Title

Aggregation level in stress testing models

Authors

Hale, Galina
Krainer, John
McCarthy, Erin

Publication Date

2023-10-29

DOI

10.24148/wp2015-14

Aggregation Level in Stress Testing Models*

Galina Hale†
Federal Reserve Bank of San Francisco
John Krainer‡
Federal Reserve Board of Governors
Erin McCarthy§
KPMG

March 29, 2019

Abstract

We explore the question of optimal aggregation level for stress testing models when the stress test is specified in terms of aggregate macroeconomic variables, but the underlying performance data are available at a loan level. We ask whether it is better to formulate models at a disaggregated level and then aggregate the predictions in order to obtain portfolio loss values or is it better to work directly with aggregated data to forecast losses. The answer to this question depends on the data structure. Therefore, we study this question empirically, using as our laboratory a large portfolio of home equity lines of credit. All the models considered produce good in-sample fit. In out-of-sample exercises, loan-level models have large forecast errors and underpredict default probability. Average out-of-sample performance is best for county-level models. This result illustrates that aggregation level is important to consider in the loss modeling process.

JEL classification: G21, G28, C18

Keywords: bank stress testing, forecasting, portfolio granularity, probability of default, home equity.

*The views expressed here are those of the authors and not necessarily those of the Federal Reserve System. The authors thank Michael Bauer, Paul Calem, Kris Gerardi, Jose Lopez, Joe Tracy, Paul Willen, as well as participants at the Central Bank of Ireland, the Interagency Risk Quantification Forum, the 2015 IBEFA conference, the 2015 FRB-Atlanta policy conference for very helpful discussions. All errors are our own.

†galina.b.hale@sf.frb.org

‡john.r.krainer@frb.gov

§mccarthy.erink@gmail.com

1 Introduction

Under the Dodd-Frank Act the Federal Reserve is required to conduct annual stress tests of the systemically-important U.S. banking institutions.[1] The centerpiece of the supervisory stress tests is a calculation of expected losses for banks under a set of economic scenarios. A scenario consists of explicit 9-quarter paths for macroeconomic and financial market variables such as interest rates, asset prices, unemployment, inflation, and GDP growth. The scenarios are not necessarily considered likely, but are meant to be coherent in the sense that, even though some variables, such as unemployment, may move to extreme values, other variables in the scenario, such as credit spreads, should comove with these extreme changes in historically consistent ways.

Given the distinct structure of supervisory stress tests, our research question centers on which risk modeling and forecasting approaches may prove to be most useful for the task at hand. Specifically, if the goal is to predict total losses at the bank level, and the inputs to the stress test are given at a certain level of aggregation, what is the desired level of aggregation of underlying loan portfolio variables in the risk modeling? In this paper we investigate whether different levels of portfolio aggregation yield different model performance characteristics, such as size of forecasting error, model stability, and sensitivity to macroeconomic variables.[2]

Our application is to model the risk of credit default for a portfolio of home equity lines of credit

[1]The Federal Reserve has conducted several distinct rounds of stress testing since the financial crisis in 2008. Both the Supervisory Capital Assessment Program of 2009, and the Comprehensive Capital Analysis and Review (CCAR) are very similar in terms of format. The principal change in supervisory stress testing over the past several years has been changes in the number of institutions included in the exercise.

[2]By model stability we mean low sensitivity of estimates to changes in data sample or small perturbations of the model specification.

(HELOCs) observed over the 2002-2013 period. We choose to model HELOCs instead of, say, first-lien residential mortgages or commercial loans, because they are relatively more homogeneous in terms of their contract structure and are less-dependent on private information. We consider "bottom-up" loan-level models, where we incorporate very detailed information on individual loan characteristics as well as local and aggregate economic variables. We also consider "top-down" time-series models at the portfolio level.[3] Finally, we evaluate hybrid approaches where we aggregate the data into buckets by county or by deciles of the risk factors. We evaluate out-of-sample forecast errors of these models and assess which levels of aggregation work best in terms of average forecast error as well as under a loss function that more heavily penalizes under-prediction of the default rate.

A number of theoretical considerations weigh on the choice of how much to aggregate the data. At one extreme, a top-down approach would have us use fairly simple specifications to capture the time-series dynamics of home equity default rates. This level of aggregation also fits with the loss function of a bank regulator, which would emphasize default rates or losses at an aggregate or firm level rather than at the individual loan level. The disadvantage of using highly aggregated data, of course, is that these models are almost always misspecified. Models based on aggregate data will perform poorly if the composition of loans is changing over time, as was recently the case in U.S. housing markets when a period of easing underwriting standards led to a large expansion of credit to previously constrained borrowers.[4] Moreover, if there is cross-observation correlation

[3]See Hirtle, Kovner, Vickery, and Bhanot (2016) for comparison of overall stress-testing results produced by "top-down" and "bottom-up" models.

[4]See Mian and Sufi (2009) for evidence of the expansion of credit to previously underserved markets during the early 2000s. Over roughly the same time period, Frame, Gerardi, and Willen (2015) show how changing loan composition led to large errors in the models used by the Office of Federal Housing Enterprise Oversight to stress

between loss rates or explanatory variables, the aggregation process introduces aggregation bias, where parameters estimated at a macro-level deviate from the true underlying micro parameters.[5]

At the other extreme, loan-level models offer solutions to the aforementioned problems associated with the "top-down" approach. However, their use presents different challenges. Traditionally, two main obstacles to estimating loan-level default models have been the lack of reliable loan-level data and computational limits. This is much less of a problem in the current day, given recent improvements in data collection in the banking and financial sector and computing technologies. However, some risk factors that enter into loan-level default models are not observable and are replaced with market-level proxies. For example, housing values (i.e., the value of the collateral for the loans) are not updated regularly at the individual borrower level. In our analysis we proxy for changes in individual house prices by using changes in a county or zip code-level house price index. We encounter the same measurement problem with the unemployment rate, which we proxy for with a county-wide unemployment rate. The home-owning and home equity borrowing population may be different from the population in a county most exposed to unemployment shocks. Indeed, for the case of unemployment, there is a further complication. Ideally, we would have a variable telling us whether the borrower him or herself is unemployed.[6] But what in fact we have is a population average probability that the borrower is unemployed (see also Gyourko and Tracy (2013)). Such use

Fannie Mae and Freddie Mac's exposure to mortgage default risk.

[5]Going back to Theil (1954), linear models that are perfectly specified at the micro-level were known to be susceptible to aggregation bias. Grunfeld and Griliches (1960) showed that once this assumption of a perfectly specified micro model is relaxed, then aggregation could produce some potential gains in the form of aggregating out specification or other types of measurement error. Also, Granger (1980) shows that time series constructed from aggregated series can have substantially different persistence properties than is present in the underlying disaggregated data.

[6]There is evidence that borrowers are not solely strategic in their default behavior and require a "double trigger" of house price declines and unemployment (Gerardi, Herkenhoff, Ohanian, and Willen (2015)).

of proxies introduces measurement error into the estimation, which may lead to biased estimates of the effect of the variable in question. A downward bias would be particularly worrisome given the design of the Federal Reserve's stress tests which are cast in terms of exactly these variables where we have measurement difficulties at loan level.

There is no general answer to which level of aggregation produces best results. If at disaggregate level the variables are independent and identically distributed, the level of aggregation would not matter. However, this is almost never the case in the data. Depending on the data structure, biases of different sizes may arise at different levels of aggregation. Since the optimal level of aggregation is essentially an empirical question, we do not limit ourselves to the extreme cases of totally aggregated "top-down" and totally disaggregated "bottom-up" approaches. We also consider intermediate levels of aggregation for our analysis of HELOC default rates. We consider county-level models, the level at which data on macroeconomic variables of interest is usually available.[7] This level of aggregation addresses the potential problem with measurement error in loan-level models. We also consider models in which data are aggregated to portfolio segments by FICO score, loan-to-value ratio (LTV), and debt-to-income ratio for the borrower (DTI), all measured at the time of loan origination. This approach helps to address the problem of changing portfolio composition that "top-down" models face.

In order to evaluate forecast quality of different levels of aggregation, we need to choose an empirical specification for each. It turns out that the set of model specifications with good fit is

[7]Throughout we will refer to economic variables observed at a higher level of aggregation that the individual borrower or loan as macroeconomic variables.

different for different aggregation levels.[8] For this reason we proceed in three steps: first we screen a very large number of specifications that include all potential risk drivers at various lags, as well as their interactions, for statistical significance, intuitive signs for the coefficient estimates, as well as in-sample and out-of-sample fit. This is done using some judgement (e.g., house prices should enter into any model of home equity loan default), as well R^2, information criteria, and forecast error. Then we focus on a smaller number of reasonable specifications that pass this screening test. For each of these specifications, we estimate regressions using data ending in each of 12 months from June 2008 through July 2009.[9] In each case, we construct the forecasted default frequency for the following 9 quarters, in the spirit of CCAR exercises, and compute an average forecast error as well as measures of how conservative is the forecast.

We find that county-level regressions tend to have lower forecast errors, produce reasonably conservative results, and are quite stable across different specifications and forecast windows. Loan-level regressions show attenuation bias and therefore tend to have the highest forecast errors and the least conservative predictions, while aggregate regressions perform well on average but are not very robust to specification changes. Models aggregated by risk factor deciles also perform quite well and are relatively stable across specifications. They are, however, inferior to county-level regressions in terms of the forecast error. Thus, in the case of HELOC default projection, neither loan-level nor top-down aggregate models are best in a stress testing environment. In our exercise

[8]We demonstrate that our conclusions are not driven by specification choice, they are the same if we use identical specifications across aggregation levels.

[9]One drawback of our data is that we have a relatively short sample period extending back to just 2002. This may appear to be a severe handicap for the loan-level models. However, the short sample period is also a constraint on the performance of the more aggregated models. Moreover, as we note in the methodology section, the types of biases that can arise in this type of credit risk modeling stem from cross-sectional correlations in the data and do not go away with longer time series.

the best approach is to aggregate the data to some extent — most meaningfully, to the level at which macroeconomic variables used in the scenarios are available.

While we do not presume that this specific result immediately generalizes to other portfolios, or that we have chosen the best possible specification for each aggregation level, our evidence supports the conclusion that the level of aggregation is an important factor to consider in any stress-testing exercise. The analysis of optimal level of aggregation only became possible with the increased availability of loan- or other micro-level data. Too often, however, the modelers simply assume that the most disaggregated level of analysis is the best one. The main takeaway from our paper is that the aggregation level should be taken as one of the parameters in the modeling process.

The paper is organized as follows. In section 2 we demonstrate that econometric theory does not provide a clear guide as to which level of aggregation will result in the lowest forecasting error. In section 3 we describe the home equity data set used in the empirical application. We detail the specifics of our forecasting exercise and results in Section 4. Section 5 concludes.

2 Econometric framework

Our goal is to predict the default rate $y \in [0, 1]$ for the entire portfolio given a macroeconomic scenario. The individual macroeconomic variables, x, do not vary by loan in portfolio, although some macroeconomic variables might vary by geographical segments within the portfolio. For simplicity of notation, suppose we are only predicting one period forward, that is predicting y_{T+1} given x_{T+1} and observed history of y's and x's up to period T. Suppose the data generating process

(DGP) is such that

$$y = X'\beta + \varepsilon,$$

where y is a vector of observed default rates (or, in case of individual loans, default indicators) over time, and X is a matrix of observed covariates, x, including a constant term. The unobserved disturbance ε is distributed $N(0, \sigma^2)$. We can use linear regression to estimate b, the estimator for β, and $\hat{\sigma}$, the estimator for σ. Denote the residuals from the regression e.

Suppose y and X are observed at the individual loan level, and there are N loans observed for T time periods. Therefore, we have a choice of whether to estimate b and $\hat{\sigma}$ on individual loan data (using linear probability regression for the ease of exposition), on average values of y and X for sub-portfolios of any type (using pooled or fixed effects panel regression), or on overall portfolio averages (using time series regression). Given that our goal is to predict aggregate y, we want to determine which method is preferable. Regardless of the regression estimated, the conditional forecast can be constructed by substituting b for β in the DGP equation above.

If individual observations are i.i.d., we can obtain an unbiased estimate of β regardless of the aggregation level. In this case, aggregation level will not affect the expected forecast mean and we would only be concerned with the precision of forecast. One can show that precision of conditional forecast will be determined by differences in estimated variance of the disturbance term $\hat{\sigma}$, the number of loans and sub-portfolios, and differences in the variance of covariates. If the observations are i.i.d., different aggregation levels will give the same results in the limit. However, in finite samples, even if observations are drawn from i.i.d. distributions, there will be differences in forecast errors, that vary by sample. The forecast errors will generally be larger the more aggregated the regression sample is. See Appendix 1 for derivations.

There are two main reasons, however, to believe that the observations in the analysis are not i.i.d. and therefore the estimates of β are not necessarily unbiased. First, there may be measurement error resulting from the correlation between observed and unobserved components of some of the variables. Second, there may be aggregation bias resulting from cross-observation correlation of both dependent and independent variables. We now consider these cases.

2.1 Individual-level measurement error

One issue that arises in loan-level analysis is that macroeconomic variables are not measured at a loan level. For example, while a borrower's unemployment status or home price have direct effects on his or her probability to default on the home equity loan, it is common to proxy for these variables with state or MSA-level unemployment rate and home price indexes. This introduces a measurement error problem in loan-level regressions. With sufficient number of observations per state or MSA, these individual errors cancel out when computing averages for the state or MSA-level regressions, so the problem is specific to loan-level regressions.

The measurement problem that arises from the use of an aggregate variable can be expressed as an omitted variable problem, where the omitted variable is the difference between the individual measure and aggregate measure used. The omitted variable is not observable by an econometrician, and, in general, will be correlated with the aggregate variable. Therefore, we would not be able to obtain an unbiased estimator of β. If the correlation and the effect of the omitted variable on the outcome are of the same sign (e.g., both positive, or both negative), there will be an upward bias. Otherwise, the bias will be downward. Since the omitted variable and its effect are not observed, it is generally not possible to evaluate on a purely econometric basis whether the bias is positive or negative. Note that an attenuation bias would be particularly harmful if the estimates are used for scenario analysis, because lower coefficients on macroeconomic variables would lead to an underestimate of the stress scenario losses. See Appendix 1 for the formal argument.

Given that the measurement error is likely to cancel out as we start aggregating the data, this problem will only arise for loan-level regressions, and not for portfolio or subportfolio-level regressions, if macroeconomic variables are observed at the subportfolio level.

2.2 Aggregation bias

The measurement error problem has to be weighed against the aggregation bias problem. In the prior discussion we assumed that explanatory variables and the error term are uncorrelated across individual loans. This assumption is necessary to obtain unbiased estimates of β in the regressions using aggregate data. In practice, this is unlikely to be the case and therefore aggregation bias will arise. Intuitively, cross-observation correlations will pollute estimates obtained at aggregate levels.

To see this, assume that the loan-level estimate is unbiased:

$$b_L = (X_L' X_L)^{-1} X_L' y_L, \quad E(b_L) = \beta,$$

where X_L is a matrix of independent variables at loan level and y_L is a loan default indicator.

One can show that for the portfolio level the estimate of β can be expressed as

$$b_P = \beta + (X_L' X_L + X_L'(U_N - I)X_L)^{-1}(X_L' e + X_L'(U_N - I)e)$$

where U_N is the $(NT \times NT)$ block-diagonal matrix of T $(N \times N)$ matrices of ones in the diagonal and zeros elsewhere, and I is the identity matrix, e is the regression residual.

One can see that for the estimator to be unbiased, we need $E(X_L' \varepsilon | X_L) = 0$ and $E(X_L'(U_N - I)\varepsilon | X_L) = 0$. The first condition is satisfied because disaggregate estimator is unbiased. However, for the second condition will be true only if there is no within-time cross-individual correlation between each variable x in X and e.[10][11] Within-time cross-individual correlation is likely to arise in default probability models because shocks to housing prices and job loss rates tend to be clustered in specific regions.

The same problem arises for a sub-portfolio level regressions. However, given that sub-portfolio level regressions allow for more non-zero terms in the variance-covariance matrix than fully aggregate regression, the bias will be smaller in magnitude the less aggregated the variables are.

To summarize, there is no sure way to tell what level of aggregation is going to produce the best conditional projection — both in terms of bias and in terms of forecast precision.[12] We have

[10]Note, however, that the standard estimate of the variance of b_L will not be unbiased, and the cluster-robust standard errors will need to be computed. See, for example, Arellano (1987).

[11]See Appendix 1 for details.

[12]This result is also demonstrated, for in-sample fit, in Pesaran, Pierse, and Kumar (1989).

illustrated that, depending on the structure of the data, projection accuracy can be better or worse for more aggregated models compared to disaggregated models by giving examples in which measurement error bias is likely to arise in individual loan level regressions, while aggregation bias is likely in the aggregate regressions. Thus, what level of aggregation is the best for predicting aggregate outcomes remains an empirical question and the answer depends on the specific data set being analyzed. In the rest of the paper we present an empirical exercise for HELOCs, in which we compare out-of-sample performance of models estimated at different levels of aggregation. The optimal level of aggregation, however, may be different for other types of loans and sample characteristics.

3 Data description

We implement our model performance tests using a data set of home equity lines of credit (HELOC). A HELOC is a commitment by a lender (usually a commercial bank) to lend up to a specified amount over a specified period of time called the draw period. HELOCs are collateralized by a claim to the equity in the borrowers' house. The lender is typically in a second-lien position in the cases where there is already a first mortgage on the property.[13] Borrowers may draw down on the commitment at any time during the draw period, at which point the draw will be added to the outstanding balance. The borrower is under no obligation to draw down the full amount on the line during the draw period, or even draw anything at all. Indeed, many borrowers reputedly put HELOCs in

[13] States have different rules on whether lien status changes for home equity when a borrower refinances the first-lien mortgage. See Bond, Elul, Garyn-Tal, and Musto (2014).

place as insurance to smooth consumption in case of future income shocks (see Hurst and Stafford (2004)). Borrowers will typically only need to make interest payments on the outstanding balance during the draw period. Interest payments are based off of a short-term benchmark interest rate such as the Prime Rate or LIBOR. After the draw period ends, further draws are not permitted and the borrower pays down the balance–either in the form of a balloon payment or over time as an amortizing loan.[14]

The data set is constructed from a five-percent random sample of HELOCs from the CoreLogic LP Home Equity Database. We draw only from the set of HELOCs with adjustable rates that are in the second lien position. The choice to limit the sample to just HELOCs and just HELOCs with adjustable-rate interest payments reflects a desire to simplify our study of aggregation bias. Consumer loan products like HELOCs are not thought to be particularly informationally sensitive. Statistical scoring models have long been an important part of the HELOC underwriting process and lenders are not thought to make great use of private, relationship capital in their lending decisions.[15] Given these factors, a model of HELOC default risk is likely to be characterized by a fairly small set of observable risk factors relative to more informationally-sensitive products such as a small business loan, or a construction loan. Thus, we have no a priori reasons to suspect that any particular model aggregation level will perform best in our statistical tests of forecasting accuracy.

Our resulting sample contains monthly observations on 454,724 unique HELOCs for a total of

[14]Our sample period stops short of any waves of borrower end of draw, and thus the empirical default rates do not appear to be related to this state. See Epouhe and Hall (2015) and Johnson and Sarama (2015) for analysis of home equity default dynamics around the end of draw period.

[15]See Rajan, Seru, and Vig (2015) for evidence of increased use of statistical underwriting models in consumer finance.

more than 20 million observations. The sample ranges from 2002 through 2013. Delinquency is defined as the event of reaching 90-days past due. Once this event takes place the loan history in our sample is terminated, meaning that we abstract away from cures and the actual transitions from delinquency to default, foreclosure, and ultimately to loss. Note that our measure of the delinquency or default rate is the transition rate from current into default rather than the stock of all outstanding loans in default. We adopt this convention for default because it matches up most cleanly with the probability of default construct that would go into an actual model used for stress testing and risk management.

All the specifications explored in our different risk models contain a grouping of observable economic factors. The main economic risk factors used in the analysis include the trailing 12-month house price depreciation from the CoreLogic monthly house price indexes (HPI). In this paper we use house depreciation (i.e., the negative of house price appreciation) because in some specifications this transformation lends itself more easily to interpreting interaction effects between house price changes and other variables. Whenever possible we link each HELOC to a zip code level HPI. When this level of precision is not possible we revert back to the county-level HPI. If the loan is situated in a county where CoreLogic has no coverage at all, we drop the history from the sample. Also included in the set of economic risk factors are the county-level unemployment rate from the BLS and a year-over-year real GDP growth series constructed from the monthly estimates provided by Macroeconomic Advisors.

To a rough approximation, the theory of default underlying the empirical specifications is that changes in economic conditions can impact a borrower's ability or willingness to stay current on the

HELOC account. Importantly, aggregated versions of all of the economic proxies for default risk that we use in our models are also specified in the CCAR stress tests. The local unemployment rate and economic growth measures speak to the probability that the borrower has the income or wealth to avoid default. House prices capitalize the value of the land in a location. Thus, fluctuations in house prices can serve as summary measures for changes in local conditions that may influence borrower ability to repay debt. Additionally, HELOCs are collateralized by a claim to the value of the underlying house. If house prices fall enough, homeowner equity can be wiped out, reducing the incentive to repay either the first or the second lien mortgage. As described above, we proxy for this source of default risk with the 12-month decline in house prices (or house price depreciation).

One drawback of the HELOC data is that we do not have consistently-reported loan-to-value information on the borrower's first-lien mortgage. Thus, given our use of house price changes, our models of this source of default risk stemming from the risk to collateral values are strictly linear. We will not capture any nonlinearities in the data if, for example, default risk rises at a faster pace once house price declines cause the total leverage of the borrower (first and second liens combined) to breach a critical threshold.[16]

In general we adopt a parsimonious approach to model selection on the grounds that these specifications tend to do better out-of-sample. As we proceed to lower levels of aggregation, however, we include increasingly more variables as a way of giving the disaggregated models the fullest opportunity to exploit the rich data at our disposal. Thus, we include commonly-used variables

[16]By and large, the mortgage default literature from our sample period has not found strong support for nonlinearities. For example Gerardi, Herkenhoff, Ohanian, and Willen (2015), find that household-level employment and financial shocks are of first-order importance in explaining mortgage default. The authors find only a limited role for strategic default, where defaults rise sharply once the household is underwater.

such as the FICO credit score (FICO), the borrower debt-to-income ratio at origination (DTI), the reported loan-to-value ratio at origination (LTV), and the loan amount.[17] In the loan-level model we also consider a host of other variables that speak to underwriting standards and other factors that might be correlated with unobservable borrower characteristics, such as loan age, the spread of the loan rate over the reference interest rate (i.e., the margin rate), the utilization rate (i.e., the ratio of current balance to committed balance), the appraised home value at origination, as well as dummy variables for whether the HELOC was made with full documentation, is an interest-only HELOC, and if the HELOC was originally set up in connection with a home purchase transaction. Note that the origination LTV, FICO scores, and DTI have all been winsorized at the 1st and 99th percentiles of the raw empirical distribution. The summary statistics for the variables used in the risk models for our sample are in Table 1.

For a first glimpse at the time-series behavior of some of the variables in our HELOC database, we see in Figure 1 that there was a steady increase in new HELOC origination through the housing boom years. New originations abruptly dried up once house prices leveled off and began falling in 2006. Essentially all of our HELOCs were originated during the 2002-2006 period. We did not include any HELOCs that were originated prior to 2002–the starting point for our sample–for fear of introducing survivorship bias.

[17] The DTI ratio is the "back-end" DTI, or the borrower's total monthly debt payments, including the payments on the subject property, divided by the borrower's monthly income.

4 Empirical analysis

Using the data described above, we now conduct analysis that is quite similar to the stress-test modeling process of many banking institutions. We search for specifications that have good fit, include stress scenario variables, and have good out-of-sample forecasting properties. We repeat this exercise at different levels of data aggregation and investigate how well model projections are able to capture turning points of HELOC delinquency rate in the sample using both in-sample and out-of-sample analysis. We also evaluate predicted delinquency rates in the 2014 supervisory severely adverse stress scenario.

4.1 Top-down model

We start our empirical analysis at the highest level of aggregation: the national level, which we also refer to as "portfolio", or "aggregate" level. We estimate variations of the model,

$$y_t = \theta_0 + X_t\beta + \varepsilon_t, \tag{1}$$

where $y_t = \frac{1}{N}\sum_{i=1}^{N} y_{it}$ is the aggregate default rate from our sample of N loans in month t, the matrix X_t contains the averages of loan-level and macroeconomic risk covariates and their lags, ε_t is an error term. The model in equation (1) is truly aggregated in the sense that both left and right-hand side variables are constructed as averages from the underlying loan-level data. The regressions are estimated using OLS.[18]

[18] The results do not change materially when we estimate equation (1) using a tobit specification, which limits predicted delinquency rates to non-negative space.

The specification search for the aggregate model takes the following form. As a first pass, we consider all possible combinations of models formed with the variables listed in Table 1. That is, we consider all possible models with one risk factor, with two risk factors, and so on. We augment this set of models with specifications that add interactions between the house price depreciation variable and unemployment. This was the only interaction term we considered in the model specification phase. This particular interaction term was chosen because of its intuitive link to the default literature that emphasize how both collateral value declines and cash flow constraints combine to accentuate defaults. Finally, we add in up to four quarters of lags of each of the variables, including interaction terms, as a way to capture dynamics in the observed default rate.

Before subjecting the models to performance-based comparisons, we drop all models in which the signs of the estimated coefficients are contrary to our expectations (for example, we expect positive effect of unemployment and negative effect of home prices on default rates). We do so, because such models would predict lower losses in more stressful economic and financial conditions. In models with no interactions, we drop specifications where the coefficients on the macroeconomic variables are insignificant at the 90th percent confidence level. We then rank order the remaining specifications by in-sample fit and select the 25 best models in terms of out-of-sample fit (one-step-ahead forecast error).[19]

For each of these 25 specifications, we estimate the models on a full sample and on a series of 12 rolling samples. That is, starting with a sample ranging from January 2002 to July 2008, we

[19]We chose 25 models because 25 models showed similarly accurate out-of-sample performance. For other levels of aggregation the number of selected models may differ. The discarded specifications would not perform better than included ones in our test because they either have poor fit both in- and out-of-sample, or because they would not produce any losses in stress scenarios due to insignificant or "wrong sign" coefficient on macroeconomic variables included in the scenarios.

estimate each model and then construct out-of-sample predictions from July 2008 for the next 9 quarters.[20] We then repeat the exercise on an estimation sample ranging from January 2002 to August 2008, and then perform a 9-quarter ahead forecast. We proceed in this fashion (12 times) so that the estimation sample gradually increases in size. In our longest subsample, we estimate the model up through June 2009. This set of rolling windows allows us to see how the model performs as it gradually learns about the dynamics of home equity defaults during the financial crisis and its aftermath. As we outlined in section 2, the short samples used in the forecasting exercises should not unduly handicap any particular model type. Both types of biases (measurement error and aggregation bias) arise from cross-sectional correlations in unobservables that do not depend on the length of the available time series.

We select the specification that performs best out-of-sample across the 12 rolling windows in terms of the mean squared forecast error and does not underpredict default probabilities, to be consistent with the goal of stress testing exercise to produce high losses in stress scenarios. We refer to such model as a winning specification and report all 12 regressions for this specification in Table 6 in the Appendix. We also retain information on the performance of the remaining 24 specifications so that we can assess the volatility of the cross-model performance measures at different aggregation levels.

We follow this same basic model selection procedure for all levels of aggregation: portfolio or aggregate level, the intermediate sub-portfolio level, as well as the loan-level.

[20]We chose 9 quarters because this is the time horizon for the loss projections in the supervisory stress testing.

4.2 Intermediate levels of aggregation

With a fully aggregated model in hand the next step is to break the delinquency rate down into subportfolios. We consider four different schemes for aggregation: by county, DTI decile, LTV decile, and FICO score decile.[21] The estimated model is now a fixed-effect panel model,

$$y_{jt} = \theta_j + X_{jt}\beta + \eta_{jt}, \tag{2}$$

where the θ_j is a subportfolio-specific fixed-effect, $y_{jt} = \frac{1}{J}\sum_{i=1}^{N_J} I(i \in j)y_{it}$ is the average default rate for all loans in sub-portfolio j in month t, X_{jt} is the set of average values of covariates for each sub-portfolio, and η_t is the error term. Each subportfolio boundary is static; these are based on the entire sample. When we aggregate to the portfolio level, each sub-portfolio, j, is then weighted according to the number of observed loans within these boundaries.

The subportfolio approach preserves some of the potential for aggregating out the measurement error problem, while also offering flexibility to introduce more portfolio-specific information to the regressions. In the disaggregated models we make predictions of the disaggregated delinquency rates and then aggregate these predictions to compare to the aggregate outcomes. That is, when we forecast default at times $t = 1, ..., T$ for subportfolio j, the forecast error that we use for forecast evaluation is not the average difference between predicted and actual subportfolio default rates. Rather, it is the difference between the average aggregate default prediction and the aggregated

[21] Only top 25 percent of counties by the share in the stock of loans as of 2005 comprise the county data set. This helps to improve the fit of the model by eliminating noise from counties with too few observations.

portfolio default rate,

$$FE = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - \frac{1}{J}\sum_{j=1}^{J}\frac{N_J}{N}\hat{y}_{jt}\right)^2. \tag{3}$$

We feel that this approach more closely mimics what a bank would do when confronted with a problem of predicting total portfolio defaults or losses. If the object of interest is the portfolio default rate or loss rate, then the appropriate measure of out-of-sample fit is one where forecast error is computed at the portfolio level.

For these subportfolio models we conduct the same procedure as for aggregate model in terms of specification selection. After extensive pre-testing in which we experiment with variety of lags, interaction terms, inclusion and exclusion of different variables, as well as subportfolio-level fixed effects, we end up with 22 reasonable specifications for each type of aggregation: by county as well as by deciles of DTI, LTV, FICO score.[22] The specifications are different for each type of aggregation, but most of them include subportfolio or county fixed effects. We estimate each of these models for the same 12 rolling windows as the aggregate model and select, for each type of aggregation, the model that performs best on average across the rolling windows, the winning model. Winning models for each of the four subportfolio levels and for each sample window are reported in Appendix Tables 6 through 9. We retain information on the performance of the remaining models.

[22]We selected 22 specifications for the last stage of model selection because there was a fairly pronounced cluster of 22 models with similar performance, as opposed to the 25 similarly-performing models that we observed in the aggregate case.

4.3 Bottom-up model

Finally, we consider fully disaggregated loan-level models. These models are estimated as logit regressions,

$$Pr(y_{it} = 1) = \frac{\exp(\alpha + X_{it}\beta)}{1 + \exp(\alpha + X_{it}\beta)} \tag{4}$$

where i is the index for individual loans or borrowers and y_{it} is a 0-1 indicator of whether loan i defaulted in month t. Some variables in the vector of covariates X_{it} , such as unemployment rate and home price depreciation, do not vary by loan but are repeated for all loans in the same county or zip code in the same month. We cluster standard errors by county to account for resulting correlation in errors.[23]

We select 18 reasonable models among all specification permutations, in which we experiment with lags of macroeconomic and loan-level variables, interaction combinations, and fixed effects. It turns out that models with loan fixed effects do not perform well. On the other hand, models that include splines for FICO scores, LTV, and loan age, as well as vintage fixed effects, do better. For a wide variety of specification, which we discard, the coefficients on macroeconomic variables are either negligibly small or have the wrong sign. Such specifications will not produce losses in stress scenarios and are therefore not useful for our analysis. The remaining 18 specifications have reasonable in- and out-of-sample fit and expected signs on macroeconomic variables. These

[23] Logit specification used here is consistent with what is most commonly used in the literature. As a robustness test, we conducted the following analysis using linear probability model instead of logit. The results are not materially different and model performance is slightly worse out of sample with linear probability approach.

specifications include a variety of loan-specific controls that improve precision of these models.[24] For the 18 retained specifications, we evaluate forecasting performance over 12 rolling regressions ending in July 2008 through June 2009. As with the subportfolio models, the forecast error is calculated as the average deviation of the aggregated fitted default probabilities compared to the aggregate default rate,

$$FE = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - \frac{1}{N}\sum_{i=1}^{N}\hat{y}_{it}\right)^2. \tag{5}$$

Once again we select the winning model that has the best in- and out-of-sample fit across 12 estimation windows among the 18 models and report the results for all 12 rolling windows in the Appendix Table 10. We retain performance data for the remaining 17 models.

4.4 Comparison across levels of aggregation

The winning models for each aggregation approach are reported in Table 2. In order to not overwhelm the reader with the results, we only report the regressions that are estimated through January 2009, the middle of our rolling window set. We can see that the best specification varies with the aggregation approach, but the effects of included variables are mostly stable across specifications.[25] We find, as one would expect, that defaults on home equity loans are more likely when unemployment rate (UR) is higher, or when the rate of home prices depreciation (HPD) is

[24] See Appendix for the specification of the winning model that lists such variables.

[25] Note that we did not explicitly rule out that the same specification could be best for all aggregation levels — this is an empirical result. Running a comparison across aggregation level of the same specification will disadvantage less aggregated model because of necessary parsimony needed for top-down models.

higher.[26] The combination of these factors seems to lead to an additional increase in default rates. We also find in most specifications that higher debt-to-income ratios and lower FICO scores of the borrowers are associated with higher default rates.

4.5 Model performance tests

Before turning to the out-of-sample results we first examine the in-sample performance of our models when estimated over the full sample of data. These plots of in-sample fitted default probabilities can be found in Figure 2. The models profiled here are actually selected on the basis of out-of-sample performance. However, it is useful to demonstrate from the beginning that all aggregation levels demonstrate a similar capability of fitting the data in-sample. All four model categories (loan level, county level, segment level, and aggregate) can roughly match the timing of the turning point in delinquency series. None of the four models is quite able to match the peak in defaults observed. We see from Table 2 that the county-level model performs slightly better than the others in terms of forecast error. But, from a visual perspective at least, there is little a priori reason to expect one particular aggregation level to dominate in the out-of-sample predictions.

With this starting point we can proceed to the out-of-sample comparisons. Figures 3-5 show the performance of the out-of-sample forecast of our winning models. Figure 3 shows the forecasts of the regressions with the estimation window through July 2008, Figure 4 through January 2009 (regressions reported in Table 2), and Figure 5 through June 2009. We can compare the forecasts resulting from each approach with the data. We find that the loan-level forecast consistently under-

[26]Note that in the regression with negative effect of the first lag of HPD (the "wrong" sign) the full effect includes additional lags or interaction terms and is overall positive, as expected.

predicts default frequency in the aggregate, while aggregate forecasts over-predicts default frequency in the beginning of our rolling window. County-level aggregation models produce forecasts that are quite accurate and stable, with the exception of the second half of the forecast horizon in the regressions that end before the crisis (top panel).[27]

We can formalize these observations by comparing average forecast errors of all reasonable models we estimated across all 12 rolling forecast windows for each aggregation approach. Table 3 presents summary statistics for all resulting forecast errors, by aggregation approach, as well as average forecast error for each winning model across 12 rolling forecast windows. We find that on average, county-level models have the smallest forecast errors, which also don't vary much across specifications. This is consistent with our expectations, because macroeconomic information that enters the regressions varies by county and is therefore fully explored in these regressions, while not generating measurement errors as in loan-level regressions. While the forecast error of the winning model of the portfolio-level approach is smaller than that of the county-level, we can see that there is high variation in the precision of the aggregate model projection resulting from small perturbations in model specification.

In the stress testing exercise accuracy of projection is the primary, but not the only, goal. Given model uncertainty it is also important that the errors of the forecast are more likely to be on the conservative side. In particular, this would be true from a financial stability perspective where underpredictions may be viewed more seriously than overpredictions. Thus, we construct a "con-

[27]To reduce clutter in the charts we do not plot the result for LTV, DTI, and FICO score buckets.

servative loss" measure

$$CL = \frac{1}{T} \sum_{t=1}^{T} \exp(y_t - \hat{y}_t), \tag{6}$$

where $\hat{y}_t$ for disaggregated models is computed as average projections. This measure is equal to 1 if there is no error, is below 1 if the error is on the side of over-predicting default frequency, and is above 1 if the model is under-predicting defaults. Summary statistics for this loss measure are presented for all reasonable models across all 12 forecast windows for each of our aggregation approaches in Table 4. We find that the aggregate model produces more conservative forecasts on average, as we saw in Figure 1, but that the loss measure varies substantially across specifications. The sub-portfolio models all have similar loss measures on average with county-level loss measures being the most stable across regression specifications. The loan-level models have the least variation across forecast windows according to the loss measure. However, this low variation in the loss measure comes at the expense of a very high mean loss. This appears to be one of the most robust results in the out-of sample analysis. Across all forecast windows and loss functions, the loan level models have a strong tendency to underpredict defaults.

It is important to remember that each winning specification used in Tables 3 and 4 is different across each of our aggregation levels. Given our ultimate objective of comparing models on the basis of measures such as in- and out-of-sample fit and stability, it is natural to allow each aggregation level to be represented by its best candidate model, whatever that may be. But another way to compare models that focuses squarely on the effect of aggregation is to compare model performance across different aggregation levels with the *same* specification.

These comparisons can be seen in Table 5, which reports a series of performance measures for models estimated over the January 2002 through June 2008 window and then used to project default rates over the rest of the sample period. In the top panel we show the mean-squared errors of the model predictions. In the bottom panel we use the conservative loss measure from equation 6. Each row of the Table 5 shows the model performance while holding the regression specification fixed. For example, in the top row of the top panel, the specification is the one selected as the best aggregate model from Table 2. This model has a mean-squared forecast error of 0.005. The other columns of the first row are the mean-squared forecast errors associated with the best aggregate model estimated and implemented over the different aggregation levels. For example, when the best aggregate model is disaggregated to the loan level, the mean-squared forecast error is 0.007.

Table 5 shows that for most levels of aggregation, the mean-squared errors and loss measures are lower when the winning models are applied as opposed to models selected for other aggregation levels (within-column comparison). Similarly, each specification performs best when applied at the level of aggregation it was chosen for (within-row comparison). The big exception to this rule is for the loan-level model. For both mean-squared error and conservative loss, the winning loan-level model performs worse than when the same model is applied to each of the other levels of model aggregation considered in this paper.[28] Moreover, using more parsimonious models without loan-level controls chosen for aggregate analysis also produces better performance than the winning loan-level model.[29] A robust conclusion from this analysis, however, is that there is no specification

[28]Including county fixed effects in the loan-level model, a specification not chosen because of poor in-sample performance, marginally improves loan-level model results, but they remain still substantially worse than aggregated models.

[29]These parsimonious models were not considered in our selection process because effectively they are the same as aggregated models, but with repeated observations, making them less efficient and their standard errors biased.

in which loan-level model performs best out of sample in our data. This is consistent with the measurement error bias that we discussed previously.

4.6 Stress test projections

As a final exercise we compare projections of the different champion models using the macroeconomic scenarios deployed in the CCAR 2014 stress tests. The Federal Reserve releases three different economic scenarios that financial institutions use to stress their balance sheet exposures. We focus here on the 2014 "severely adverse" scenario which features a deep recession much like the one experienced by the U.S. starting in 2007. The scenario consists of a nine-quarter path for a large set of variables measuring economic activity.[30] In Figure 6 we plot the paths for the key variables that go into the home equity default models. House prices were assumed to fall through the scenario 10 percent year-over-year rates before leveling off at the end of the scenario horizon in 2016. The unemployment shock was particularly severe in the CCAR 2014 exercise, with the rate climbing to about 11 percent at the peak.

In the specification searches described earlier, we allow the models at each aggregation level to make use of the real GDP growth measure, house prices changes, and unemployment. However, none of the champion specifications at any aggregation level selected GDP growth. Similarly, the current interest rate on the home equity line of credit was among the set of variables available for the loan-level specifications. All of the home equity lines in our sample have payments based on

Additionally, the parsimonious models applied at the loan level often have worse in-sample fit and did not pass through our model selection procedure.

[30] See http://www.federalreserve.gov/bankinforeg/bcreg20131101a1.pdf for the exact specification of the scenarios and variables included.

short-term interest rates that change over time (i.e., adjustable-rate lines of credit), and proxies for short-term rates are included in the macroeconomic scenarios. But again, this variable was not selected in the champion loan-level model. Thus, the stress projections from this set of models are wholly determined by the assumed scenarios for house prices and the unemployment rate.

For the aggregate model and most of the segmented models, it is fairly routine to simulate out projected default rates. The first thing we do is convert the monthly data of our default models into the quarterly frequency of the CCAR scenarios. We do this by keeping the last monthly-observation of each loan in every quarter. Next, the champion model specifications listed in Table 2 are estimated using data through the third quarter of 2013, as the fourth quarter of 2013 is the start of the CCAR 2014 exercise. The loans in the portfolio and their relevant characteristics at origination are held constant as of this date. We then use the estimated model coefficients and the assumed paths of house prices and unemployment to project out aggregate and segmented portfolio default rates. In the case of the segmented portfolios, these default projections are then aggregated to produce projected aggregate default rates under the CCAR scenarios. In the case of the loan-level and the county-level models, we must perform an extra step to map the aggregate path for house prices and unemployment down to the local level. We do this using a simple regression model. Specifically, we model the relationships between aggregate house price depreciation (HPD) and unemployment (UR) and their zip-code and county-level counterparts, respectively, as follows:

$$HPA_{kt} = \alpha_j + \beta_j HPA_t^{agg} + \epsilon_{kt},$$

$$UR_{jt} = \theta_j + \gamma_j UR_t^{agg} + \eta_{jt},$$

for each zip code k or county j in our sample. With the estimated coefficients from these models in hand, we simulate out county-level variables, as required by the champion county-level and loan-level models.

The results from the stress test simulation are in Figure 7. In this chart we follow the CCAR stress test convention and report default rates at the quarterly frequency. Turning to results, we note the broad contours of the aggregate, county-level, and loan-level projections are quite similar. This is not surprising given that these same macro risk factors enter into these models, just at different levels of aggregation. Interestingly, we still see that the loan-level model has slightly less sensitivity to shocks than the aggregate and the county-level models. Of course, we cannot say for sure whether this is a virtue or a shortcoming. We are comparing default projections in response to hypothetical shocks that have not actually occurred. Even so, the attenuation that we see in the risk factor sensitivities in the loan-level model still appears to be present in these simulations that make use of the full sample of available data.

4.7 Discussion

We observed in Figure 2 that winning models for all aggregation levels perform quite well in sample. Why then do loan-level models perform substantially worse out of sample? Our primary explanation is that this effect is due to attenuation bias in the coefficients on macroeconomic variables in loan-level regressions. This bias is clearly visible in the dynamics of models out-of-sample fit as we roll the window forward. For example, as we see in Figure 3, both loan-level and county-level models underpredict default rates out of sample. However, as we roll the window forward, subportfolio models become more sensitive to macroeconomic variables, as seen by increasing coefficients on unemployment rate (UR), house price depreciation (HPD), as well as their interactions and lags, reported in the Appendix. This means that in these models, as we roll the window forward, not only worsening macroeconomic conditions, but also increased sensitivity of the default rate to these conditions lead to higher predicted default rates. For loan-level models, there is no such increase in sensitivity over time, resulting in a much more muted increase in predicted default rate, even in-sample, as seen most clearly in Figure 5.

This attenuation bias is also apparent in Figure 7: the only difference between projected default rate dynamics across aggregation levels in this Figure is due to different estimated sensitivity of default rates to macroeconomic characteristics. We can see that loan-level regressions project lower losses for the same macroeconomic scenario, indicating that they are less sensitive to macroeconomic variables. This is not specific to the winning model for the loan level regressions. We can see from Table 4 that even the lowest loss statistic across loan-level models (1.13) is greater than one and is higher than the loss statistics for winning models at all levels of aggregation, meaning that even the model that produces the highest default rates across all loan-level models underpredicts defaults relative to data and relative to winning models at all levels of aggregation. On the other hand, the most conservative of aggregated model and all intermediate level of aggregation models overpredict default rates relative to the data.

5 Conclusion

Modern applications of bank stress testing are characterized by the specification of adverse macroeconomic and financial market scenarios. These aggregate scenarios are then paired with less aggregated data on risk factors—sometimes available at the loan level—to produce loss estimates for individual banks under the scenarios. The question that arises is whether it is better to aggregate the scenario and risk factor data first and then estimate the risk model, or estimate the loan-level model and then aggregate the individual loan loss projections, or estimate some intermediate aggregation level model. We demonstrate that different aggregation levels may give rise to different econometric problems. In principle, loan-level models may be subject to measurement errors that arise from the use of economic risk factors that are not available at the individual or loan level. Aggregate models may be subject to aggregation bias. The optimal level of aggregation, then, is an empirical question.

We use a large portfolio of home equity lines of credit to assess which model aggregation levels display the best performance in predicting default rates out of sample. We compare models estimated at an aggregate portfolio level, at the loan level, and also hybrid approaches where we model default probabilities for different segments of a portfolio, such as buckets of debt-to-income ratios, loan-to-value ratios, FICO risk scores, and with loans aggregated by county. In our empirical exercise we find that this tension is best resolved at the intermediate level of aggregation. In particular, county-level models, where macroeconomic variables at county level are used, appear to perform best for the purpose of stress testing. We measure model performance using two different criteria appropriate for the stress testing exercise. We use a forecast error approach that puts equal weight on positive and negative forecast errors. Policymakers and bank supervisors, however, are often thought to have preferences that put more weight on downside risks than upside risks. For this reason, we also employ a "conservative loss" measure which punishes model underpredictions. In this context, loan-level models appear to perform particularly poorly, given their persistent underprediction of home equity default rates. While aggregate models are quite conservative on average, their predictions are not robust to model specification and can at times produce very low default rates.

To be clear, our goal is not to recommend one specific level of aggregation for risk modeling. The purpose of our exercise is to illustrate that the choice of aggregation level matters. In some cases, intermediate levels of aggregation might be a best approach to modeling default probabilities or loss rates on banks' loan portfolios. We also provide an econometric argument to provide some intuition why this might be the case. Our hope is that researchers, regulators, and practitioners alike devote due attention to the implications of the aggregation level of models used for stress testing. Availability of the loan-level data allows institutions to consider aggregation level as a key factor in model design.

References

Arellano, M. (1987): "Computing Robust Standard Errors for Within-Group Estimators," *Oxford Bulletin of Economics and Statistics*, 49, 431–434.

Bond, P., R. Elul, S. Garyn-Tal, and D. K. Musto (2014): "Does Junior Inherit? Refinancing and Blocking Power of Second Mortgages," Federal Reserve Bank of Philadelphia working paper No. 13-3.

Epouhe, O., and A. Hall (2015): "Payment Shock in HELOCS at the End of the Draw Period," Federal Reserve Bank of Philadelphia Working Paper.

Frame, W. S., K. Gerardi, and P. Willen (2015): "The Failure of Supervisory Stress Testing: Fannie Mae, Freddie Mac, and OFHEO," Federal Reserve Bank of Atlanta working paper.

Gerardi, K., K. Herkenhoff, L. Ohanian, and P. Willen (2015): "Can't pay or won't pay? Unemployment, negative equity, and strategic default," NBER working paper No. 21630.

Granger, C. (1980): "Long memory relationships and the aggregation of dynamic models," *Journal of Econometrics*, 14, 227–238.

Grunfeld, Y., and Z. Griliches (1960): "Is Aggregation Necessarily Bad?," *Review of Economics and Statistics*, 42(1), 1–13.

Gyourko, J., and J. Tracy (2013): "Unemployment and Unobserved Credit Risk in the FHA Single Family Mortgage Insurance Fund," NBER Working Paper No 18880.

Hirtle, B., A. Kovner, J. Vickery, and M. Bhanot (2016): "Assessing Financial Stability: The Capital and Loss Assessment under Stress Scenarios (CLASS) Model," .

Hurst, E., and F. Stafford (2004): "Home is Where the Equity Is: Liquidity Constraints, Refinancing, and Consumption," *Journal of Money, Credit, and Banking*, 36(6), 985–1014.

Johnson, K., and R. Sarama (2015): "End of the Line: Behavior of Credit-Constrained HELOC Borrowers," Federal Reserve Board of Governors Working Paper.

Mian, A. R., and A. Sufi (2009): "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis," *Quarterly Journal of Economics*, 124(4).

Pesaran, M., R. Pierse, and M. Kumar (1989): "Econometric Analysis of Aggregation in the Context of Linear Prediction Models," *Econometrica*, 57(4), 861–888.

Rajan, U., A. Seru, and V. Vig (2015): "The failure of models that predict failure: distance, incentives, and defaults," *Journal of Financial Economics*, 115(2), 237–260.

Theil, H. (1954): *Linear Aggregation of Economic Relations*. Amsterdam.

Table 1: Summary Statistics

	Mean	Standard Deviation	25th Percentile	75th Percentile
Loan-to-Value at origination (%)	37.765	26.073	15.57	60.11
FICO at origination	739.451	51.109	703	779
Debt-to-Income at origination (%)	35.667	14.527	26.8	44.7
Unemployment Rate, County (%)	6.961	2.884	4.7	8.9
GDP growth, yearly (%)	1.614	2.072	1.285	2.856
House Price Depreciation, yearly (%)	-0.118	12.397	-7.669	6.892
Loan Origination Balance, log	11.044	1.008	10.463	11.575
Margin Rate (spread over benchmark, %)	0.548	0.995	0	.875
Current Interest Rate (on HELOC, %)	5.664	2.310	3.74	7.625
Loan length (months)	222.136	154.819	0	360
Current Utilization Rate	0.686	.613	0.154	0.989
Home appraisal at origination ($000)	465	11,400	216	560
Full documentation	0.326			
Interest-only loan	0.649			
Loan purpose: purchase	0.055			
Observations	20,757,776			

Note: for loan-level models the non-missing value sample size is 13,920,237 observations. Summary statistics for current utilization rate, home appraisal value at origination, full doc dummy, interest-only dummy, and purchase loan dummy are based on this smaller number of observations.

Table 2: Best Models: Regression through January 2009

	Aggregate	DTI	LTV	FICO	County	Loan Level
HPD, lag(1)	-0.0025	0.0273***	-0.0128***	0.0269*	-0.00527***	0.0489***
	(0.0039)	(0.0070)	(0.0032)	(0.0127)	(0.00198)	(0.0018)
UR, lag(1)	0.0330***	0.0127**	0.0059	0.0201***	0.02091***	0.0651***
	(0.0069)	(0.0042)	(0.0086)	(0.0031)	(0.0044)	(0.0150)
HPD*UR, lag(1)	0.0020***		0.0041***		0.0016***	
	(0.0007)		(0.0008)		(0.0004)	
HPD, lag(2)		-0.0564***		-0.0368	0.0058***	
		(0.0140)		(0.0211)	(0.0021)	
UR, lag(2)		0.0200***		0.0258***	0.0049	
		(0.0042)		(0.0054)	(0.0045)	
HPD*UR, lag(2)					-0.0003	
					(0.0004)	
HPD, lag(3)		0.0350***		0.0194		
		(0.0078)		(0.0107)		
DTI, lag(1)	0.0109***	-0.0102**	0.0001	0.0208***	0.00428***	-0.0031***
	(0.0011)	(0.0044)	(0.0021)	(0.0059)	(0.00054)	(0.0010)
FICO, lag(1)		0.0073***		-0.0122**		spline
		(0.0019)		(0.0051)		
LTV, lag(1)	0.0049**					spline (imputed)
	(0.0020)					
Loan Amount, lag(1)		0.1514***		-0.3764		
		(0.0385)		(0.2264)		
Constant	-0.5959***					-8.4449***
	(0.1007)					(1.6041)
Additional Loan Characteristics[a]	No	No	No	No	No	Yes
Fixed Effects	No	Yes	Yes	Yes	Yes	No
Observations	83	810	830	810	24854	6208287
Adjusted R^2	0.9300	0.8557	0.6387	0.6363	0.3655	0.1961
Forecast error	0.0046	0.0070	0.0135	0.0046	0.0041	0.0617
Loss	0.9766	0.9832	1.0576	1.0346	1.0293	1.2754

Notes: [a] Vintage fixed effects, loan age spline, additional variables as listed in Appendix.
Standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$. Robust standard errors.

Table 3: Forecast error summary: Based on 9-quarter forecast

	Aggregate	FICO	LTV	DTI	County	Loan Level
Mean	0.0325	0.0416	0.0417	0.0356	0.0158	0.0845
Std Dev	0.0478	0.0759	0.0784	0.0605	0.0141	0.0130
Min.	0.0035	0.0035	0.0038	0.0053	0.0031	0.0180
Max.	0.2683	0.6892	0.695	0.4984	0.0709	0.1153
Winning Model (average forecast error)	0.0064	0.0073	0.0197	0.0221	0.0080	0.0570

All summary statistics refer to forecast error defined in equation 5 from all models deemed reasonable across all 12 rolling windows. For the winning model the average forecast error is computed across 12 rolling windows. See Appendix for winning model regression.

Table 4: Loss Summary

	Aggregate	FICO	LTV	DTI	County	Loan Level
Mean	0.9655	1.020	1.108	1.0721	1.0541	1.3178
Std Dev	0.1598	0.1813	0.187	0.1795	0.1015	0.0355
Min.	0.6709	0.4922	0.6927	0.6171	0.8463	1.1288
Max.	1.6425	1.7003	2.282	1.9814	1.2901	1.3978
Winning Model (average CL)	0.9837	1.0412	1.0750	1.1192	1.0478	1.2494

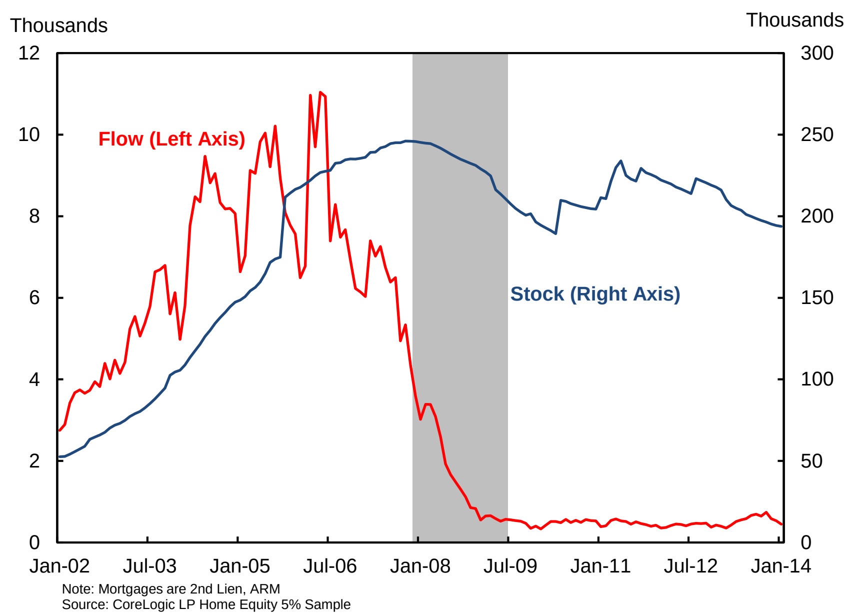
All summary statistics refer to CL defined in equation 6 from all models deemed reasonable across all 12 rolling windows. For the winning model the average CL is computed across 12 rolling windows. See Appendix for winning model regression.

Table 5: Forecast error summary for the same specification for all models

	Aggregate	FICO	DTI	LTV	County	Loan Level
			MSE			
Best Agg	**0.005**	0.074	0.017	0.010	0.005	0.007
Best FICO	0.009	**0.004**	0.007	0.045	0.016	0.020
Best DTI	0.009	0.004	**0.007**	0.045	0.016	0.020
Best LTV	0.007	0.065	0.015	**0.010**	0.005	0.008
Best County*	0.009	0.099	0.014	0.009	**0.005**	0.009
Best LL**	0.033	0.006	0.022	0.050	0.005	**0.065**
			Loss			
Best Agg	**0.98**	0.78	1.09	1.05	1.03	1.06
Best FICO	0.98	**1.01**	1.05	1.23	1.12	1.14
Best DTI	0.98	1.01	**1.05**	1.23	1.12	1.14
Best LTV	0.95	0.79	1.08	**1.05**	1.03	1.06
Best County*	0.93	0.75	1.07	1.04	**1.03**	1.07
Best LL**	0.89	0.97	0.92	0.86	1.02	**1.28**

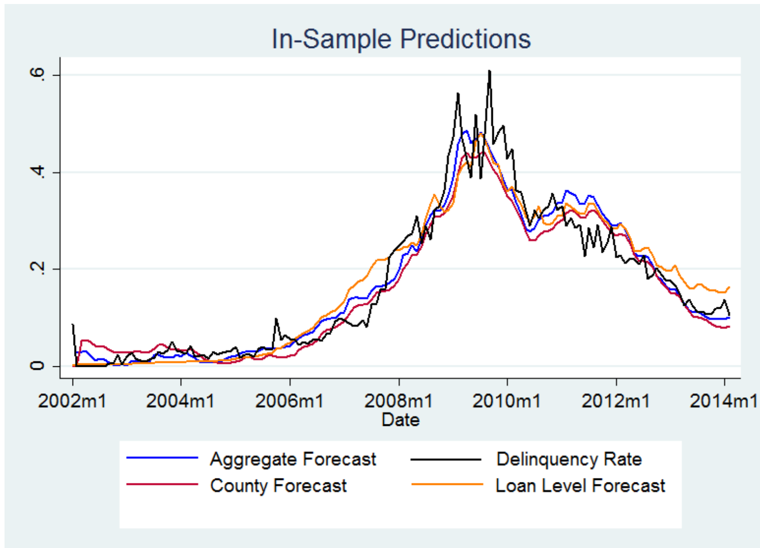
Each row of the table shows mean-squared error or conservative loss measures while holding the specification fixed and then re-estimating the models over different aggregation levels. All models are estimated over the January 2002 to June 2008 window. The mean-squared error is calculated as in equation 5. The conservative loss estimate is calculated as in equation 6. Bold font denotes the performance measure associated with the winning model from an aggregation level estimated and implemented over the intended aggregation level. * No county fixed effects in aggregate or sub-portfolio models. ** Weighted average of loan characteristic is used for aggregate models.

Figure 1: Flow and Stock of Loans in Sample



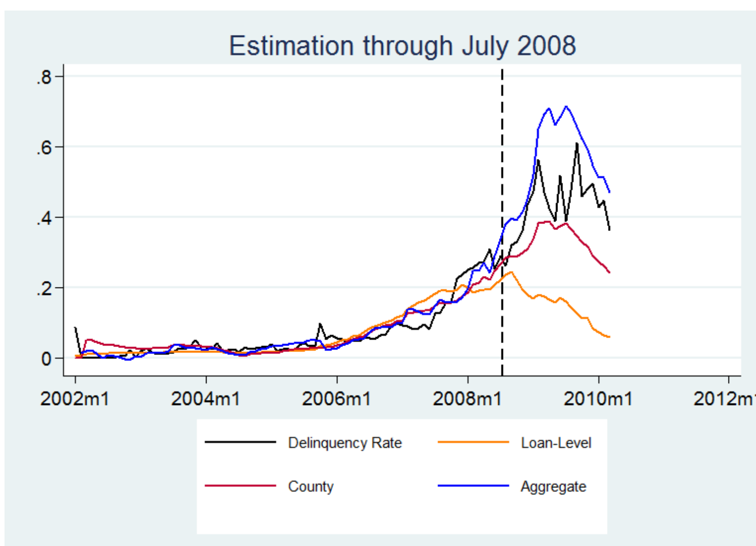
Note: Flow is the number of new HELOCs originated in a given month in our sample. Stock is the number of loans outstanding in a given month in our sample.

Figure 2: In-sample Model Predictions



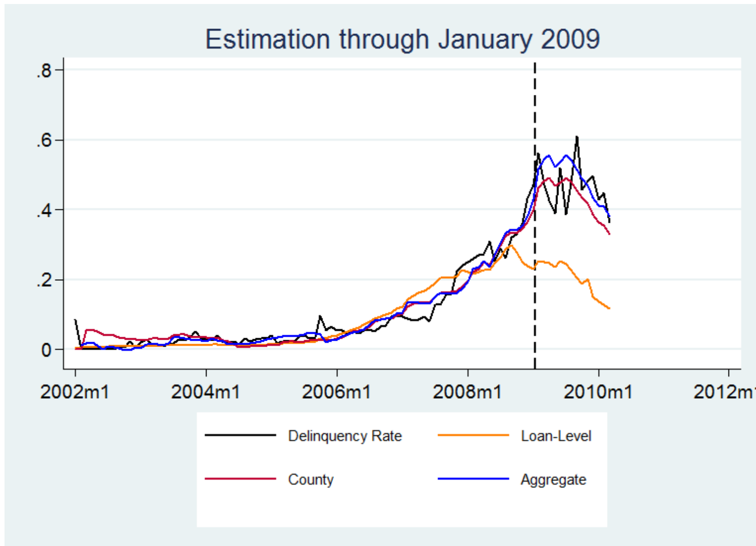
Note: Actual and predicted delinquency rates (in percent) are plotted.
Winning models for each level of aggregation estimated on the full sample are used.

Figure 3: Out-of-Sample Model Predictions



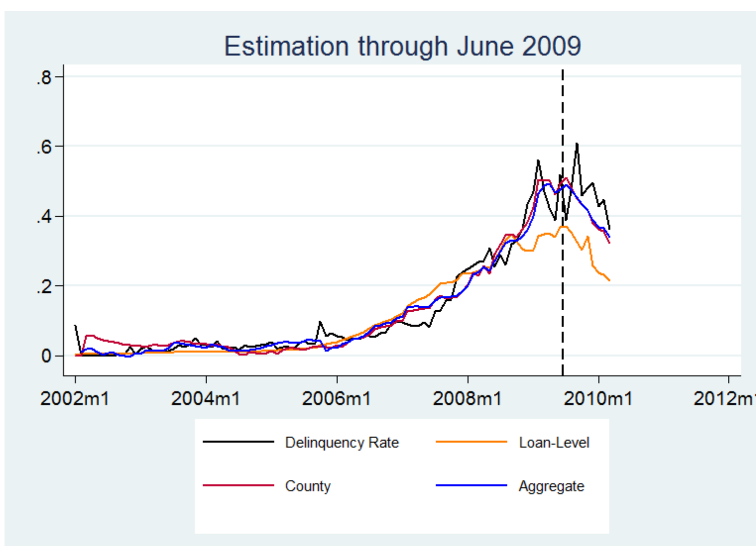
Note: dashed line indicates the last month of estimation sample.
Actual and predicted delinquency rates (in percent) are plotted.

Figure 4: Out-of-Sample Model Predictions



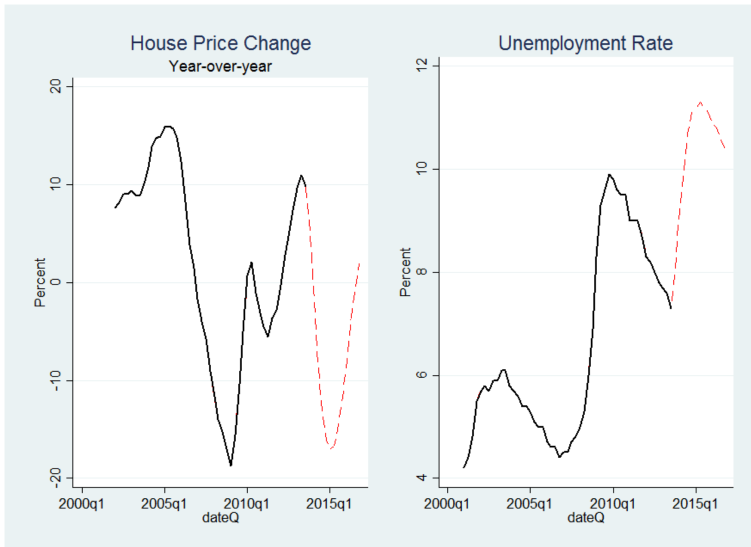
Note: dashed line indicates the last month of estimation sample.
Actual and predicted delinquency rates (in percent) are plotted.

Figure 5: Out-of-Sample Model Predictions



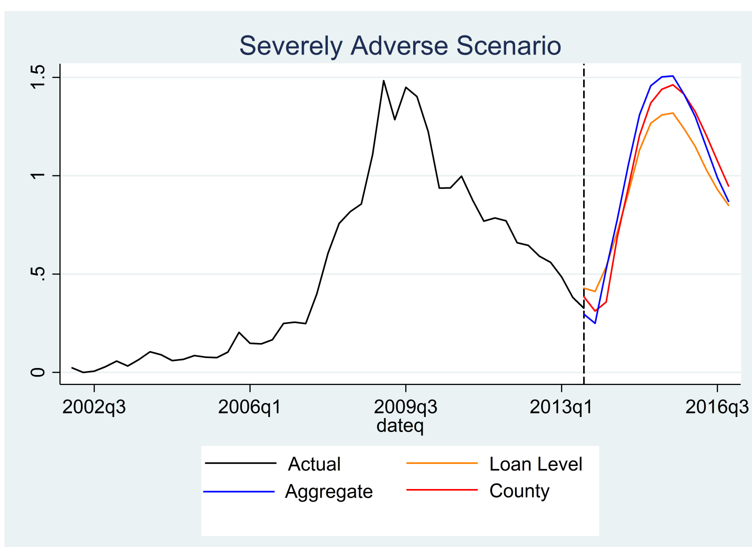
Note: dashed line indicates the last month of estimation sample.
Actual and predicted delinquency rates (in percent) are plotted.

Figure 6: CCAR 2014 Stress Scenarios



Source: Federal Reserve Board.

Figure 7: Stress Scenario Predictions



Notes: Projections are constructed from winning models for each level of aggregation estimated on the sample ending in 2013:Q3.

6 Appendix 1

Here we present the proof for the statements made in the Econometrics Framework section.

6.1 Conditional forecast precision

Our goal is to predict default rate $y \in [0, 1]$ on the entire portfolio given macroeconomic scenarios. The macroeconomic variables x do not vary by loan in portfolio, although some macroeconomic variables might vary by geographical segments of portfolio. For simplicity of notation, suppose we are only predicting one period forward, that is predicting y_{T+1} given x_{T+1} and observed history of y's and x's up to period T. Suppose the data generating process (DGP) is such that

$$y_t = X'\beta + \varepsilon,$$

where y is a vector of observed default rates (or, in case of individual loans, default indicators) over time, X is a matrix of observed covariates, including constant term, unobserved disturbance ε is distributed $N(0, \sigma^2)$. We can use linear regression to estimate b, the estimator for β, and $\hat{\sigma}$, the estimator for σ.

Suppose y and X are observed at individual loan level, and there are N loans observed for T time periods. Therefore, we have a choice of whether to estimate b and $\hat{\sigma}$ on individual loan data (using linear probability regression for the ease of exposition), on average values of y and X for sub-portfolios of any type (using pooled or fixed effects panel regression), or on overall portfolio averages (using time series regression). Regardless of the regression estimated, the conditional forecast can be constructed by substituting b for β in the DGP equation above.

First, consider the case where, that regardless of the aggregation level, we can obtain an unbiased estimate of β. Therefore, aggregation level will not affect the expected forecast mean. That is, assume that all the individual observations are i.i.d. In case of unbiased estimates we are only concerned with the precision of our forecast. Below we show that precision of conditional forecast will be determined by differences in estimated variance of the disturbance term $\hat{\sigma}$, the number of loans and sub-portfolios, and differences in the variance of covariates.

Denote y_L and X_L the observables measured at loan level, y_P and X_P those at portfolio level, and y_B and X_B those at sub-portfolio, or bucket, level. Portfolio and sub-portfolio variables can be expressed as averages of loan-level data:

$$y_P = \frac{1}{N} S'_N y_L, \;\; X_P = \frac{1}{N} S'_N X_L,$$

where S_N is an $(NT \times T)$ summation matrix such that each element of y_P and each row of X_P are sums of elements in a given time t.[31] Similarly,

$$y_B = \frac{1}{J} S'_J y_L, \;\; X_B = \frac{1}{J} S'_J X_L,$$

where $1 < J < N$ is the number of sub-portfolios, S_J is an $(NT \times JT)$ summation matrix such that each element of y_B and each row of X_B is the sum for a given t of all the elements of subportfolio

[31] $S_N = I_T \otimes 1_N$, where I_T is a $(T \times T)$ identity matrix and 1_N is a vector of N ones.

j.[32]

One can show that differences in Brier score (measuring variance of conditional forecast) for predicting y_{T+1} using regressions with different level of aggregation is

$$BS_P = \hat{\sigma_P}^2(1 + N^2 x'_{t+1}(X'_L U_N X_L)^{-1} x_{T+1}),$$

where x_{T+1} is a $(K \times 1)$ vector of covariates given for time $T+1$, the scenario. Using loan-level regression, predicting individual default probabilities, and aggregating them up would produce squared forecast error

$$BS_L = \hat{\sigma_L}^2(\frac{1}{N^2} + x'_{t+1}(X'_L X_L)^{-1} x_{T+1}).$$

Similarly, at the sub-portfolio level, predicting sub-portfolio default rates and aggregating them up to portfolio level would produce

$$BS_B = \hat{\sigma_B}^2(\frac{1}{J^2} + J^2 x'_{t+1}(X'_L U_J X_L)^{-1} x_{T+1}),$$

Unbiased estimates of σ^2 in each of the aggregation cases are given as follows

$$\hat{\sigma_P}^2 = \frac{1}{T-K} y'_P(I - X_P Q_P^{-1} X'_P) y_P = \frac{1}{T-K}\frac{1}{N^2} y'_L S_N (I - S'_N X_L (X'_L U_N X_L)^{-1} X'_L S_N) S'_N y_L.$$

[32] $S_J = I_T \otimes I_J \otimes 1_{N_J}$ in a special case of all subportfolios having the same number of observations, so that $J * N_J = N$.

For the individual level,

$$\hat{\sigma_L}^2 = \frac{1}{NT-K}\, y_L'(I - X_L(X_L'X_L)^{-1}X_L')y_L.$$

For the sub-portfolio level,

$$\hat{\sigma_B}^2 = \frac{1}{JT-K}\, y_B'(I - X_B Q_B^{-1} X_B')y_B = \frac{1}{JT-K}\frac{1}{J^2}\, y_L' S_J(I - S_J' X_L(X_L'U_J X_L)^{-1}X_L'S_J)S_J'y_L.$$

Thus, the differences in the forecast errors will be determined by differences in estimated variance of the disturbance term $\hat{\sigma}$, the number of loans and sub-portfolios, and differences in inverse sum of squared covariates. If the observations are i.i.d., different aggregation levels will give the same results in the limit. However, in finite samples, even if observations are drawn from i.i.d. distributions, there will be differences in forecast errors, depending on a sample. They will generally be larger the more aggregated the regression sample is.

6.2 Measurement error

Define observables Z, a subset of X, that is only observable at aggregation level of sub-portfolios. Thus, for loan i in subportfolio j and time t, the true covariate Z_{ijt} is

$$Z_{ijt} = \overline{Z}_{jt} + \zeta_{ijt},$$

where ζ is unobserved and is distributed with mean zero and variance ς^2. When $\overline{Z}_{jt}$ is used instead of unobserved Z_{ijt} in the regressions, they suffer from an omitted variable bias, due to correlation between the regressor $\overline{Z}_{jt}$ and the error term, which now is $\varepsilon_{ijt} + \zeta_{ijt}$. Thus, the estimator b_L is no longer unbiased. To see this, denote as $\tilde{X}$ the subset of regressors in X that is not Z combined with observable $\overline{Z}$. The unbiased estimators would be produced by the regression

$$y = \tilde{X}'b + \zeta'c + e.$$

Since ζ is unobserved, we estimate instead the regression

$$y = \tilde{X}'\tilde{b} + \tilde{e}.$$

We can show that

$$E(\tilde{b}) = \beta + (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\zeta c.$$

Given that in general, $\overline{Z}$ and ζ are correlated, $\tilde{X}$ and ζ are correlated, and therefore $\tilde{b}$ will not be an unbiased estimator of β. If the correlation is positive and $c > 0$ or if correlation is negative and $c < 0$, $E(\tilde{b}) > \beta$, otherwise, $E(\tilde{b}) < \beta$. Since c and ζ are not observed, it is generally not possible to evaluate on pure econometric basis whether the bias is positive or negative. Note that an attenuation bias would be particularly harmful if the estimates are used for scenario analysis, because lower coefficients on macroeconomic variables would lead to an underestimate of the stress scenario losses.

Moreover, one can show that

$$E(\tilde{e}'\tilde{e}) - E(e'e) = \hat{\sigma_L}^2 * k_2 + c'\zeta'\zeta c,$$

that is, the sum of squared errors and therefore the forecast error will always be larger in the presence of measurement error.

Given that we assumed that the measurement error is zero on average for each j, the level of aggregation at which Z is observed, such a problem will only arise for loan-level regressions, not for portfolio, or j-level regressions.

6.3 Aggregation bias

Assume that the loan-level estimate is unbiased:

$$b_L = (X_L' X_L)^{-1} X_L' y_L, \quad E(b_L) = \beta.$$

One can show that for the portfolio level the estimate of β can be expressed as

$$b_P = (X_L' X_L + X_L'(U_N - I)X_L)^{-1}(X_L' y_L + X_L'(U_N - I)y_L),$$

where U_N is the $(NT \times NT)$ block-diagonal matrix of T $(N \times N)$ matrices of ones in the diagonal and zeros elsewhere.

One can see that only if there is no within-time cross-individual correlation in x and e, the cross-product terms will be zero in expectation, that is $U_N = I$, and therefore $E(b_P) = E(b_L) = \beta$, estimate of β is unbiased. Otherwise $E(b_P) \neq E(b_L)$, reflecting the aggregation bias.

If $\forall t\, E(x_{it}x_{jt}) \neq 0$, $E(e_{it}e_{jt}) \neq 0$, for $i \neq j$, aggregate regression estimates will not be unbiased. More specifically,

$$b_P = (X_P'X_P)^{-1}X_P'y_P = (X_L'U_NX_L)^{-1}X_L'U_Ny_L$$

$$= (X_L'X_L + X_L'(U_N - I)X_L)^{-1}(X_L'y_L + X_L'(U_N - I)y_L)$$

$$= \beta + (X_L'X_L + X_L'(U_N - I)X_L)^{-1}(X_L'e + X_L'(U_N - I)e),$$

where $U_N = S_NS_N'$, the $(NT \times NT)$ block-diagonal matrix of T $(N \times N)$ matrices of ones in the diagonal and zeros elsewhere, e is the residual. Similarly $U_J = S_JS_J'$ is an $(NT \times NT)$ block-diagonal matrix of JT $(N_j \times N_j)$ matrices of ones in the diagonal, where N_j is the number of individual loans in sub-portfolio j.

One can see that for the estimator to be unbiased, we need $E(X_L'\varepsilon|X_L) = 0$ and $E(X_L'(U_N - I)\varepsilon|X_L) = 0$. The first condition is satisfied because disaggregate estimator is unbiased. However, for the second condition will be true only if there is no within-time cross-individual correlation between each variable x in X and e.[33] The same problem arises for a sub-portfolio level regressions. However, given that fewer cross-product terms appear in sub-portfolio level regressions, the bias is

[33] Note, however, that the standard estimate of the variance of b_L will not be unbiased, and the cluster-robust standard errors will need to be computed. See, for example, Arellano (1987).

smaller in magnitude the less aggregated the variables are.

7 Appendix 2

Table 6: Portfolio Level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
DTI, lag(1)	0.0118***	0.0113***	0.0106***	0.0106***	0.0107***	0.0108***	0.0109***	0.0110***	0.0109***	0.0110***	0.0112***	0.0112***
	(0.0013)	(0.0012)	(0.0012)	(0.0012)	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0012)	(0.0012)
HPD, lag(1)	-0.0084	-0.0044	0.0015	0.0013	0.0009	-0.0000	-0.0025	-0.0040	-0.0048**	-0.0020	0.0002	0.0011
	(0.0064)	(0.0055)	(0.0053)	(0.0043)	(0.0040)	(0.0038)	(0.0039)	(0.0032)	(0.0023)	(0.0029)	(0.0028)	(0.0026)
UR, lag(1)	0.0447***	0.0374***	0.0275***	0.0278***	0.0284***	0.0298***	0.0330***	0.0347***	0.0354***	0.0325***	0.0296***	0.0271***
	(0.0114)	(0.0097)	(0.0095)	(0.0079)	(0.0074)	(0.0070)	(0.0069)	(0.0061)	(0.0055)	(0.0059)	(0.0061)	(0.0061)
HPD*UR, lag(1)	0.0032**	0.0024**	0.0012	0.0012	0.0013*	0.0015**	0.0020***	0.0023***	0.0024***	0.0019***	0.0015***	0.0014***
	(0.0013)	(0.0011)	(0.0010)	(0.0008)	(0.0007)	(0.0007)	(0.0007)	(0.0006)	(0.0004)	(0.0005)	(0.0005)	(0.0005)
LTV, lag(1)	0.0058***	0.0052**	0.0044**	0.0045**	0.0045**	0.0047**	0.0049**	0.0051**	0.0051**	0.0051**	0.0053**	0.0056**
	(0.0021)	(0.0021)	(0.0021)	(0.0021)	(0.0021)	(0.0021)	(0.0020)	(0.0020)	(0.0020)	(0.0021)	(0.0022)	(0.0023)
Observations	77	78	79	80	81	82	83	84	85	86	87	88
Adjusted R^2	0.8876	0.8961	0.8947	0.9060	0.9153	0.9237	0.9300	0.9391	0.9500	0.9489	0.9447	0.9425
MSE	0.0196	0.0069	0.0045	0.0043	0.0039	0.0037	0.0046	0.00671	0.0080	0.0056	0.0044	0.0045
Loss	0.8948	0.9513	1.0413	1.0382	1.0312	1.0134	0.9766	0.9542	0.9433	0.9665	0.989	1.004

Notes: Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$.
Variables are based on the mean value, by date.

Table 7: DTI Level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
FICO, lag(1)	0.0069***	0.0069***	0.0066***	0.0067***	0.0067***	0.0069***	0.0073***	0.0076***	0.0079***	0.0082***	0.0085***	0.0087***
	(0.0016)	(0.0016)	(0.0016)	(0.0017)	(0.0017)	(0.0018)	(0.0019)	(0.0021)	(0.0021)	(0.0020)	(0.0018)	(0.0017)
DTI, lag(1)	-0.0109**	-0.0107**	-0.0102**	-0.0098**	-0.0095**	-0.0096*	-0.0102**	-0.0101*	-0.0098*	-0.0105*	-0.0119**	-0.0129**
	(0.0037)	(0.0038)	(0.0038)	(0.0040)	(0.0042)	(0.0044)	(0.0044)	(0.0048)	(0.0049)	(0.0048)	(0.0044)	(0.0040)
HPD, lag(1)	0.0308***	0.0292***	0.0282***	0.0274***	0.0283***	0.0267***	0.0273***	0.0361***	0.0311***	0.0335***	0.0301***	0.0313***
	(0.0080)	(0.0074)	(0.0075)	(0.0072)	(0.0073)	(0.0071)	(0.0070)	(0.0083)	(0.0073)	(0.0068)	(0.0066)	(0.0067)
UR, lag(1)	0.0068	0.0086	0.0066	0.0073	0.0069	0.0085*	0.0127**	0.0182***	0.0362***	0.0348***	0.0361***	0.0415***
	(0.0053)	(0.0048)	(0.0043)	(0.0042)	(0.0043)	(0.0044)	(0.0042)	(0.0043)	(0.0057)	(0.0062)	(0.0064)	(0.0060)
HPD, lag(2)	-0.0596***	-0.0571***	-0.0538***	-0.0528***	-0.0550***	-0.0532***	-0.0564***	-0.0728***	-0.0564***	-0.0622***	-0.0554***	-0.0565***
	(0.0152)	(0.0142)	(0.0143)	(0.0139)	(0.0142)	(0.0140)	(0.0140)	(0.0162)	(0.0136)	(0.0132)	(0.0127)	(0.0128)
UR, lag(2)	0.0165***	0.0160***	0.0147***	0.0162***	0.0177***	0.0187***	0.0200***	0.0212***	0.0161***	0.0146**	0.0059	-0.0040
	(0.0039)	(0.0040)	(0.0041)	(0.0041)	(0.0042)	(0.0042)	(0.0042)	(0.0044)	(0.0047)	(0.0046)	(0.0041)	(0.0033)
HPD, lag(3)	0.0344***	0.0336***	0.0312***	0.0311***	0.0325***	0.0323***	0.0350***	0.0424***	0.0305***	0.0340***	0.0307***	0.0307***
	(0.0079)	(0.0075)	(0.0076)	(0.0075)	(0.0077)	(0.0077)	(0.0078)	(0.0087)	(0.0071)	(0.0072)	(0.0070)	(0.0069)
Loan Amt, lag(1)	0.1226**	0.1283**	0.1194**	0.1263**	0.1297**	0.1373***	0.1514***	0.1613***	0.1822***	0.1690***	0.1435***	0.1233***
	(0.0451)	(0.0446)	(0.0424)	(0.0411)	(0.0404)	(0.0398)	(0.0385)	(0.0400)	(0.0394)	(0.0381)	(0.0361)	(0.0346)
Observations	750	760	770	780	790	800	810	820	830	840	850	860
Adjusted R^2	0.7870	0.8022	0.8109	0.8254	0.8384	0.8467	0.8557	0.8631	0.8688	0.8759	0.8781	0.8781
MSE	0.00868	0.0077	0.0099	0.0083	0.0077	0.0066	0.007	0.0111	0.0249	0.0226	0.0155	0.0119
Loss	1.0611	1.0491	1.0723	1.0574	1.0476	1.0262	0.9832	0.9410	0.8759	0.8811	0.9103	0.9313

Notes: Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$.
Variables are based on the mean value, by date.

Table 8: LTV Level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
DTI, lag(1)	0.0008	0.0007	0.0006	0.0005	0.0004	0.0003	0.0001	-0.0001	-0.0001	-0.0001	-0.0002	-0.0004
	(0.0018)	(0.0018)	(0.0018)	(0.0019)	(0.0019)	(0.0020)	(0.0021)	(0.0021)	(0.0023)	(0.0024)	(0.0024)	(0.0025)
HPD, lag(1)	-0.0073*	-0.0083***	-0.0061**	-0.0078**	-0.0086**	-0.0102***	-0.0128***	-0.0143***	-0.0172***	-0.0144***	-0.0119***	-0.0111***
	(0.0037)	(0.0024)	(0.0019)	(0.0024)	(0.0029)	(0.0029)	(0.0032)	(0.0034)	(0.0025)	(0.0023)	(0.0017)	(0.0015)
UR, lag(1)	0.0015	0.0024	-0.0018	0.0004	0.0010	0.0029	0.0059	0.0070	0.0097	0.0057	0.0020	-0.0008
	(0.0159)	(0.0117)	(0.0103)	(0.0101)	(0.0092)	(0.0092)	(0.0086)	(0.0077)	(0.0110)	(0.0109)	(0.0102)	(0.0098)
HPD*UR, lag(1)	0.0030**	0.0032***	0.0027***	0.0031***	0.0032***	0.0036***	0.0041***	0.0044***	0.0049***	0.0044***	0.0039***	0.0038***
	(0.0009)	(0.0007)	(0.0005)	(0.0006)	(0.0007)	(0.0007)	(0.0008)	(0.0008)	(0.0008)	(0.0007)	(0.0006)	(0.0006)
Observations	770	780	790	800	810	820	830	840	850	860	870	880
Adjusted R^2	0.5324	0.5517	0.5660	0.5828	0.5996	0.6159	0.6387	0.6614	0.6590	0.6615	0.6641	0.6609
MSE	0.0158	0.0152	0.0226	0.0192	0.0184	0.0159	0.0135	0.0135	0.0146	0.0136	0.015	0.0173
Loss	1.0947	1.0888	1.1355	1.1170	1.1119	1.0925	1.0576	1.0427	1.0139	1.0508	1.0838	1.1076

Notes: Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$.
Variables are based on the mean value, by date.

Table 9: FICO Level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
FICO, lag(1)	-0.0104*	-0.0106*	-0.0106**	-0.0108**	-0.0110**	-0.0115**	-0.0122**	-0.0130**	-0.0138**	-0.0138**	-0.0135**	-0.0132**
	(0.0047)	(0.0047)	(0.0047)	(0.0048)	(0.0048)	(0.0049)	(0.0051)	(0.0053)	(0.0057)	(0.0056)	(0.0054)	(0.0054)
DTI, lag(1)	0.0161***	0.0169***	0.0171***	0.0179***	0.0187***	0.0196***	0.0208***	0.0219***	0.0235***	0.0243***	0.0249***	0.0256***
	(0.0045)	(0.0046)	(0.0047)	(0.0049)	(0.0052)	(0.0055)	(0.0059)	(0.0063)	(0.0069)	(0.0072)	(0.0074)	(0.0076)
HPD, lag(1)	0.0282**	0.0271*	0.0265*	0.0261*	0.0272*	0.0264*	0.0269*	0.0355**	0.0277**	0.0325**	0.0294**	0.0314**
	(0.0119)	(0.0125)	(0.0120)	(0.0120)	(0.0124)	(0.0124)	(0.0127)	(0.0143)	(0.0118)	(0.0134)	(0.0123)	(0.0125)
UR, lag(1)	0.0160***	0.0172***	0.0143***	0.0148***	0.0141***	0.0157***	0.0201***	0.0257***	0.0428***	0.0402***	0.0411***	0.0457***
	(0.0034)	(0.0031)	(0.0019)	(0.0021)	(0.0021)	(0.0020)	(0.0031)	(0.0039)	(0.0101)	(0.0091)	(0.0096)	(0.0104)
HPD, lag(2)	-0.0385*	-0.0363	-0.0334	-0.0326	-0.0351	-0.0344	-0.0368	-0.0524*	-0.0305	-0.0415*	-0.0353*	-0.0378*
	(0.0201)	(0.0209)	(0.0194)	(0.0195)	(0.0202)	(0.0202)	(0.0211)	(0.0239)	(0.0174)	(0.0199)	(0.0178)	(0.0181)
UR, lag(2)	0.0225***	0.0220***	0.0199***	0.0210***	0.0225***	0.0239***	0.0258***	0.0272***	0.0217***	0.0188***	0.0100*	0.0007
	(0.0058)	(0.0056)	(0.0048)	(0.0050)	(0.0049)	(0.0051)	(0.0054)	(0.0056)	(0.0044)	(0.0041)	(0.0047)	(0.0046)
HPD, lag(3)	0.0187	0.0179	0.0155	0.0153	0.0169	0.0172	0.0194	0.0264*	0.0121	0.0189*	0.0162*	0.0170*
	(0.0104)	(0.0107)	(0.0096)	(0.0096)	(0.0100)	(0.0101)	(0.0107)	(0.0121)	(0.0082)	(0.0092)	(0.0082)	(0.0084)
Loan Amt, lag(1)	-0.2401	-0.2652	-0.2959	-0.3193	-0.3437	-0.3600	-0.3764	-0.3917	-0.4156	-0.4628	-0.5177	-0.5646*
	(0.1557)	(0.1675)	(0.1775)	(0.1879)	(0.1990)	(0.2119)	(0.2264)	(0.2435)	(0.2618)	(0.2777)	(0.2931)	(0.3029)
Observations	750	760	770	780	790	800	810	820	830	840	850	860
Adjusted R^2	0.5628	0.5772	0.5894	0.6028	0.6169	0.6300	0.6363	0.6505	0.6531	0.6663	0.6729	0.6776
MSE	0.0084	0.0079	0.0127	0.0115	0.0109	0.0084	0.0046	0.0038	0.0069	0.0044	0.0035	0.0048
Loss	1.069	1.0655	1.1009	1.0953	1.0925	1.0756	1.0346	0.9973	0.9447	0.9694	1.0105	1.0393

Notes: Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$.
Variables are based on the mean value, by date.

Table 10: County Level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HPD, lag(1)	-0.0016	-0.0038*	-0.0031	-0.0040*	-0.0039*	-0.0041**	-0.0053***	-0.0068***	-0.0089***	-0.0102***	-0.0122***	-0.0139***
	(0.0023)	(0.0023)	(0.0022)	(0.0022)	(0.0022)	(0.0020)	(0.0020)	(0.0022)	(0.0025)	(0.0024)	(0.0024)	(0.0025)
UR, lag(1)	0.0159***	0.0198***	0.0178***	0.0194***	0.0190***	0.0189***	0.0209***	0.0257***	0.0292***	0.0268***	0.0276***	0.0302***
	(0.0053)	(0.0054)	(0.0047)	(0.0048)	(0.0048)	(0.0046)	(0.0044)	(0.0053)	(0.0054)	(0.0054)	(0.0053)	(0.0053)
HPD*UR, lag(1)	0.0014***	0.0017***	0.0016***	0.0017***	0.0016***	0.0015***	0.0016***	0.0020***	0.0023***	0.0022***	0.0025***	0.0028***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
HPD, lag(2)	0.0035	0.0055**	0.0050**	0.0056**	0.0052**	0.0050**	0.0058***	0.0070***	0.0085***	0.0103***	0.0129***	0.0149***
	(0.0023)	(0.0022)	(0.0023)	(0.0022)	(0.0022)	(0.0021)	(0.0021)	(0.0023)	(0.0027)	(0.0026)	(0.0026)	(0.0025)
UR, lag(2)	-0.0012	-0.0037	-0.0020	-0.0015	0.0009	0.0034	0.0049	0.0038	0.0039	0.0041	-0.0001	-0.0048
	(0.0046)	(0.0049)	(0.0052)	(0.0050)	(0.0044)	(0.0043)	(0.0045)	(0.0052)	(0.0051)	(0.0047)	(0.0043)	(0.0036)
HPD*UR, lag(2)	-0.0005	-0.0008*	-0.0007*	-0.0007*	-0.0005	-0.0004	-0.0003	-0.0006	-0.0008	-0.0008	-0.0012**	-0.0015***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
DTI, lag(1)	0.0036***	0.0037***	0.0037***	0.0038***	0.0039***	0.0040***	0.0043***	0.0045***	0.0047***	0.0047***	0.0044***	0.0043***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Observations	23018	23324	23630	23936	24242	24548	24854	25160	25466	25772	26078	26384
Adjusted R^2	0.2310	0.2546	0.2672	0.2904	0.3119	0.3356	0.3655	0.3945	0.4289	0.4460	0.4554	0.4620
MSE	0.01549	0.0132	0.0145	0.0116	0.00893	0.0066	0.0041	0.0036	0.0044	0.0045	0.0041	0.0046
Loss	1.1141	1.103	1.114	1.0980	1.0815	1.0613	1.0294	1.0020	0.9741	0.9806	1.0020	1.0142

Notes: Standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$. Standard errors clustered by county. Fixed Effects included. Variables are based on the mean value, by date.

Table 11: Loan Level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
delinquent												
HPD, lag(1)	0.0550***	0.0543***	0.0536***	0.0528***	0.0523***	0.0512***	0.0489***	0.0467***	0.0447***	0.0441***	0.0434***	0.0429***
	(0.0020)	(0.0019)	(0.0019)	(0.0018)	(0.0018)	(0.0018)	(0.0018)	(0.0018)	(0.0019)	(0.0020)	(0.0020)	(0.0020)
UR, lag(1)	0.0453***	0.0467***	0.0432***	0.0478***	0.0517***	0.0559***	0.0651***	0.0768***	0.0879***	0.0875***	0.0901***	0.0917***
	(0.0171)	(0.0167)	(0.0155)	(0.0149)	(0.0148)	(0.0148)	(0.0150)	(0.0150)	(0.0152)	(0.0147)	(0.0144)	(0.0142)
win	0.0217	0.0321	0.0491	0.0547	0.0385	0.0242	-0.0105	-0.0165	-0.0292	-0.0385	-0.0511	-0.0615
	(0.0797)	(0.0809)	(0.0765)	(0.0754)	(0.0731)	(0.0726)	(0.0686)	(0.0679)	(0.0684)	(0.0672)	(0.0687)	(0.0677)
Appraisal Amt	0.0004***	0.0004***	0.0005***	0.0005***	0.0005***	0.0005***	0.0005***	0.0005***	0.0005***	0.0005***	0.0005***	0.0005***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Full Doc. (D=1)	-0.3492***	-0.3728***	-0.3832***	-0.4018***	-0.4213***	-0.4504***	-0.4504***	-0.4534***	-0.4662***	-0.4657***	-0.4695***	-0.4642***
	(0.0515)	(0.0513)	(0.0518)	(0.0512)	(0.0509)	(0.0510)	(0.0508)	(0.0511)	(0.0519)	(0.0511)	(0.0501)	(0.0511)
Interest Only(D=1)	-1.8574***	-1.9012***	-1.9316***	-1.9770***	-2.0378***	-2.0991***	-2.1729***	-2.2478***	-2.3187***	-2.3602***	-2.3885***	-2.4405***
	(0.0721)	(0.0752)	(0.0763)	(0.0775)	(0.0792)	(0.0797)	(0.0802)	(0.0828)	(0.0851)	(0.0842)	(0.0801)	(0.0821)
Margin Rate	-0.0104	-0.0154	-0.0166	-0.0138	-0.0080	0.0018	0.0107	0.0168	0.0149	0.0147	0.0154	0.0117
	(0.0153)	(0.0158)	(0.0156)	(0.0152)	(0.0149)	(0.0147)	(0.0152)	(0.0150)	(0.0150)	(0.0150)	(0.0148)	(0.0148)
Loan Amt	0.2441***	0.2423***	0.2450***	0.2442***	0.2456***	0.2420***	0.2384***	0.2339***	0.2400***	0.2384***	0.2374***	0.2375***
	(0.0215)	(0.0191)	(0.0175)	(0.0175)	(0.0166)	(0.0160)	(0.0157)	(0.0145)	(0.0135)	(0.0124)	(0.0116)	(0.0112)
Purchase(D=1)	0.2603***	0.3417***	0.3652***	0.4032***	0.4085***	0.4225***	0.4166***	0.4536***	0.4847***	0.4674***	0.4763***	0.5016***
	(0.0653)	(0.0632)	(0.0620)	(0.0603)	(0.0563)	(0.0495)	(0.0488)	(0.0491)	(0.0463)	(0.0454)	(0.0466)	(0.0448)
Loan Term	-0.0019***	-0.0021***	-0.0021***	-0.0022***	-0.0023***	-0.0024***	-0.0025***	-0.0026***	-0.0028***	-0.0029***	-0.0030***	-0.0031***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Current IR	0.2514***	0.2472***	0.2502***	0.2478***	0.2455***	0.2366***	0.2197***	0.2049***	0.1976***	0.1931***	0.1928***	0.1908***
	(0.0090)	(0.0092)	(0.0092)	(0.0094)	(0.0089)	(0.0088)	(0.0089)	(0.0090)	(0.0091)	(0.0092)	(0.0093)	(0.0093)
DTI	-0.0023**	-0.0026**	-0.0029**	-0.0029***	-0.0031***	-0.0031***	-0.0031***	-0.0029***	-0.0027***	-0.0026***	-0.0027***	-0.0026***
	(0.0012)	(0.0012)	(0.0012)	(0.0011)	(0.0011)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0009)	(0.0009)	(0.0009)
Utilization	0.6307***	0.6370***	0.6449***	0.6481***	0.6565***	0.6619***	0.6605***	0.6654***	0.6697***	0.6740***	0.6760***	0.6788***
	(0.0179)	(0.0177)	(0.0174)	(0.0172)	(0.0167)	(0.0165)	(0.0162)	(0.0169)	(0.0165)	(0.0155)	(0.0156)	(0.0151)
Constant	-9.0459***	-8.8904***	-8.9194***	-8.8802***	-8.4734***	-8.8679***	-8.4449***	-8.8562***	-9.2848***	-9.2778***	-9.3871***	-9.4173***
	(1.6328)	(1.6499)	(1.5915)	(1.5381)	(1.5071)	(1.5370)	(1.6041)	(1.5397)	(1.5333)	(1.4969)	(1.4896)	(1.4465)
Observations	5438690	5568757	5698048	5826645	5954505	6081629	6208287	6333879	6458268	6581310	6702895	6818612
Pseudo R^2	0.1744	0.1785	0.1805	0.1843	0.1887	0.1919	0.1961	0.2002	0.2041	0.2063	0.2079	0.2100
MSE	0.0855	0.0842	0.0900	0.0882	0.0858	0.0779	0.0617	0.0426	0.0290	0.0246	0.0201	0.01804
Loss	1.3202	1.3200	1.3381	1.3370	1.334	1.316	1.2754	1.2212	1.1759	1.1577	1.1390	1.1288

Notes: County clustered standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$.
Vintage Fixed Effects, FICO/LTV/age splines