg.	0.06	0.07	0.07	0.07	0.07	0.08	0.08	0.08	0.08	0.08
4b. Observed Factors, VAR	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
5. DFM	0.07	0.08	0.08	0.08	0.09	0.09	0.10	0.09	0.10	0.10
6a. PCR, Multiv. Reg.	0.06	0.07	0.07	0.07	0.08	0.08	0.08	0.08	0.08	0.08
6b. PCR, VAR	0.06	0.06	0.06	0.07	0.07	0.07	0.07	0.07	0.07	0.07
7. PLS	0.07	0.09	0.09	0.10	0.11	0.11	0.11	0.11	0.11	0.11
Out-of-Sample RMSEs										
1. No-Change Forecast	0.08	0.12	0.15	0.17	0.20	0.22	0.24	0.25	0.27	0.29
2. Yields, F. Combination	0.08	0.12	0.15	0.18	0.22	0.25	0.28	0.31	0.35*	0.38*
3a. 3M and 10Y, Multiv. Reg.	0.08	0.10	0.11	0.14	0.16	0.19	0.22	0.25	0.28	0.32
3b. 3M and 10Y, VAR	0.09	0.10	0.11	0.14	0.16	0.18	0.20	0.22	0.25	0.28
4a. Observed Factors, Multiv. Reg.	0.08	0.10	0.11	0.14	0.16	0.19	0.22	0.25	0.28	0.32
4b. Observed Factors, VAR	0.09	0.10	0.11	0.14	0.16	0.18	0.20	0.22	0.25	0.28
5. DFM	0.09	0.11	0.14	0.16	0.20	0.23	0.26	0.28	0.31	0.35
6a. PCR, Multiv. Reg.	0.08	0.11	0.12	0.15	0.17	0.20	0.23	0.25	0.29	0.32
6b. PCR, VAR	0.09	0.11	0.13	0.15	0.17	0.19	0.21	0.23	0.26	0.29
7. PLS	0.09	0.12	0.15	0.18	0.21	0.24	0.27	0.30	0.33	0.37

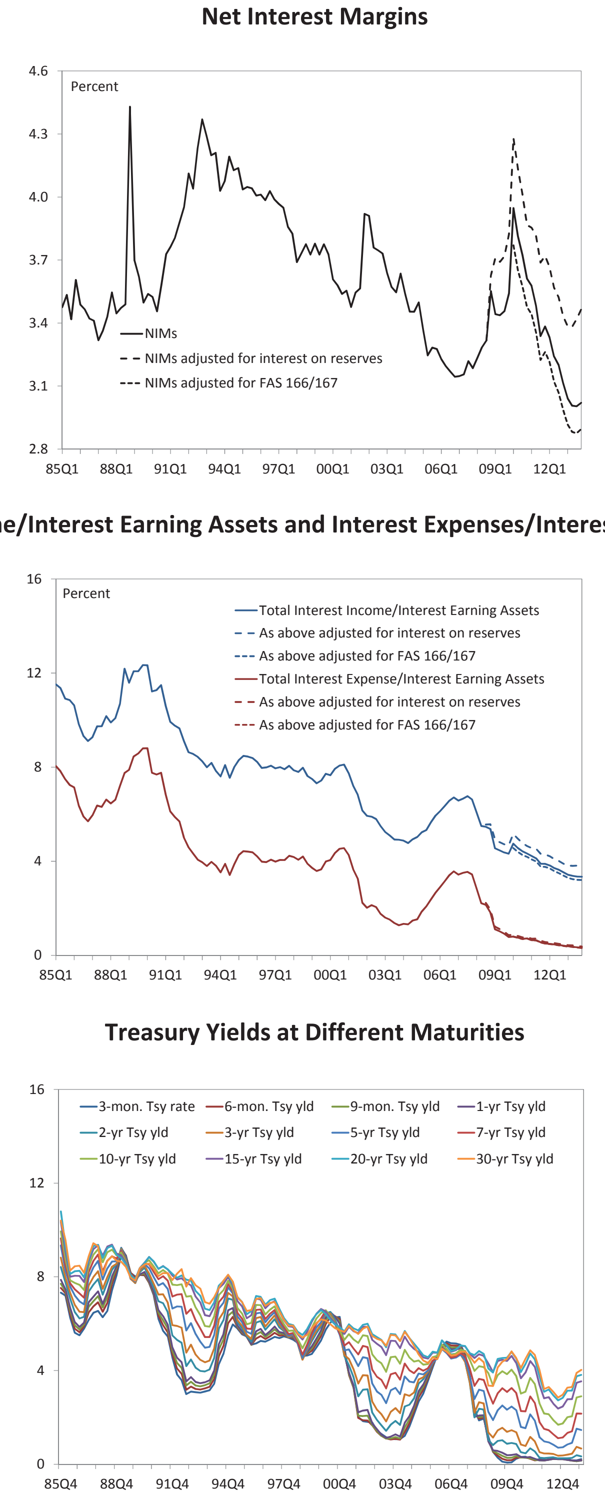
Models 1 through 7 are described in subsection 4.2. An asterisk '*' denotes significance at the 5% level of the difference in RMSEs relative to the no-change forecast using the Diebold-Mariano-West (DMW) test.

Table 2: RMSEs of Iterative Forecasts, Changes on Changes, and 2000q1 to 2008q3 Evaluation Window

	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Step 10
In-Sample RMSEs										
1. No-Change Forecast	0.08	0.12	0.14	0.17	0.20	0.22	0.23	0.24	0.25	0.27
2. Yields, F. Combination	0.08	0.10	0.12	0.13	0.15	0.16	0.16	0.16	0.17	0.18
3a. 3M and 10Y, Multiv. Reg.	0.07	0.09	0.09	0.10	0.11	0.13	0.13	0.13	0.14	0.16
3b. 3M and 10Y, VAR	0.07	0.08	0.08	0.09	0.10	0.12	0.12	0.12	0.12	0.13
4a. Observed Factors, Multiv. Reg.	0.07	0.09	0.09	0.10	0.11	0.13	0.13	0.13	0.14	0.16
4b. Observed Factors, VAR	0.07	0.08	0.08	0.09	0.10	0.12	0.12	0.12	0.12	0.13
5. DFM	0.07	0.09	0.09	0.10	0.12	0.13	0.14	0.14	0.15	0.17
6a. PCR, Multiv. Reg.	0.07	0.09	0.09	0.10	0.11	0.13	0.14	0.14	0.14	0.16
6b. PCR, VAR	0.08	0.11	0.12	0.14	0.16	0.18	0.19	0.19	0.20	0.20
7. PLS	0.08	0.09	0.09	0.10	0.12	0.13	0.14	0.14	0.15	0.16
Out-of-Sample RMSEs										
1. No-Change Forecast	0.08	0.12	0.15	0.17	0.20	0.22	0.24	0.25	0.27	0.29
2. Yields, F. Combination	0.08	0.11	0.13*	0.15*	0.17*	0.18*	0.20	0.20	0.22	0.24
3a. 3M and 10Y, Multiv. Reg.	0.08	0.11	0.11*	0.12*	0.14*	0.15*	0.16*	0.16*	0.19	0.21
3b. 3M and 10Y, VAR	0.09	0.10	0.10*	0.12*	0.14*	0.17*	0.19	0.20	0.23	0.25
4a. Observed Factors, Multiv. Reg.	0.08	0.11	0.11*	0.12*	0.14*	0.15*	0.16*	0.16*	0.18	0.21
4b. Observed Factors, VAR	0.09	0.10	0.10*	0.12*	0.15*	0.17*	0.19	0.21	0.24	0.26
5. DFM	0.08	0.11	0.12	0.13*	0.14*	0.15*	0.16*	0.16*	0.18	0.20
6a. PCR, Multiv. Reg.	0.08	0.11	0.12	0.13*	0.14*	0.15*	0.16*	0.16*	0.18	0.20
6b. PCR, VAR	0.10	0.12	0.15	0.17	0.18	0.20	0.21	0.21	0.22	0.24
7. PLS	0.08	0.10	0.11*	0.12*	0.14*	0.16*	0.18*	0.19	0.21	0.24

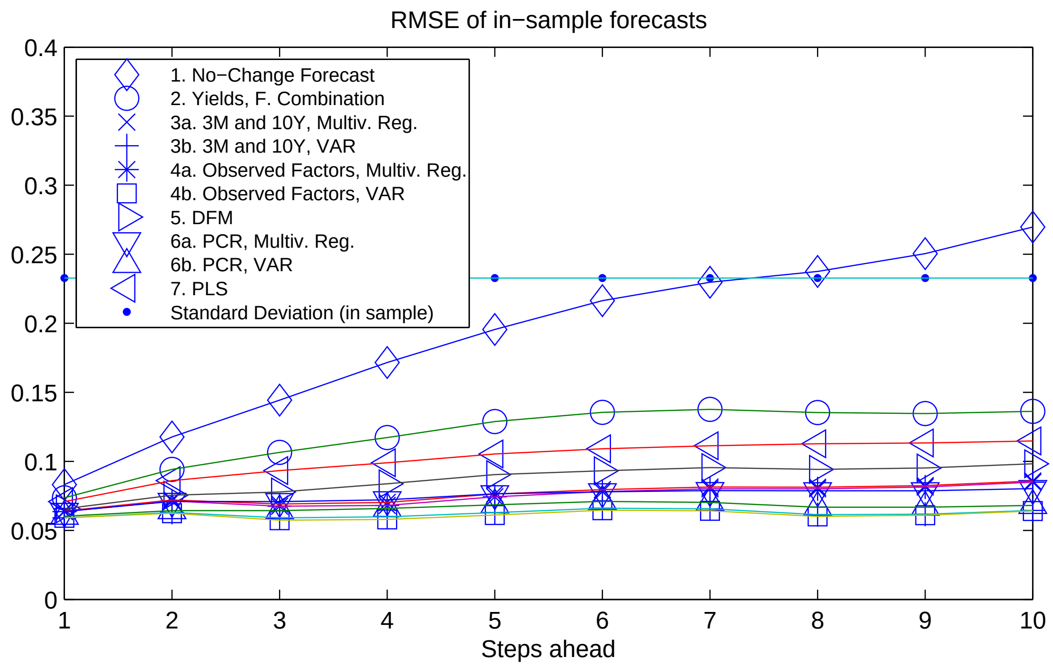
Models 1 through 7 are described in subsection 4.2. An asterisk '*' denotes significance at the 5% level of the difference in RMSEs relative to the no-change forecast using the Diebold-Mariano-West (DMW) test.

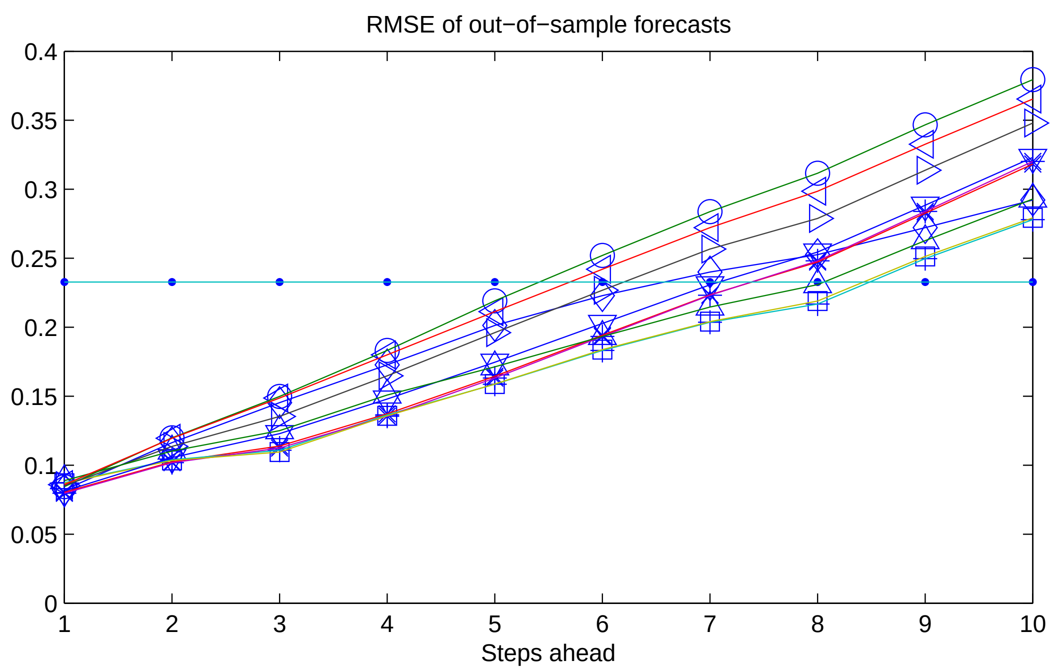
Figure 1: Interest Income, Interest Expenses, and Short-term Treasury Rates



Interest income and interest expenses are an average for the top 25 bank holding companies by assets.

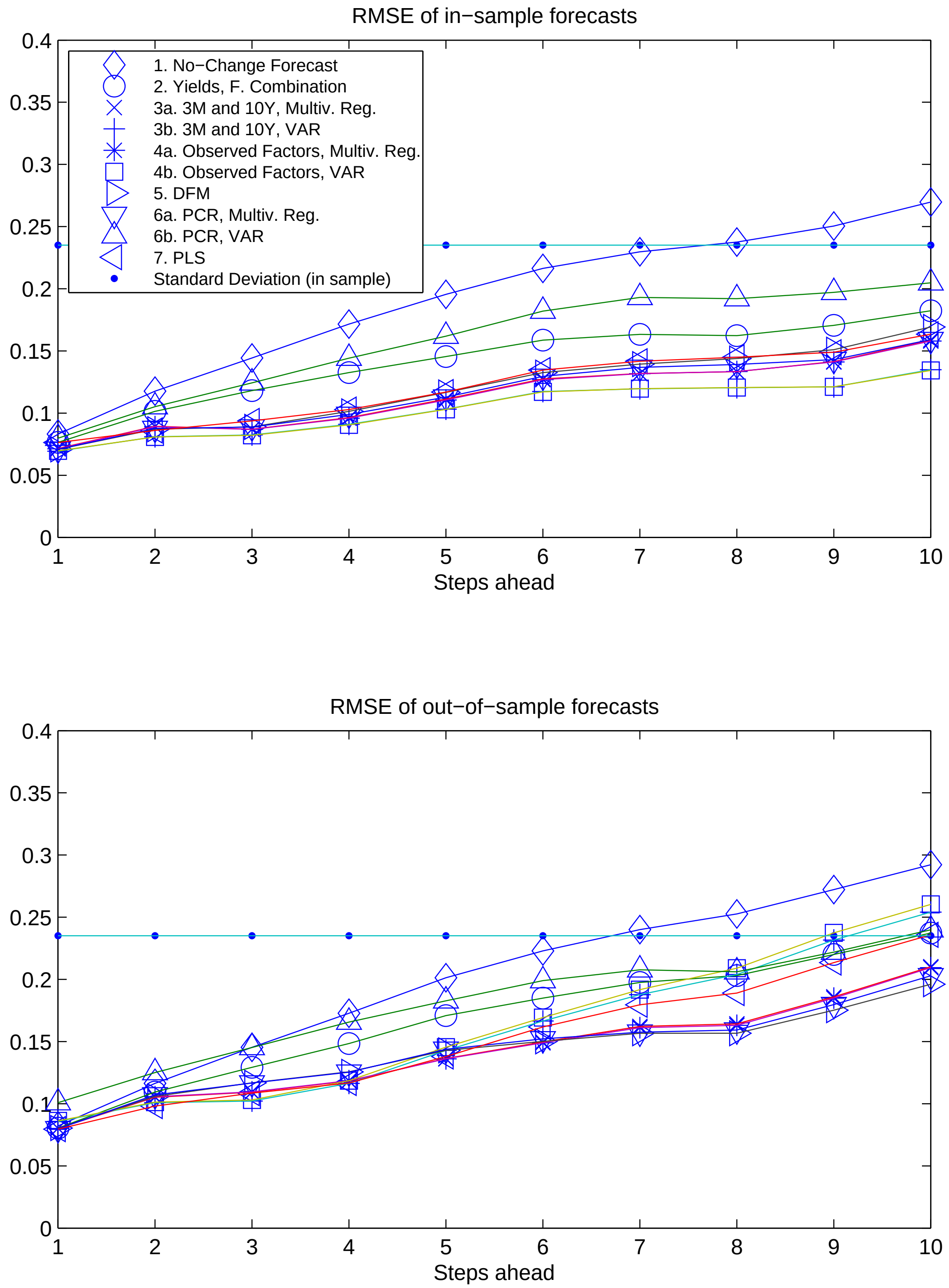
Figure 2: RMSEs of Iterative Forecasts: Levels on Levels, 2000q1 to 2008q3





Models 1 through 7 are described in subsection 4.2.

Figure 3: RMSEs of Iterative Forecasts: Changes on Changes, 2000q1 to 2008q3



Models 1 through 7 are described in subsection 4.2.

Figure 4: Three-month and Ten-year Yields in the 2013 DFA/CCAR Stress Test Scenarios

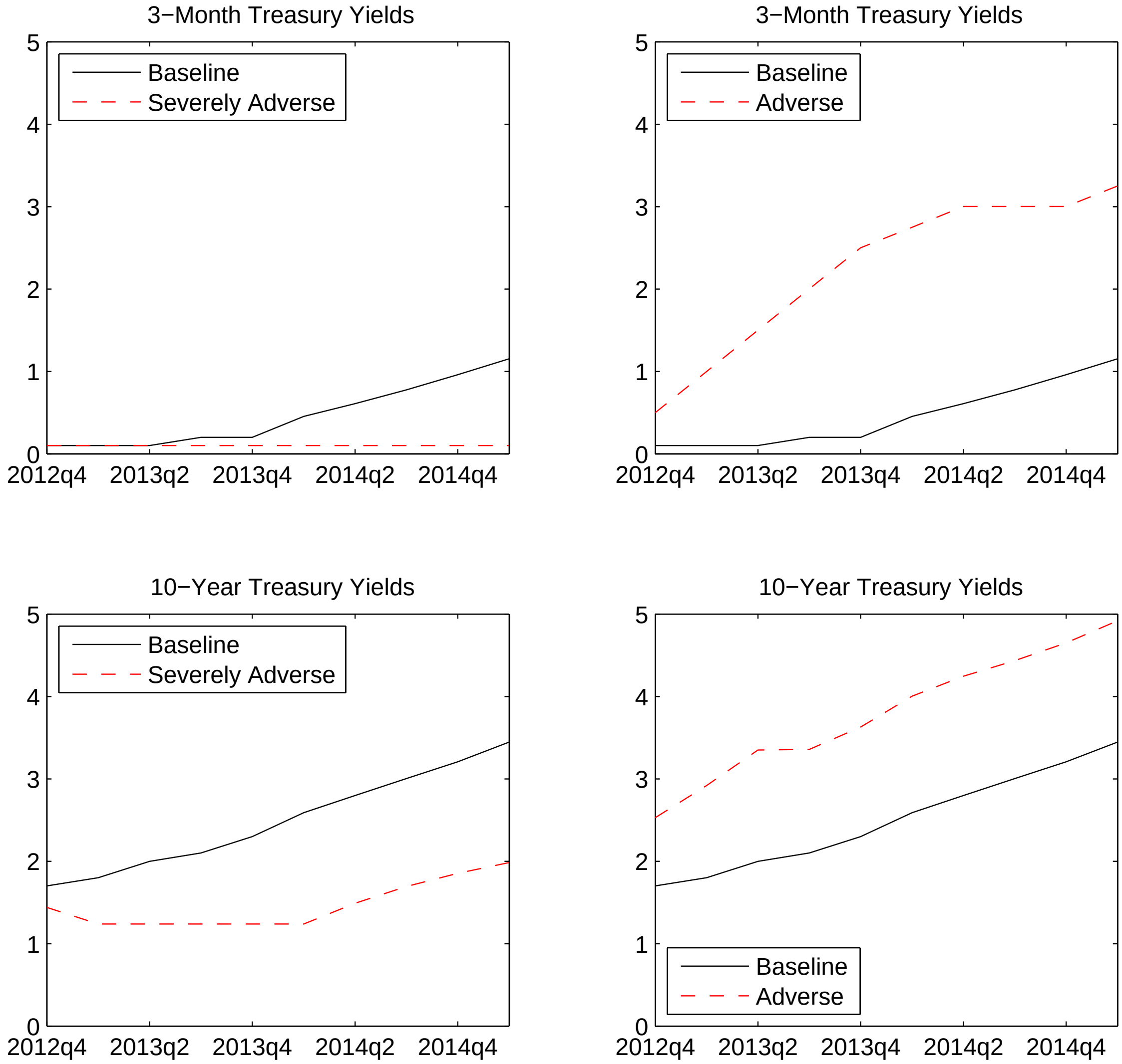
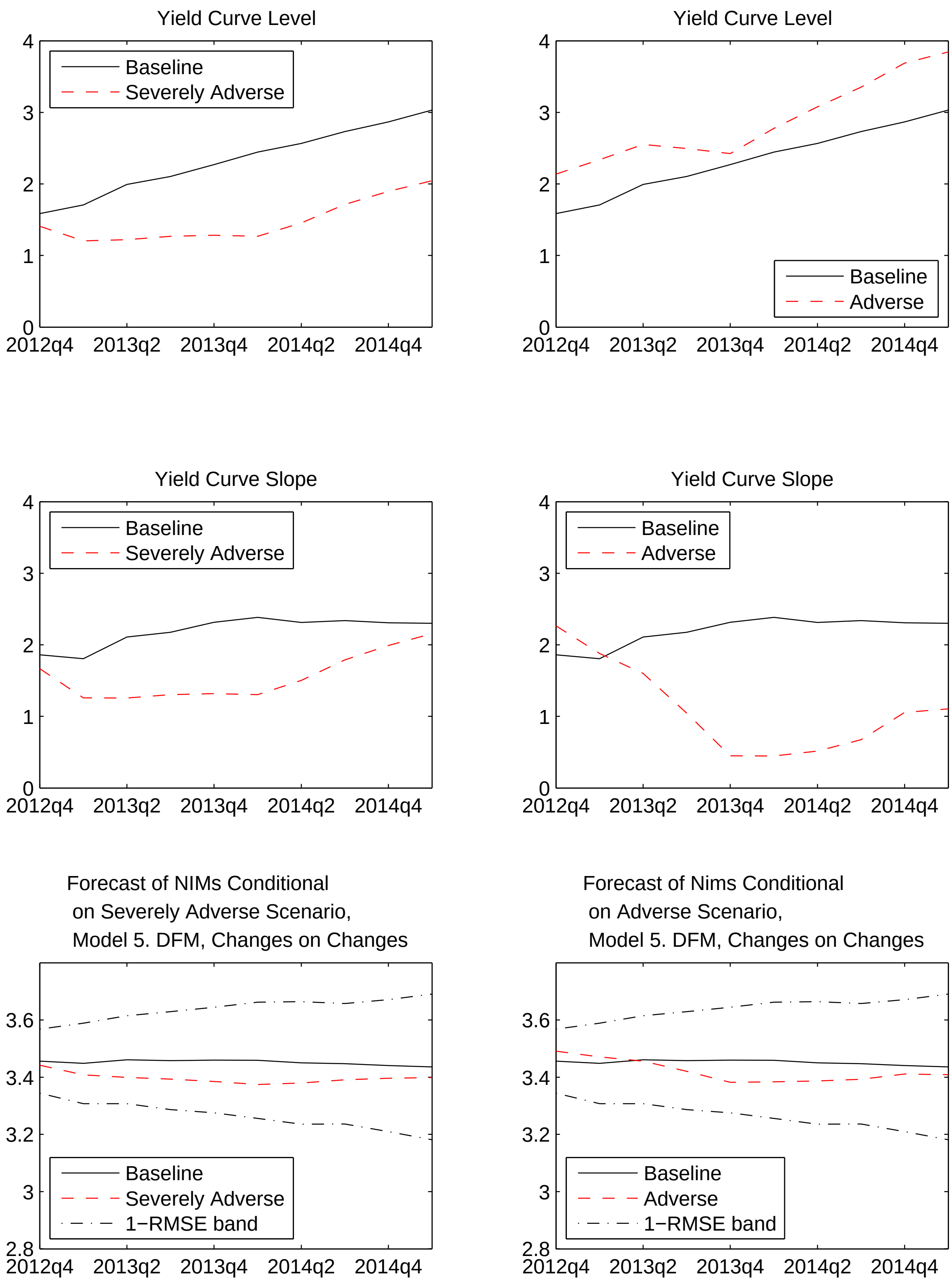


Figure 5: Forecast for NIMs Conditional on the 2013 DFA/CCAR Stress Test Scenarios



Appendix for Online Publication

A Performing the Rossi-Sekhposyan decomposition

A.1 The Rossi-Sekhposyan decomposition

Comparing forecast errors across models tells us *whether* one model forecasts NIMs better or worse than another out-of-sample, but does not tell us *why*. There are two main reasons why a model may forecast better or worse out-of-sample. The first is that the model that forecasts better out-of-sample captures the data over history better and thereby has better in-sample predictive content. The second is that the model that forecasts better out-of-sample suggers less from overfitting in-sample. We can get some sense of the relative importance of these explanations by simple ocular comparisons of relative in-sample forecast performance across models with relative out-of-sample forecast performance, although doing so is imprecise. A more formal approach can be taken by following the methodology developed by Rossi and Sekhposyan (2011). This approach decomposes the out-of-sample loss differential of the DMW test into three independent and interpretable components: marginal predictive content, over-fitting, and time-variation.

To explain how this is done (in loose terms), let $\epsilon^2_{out,m,t+h|t,t+h-1}$ denote the squared h-step ahead conditional forecast error for model m, which is estimated on data extending out to quarter t and for which the forecast is conditional on NIM data out to quarter t, and yields data to quarter $t+h-1$. Also, let $\delta_{out,X,Y,t+h|t,t+h-1} = \epsilon^2_{out,X,t+h|t,t+h-1} - \epsilon^2_{out,Y,t+h|t,t+h-1}$ denote the time series of the out-of-sample h-period ahead conditional forecast loss differential for models X and Y, where $\delta_{out,X,Y,t+h|t,t+h-1}$ is recognizable as the numerator of the DMW test-statistic and is what Rossi and Sekhposyan (2011) seek to understand better. When reporting our results, we call this term $DOSSE$, an abbreviation for difference in out-of-sample squared errors.

Employing the same models that were used to generate each $t+h$-step ahead NIM forecast (from which we then calculated $\epsilon_{out,m,t+h|t,t+h-1}$) we can generate in-sample forecasts, also h periods ahead. We can then calculate the squared h-period ahead conditional forecast error for the *very last* h-quarters of the estimation period—that is, $\epsilon^2_{in,m,t|t-h,t-1}$—and from this we calculate a time-series of in-sample h-step ahead loss differentials: $\delta_{in,X,Y,t|t-h,t-1} = \epsilon^2_{in,X,t|t-h,t-1} - \epsilon^2_{in,Y,t|t-h,t-1}$. When reporting our results we call this term $DISSE$, an abbreviation for difference in in-sample squared errors.

One part of the Rossi and Sekhposyan (2011) decomposition involves (in loose terms) regressing $\delta_{out,X,Y,t+h|t,t+h-1}$ ($DOSSE$) on $\delta_{in,X,Y,t|t-h,t-1}$ ($DISSE$), saving the estimated coefficient, β, from the regression, and then decomposing the full difference in out-of-sample errors, $E[\delta_{out,X,Y,t+h|t,t+h-1}]$—that is, $DOSSE$ and the numerator of the DMW test statistic—into $B = \beta \cdot E[\delta_{in,X,Y,t|t-h,t-1}] = \beta \cdot DISSE$ and $U = E[\delta_{out,X,Y,t+h|t,t+h-1}] - \beta \cdot E[\delta_{in,X,Y,t|t-h,t-1}] = DOSSE - \beta \cdot DISSE$. Rossi and Sekhposyan (2011) show that B and U are independent and also develop test statistics for these two terms—denoted $\Gamma_P^{(B)}$ and $\Gamma_P^{(U)}$—that examine these terms' statistical significance in the same way that the DMW test-statistic examines the significance of $E[\delta_{out,X,Y,t+h|t,t+h-1}]$ ($DOSSE$). These two terms allow one to understand two possible reasons *why* one model forecasts better or worse than the other, which we explain in the context of our results in appendix A.3.

Rossi and Sekhposyan (2011) also consider the time-variation in the out-of sample loss differentials, $\delta_{out,X,Y,t+h|t,t+h-1}$ ($DOSSE$). They also look at averages of $\delta_{out,X,Y,t+h|t,t+h-1}$ ($DOSSE$) over rolling windows and—via the test statistic $\Gamma_P^{(A)}$ that they formulate—provide a method to test the significance of any observed time variation in these rolling-window averages. This part of the decomposition is informative as to whether time variation accounts for relative forecast performance.

Note again that the Rossi and Sekhposyan (2011) decomposition can only be used to compare forecasting performance across direct forecasting models. For iterative models, understanding why one model forecasts better than another still requires the ocular comparison of in-sample and out-of-sample, as we perform in section 4.

A.2 Models for performing the Rossi-Sekhposyan methodology

Generating direct forecasts requires a separate equation to be estimated for each step-ahead forecast. For the levels specification of our yields with forecast combination model, we would have ten equations for each of our twelve yield-maturity equations; that is,

$$NIM_t = c_{\tau,1} + \rho_{\tau,1} \cdot NIM_{t-1} + \sum_{j=0}^{1} \gamma_{\tau,1,j} \cdot Y^{\tau}_{t-j-1} + \eta_t, \text{ for the one-step ahead forecast,}$$

$$NIM_t = c_{\tau,2} + \rho_{\tau,2} \cdot NIM_{t-2} + \sum_{j=0}^{1} \gamma_{\tau,2,j} \cdot Y^{\tau}_{t-j-1} + \eta_t, \text{ for the two-step ahead forecast,}$$

$$\ldots$$

$$NIM_t = c_{\tau,10} + \rho_{\tau,10} \cdot NIM_{t-10} + \sum_{j=0}^{1} \gamma_{\tau,10,j} \cdot Y^{\tau}_{t-j-1} + \eta_t, \text{ for the ten-step ahead forecast,}$$

and likewise for all our other models. Note also that in the above equations we include only lagged variables. This is also a specification requirement for us to be able to perform the Rossi and Sekhposyan (2011) decomposition.

A.3 Rossi-Sekhposyan results

Figure A.1 shows RMSEs for in-sample forecasts (upper panel) and out-of-sample forecasts (lower panel) for our direct forecasting models specified in levels and table A.1 reports the numbers underlying these charts. These figures and tables show fewer model results and, in particular, do not show our VAR forecast models for which we can not generate direct forecasts. For the most part, our direct forecasting model results are similar to those obtained for the iterative forecasting models. In-sample all of our direct levels models have lower RMSEs across all forecast horizons than Model 1—the no-change forecast—while out-of-sample their forecast performance deteriorates. Likewise, Models 2 and 7—like their iterative counterparts—forecast worse out-of-sample than Model 1 over all horizons, while Models 3a, 4a, 5, and 6a—like their iterative counterparts—forecast better at short horizons and worse at long horizons. Among our direct forecasting models, there are none that forecast better out-of-sample over all horizons than Model 1. This, however, reflects the fact that our best performing out-of-sample iterative forecast models were our VAR models that we are unable to use when generating direct forecasts.

As was the case with our iterative models, the fact that our direct models forecast better than Model 1 in-sample but worse out-of-sample suggests that these level models suffer from in-sample overfitting. We can employ Rossi and Sekhposyan's decomposition of relative forecast performance to understand more rigorously why Models 2 through 7 generally forecast worse out-of-sample than Model 1. These results are shown in tables A.2 and A.3.

For each step-ahead forecast and each comparison that we consider, we first report in table A.2 the mean of the difference in in-sample squared errors ($DISSE$) and the mean of the difference in the out-of-sample squared errors ($DOSSE$). When the mean of $DISSE$ and mean of $DOSSE$ are negative, the model listed at the top of the columns in table A.2 has smaller average forecast errors than Model 1. These numbers convey the same information as a comparison of in-sample and out-of-sample RMSEs in table A.1 and figure A.1, the only difference being that the information in table A.1 and figure A.1 shows square-roots of squared errors, whereas $DISSE$ and $DOSSE$ in table A.2 take differences across squared errors. Recall that the term we call $DOSSE$ is the numerator of the DMW test statistic, which is insignificant for all but the comparison between Models 1 and 2 at the 6, 8, and 10 step-ahead forecast horizons, for which Model 1 forecasts (statistically significantly) better. (Table A.3 reports all test statistics.)

As can be seen from table A.2 (as well as table A.1 and figure A.1), Models 3a, 4a, 5, and 6a all forecast better in-sample and out-of-sample than Model 1 at the 2, 4, and 6 step horizons. However, since the mean of B is positive in all these cases—and usually significantly so—while the mean of $DISSE$ is negative, we will get a negative coefficient β when we regress the time series of $DOSSE$ on the time series of DISSE. That is, in-sample performance is a statistically-significant, predictive, *but misleading guide* to out-of-sample performance and therefore it is not because of better in-sample predictive ability of the data that Models 3a, 4a, 5, and 6a forecast better out-of-sample than Model 1. Interestingly, the mean of B is always the opposite sign to the mean of $DISSE$ in the comparisons that we look at in table A.2. The results would seem to suggest that overfitting in-sample is the main reason why one model forecasts better than the other in all of the out-of-sample comparisons that we consider. In general, however, the mean of U—like the mean of $DOSSE$—is not significantly different from zero. Figure A.2 shows the time series and mean of $DOSSE$, B, and U for a comparison of the 2 and 10 step ahead forecasts from Models 1 and 4a. The time series of U clearly accounts for much of the contour of $DOSSE$, but because it swings around sharply it is never significantly different from zero.

Finally, recall that Rossi and Sekhposyan (2011) also test for the significance of any time-variation in $DOSSE$ through a test statistic $\Gamma_P^{(A)}$ that they formulate. This statistic, which is reported in table A.3, finds no significant time variation in relative forecast performance for the combinations of models that we consider.

The general conclusion from the Rossi-Sekhposyan decomposition is that in-sample forecast performance is never unmisleadingly predictive of out-of-sample performance. That is, in-sample forecast performance is either not predictive of out-of-sample performance or it is predictive but in the opposite direction to actual out-of-sample forecast performance.

Table A.1: RMSEs of Direct Forecasts, Levels on Levels, and 2000q1 to 2008q3 Evaluation Window

	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Step 10
In-Sample RMSEs										
1. No-Change Forecast	0.08	0.12	0.14	0.17	0.20	0.22	0.23	0.24	0.25	0.27
2. Yields, F. Combination	0.07	0.10	0.12	0.14	0.16	0.18	0.19	0.19	0.19	0.19
3a. 3M and 10Y, Multiv. Reg.	0.07	0.08	0.09	0.08	0.10	0.10	0.11	0.11	0.11	0.11
4a. Observed Factors, Multiv. Reg.	0.07	0.08	0.09	0.08	0.10	0.10	0.11	0.11	0.11	0.11
5. DFM	0.07	0.08	0.09	0.10	0.12	0.13	0.13	0.12	0.12	0.11
6a. PCR, Multiv. Reg.	0.07	0.08	0.09	0.09	0.10	0.10	0.11	0.10	0.10	0.10
7. PLS	0.07	0.10	0.11	0.12	0.12	0.13	0.13	0.13	0.13	0.13
Out-of-Sample RMSEs										
1. No-Change Forecast	0.08	0.12	0.15	0.17	0.20	0.22	0.24	0.25	0.27	0.29
2. Yields, F. Combination	0.09	0.12	0.16	0.20	0.25	0.29	0.31	0.32*	0.35*	0.37*
3a. 3M and 10Y, Multiv. Reg.	0.09	0.11	0.13	0.15	0.18	0.22	0.26	0.29	0.34	0.36
4a. Observed Factors, Multiv. Reg.	0.09	0.11	0.13	0.15	0.18	0.21	0.26	0.30	0.35	0.38
5. DFM	0.09	0.11	0.14	0.17	0.21	0.25	0.27	0.29	0.33	0.36
6a. PCR, Multiv. Reg.	0.09	0.11	0.14	0.16	0.19	0.22	0.25	0.29	0.33	0.35
7. PLS	0.09	0.13	0.17	0.20	0.23	0.25	0.28	0.31	0.35	0.37

Models 1 through 7 are described in subsection 4.2. An asterisk '*' denotes significance at the 5% level of the difference in RMSEs relative to the no-change forecast using the Diebold-Mariano-West (DMW) test.

Table A.2: Rossi-Sekhposyan Decomposition Against Model 1 and 2000q1 to 2008q3 Evaluation Window

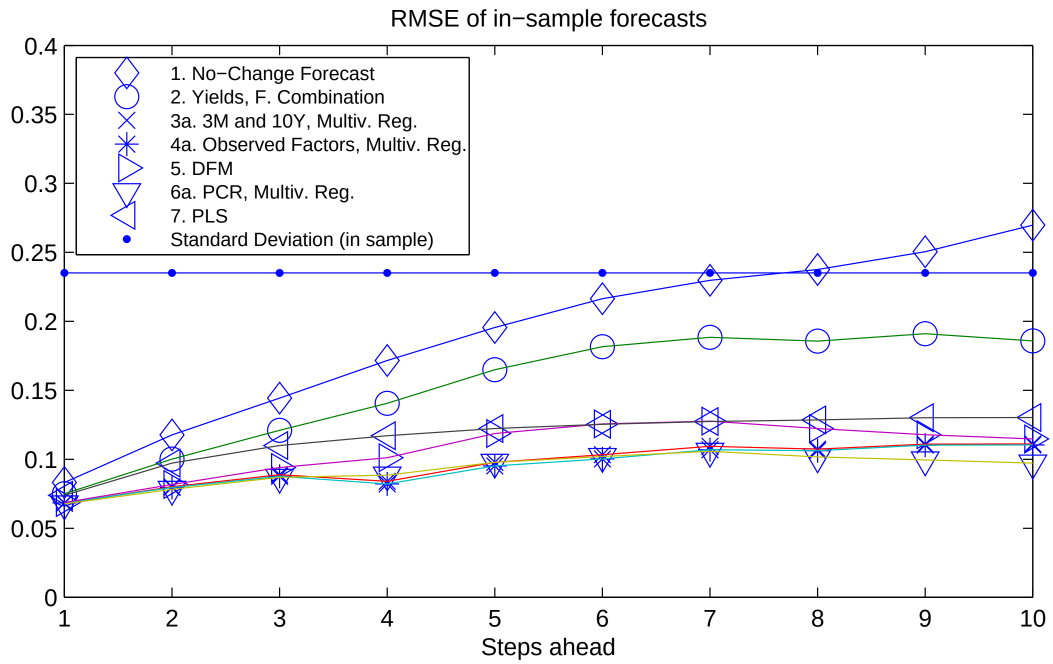
	Mean of	Model 2	Model 3a	Model 4a	Model 5	Model 6a	Model 7
2 steps ahead	DISSE	-0.0037	-0.0074	-0.0074	-0.0071	-0.0076	-0.0043
	DOSSE	0.0017	-0.0022	-0.0022	-0.0011	-0.0017	0.0031
	B	0.0003	0.0035*	0.0034*	0.0035*	0.0028*	0.0009
	U	0.0014	-0.0057*	-0.0056*	-0.0046	-0.0045	0.0022
4 steps ahead	DISSE	-0.0113	-0.0240	-0.0243	-0.0208	-0.0232	-0.0173
	DOSSE	0.0113	-0.0081	-0.0080	-0.0015	-0.0045	0.0096
	B	0.0078	0.0063*	0.0066*	0.0078*	0.0077*	0.0120*
	U	0.0035	-0.0143	-0.0146	-0.0093	-0.0121	-0.0024
6 steps ahead	DISSE	-0.0202	-0.0425	-0.0431	-0.0374	-0.0427	-0.0374
	DOSSE	0.0330	-0.0034	-0.0041	0.0112	-0.0008	0.0152
	B	0.0183*	0.0035*	0.0025*	0.0154*	0.0057*	0.0175*
	U	0.0147	-0.0069	-0.0066	-0.0042	-0.0066	-0.0023
8 steps ahead	DISSE	-0.0284	-0.0513	-0.0516	-0.0479	-0.0525	-0.0463
	DOSSE	0.0414*	0.0227	0.0263	0.0210	0.0186	0.0312
	B	0.0305*	0.0115*	0.0126*	0.0202*	0.0118*	0.0168*
	U	0.0109	0.0113	0.0136	0.0008	0.0069	0.0144
10 steps ahead	DISSE	-0.0339	-0.0561	-0.0563	-0.0553	-0.0590	-0.0515
	DOSSE	0.0484*	0.0472	0.0557	0.0409	0.0364	0.0510
	B	0.0226	0.0118*	0.0134*	0.0126*	0.0070*	0.0114*
	U	0.0259*	0.0354	0.0423	0.0284	0.0295	0.0396

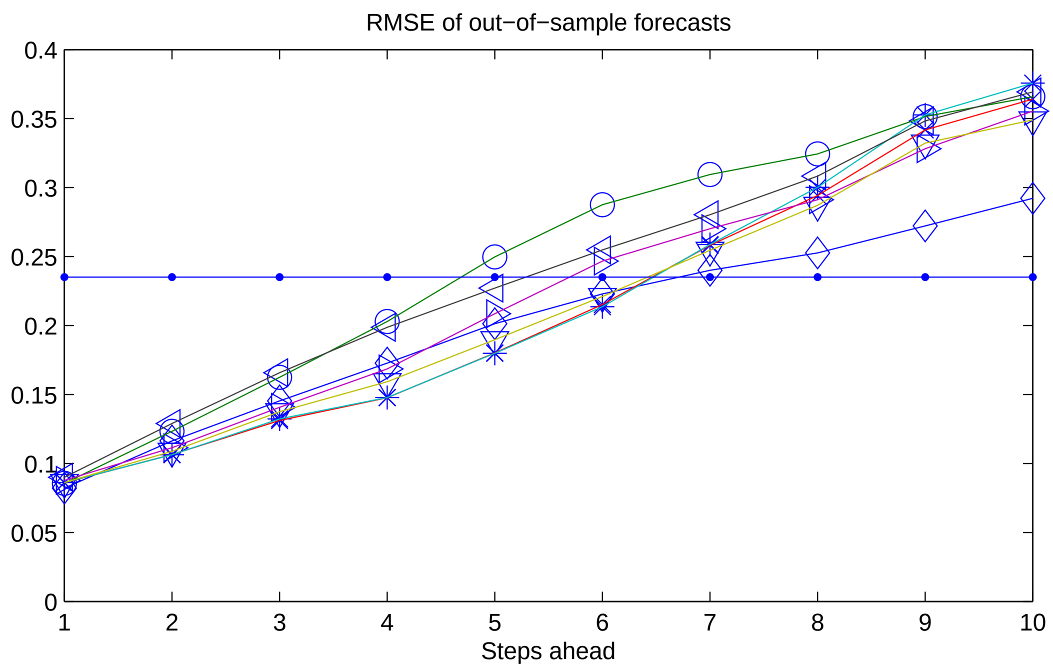
DISSE refers to the difference in in-sample squared errors for each model considered, relative to Model 1. DOSSE refers to the difference in out-of-sample squared errors for each model considered, relative to Model 1. The statistic denoted as B captures the predictable component of the difference in out-of-sample squared errors, DOSSE, based on the difference in in-sample squared errors, DISSE. The statistic denoted as U captures the component of DOSSE orthogonal to DISSE. In the rows that report the mean of DOSSE, the symbol "*" indicates significance at the 5% level of the Diebold-Mariano-West (DMW) test. In the rows that report the mean of B and mean of U, the symbol "*" indicates significance at the 5% level based on the tests in Rossi and Sekhposyan (2011).

Table A.3: Test Statistics for Rossi-Sekhposyan Decomposition Against Model 1 and 2000q1 to 2008q3 Evaluation Window

		Model 2	Model 3a	Model 4a	Model 5	Model 6a	Model 7
2 steps ahead	DMW	0.623	-0.877	-0.868	-0.380	-0.554	0.602
	$\Gamma_P^{(A)}$	7.200	7.840	7.518	7.185	5.865	7.363
	$\Gamma_P^{(B)}$	1.406	2.483*	2.529*	2.194*	2.399*	1.180
	$\Gamma_P^{(U)}$	0.528	-2.698*	-2.545*	-1.819	-1.574	0.431
4 steps ahead	DMW	1.052	-0.882	-0.888	-0.126	-0.453	0.678
	$\Gamma_P^{(A)}$	7.714	7.551	7.649	6.665	7.586	6.336
	$\Gamma_P^{(B)}$	1.699	3.731*	3.812*	3.143*	3.513*	2.364*
	$\Gamma_P^{(U)}$	0.358	-1.595	-1.650	-0.815	-1.259	-0.177
6 steps ahead	DMW	1.686	-0.187	-0.230	0.511	-0.045	0.687
	$\Gamma_P^{(A)}$	6.139	5.885	5.864	5.292	5.694	5.702
	$\Gamma_P^{(B)}$	2.557*	6.087*	6.227*	4.931*	6.098*	5.396*
	$\Gamma_P^{(U)}$	0.808	-0.382	-0.373	-0.194	-0.354	-0.104
8 steps ahead	DMW	2.490*	0.689	0.760	0.691	0.576	0.957
	$\Gamma_P^{(A)}$	5.170	5.808	5.850	5.491	5.694	5.492
	$\Gamma_P^{(B)}$	2.398*	3.981*	4.007*	3.593*	3.892*	3.833*
	$\Gamma_P^{(U)}$	1.018	0.343	0.397	0.027	0.214	0.445
10 steps ahead	DMW	3.188*	0.988	1.097	0.838	0.745	1.074
	$\Gamma_P^{(A)}$	4.552	5.421	5.454	5.213	5.393	4.910
	$\Gamma_P^{(B)}$	1.686	2.495*	2.507*	2.397*	2.461*	2.475*
	$\Gamma_P^{(U)}$	3.604*	0.744	0.837	0.584	0.603	0.837

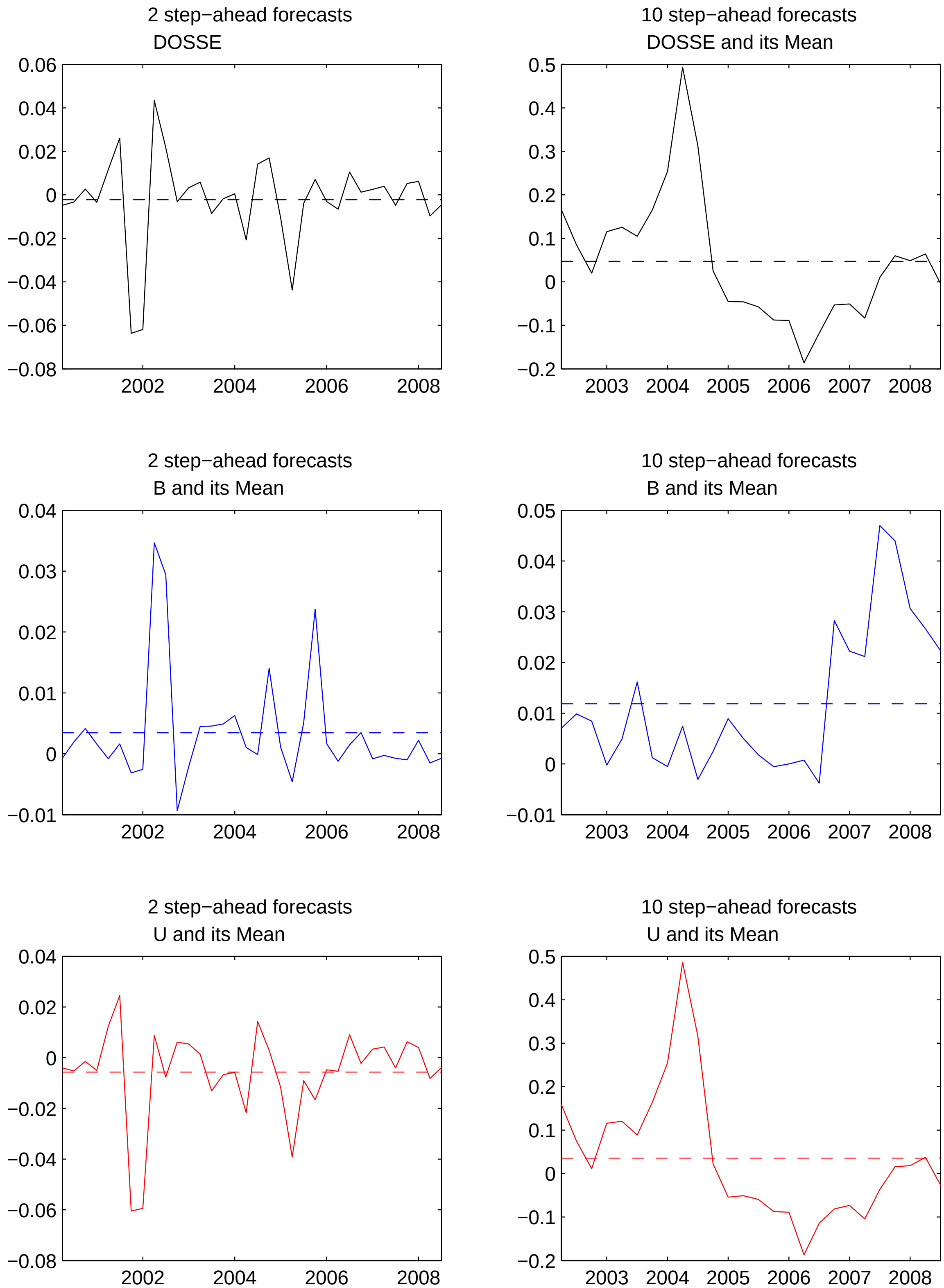
Figure A.1: RMSEs of Direct Forecasts: Levels on Levels, 2000q1 to 2008q3





Models 1 through 7 are described in subsection 4.2.

Figure A.2: Rossi-Sekhposyan Decomposition: Models 4a against Model 1, Levels on Levels, 2000q1 to 2008q3



Model 4a is described in subsection 4.2.

B Robustness analysis: Extending the evaluation period to 2013q4

For the paper's baseline results we used a forecast evaluation period that ended in 2008q3. This was due to the large number of complications posed by using the last half-decade of data, including the zero interest rate environment that began in 2008q4, the Fed starting to pay interest on excess reserves in 2008q4, Financial Accounting Statements (FAS) 166/167 coming into effect at the beginning of 2010, and the full repeal of Reg. Q by the DFA. In this robustness appendix, we re-examine the results of section 4 for a longer evaluation period, attempting where possible to control for the complications of the last half decade.

B.1 Data for the evaluation period extending to 2013q4

The aggregate NIM data that we use for our analysis in which the evaluation period extends to 2013q4 comes from the same source as the data used for the analysis extending up to 2008q3 described in subsection 4.1.1. However, where possible, we try to account for the important structural changes that occurred in the banking industry over the last half-decade that have implications for measurement of NIMs. When performing our forecast analysis, however, we consider results when we both control and do not control for these changes in the banking industry.

The first important structural change in the longer sample period was the Fed starting to pay, in the fourth quarter of 2008, interest on excess reserves (IOER). This change meant that excess reserves, which were previously not part of interest earning assets, became part of interest earning assets, so increasing the denominator of NIMs, while barely altering the numerator given the very low interest paid on excess reserves. The amount by which this structural change pushes down NIMs can be seen by comparing the black solid and black dashed lines in the top panel of figure 1. We also show in the middle panel, how the interest income and expense components of NIMs change as a result of this adjustment.

The second relevant change over the latter part of our sample was Financial Accounting Statements (FAS) 166/167 coming into effect at the beginning of 2010. These new standards resulted in banks needing to bring nearly \$362 billion in assets and liabilities back onto their balance sheets, of which nearly 90 percent were consumer loans (see El-Ghazaly and Gopalan 2010). Of these consumer loans, a sizable fraction were credit cards, which have higher interest margins than other types of loans. As a result, NIMs jumped notably in the first quarter of 2010, as can be seen by comparing the black solid and black dotted lines in the top panel of figure 1. We also show in the middle panel how the interest income and expense components of NIMs change as a result of the FAS 166/167 adjustment. The NIM series that we use when we extend our analysis out to the end of 2013 combines the two adjustments shown in the top panel of figure 1.

The final change that we note that occurred in the latter part of our sample was the full repeal of Reg. Q by the DFA. While Reg. Q celings for savings accounts and most other types of deposit accounts were phased out during the early- to mid-1980s, they remained in effect for demand deposit accounts until they were repealed by Section 627 of the DFA. This change, for which we do not make any adjustments, could be a complication in the modeling of NIMs. That said, given the low interest rate environment, it likely represents a relatively small complication to date.

The Treasury yield data that we use in this appendix comes from the same source as the data used for the analysis extending up to 2008q3 described in subsection 4.1.2.

B.2 Models for the evaluation period extending to 2013q4

The models for forecasting NIMs used in this appendix are the same as those used for the analysis in which the assessment window extended to 2008q3 and described in subsection 4.2. To be sure, the zero interest rate environment does pose challenges to models of interest rates that extract factors from a large set of yields; in our case, Models 5 and 6. We do not make any specific adjustments to our models to address the zero interest-rate environment, though we remain mindful of these challenges when interpreting our results.

B.3 Results for the evaluation period extending to 2013q4

Tables B.1 and B.2 are the 2000q1 to 2013q4 counterparts for tables 1 and 2 when we use our IOER and FAS 166/167 adjusted NIM series. Our results are, however, not greatly affected by these adjustments. As can be seen from tables B.3 and B.4, when we do not adjust for these changes our results are barely changed.

Results for the longer evaluation period are broadly similar to those obtained for the shorter evaluation period. That is, forecast performance deteriorateswhen we go from the in-sample to the out-of-sample results for both our levels and changes specifications, out-of-sample forecast performance is better for our changes specifications relative to our levels specifications (though the difference in forecast performance over the longer sample is less marked), and the forecast errors from even our best performing models are large relative to the variation in NIMs, indicating that over the extended evaluation period NIM forecasts continue to explain little of the variation in NIMs.

Table B.1: RMSEs of Iterative Forecasts, Levels on Levels, and 2000q1 to 2013q3 Evaluation Window

	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Step 10
In-Sample RMSEs										
1. No-Change Forecast	0.09	0.13	0.16	0.20	0.23	0.26	0.28	0.30	0.32	0.34
2. Yields, F. Combination	0.08	0.11	0.12	0.13	0.14	0.15	0.15	0.15	0.15	0.15
3a. 3M and 10Y, Multiv. Reg.	0.07	0.08	0.09	0.10	0.10	0.10	0.10	0.10	0.11	0.11
3b. 3M and 10Y, VAR	0.06	0.08	0.08	0.09	0.09	0.09	0.09	0.09	0.09	0.09
4a. Observed Factors, Multiv. Reg.	0.07	0.08	0.09	0.09	0.10	0.10	0.10	0.10	0.10	0.10
4b. Observed Factors, VAR	0.06	0.08	0.08	0.08	0.09	0.09	0.09	0.09	0.09	0.09
5. DFM	0.06	0.08	0.09	0.09	0.10	0.10	0.10	0.10	0.10	0.10
6a. PCR, Multiv. Reg.	0.07	0.09	0.10	0.10	0.11	0.11	0.11	0.11	0.11	0.11
6b. PCR, VAR	0.07	0.08	0.09	0.09	0.10	0.10	0.10	0.10	0.10	0.10
7. PLS	0.07	0.10	0.11	0.12	0.12	0.13	0.13	0.13	0.13	0.13
Out-of-Sample RMSEs										
1. No-Change Forecast	0.09	0.13	0.16	0.20	0.23	0.27	0.30	0.32	0.35	0.37
2. Yields, F. Combination	0.09	0.13	0.17	0.21	0.25	0.28	0.32	0.35	0.38	0.41
3a. 3M and 10Y, Multiv. Reg.	0.08	0.11	0.13*	0.16*	0.19*	0.21	0.24	0.26	0.29	0.32
3b. 3M and 10Y, VAR	0.09	0.12	0.14	0.17*	0.19*	0.22	0.24	0.26	0.29	0.32
4a. Observed Factors, Multiv. Reg.	0.08	0.11	0.13*	0.16*	0.19*	0.21	0.24	0.26	0.29	0.32
4b. Observed Factors, VAR	0.09	0.12	0.14*	0.17*	0.19*	0.22	0.24	0.26	0.29	0.32
5. DFM	0.08	0.12	0.14	0.17	0.19	0.22	0.24	0.26	0.29	0.31
6a. PCR, Multiv. Reg.	0.09	0.12	0.15	0.18	0.20	0.23	0.25	0.27	0.30	0.32
6b. PCR, VAR	0.09	0.13	0.16	0.19	0.21	0.23	0.25	0.27	0.30	0.32
7. PLS	0.09	0.14	0.17	0.21	0.24	0.26	0.29	0.31	0.34	0.36

Models 1 through 7 are described in subsection 4.2. An asterisk '*' denotes significance at the 5% level of the difference in RMSEs relative to the no-change forecast using the Diebold-Mariano-West (DMW) test.

Table B.2: RMSEs of Iterative Forecasts, Changes on Changes, and 2000q1 to 2013q3 Evaluation Window

	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Step 10
In-Sample RMSEs										
1. No-Change Forecast	0.09	0.13	0.16	0.20	0.23	0.26	0.28	0.30	0.32	0.34
2. Yields, F. Combination	0.08	0.12	0.14	0.17	0.19	0.21	0.21	0.22	0.23	0.25
3a. 3M and 10Y, Multiv. Reg.	0.08	0.11	0.12	0.13	0.15	0.16	0.17	0.18	0.19	0.20
3b. 3M and 10Y, VAR	0.07	0.09	0.11	0.12	0.14	0.15	0.16	0.17	0.18	0.19
4a. Observed Factors, Multiv. Reg.	0.07	0.10	0.12	0.13	0.15	0.16	0.17	0.18	0.19	0.20
4b. Observed Factors, VAR	0.07	0.09	0.11	0.12	0.13	0.15	0.16	0.17	0.18	0.19
5. DFM	0.07	0.10	0.12	0.13	0.15	0.16	0.18	0.19	0.20	0.21
6a. PCR, Multiv. Reg.	0.08	0.11	0.12	0.14	0.15	0.17	0.18	0.19	0.19	0.20
6b. PCR, VAR	0.08	0.12	0.15	0.18	0.21	0.24	0.25	0.26	0.28	0.29
7. PLS	0.08	0.10	0.12	0.13	0.14	0.16	0.17	0.18	0.18	0.20
Out-of-Sample RMSEs										
1. No-Change Forecast	0.09	0.13	0.16	0.20	0.23	0.27	0.30	0.32	0.35	0.37
2. Yields, F. Combination	0.09	0.13	0.16	0.19	0.23	0.26	0.29	0.32	0.35	0.38
3a. 3M and 10Y, Multiv. Reg.	0.08	0.12	0.14*	0.16*	0.18*	0.21*	0.24*	0.26*	0.28*	0.31*
3b. 3M and 10Y, VAR	0.09	0.11	0.13*	0.15*	0.18*	0.20*	0.22*	0.24*	0.27	0.30
4a. Observed Factors, Multiv. Reg.	0.08	0.12	0.14*	0.16*	0.18*	0.21*	0.23*	0.25*	0.28*	0.30*
4b. Observed Factors, VAR	0.09	0.11	0.13*	0.15*	0.18*	0.20*	0.23*	0.25*	0.27	0.30
5. DFM	0.08	0.12	0.14*	0.16*	0.18*	0.20*	0.23*	0.24*	0.27*	0.29*
6a. PCR, Multiv. Reg.	0.08	0.12	0.15*	0.17*	0.19*	0.22*	0.24*	0.26*	0.28*	0.31*
6b. PCR, VAR	0.10	0.14	0.17	0.21	0.25	0.29	0.32	0.35	0.39	0.42
7. PLS	0.09	0.12*	0.14*	0.16*	0.18*	0.21*	0.24*	0.26*	0.28*	0.31*

Models 1 through 7 are described in subsection 4.2. An asterisk '*' denotes significance at the 5% level of the difference in RMSEs relative to the no-change forecast using the Diebold-Mariano-West (DMW) test.

Table B.3: No Adjustments for Interest on Reserves and FAS 166/167: RMSEs – Iterative Forecasts, Regressions of Levels on Levels, and 2000q1 to 2013q3 Evaluation Window

	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Step 10
In-Sample RMSEs										
1. No-Change Forecast	0.09	0.13	0.17	0.20	0.22	0.24	0.26	0.27	0.29	0.30
2. Yields, F. Combination	0.08	0.11	0.12	0.13	0.13	0.14	0.13	0.14	0.14	0.14
3a. 3M and 10Y, Multiv. Reg.	0.07	0.08	0.09	0.09	0.09	0.10	0.10	0.10	0.10	0.10
3b. 3M and 10Y, VAR	0.07	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
4a. Observed Factors, Multiv. Reg.	0.07	0.08	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.10
4b. Observed Factors, VAR	0.07	0.07	0.07	0.08	0.08	0.08	0.08	0.08	0.08	0.08
5. DFM	0.07	0.08	0.09	0.09	0.09	0.09	0.10	0.10	0.10	0.10
6a. PCR, Multiv. Reg.	0.07	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.10
6b. PCR, VAR	0.07	0.08	0.08	0.08	0.09	0.09	0.09	0.09	0.09	0.09
7. PLS	0.08	0.09	0.10	0.11	0.11	0.11	0.11	0.11	0.11	0.12
Out-of-Sample RMSEs										
1. No-Change Forecast	0.09	0.13	0.17	0.20	0.22	0.25	0.27	0.29	0.31	0.33
2. Yields, F. Combination	0.10	0.14	0.17	0.20	0.22	0.25	0.28	0.30	0.33	0.35
3a. 3M and 10Y, Multiv. Reg.	0.09	0.12	0.14*	0.16*	0.18*	0.21	0.23	0.25	0.28	0.31
3b. 3M and 10Y, VAR	0.10	0.13	0.14	0.16*	0.18*	0.21*	0.23	0.25	0.28	0.30
4a. Observed Factors, Multiv. Reg.	0.09	0.12	0.14*	0.16*	0.18*	0.20	0.23	0.25	0.28	0.31
4b. Observed Factors, VAR	0.10	0.12	0.14*	0.16*	0.18*	0.21*	0.23	0.25	0.28	0.30
5. DFM	0.09	0.13	0.15	0.17	0.19	0.22	0.24	0.26	0.28	0.31
6a. PCR, Multiv. Reg.	0.09	0.12	0.14	0.16*	0.18	0.20	0.22	0.24	0.27	0.29
6b. PCR, VAR	0.10	0.13	0.14	0.17*	0.18*	0.20*	0.22	0.23	0.26	0.29
7. PLS	0.10	0.13	0.16	0.19	0.21	0.23	0.26	0.28	0.31	0.33

Models 1 through 7 are described in subsection 4.2. An asterisk '*' denotes significance at the 5% level of the difference in RMSEs relative to the no-change forecast using the Diebold-Mariano-West (DMW) test.

Table B.4: No Adjustments for Interest on Reserves and FAS 166/167: RMSEs – Iterative Forecasts, Regressions of Changes on Changes, and 2000q1 to 2013q3 Evaluation Window

	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Step 10
In-Sample RMSEs										
1. No-Change Forecast	0.09	0.13	0.17	0.20	0.22	0.24	0.26	0.27	0.29	0.30
2. Yields, F. Combination	0.09	0.12	0.15	0.17	0.18	0.19	0.19	0.19	0.20	0.21
3a. 3M and 10Y, Multiv. Reg.	0.09	0.12	0.13	0.14	0.15	0.16	0.16	0.16	0.17	0.17
3b. 3M and 10Y, VAR	0.08	0.11	0.12	0.13	0.14	0.15	0.15	0.16	0.16	0.17
4a. Observed Factors, Multiv. Reg.	0.09	0.11	0.13	0.14	0.14	0.16	0.16	0.16	0.16	0.17
4b. Observed Factors, VAR	0.08	0.11	0.12	0.13	0.14	0.15	0.15	0.15	0.16	0.17
5. DFM	0.08	0.11	0.13	0.14	0.14	0.15	0.16	0.16	0.17	0.18
6a. PCR, Multiv. Reg.	0.09	0.11	0.13	0.14	0.14	0.15	0.16	0.16	0.16	0.17
6b. PCR, VAR	0.09	0.12	0.15	0.18	0.19	0.21	0.22	0.23	0.24	0.25
7. PLS	0.09	0.11	0.13	0.14	0.14	0.16	0.16	0.16	0.16	0.17
Out-of-Sample RMSEs										
1. No-Change Forecast	0.09	0.13	0.17	0.20	0.22	0.25	0.27	0.29	0.31	0.33
2. Yields, F. Combination	0.09	0.13	0.16	0.19	0.21	0.24	0.26	0.28	0.30	0.32
3a. 3M and 10Y, Multiv. Reg.	0.09	0.13	0.15	0.17*	0.18*	0.20*	0.22*	0.23*	0.25*	0.27
3b. 3M and 10Y, VAR	0.10	0.13	0.14	0.16*	0.17*	0.19*	0.21*	0.22	0.25	0.27
4a. Observed Factors, Multiv. Reg.	0.09	0.13	0.15	0.17*	0.18*	0.20*	0.21*	0.23*	0.25*	0.27
4b. Observed Factors, VAR	0.10	0.12	0.14	0.16*	0.18*	0.20*	0.21*	0.23	0.26	0.28
5. DFM	0.09	0.13	0.16	0.17	0.18*	0.19*	0.20*	0.21*	0.24*	0.26
6a. PCR, Multiv. Reg.	0.09	0.13	0.15	0.17*	0.18*	0.20*	0.21*	0.22*	0.25*	0.26
6b. PCR, VAR	0.11	0.14	0.17	0.20	0.23	0.26	0.29	0.31	0.34	0.36
7. PLS	0.10	0.12	0.14*	0.16*	0.17*	0.20*	0.21*	0.23*	0.25*	0.27

Models 1 through 7 are described in subsection 4.2. An asterisk '*' denotes significance at the 5% level of the difference in RMSEs relative to the no-change forecast using the Diebold-Mariano-West (DMW) test.

C Robustness analysis: Adding auxiliary variables

For the paper's main results, we focused on models that only use Treasury yields because those are the variables emphasized by the macro-banking NIM literature. Of course, the poor forecast performance of those models naturally raises the question of whether variables—such as those emphasized in the micro-banking NIM literature or, indeed, any other plausible variables—might be able to improve on our forecasts of NIMs. In this robustness appendix, we consider this question over the evaluation period 2000q1 to 2008q3.

C.1 Data for the auxiliary variables

The first two variables that we add to our forecasting models in our auxiliary variables analysis are those emphasized by the micro-banking literature; in particular, the degree of competition facing banks in deposit and loan markets and the volatility of interest rates.

The degree of competition that banks face is captured by the relative size of the shadow banking industry. This is measured by the volume of assets in the shadow banking industry relative to the size of the total banking industry, that is, both traditional and shadow banking (as measured by the combined assets of these two industries). We prefer, however, to use the size of the traditional banking industry as our measure of competition, rather than measures of competition within the banking sector itself, like Herfindahl-Hirschman indices. This is for the reason that over the time period that we consider competition from the shadow banking sector is the main source of increased competition faced by the traditional banking sector that has compressed interest margins. Importantly, competition from the shadow banking sector adversely affects both sides of traditional banks' balance sheets. That is, it puts upward pressure on the deposit rate that traditional banks must pay to attract deposits—since the shadow banking sector offers alternatives to deposits, like money market mutual funds—and it puts downward pressure on the lending rate that traditional banks can earn—since the shadow banking sector offers alternatives to bank loans, like loans from finance companies. The relative size of the shadow banking industry can be calculated from the U.S. Financial Accounts, where the traditional banking sector consists of commercial banks, savings institutions, and credit unions, while the shadow banking sector consists of broker-dealers, ABS issuers, finance companies, mortgage pools, and funding corporations. The shadow banking share of the overall banking sector is plotted in the top panel of figure C.1.

We capture the volatility of interest rates by the Merrill Lynch Option Volatility Estimate (MOVE) index, which is a market-based estimate of future Treasury bond yield volatility. The MOVE Index reports the average implied volatility across a wide range of outstanding options (with expiry dates of approximately one month) on two-year, five-year, 10-year, and 30-year U.S. Treasury securities. This series is plotted in the middle panel of figure C.1.

Two other variables that we add to our forecasting models are the GZ credit-spread measure, developed by Gilchrist and Zakrajsek (2012), and the Freddie Mac median age of a refinanced loan. We include the GZ credit spread because this spread reflects both a higher anticipated level of private-sector default rates and a higher price of default risk, both of which are factors that could raise loan rates relative to deposit rates. We include an indicator of refinancing activity, in this case the Freddie Mac median age of a refinanced loan, to reflect revenues from mortgage points that enter NIMs and have the potential to alter the relationships that we might expect between higher long rates and NIMs.[21] In particular, because higher long interest rates decrease mortgage refinancing, a steeper yield curve that is caused by higher rates has the potential to reduce NIMs, contrary to what might be expected from the fact that banks are earning higher returns from maturity transformation.

Like Guerrieri and Welch (2012) and Grover and McCracken (2014), we also consider in our NIM models the variables that comprise the Federal Reserve Board's DFA/CCAR stress test scenarios (see FRB 2013). Here, we also only consider the U.S. variables, which include six measures of activity and prices (specifically, real and nominal GDP, the unemployment rate, real and nominal disposable personal income, and the CPI), four aggregate measures of asset prices or financial conditions (specifically, indexes of house prices, commercial property prices, equity prices, and U.S. stock market volatility), and six measures of interest rates (specifically, the rate on the 3-month Treasury bill, the yield on the 5-year Treasury bond, the yield on the 10-year Treasury bond, the yield on a 10-year BBB corporate security, the prime rate, and the conventional 30-year mortgage rate).

[21] Revenues from mortgage points are included in NII because they are considered to be upfront interest payments. Similarly, they are also tax deductible.

C.2 Models that include auxiliary variables

We add variables like competition, interest-rate volatility, credit spreads, and mortgage refinancing one by one to our Treasury-yield models, as well as all together. We add the variables featured in the U.S. DFA/CCAR stress tests scenarios together to our Treasury-yield models, though we do so by extracting principal components from these variables. How we add additional variables and stress test scenario factors to our models depends on the model. For our forecast combination models we add a bi-variate equation consisting of NIMs and the auxiliary variable in question, and we add the forecasts from that equation to the set that we average across to obtain the overall forecast. For the models that include either two yields or two yield-curve factors we add the auxiliary variable in question to that equation. For the PLS models we add the auxiliary variable or variables in question to the set of Treasury yields from which we want to extract the PLS factors and then allow for additional factors in the model.

C.3 Results with auxiliary variables

Tables C.1 and C.2 report our forecast results for when we include the above-mentioned additional variables in our levels and changes models. Our point of comparison in these tables is the best performing specification among the type of model considered. In particular, for the levels specification, Model 3b is our point of comparison, while for the changes specification, Model 5 is our point of comparison.

For our levels model with the best forecasting performance, the shadow banking sector's share of total banking sector assets is the only additional variable that improves forecast performance, albeit not economically or statistically significantly so. For our changes model with the best forecasting performance, the (change in the) shadow banking sector's share of total banking sector assets again improves forecast performance and in this case at most forecast horizons, where this improvement is economically but not statistically significantly. A couple of other additional variables also improve forecast performance, in particular the Freddie Mac median age of a refinanced loan and the GZ credit spread. Of these variables, only the Freddie Mac median age of a refinanced loan improves forecast performance signficiantly, and in this case only for a couple of quarters.

Since none of the additional variables that we consider in this appendix are part of the set of variables that the Federal Reserve published in its DFA/CCAR stress test scenarios, our results that find that there is some scope to being able to lower RMSEs in our changes models suggest that there could be some benefit from including some of these variables in the DFA/CCAR stress test scenarios. That being said, the results still show that the forecast errors from our best performing models are large relative to the variation in NIMs.

Table C.1: RMSEs of Iterative Forecasts with Additional Variables, Levels on Levels, and 2000q1 to 2008q3 Evaluation Window

	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Step 10
In-Sample RMSEs										
Benchmark: 3b	0.06	0.06	0.06	0.06	0.06	0.07	0.07	0.06	0.06	0.06
Shadow Banking Share: 4a	0.06	0.07	0.07	0.07	0.07	0.08	0.08	0.08	0.08	0.08
Index of Bond Yield Volatility: 5	0.07	0.08	0.08	0.08	0.09	0.09	0.10	0.09	0.10	0.10
GZ Spread: 5	0.07	0.08	0.08	0.08	0.09	0.09	0.10	0.09	0.10	0.10
Freddie Mac Age of Loan Index: 6a	0.06	0.07	0.06	0.06	0.05	0.05	0.05	0.04	0.04	0.04
CCAR Variables: 6a	0.06	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07
All Additional Variables: 6b	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.06	0.06	0.07
Out-of-Sample RMSEs										
Benchmark: 3b	0.09	0.10	0.11	0.14	0.16	0.18	0.20	0.22	0.25	0.28
Shadow Banking Share: 4a	0.09	0.11	0.12	0.14	0.15	0.17	0.20	0.21	0.25	0.28
Index of Bond Yield Volatility: 5	0.09	0.11	0.14*	0.16*	0.20*	0.23	0.26	0.28	0.31	0.35
GZ Spread: 5	0.09	0.11	0.14	0.16	0.20	0.23	0.26	0.28	0.31	0.35
Freddie Mac Age of Loan Index: 6a	0.08	0.12	0.14	0.16	0.18	0.20	0.21	0.23	0.26	0.28
CCAR Variables: 6a	0.08	0.11	0.13*	0.15*	0.17*	0.19*	0.22*	0.24*	0.28	0.31
All Additional Variables: 6b	0.10	0.14	0.16*	0.18	0.20	0.21	0.24	0.25	0.27	0.29

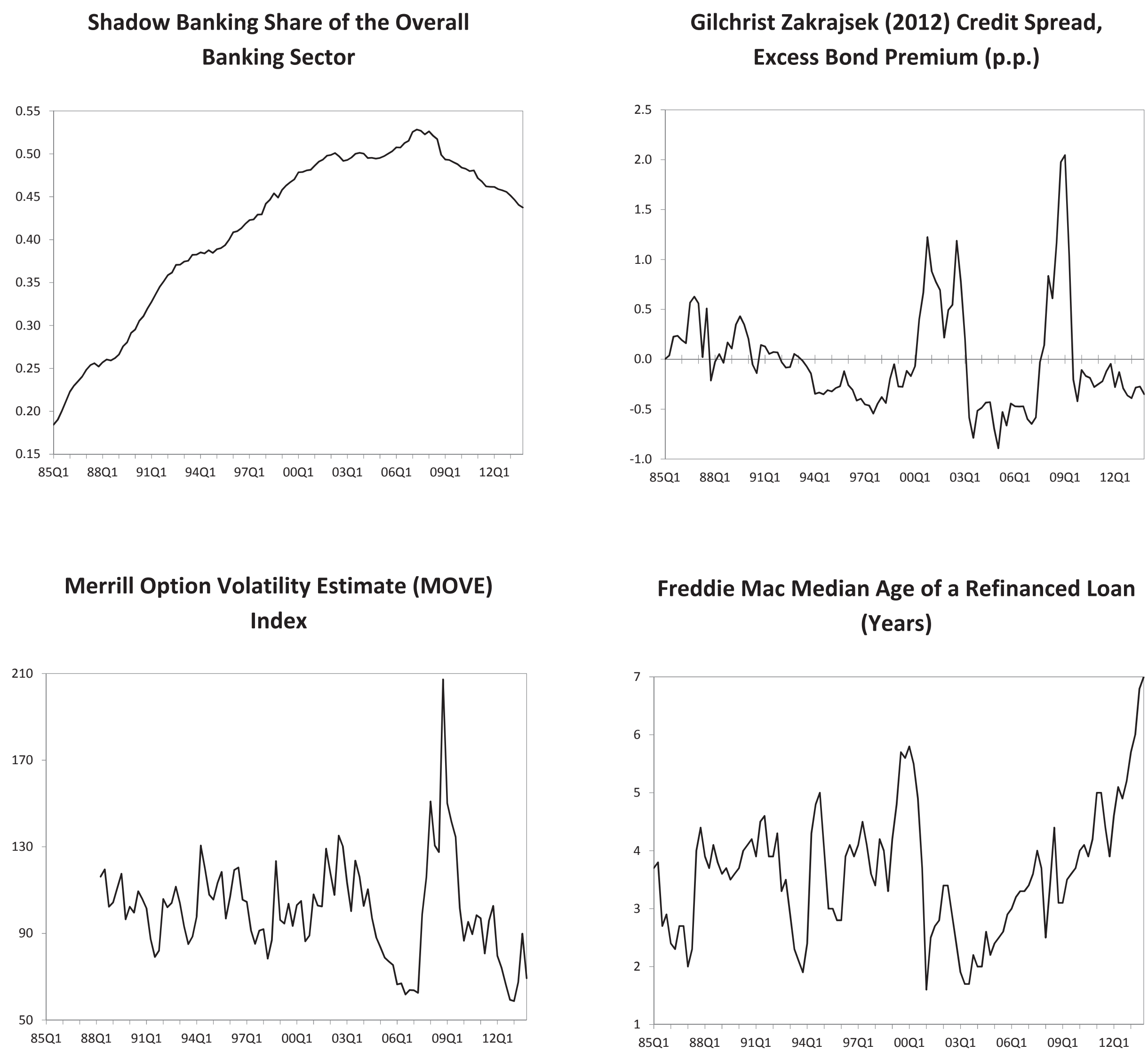
Models 1 through 7 are described in subsection 4.2. An asterisk '*' denotes significance at the 5% level of the difference in RMSEs relative to the no-change forecast using the Diebold-Mariano-West (DMW) test.

Table C.2: RMSEs of Iterative Forecasts with Additional Variables, Changes on Changes, and 2000q1 to 2008q3 Evaluation Window

	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Step 10
In-Sample RMSEs										
Benchmark: 5	0.071	0.087	0.089	0.101	0.117	0.133	0.139	0.144	0.151	0.169
Shadow Banking Share: 7	0.077	0.084	0.090	0.094	0.108	0.127	0.127	0.130	0.130	0.144
Index of Bond Yield Volatility: 4a	0.070	0.087	0.085	0.094	0.107	0.122	0.128	0.130	0.138	0.155
GZ Spread: 4a	0.070	0.089	0.091	0.101	0.117	0.137	0.152	0.168	0.186	0.210
Freddie Mac Age of Loan Index: 4a	0.063	0.076	0.076	0.080	0.090	0.103	0.105	0.109	0.114	0.124
CCAR Variables: 6a	0.065	0.087	0.089	0.103	0.119	0.137	0.150	0.158	0.170	0.185
All Additional Variables: 7	0.104	0.099	0.118	0.121	0.126	0.141	0.133	0.137	0.133	0.143
Out-of-Sample RMSEs										
Benchmark: 5	0.080	0.106	0.117	0.126	0.143	0.150	0.157	0.157	0.175	0.196
Shadow Banking Share: 7	0.080	0.095	0.102	0.099	0.115	0.132	0.140	0.137	0.141	0.149
Index of Bond Yield Volatility: 4a	0.081	0.105	0.107	0.117	0.134	0.144	0.156	0.158	0.179	0.202
GZ Spread: 4a	0.081	0.108	0.110	0.122	0.128	0.128	0.134	0.127	0.144	0.165
Freddie Mac Age of Loan Index: 4a	0.071	0.091	0.093	0.101	0.110*	0.127	0.141	0.151	0.163	0.174*
CCAR Variables: 6a	0.078	0.106	0.121	0.134*	0.151	0.154*	0.161*	0.161	0.177	0.200
All Additional Variables: 7	0.102	0.110	0.128	0.138	0.143	0.148	0.144	0.138	0.145	0.147

Models 1 through 7 are described in subsection 4.2. An asterisk '*' denotes significance at the 5% level of the difference in RMSEs relative to the no-change forecast using the Diebold-Mariano-West (DMW) test.

Figure C.1: Variables from the Micro-banking Literature



D Robustness analysis: BHC-level analysis

The last set of results that we consider is for NIM forecasts of individual BHCs and, in particular, for a group that is as close as possible to those that participate in the Federal Reserve's DFA/CCAR stress tests. In addition to being worthwhile for robustness analysis, considering BHC-specific NIM forecasts has a clear practical motivation, given that ultimately what supervisors require in evaluating BHCs' capital plans is individual BHCs' *pro forma* capital ratios and projections of components of their net income, including net interest income.

Another benefit of considering BHC-specific NIM forecasts is that these forecasts can be compared against earnings analysts' forecasts of NIMs. These forecasts provide a challenging point of comparison for our model forecasts, since earnings analysts have much more information available to them than we can include in our models and they also devote a great deal of resources to forecasting bank financial-statement variables. Moreover, we average across a large number of analysts' forecasts (on the order of twenty to thirty analysts, depending on the BHC and the forecast horizon), which generally tends to improve forecast performance. Our comparison to earnings analysts' forecasts of NIMs limits the period over, which, due to data ability, we perform our forecast evaluation to 2007q4 to 2013q4.

D.1 Data for the BHC-level analysis

D.1.1 BHC-level net interest margins

The data that we use to perform our BHC-specific analysis is from the quarterly Consolidated Financial Statements for bank holding companies, called the Y-9C. We use a definition of net interest margins using the items from FR Y-9C forms that correspond to the items in the Call Report forms described in subsection 4.1.1. For each BHC in our sample, mergers are accounted for using a so-called *pro forma* approach. That is, a historical time series for each BHC is constructed assuming that all of the institutions that are now part of the BHC always were part of the BHC. Our current data set starts in 1996q1, which is when the *pro forma* Y-9C data that we use for this analysis starts. Y-9C data, however, do extend back to 1991, so our sample period for the firm-level analysis should be able to be extended back to then as well.

D.1.2 Bank earnings analysts' forecasts

As noted, for our BHC-specific analysis, we compare our model-based BHC-specific NIM forecasts with bank-earnings analysts' forecasts of individual BHC NIMs. The financial-information firm SNL Financial LC has, since 2007q4, collected bank earnings analysts' forecasts for a number of key variables in BHC financial statements, including NII and NIMs. For any quarter SNL begins collecting forecasts about two years in advance, however, at this horizon only a very small number of analysts report forecasts. Forecasts are more widely reported by analysts over the year that precedes the quarter, accordingly we focus on forecasts between one and four quarters ahead.[22]

The BHCs for which we consider analyst forecasts are as close as possible to the 30 BHCs that were included in the 2014 DFA/CCAR stress tests, but data constraints limit our comparisons to 16 BHCs. The BHCs that we exclude are either those for which we do not have sufficiently long statutory BHC data (that is, Ally, American Express, Discover, Goldman Sachs, HSBC North America, Morgan Stanley, and Santander/Sovereign) or those for which we do not have SNL forecasts (that is, BBVA, BMO, RBS, and UnionBanCal). We also decided not to include custodial banks in our comparison (that is, BNY Mellon, State Street, and Northern Trust), since maturity transformation is not one of their main functions. For the 16 BHCs that we do consider, we used, in all cases, the average across all analysts for the BHC for the quarter. For all of the BHCs, the averages reported relate to about 20 different analysts per BHC, although for several BHCs the number of analysts is closer to 30.

[22]SNL Financial LC collects bank-earnings analysts' forecasts for all of the items listed in the first line of equation 1, as well as provisions, listed in the second line. It also collects bank-earnings analysts' forecasts for the first three lines together—that is, pre-tax net income—and the first four lines together—that is, post-tax net income. For many of these variables, bank-earnings analysts' forecasts are reported as both dollar volumes and, depending on the variable, as a ratio to either interest earning assets, total assets, or equity.

D.1.3 Blue Chip Treasury forecasts

Bank earnings analysts do not have realized interest rates available to them when they form their forecasts, whereas our NIM models treat realized interest rates as being available. So as not to give our forecasts an unfair advantage, we condition instead on Blue Chip (Financial) Treasury-rate forecasts. which include rates on the 3-month Treasury bill, on the 2-year Treasury note, and on the 10-year Treasury bond. Blue Chip releases these forecasts on the first day of each month. We therefore generate all of our NIM-model forecasts—conditional on Blue Chip Treasury rate forecasts—as if we were doing so on the first day of the quarter; that is, January 1, April 1, July 1, and October 1. Given this timing of our NIM-model forecasts, the SNL forecasts that we use for comparison are those that were recorded on January 4, April 4, July 4, and October 4 (or the earliest day thereafter). Choosing these dates for the SNL forecasts means that the equity analysts have a comparable information set to ours.

D.2 Models for the BHC-level analysis

We only show results for one of our models; specifically, the model specified in first differences that uses factors obtained from the dynamic factor model in a multivariate regression (Model 5). We show results for this model for the reason that when we specified our models in changes form and considered aggregate NIM data, this model resulted in the lowest RMSEs. As discussed above, we condition on the Blue Chip (Financial) Treasury-rate forecasts—not on realized Treasury-rate forecasts as we have in earlier analysis—since this gives our model the same amount of information as was available to earnings analysts when they formulated their NIM forecasts, and therefore does not unfairly advantage our model.

We should note that there are a very small number of instances in our SNL financial database of NIM forecasts (*e.g.*, on the order of about 17 out of more than 700 forecasts) for which forecasts are missing. In these cases, which are all for JP Morgan Chase and Citigroup, we use the actual observation of NIMs as the SNL financial forecast for this quarter. This assumption clearly favors the SNL forecasts in our forecast evaluation exercises, so the performance of our model-based forecasts for JP Morgan Chase and Citigroup are actually better relative to the SNL forecasts than how they appear in our results tables.

D.3 BHC-level NIM forecasting results for the evaluation period 2007q4 to 2013q4

Tables D.1 and D.2 present the results for our BHC-specific analysis. The horizon for these forecasts is one- to four-quarters ahead, which is the horizon for which SNL has forecasts from a sizable number of analysts.

Tables D.1 and D.2 differ only in terms of the data series against which the SNL forecast is evaluated. In table D.1, the SNL forecasts—along with the model-based forecasts—are evaluated relative to paths of *pro forma* NIMs, whereas in table D.2, the SNL forecasts are evaluated relative to paths of actual NIMs, while the model-based forecasts continue to be evaluated relative to *pro forma* NIMs. Conceptually, it makes more sense to compare SNL forecasts to actual NIMs rather than a *pro forma* NIM series that assumes that all institutions that are now part of the BHC were always part of it (which is clearly not what analysts were forecasting). The only drawback of comparing the model forecasts to one set of actual data and the SNL forecasts to another set of actual data is that we cannot get an idea of the significance of these differences, since it is no longer appropriate to use the Diebold-Mariano-West (DMW) test.

The results of both tables D.1 and D.2 indicate that our model-generated forecasts are competitive with the SNL forecasts that are formulated using much more information than we can include in our models and many more resources. In particular, for the four universal banks, reported in the first four lines of the table, our model-generated forecasts improve on the SNL forecasts at all horizons. For the other BHCs, our model-generated results do not perform as well and only improve on the SNL forecasts for two BHCs; specifically, Capital One and Zions. The fact that our model-generated forecasts perform better for the largest BHCs may reflect the fact that these BHCs have a much greater influence on the aggregate data and our decision to use Model 4a for our BHC-specific results was informed by our aggregate NIM results. It may also reflect the fact that the paths of NIMs for smaller banks may reflect more regional developments to which analysts pay close attention, but that our models do not capture as well. Nonetheless, our results indicate that our model-based forecasts are competitive with those made by equity analysts, particularly for the largest universal banks.

The last column of tables D.1 and D.2 reports the standard deviation of NIMs for the period over which our forecasts are being evaluated. As can be seen, in line with the results that we obtained when considering aggregate NIMs, while from a relative perspective our model forecasts are reasonable, from an absolute perspective they are quite poor and predict very little of the variation of NIMs over time.

Table D.1: Bank Specific RMSEs of Iterative Forecasts, Regressions of Changes on Changes, 2007q4 to 2013q4 Evaluation Window, and Evaluation Relative to *Pro Forma* NIMs for All Forecasts

	5. DFM, Mult. Reg.				1. No-Change Forecast				SNL Forecast†				st.dev.
	Step 1	Step 2	Step3	Step 4	Step 1	Step 2	Step3	Step 4	Step 1	Step 2	Step3	Step 4	
JPMORGAN CHASE & CO	0.15*	0.20*	0.24*	0.26	0.15*	0.21*	0.27	0.34	0.38	0.39	0.41	0.46	0.28
BANK OF AMER CORP	0.26	0.33	0.39	0.44	0.26	0.31	0.36	0.40	0.29	0.34	0.39	0.44	0.30
CITIGROUP	0.25	0.40	0.46	0.46	0.21	0.27	0.30	0.30	0.43	0.47	0.49	0.46	0.19
WELLS FARGO & CO	0.13	0.22	0.31	0.39	0.12	0.20	0.29	0.35	0.54	0.50	0.49	0.48	0.29
U S BC	0.12	0.18	0.20	0.21	0.10	0.15	0.18	0.19	0.10	0.13	0.16	0.18	0.13
PNC FNCL SVC GROUP	0.19	0.28*	0.33	0.38	0.18	0.25	0.31	0.35	0.17	0.23	0.26	0.29	0.35
CAPITAL ONE FC	0.54	0.68	0.80	0.92	0.53	0.68	0.84	0.98	0.89	1.05	1.14	1.31	0.87
BB&T CORP	0.14	0.21	0.27	0.33	0.13	0.19	0.24	0.28	0.15	0.17	0.19	0.19	0.26
SUNTRUST BK	0.19	0.27*	0.31	0.34	0.10*	0.15	0.18	0.21	0.16	0.17	0.18	0.19	0.19
FIFTH THIRD BC	0.39	0.45	0.41	0.36	0.34	0.38	0.35	0.32	0.22	0.26	0.31	0.29	0.24
REGIONS FC	0.15	0.25	0.32	0.40	0.12	0.22*	0.31	0.38	0.12	0.19	0.29	0.36	0.25
KEYCORP	0.50	0.46	0.53	0.43	0.49	0.48	0.51	0.38	0.45	0.48	0.51	0.38	0.34
M&T BK CORP	0.13	0.19	0.22	0.24	0.12	0.18	0.20	0.21	0.11	0.14	0.15	0.18	0.18
COMERICA	0.16*	0.24	0.30	0.36	0.17*	0.25*	0.32*	0.37*	0.09	0.18	0.24	0.27	0.22
HUNTINGTON BSHRS	0.19	0.23*	0.24*	0.27*	0.11	0.16	0.18	0.20	0.11	0.15	0.15	0.16	0.13
ZIONS BC	0.22	0.22	0.28	0.27*	0.20	0.20	0.27	0.28*	0.21	0.26	0.33	0.38	0.27

Table D.2: Bank Specific RMSEs of Iterative Forecasts, Regressions of Changes on Changes, 2007q4 to 2013q4 Evaluation Window, and Evaluation Relative to *Pro Forma* NIMs for Model Forecasts and Actual NIMs for SNL Forecasts

	5. DFM, Mult. Reg.				1. No-Change Forecast				SNL Forecast†				st.dev.
	Step 1	Step 2	Step3	Step 4	Step 1	Step 2	Step3	Step 4	Step 1	Step 2	Step3	Step 4	
JPMORGAN CHASE & CO	0.15*	0.20*	0.24	0.26	0.15*	0.21*	0.27	0.34	0.56	0.47	0.43	0.35	0.28
BANK OF AMER CORP	0.26	0.33	0.39	0.44	0.26	0.31	0.36	0.40	0.32	0.35	0.35	0.34	0.30
CITIGROUP	0.25*	0.40*	0.46*	0.46*	0.21*	0.27*	0.30*	0.30*	0.92	0.92	0.93	0.85	0.19
WELLS FARGO & CO	0.13	0.22	0.31	0.39	0.12	0.20	0.29	0.35	0.77	0.75	0.84	0.87	0.29
U S BC	0.12*	0.18*	0.20	0.21	0.10	0.15	0.18	0.19	0.08	0.12	0.15	0.17	0.13
PNC FNCL SVC GROUP	0.19	0.28	0.33	0.38	0.18	0.25	0.31	0.35	0.76	0.83	0.90	0.96	0.35
CAPITAL ONE FC	0.54	0.68	0.80	0.92	0.53	0.68	0.84	0.98	0.86	1.03	1.11	1.28	0.87
BB&T CORP	0.14	0.21	0.27	0.33	0.13	0.19	0.24	0.28	0.08	0.12	0.15	0.15	0.26
SUNTRUST BK	0.19	0.27*	0.31*	0.34*	0.10	0.15*	0.18*	0.21	0.09	0.10	0.12	0.14	0.19
FIFTH THIRD BC	0.39	0.45	0.41	0.36	0.34	0.38	0.35	0.32	0.25	0.28	0.33	0.31	0.24
REGIONS FC	0.15	0.25*	0.32	0.40	0.12*	0.22*	0.31	0.38*	0.09	0.17	0.27	0.34	0.25
KEYCORP	0.50	0.46	0.53	0.43	0.49	0.48	0.51	0.38	0.41	0.44	0.47	0.33	0.34
M&T BK CORP	0.13	0.19*	0.22	0.24	0.12	0.18	0.20	0.21	0.09	0.13	0.16	0.19	0.18
COMERICA	0.16*	0.24	0.30	0.36	0.17*	0.25*	0.32	0.37*	0.10	0.20	0.26	0.29	0.22
HUNTINGTON BSHRS	0.19	0.23*	0.24*	0.27*	0.11	0.16	0.18	0.20	0.12	0.15	0.15	0.16	0.13
ZIONS BC	0.22	0.22	0.28	0.27	0.20	0.20	0.27	0.28	0.20	0.24	0.31	0.36	0.27