NBER WORKING PAPER SERIES

UNEMLOYMENT AND UNOBSERVED CREDIT RISK IN THE FHA SINGLE FAMILY MORTGAGE INSURANCE FUND

Joseph Gyourko
Joseph Tracy

Working Paper 18880
http://www.nber.org/papers/w18880

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2013

These views represent those of the authors and not necessarily those of the Federal Reserve Bank of New York, the Federal Reserve System, or the National Bureau of Economic Research. John Grigsby provided excellent research assistance. We thank Andrew Haughwout and Wilbert van der Klaauw for their helpful comments. Gyourko thanks the Research Sponsor Program of the Zell/Lurie Real Estate Center at Wharton for financial support.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w18880.ack

Unemloyment and Unobserved Credit Risk in the FHA Single Family Mortgage Insurance Fund
Joseph Gyourko and Joseph Tracy
NBER Working Paper No. 18880
March 2013
JEL No. G17,R48

ABSTRACT

Empirical models of mortgage default typically find that the influence of unemployment is negligible compared to other well known risk factors such as high borrower leverage or low borrower FICO scores. This is at odds with theory, which assigns a critical role to unemployment status in the decision to stop payment on a mortgage. We help reconcile this divergence by employing a novel empirical strategy involving simulated unemployment histories to measure the severity of attenuation bias in loan-level estimations of default risk due to a borrower becoming unemployed. Attenuation bias results because individual data on unemployment status is unobserved, requiring that a market-wide unemployment rate be used as a proxy. Attenuation is extreme, with our results suggesting that the use of an aggregate unemployment rate in lieu of actual borrower unemployment status results in default risk from a borrower becoming unemployed being underestimated by a factor of 100 or more. Correcting for this indicates unemployment is more powerful than other well-known factors such as extremely high leverage or extremely low FICO scores in predicting individual borrower default. Our simulated data indicate that adding the unemployment rate as a proxy for the missing borrower-specific unemployment indicator does not improve the accuracy of the estimated model over the specification without the proxy variable included.

Joseph Gyourko
University of Pennsylvania
Wharton School of Business
3620 Locust Walk
1480 Steinberg-Dietrich Hall
Philadelphia, PA 19104-6302
and NBER
gyourko@wharton.upenn.edu

Joseph Tracy
Executive Vice President
Federal Reserve Bank of New York
33 Liberty Street
New York, NY 10045
joseph.tracy@ny.frb.org

I. Introduction

The recent announcement of a negative estimated economic value of the Federal Housing Administration's (FHA) single-family mortgage insurance portfolio, which measures the difference between expected fund revenues and costs, has increased interest in the financial viability of what is now over a $1 trillion guarantee program that amounts to about 7% of annual national output.[1] Each of us in prior research has argued that unobserved credit risk in the FHA guarantee portfolio is high (Caplin, et. al. (2010, 2012); Gyourko (2011)), with FHA's persistent overestimation of the value of each new book of insured loans indicating that over the past several years something systematic in terms of uncontrolled-for risk is being missed by program evaluators. In this paper, we focus on unemployment as a key uncontrolled for risk factor.

From 2008 until the most recent annual evaluation, FHA's outside actuarial reviewer made no attempt to control for unemployment, having discontinued use of a metropolitan area-wide unemployment rate in its default models. The explanation provided was that controlling for unemployment did not help in explaining default.[2] This decision was at odds with theories of default, which hypothesize that unemployment risk plays a key role in any decision to stop paying on a mortgage.[3] The most recent annual review of the FHA insurance program reversed this decision and included aggregate unemployment rates as controls in their underlying default (and prepayment) models. Consistent with other recent empirical evidence (e.g., Elul, et. al.

[1] The report, entitled *Actuarial Review of the Federal Housing Administration Mutual Mortgage Insurance Fund Forward Loans for Fiscal Year 2012* (IFE (2012), indicated the fund had an economic value of -$13.478 billion. It can be accessed at http://portal.hud.gov/hudportal/HUD?src=/program_offices/housing/rmra/oe/rpts/actr/actrmenu.
[2] The actuarial review for 2008 specifically noted the following: "The unemployment rate variables did not perform well in any of the preliminary models that were estimated, and have not been included in the final model specifications. No consistent pattern was observed between mortgage claims and increases in local unemployment rates."[IFE 2008, page A-13]
[3] As is discussed more fully in the literature review in the next section, the so-called 'double trigger' hypothesis suggests that it is a combination of an adverse income shock and the borrower being in negative equity that results in a default. See Gerardi, et. al. (2007), Foote, et. al. (2008), Foote, et. al. (2010), and the discussion in Section II for more on the double trigger hypothesis, as well as additional evidence on the correlation between unemployment and default.

(2010)), the 2012 actuarial review reports evidence of a positive correlation between an aggregate unemployment rate and the propensity of individual borrowers in the market to default.

However, the estimated economic influence of unemployment rates on default is very small empirically compared to other known risk factors such as borrower credit quality as reflected in FICO scores and loan-to-value (LTV) ratios. One possible reason investigated in this paper is attenuation bias. This is a potentially serious problem because the default decision is observed at the micro-household level, but the unemployment status of the borrower is not. Hence, a market-level unemployment rate is used as a proxy for the unobserved unemployment status of the borrower in empirical models of individual default, and it well could be a very noisy proxy.[4] For example, even in a market with a high unemployment rate of 10%, the vast majority of borrowers still are employed, and rises in that rate obviously do not mean that most borrowers lose their jobs and become income constrained.

We bring new empirical analysis to bear on this issue, using results from a simulation exercise involving transitions into and out of unemployment, to help gauge the true impact of unemployment on default. The analysis strongly suggests that borrower unemployment status is very important in determining default behavior, with the likely impact at least as large as that found for other well known risk factors such as FICO score and loan-to-value ratio. Hence, we are able to account for the discrepancy between the prominent role that unemployment risk is accorded in conceptual models of default and the very limited impact estimated in recent empirical work.

The plan of the paper is as follows. The next section begins by briefly reviewing the literature on the linkage between mortgage default and unemployment risk. This is followed

[4] This point was raised by Haughwout, et. al. (2010), page 17 footnote 22.

with a presentation of our analysis on how one can use labor market employment transition data to estimate the potential magnitude of the attenuation bias that inevitably results from regressing individual borrower default data on a market-wide measure of unemployment. Section III then describes the sample of mortgages and borrowers used to estimate the impact of unemployment on the propensity to default. That is followed by a presentation and discussion of our key results in Section IV. This section also addresses the implication of our analysis for whether the FHA's outside actuary's recent inclusion of the aggregate unemployment rate in its micro-level default model helps it generate more reliable forecasts of portfolio-level default risk. Unfortunately, the answer is that it likely does not in this case. Hence, unemployment largely remains an unpriced risk factor in terms of evaluating the likely cost of the FHA insurance program. Finally, there is a brief summary and conclusion.

II. Default and Unemployment Risk

a. Literature Review

There is a long academic literature on mortgage default (and prepayment). Financial economists have traditionally modeled default as a put option because the decision not to pay the contractually required future stream of interest and principal payments essentially involved the borrower 'putting' back the mortgage to the lender. Early thinking on the problem viewed the decision to default as being determined by whether the borrower had negative equity in the home, on the premise that it was rational to walk away from a house only when its value was less than the present value of the debt owed on it. Kau, et. al. (1994) showed that even this was not sufficient because it still could be optimal for a borrower to wait and default in the future. That is, the value of the put option need not be maximized when the borrower first enters negative equity.

That negative equity is not the sole factor behind the decision to default seems evident from the fact that at any point in time most borrowers with negative equity are not seriously delinquent on their mortgages. One recent industry estimate indicates that about 85% of households with a mortgage who are estimated to be in negative equity are current on their debt service payments.[5] Empirical analysis in the academic literature also concludes that the decision to default is based on more than current and prospective negative equity.[6] In particular, default has been shown to be associated with negative shocks to income, including that arising from becoming unemployed.[7] Deng, et. al.'s (2000) classic empirical paper on the competing risks of default and prepayment reports evidence consistent with negative equity and income each influencing the probability of default.

One way to summarize the literature is with the so-called 'double trigger' terminology in which negative equity and income loss are the two triggers.[8] In this framework, a borrower in negative equity is at heightened risk of default. But, we know from above that this is not a sufficient condition for default to occur. The second trigger is a large enough adverse income shock, say via losing one's job, that leaves the borrower unable to make scheduled monthly

---

[5] This particular estimate is from a study by CoreLogic in the second quarter of 2012. See the web article here (http://www.dsnews.com/articles/borrowers-in-negative-equity-slowly-declining-as-home-values-gain-report-2012-09-12) for more on this.
[6] Measurement error in determining if a borrower is in negative equity is also likely a factor. That is, some borrower's who are estimated to be in negative equity may in fact have positive equity.
[7] See Foster and Van Order (1984) and Vandell (1995) for early discussions and presentation of data on this matter.
[8] See Gerardi, et. al. (2007), Foote, et. al. (2008), and Foote, et. al. (2010) for more on the double trigger hypothesis and the most recent housing cycle.

mortgage payments. That will precipitate default because the borrower also cannot pay off the mortgage in full from sale proceeds.[9]

That unemployment risk plays a critical role in conceptual models of default through its impact on income is consistent with the FHA's own survey results from special servicers which tell us that income loss is the primary reason why the typical FHA borrower is no longer current on her mortgage payments.[10]

This then begs the question of whether one can reconcile the important role of unemployment in mortgage default models with the economically small effects found for this factor in empirical studies. If not, then a potentially important mistake is being made in terms of risk analysis. To better understand this issue, we turn next to how unemployment risk should be measured if it were to be controlled for in an empirical investigation of risk in mortgage pools guaranteed against default.

b. Measuring Unemployment Risk

To measure how unemployment-related risk would affect a portfolio of mortgages such as that insured by the FHA, one would like to include a variable in a default model that accurately reflects changes over time in the unemployment status of individual borrowers.[11] That would allow researchers to estimate the extent to which becoming unemployed is

---

[9] Even before a borrower experiences a negative income shock, negative equity makes it difficult for the borrower to pay off the mortgage either by selling the house or by refinancing (see Ferreria, et.al. (2010, 2012) and Caplin, et. al. (1997)). This means that the mortgage will be exposed to the default risk for a longer period of time, which increases the expected cumulative default probability.

[10] See Table 5 (p. 22) of *HUD's Annual Report to Congress, Fiscal Year 2011 Financial Status, FHA Mutual Mortgage Insurance Fund*, November 15, 2011.

[11] One potential alternative in which the market-level unemployment rate would be the appropriate control variable involves strategic default. Even in the absence of an income shock, borrowers who are in negative equity and perceive that they will remain so for a prolonged period of time may choose to default even when they still have the income to make the monthly payments. Such behavior has been labeled "strategic" or "ruthless" default. Foote, et. al. (2008) and Haughwout, et. al. (2010) provide more recent discussions of this concept. Experian (2011) and Bradley, et. al. (2012, 2013) provide estimates of the magnitude of this type of default. In this situation, the risk of a future job loss to a borrower is important and that may be best captured by the aggregate unemployment rate.

positively correlated with default. The problem is that mortgage servicing data that are typically used to estimate default models do not track the borrower's unemployment status. Lenders know a borrower's employment status when a mortgage is originated, but borrowers are not required to report it subsequently. With no accurate measure of unemployment status at the borrower level, one is forced to fall back upon more aggregated, market-level unemployment measures such as the metropolitan area-wide unemployment rate calculated each month by the Bureau of Labor Statistics. Clearly, this is an imperfect proxy for what is happening at the borrower level. Even in a market with a very high unemployment rate of (say) 10%, the vast majority of borrowers still are employed at any point in time. Similarly, an increase in the market average rate of unemployment does not mean that most borrowers became unemployed or otherwise suffer adverse income shocks.

More formally, the implications of using a local unemployment rate as a proxy for an indicator of the unemployment status of an individual borrower can most easily be seen if we use a simple linear probability model to estimate the determinants of mortgage default. Let D_{ijt} be an indicator that takes a value of one if borrower *i* in metro area *j* defaults in time *t*. Let I_{ijt}^{U} denote an indicator that takes a value of one if borrower *i* in metro area *j* is unemployed in period *t*. Letting all other variables one wants to control for be denoted as $x_1 \ldots x_{k-1}$, our simple default specification can be expressed as in equation (1):

$$D_{ijt} = \beta_0 + \beta_1 x_{ijt1} + \ldots + \beta_{k-1} x_{ijtk-1} + \beta_k I_{ijt}^{U} + \varepsilon_{ijt} \ , \tag{1}$$

where ε_{ijt} is the error term.

Because individual borrower employment status is not directly observed, it cannot be controlled for in estimating (1). The two options are either to drop the unemployment status variable or to add a proxy variable in its place. Assume that the borrower's local market unemployment rate, which is denoted by U_{jt}, is used as the proxy. The estimation specification then is given by equation (2).

$$D_{ijt} = \tilde{\beta}_0 + \beta_1 x_{ijt1} + \ldots + \beta_{k-1} x_{ijtk-1} + \tilde{\beta}_k U_{jt} + \varepsilon_{ijt} \ , \tag{2}$$

Pinning down the relationship between $\tilde{\beta}_k$ and β_k obviously is key. To help do so, let the relationship between an individual i's employment status at time t and her market's unemployment rate in the same period be described by the following simple linear equation:[12]

$$I^U_{ijt} = \delta_0 + \delta_1 U_{jt} + \mu_{ijt} \ . \tag{3}$$

Using these three equations, the coefficient on the market-level unemployment rate, $\tilde{\beta}_k$, in equation (2) will equal $\beta_k\delta_1$ (and $\tilde{\beta}_0$ is given by $\beta_0 + \beta_k\delta_0$). Thus, if the market unemployment rate is such a noisy proxy for individual unemployment status that δ_1 is close to zero in equation (3), then the regression results from estimating equation (2) would show that unemployment has little impact on default decisions even if individual borrower unemployment spells significantly increase the risk of default. In that case, the resulting small coefficient on the unemployment rate reflects the severe attenuation bias from measurement error of individual employment status.

[12] For simplicity, this specification ignores that I^U_{ijt} is a limited dependent variable.

This means that we need an estimate of δ_1 in order to recover the coefficient of interest, β_k. To help gauge the potential magnitude of δ_1 and evaluate the likely effectiveness of using the unemployment rate as a proxy for an individual borrower's unemployment experience, we carry out the following simulation exercise using the Bureau of Labor Statistics' (BLS) month-to-month employment transition rates.[13] These data are available only at the national level. Figure 1 plots the nine distinct transition rates reported by the BLS. Start with an individual who is employed in period *t*. The probability that individual remains employed in period *t*+1 is given by ee_t . Similarly, the probability that the same individual becomes unemployed is given by eu_t, while the probability that the individual leaves the labor market is en_t. The analogous permutations for individuals starting out either as unemployed or out of the labor market yield the nine possibilities depicted in Figure 1.

We use the BLS reported transition rates starting in February 1990 and ending in October 2012 to simulate employment paths for 150,000 individuals. We initialize everyone as employed in February 1990. We then randomly transition individuals across the three employment states through time. We use the data up to December 2004 to generate a distribution of individuals across the three labor market states as of January 2005. Figure 2 shows the aggregate unemployment rate for the nation computed by the BLS and compares it to the implied aggregate unemployment rate from our simulated data for the period covered by our FHA data to be discussed below – from January 2005 to March 2012. As expected, the two series track each other very closely.

This simulated data is then used to obtain an estimate for δ_1. More specifically, we begin by randomly merging simulated employment histories into a sample of mortgages that we will

[13] The labor flow data is produced by the Bureau of Labor Statistics. For details, see Frazis, et. al. (2005).

describe later. For each potential merge between a mortgage and an employment history, we verified that the borrower was employed in the month that the mortgage was originated to account for standard lending practice. Otherwise, we randomly selected a new employment history for that mortgage until this condition was satisfied. The final sample consists of 149,336 mortgages with associated employment histories. For each employment history, we selected the monthly observations between the time the mortgage was originated and the earlier of the date the mortgage prepaid or the borrower became 90 days delinquent for the first time. For each borrower, we create an indicator variable that takes a value of one if the borrower is unemployed in that month. We then merge in the implied aggregate unemployment rates for each month derived from our unmerged simulated employment histories.[14]

Using this data on mortgages merged with simulated employment histories, we next regress the individual borrower unemployment indicator on the aggregate unemployment rate for the month, as above in equation (3). The coefficient estimates (standard errors) from this bivariate regression are as follows:[15]

$$\underset{}{Unemployed} = \underset{(0.00052)}{0.00042} + \underset{(0.00006)}{0.00553} * Unemployment\_rate$$

$R^2 = 0.0019.$

This result highlights how noisy a proxy the aggregate unemployment rate is for any given individual's true unemployment status. Taking the estimation at face value implies that the coefficient on the unemployment rate in our default regression will understate the true effect of

[14] Here we follow the convention of measuring the unemployment rate so that a unit change equals a one percentage point change.
[15] This regression has 4.081 million simulated monthly observations.

the borrower being unemployed by a factor of about 180 ($\approx 1/0.00553$). If this simulation exercise with national data is informative about the relationship between individual employment status and unemployment rates at the metropolitan area-level, then the results help account for why recent academic research and FHA's actuarial reviewer finds that controlling for market-wide unemployment yields such a small economic impact empirically, especially compared to other well-known risk factors such as borrower leverage and FICO score at origination.[16]

Importantly, the small estimated value for δ_1 implies that the small estimated coefficient for unemployment in a default regression does not mean that unemployment status really has only a minor influence on a given borrower's decision to default. To investigate this question further, we estimate a default specification like that in equation (1) on a sample of FHA-insured loans, and then use our imputed value of δ_1 to adjust the regression coefficient on the market-level unemployment variable. The next section describes our underlying mortgage data and the key variables used in that analysis.

III. Data Description

Our primary data source on loans that likely were insured by FHA is Lender Processor Services Inc. (LPS) Applied Analytics. The LPS data form a monthly panel of loans, which is derived from the portfolios of large mortgage servicers in the U.S. Information provided includes mortgage-level data on loan payment status, documentation on the loan, mortgage terms such as length (e.g., 30 years) and coupon, loan-to-value (LTV) ratio at origination[17], whether the loan was reported to be for a home purchase or a refinance, the type and location of the property, and the borrower's FICO score at origination.

---

[16] We find similarly small impacts in our own analysis, which is discussed below.

[17] The origination LTV only reflects the first-lien mortgage, but this is less of an issue for FHA-insured loans given the very high LTVs permitted with a single loan by that guarantor.

We take a 5% random sample of loans from the LPS database, and constrain ourselves to a coverage period from January 2005 through March 2012 to combat any selection bias that might arise due to LPS's relatively sparse coverage of the U.S. mortgage market before 2005. We limit our sample to loans held by the Government National Mortgage Association (henceforth GNMA or Ginnie Mae) in order to isolate those mortgages most likely to be insured by the FHA.[18] Even this sample is not a perfect representation of the loans held by Ginnie Mae at any given time. Since loans can be traded, a loan may enter the GNMA portfolio and subsequently exit. Because we are interested primarily in the loans guaranteed by the FHA, we keep all loans in our sample that were owned by Ginnie Mae at any time, whether permanently or temporarily.

For our dependent variable, we create an indicator of whether a loan is 90 days delinquent for the first time. This is in keeping with much of the literature (e.g., see Caplin, et. al. (2012) for a recent example and discussion). Once a borrower reaches the 90-day delinquency trigger, we censor the remaining observations for that borrower. For borrowers who prepay their mortgage, we include data up to the month prior to the prepayment.

Our unemployment data come from the Bureau of Labor Statistics (BLS) Local Area Unemployment Statistics (LAUS) database, which is comprised of non-seasonally adjusted county- and state-level unemployment rates.[19] These data are used to create metropolitan area-level unemployment rates which are then matched to our sample of loans using the Zip code of the underlying properties (which is provided in the LPS files).

[18] Not only is the underlying data base large (e.g., it contained 37 million active loans at the end of 2010), but its coverage of the GNMA universe in this period is good. From 2005-2012, LPS averaged a coverage rate for GNMA loans of 77.3%, versus an average coverage rate of 41.8% for similar loans from 2000-2004.
[19] The web site describing these data in more detail is here: http://www.bls.gov/lau/.

The current LTV ratio provided in the LPS data is calculated by dividing the balance remaining on the mortgage by the sale price of the property. As such, it does not account for changes in home prices, so to better reflect a borrower's current equity position we adjust for changes in the value of their home. Since we do not directly observe the current value of any specific house, we estimate this value based on the cumulative change in the metro area house prices from the origination date to the current date.[20]

We also construct dummy variables for whether a borrower would be at risk of recourse to a deficiency judgment from the lender in the event of a default and whether the property is in a judicial foreclosure state.[21] Finally, we construct dummies for the duration of the loan that measure whether the loan has survived a given number of months since the origination date. Because we do not consider a loan to have defaulted until it is at least 90 days late, this vector of dummies begins with the fourth month after loan origination. We control for each subsequent month through the 24th month after origination. After that, we aggregate to three month periods through the 48th month due to small cohort sizes at the upper end of the duration scale.

IV. Estimation and Results

a. Linear Probability Model Results

We begin by estimating a simple linear probability model for the default indicator as described above in equation (1). The *X*'s from that specification include a number of controls traditional to the mortgage default estimation literature as described just above. These are a

[20] We use the CoreLogic metro area overall house price indices.

[21] A borrower is not at risk of recourse if he/she lives in Arizona, Iowa, Minnesota, Washington or Wisconsin, or lives in North Carolina or California and the loan is for a home purchase. Connecticut, Delaware, Florida, Iowa, Illinois, Indiana, Kansas, Kentucky, Louisiana, Massachusetts, Maryland, Maine, North Dakota, Nebraska, New Jersey, New Mexico, New York, Ohio, Pennsylvania and South Carolina are the states with a judicial foreclosure process. See Ghent and Kudlyak (2009).

measure of the degree of leverage by the borrower (e.g., loan-to-value (LTV) ratio below 80% [left-out category], between 80% and 100%, between 100% and 120%, and greater than 120%), the vintage year of the loan pool, FICO score categories at the time of loan origination (e.g., below 580, from 580-680, and 680+[left-out category]), dummies for loan type (e.g., whether the mortgage required full documentation, whether it was for a home purchase versus a refinancing, whether it had a 30 year term versus anything else), whether it was on a single unit residence versus anything else, whether it was taken out in a non-recourse state, and whether it was taken out in a judicial foreclosure state. Also included is the series of duration dummies discussed in the previous section. We also explicitly control for unemployment with the metropolitan area-level proxy discussed above.

We estimate this model on a sample of markets that had at least 50 observations on borrowers per month and use the full sample period from March 2005-March 2012 available to us. We restrict the sample geographically this way because we are concerned that including data on markets with very small samples of borrower observations in given months will amplify the measurement error problem plaguing the relationship between unemployment status and default risk. This still leaves us with fifty metropolitan areas in the sample which are listed in Appendix Table A1.[22] Summary statistics are provided in Appendix Table A3.

Our baseline results are reported in Table 1. Two specifications are estimated. The first controls for vintage (i.e., origination year) effects and year (i.e., time) effects, but not MSA effects. The second adds in MSA effects. Depending upon the specification, the results indicate that a one percentage point increase in the area unemployment rate is associated with from a 3.6 (column 1) to 6.3 (column 2) basis point increase in the monthly default risk. These effects are

[22] FHA insurance exposure is not uniformly distributed around the country, so these 50 markets are not a random sample of the nation's population. Given our focus on FHA-insured borrowers, that is not our goal in any event.

highly statistically significant, but are economically small relative to an average monthly default risk of about 50 basis points, and are consistent with findings in recent academic work (e.g., Elul, et. al. (2010)) and FHA's latest actuarial review (IFE (2012)). Other variables are much more influential. For example, moving from a high FICO score (above 680) to a low score (below 580) is associated with roughly a 120 basis point increase in the default risk, which is equivalent to at least a 20 percentage point increase in the local market unemployment rate.

However, we know from equations (1) to (3) that the point estimates reported in the first row of Table 1 are the product of $\beta_k\delta_1$ and reflect the attenuation bias from having a noisy proxy for individual borrower employment status. We can help correct for this by using the result about the relationship between individual borrower employment status and the national aggregate unemployment rate from our simulation exercise. Our estimate above that δ_1=0.00553 implies that the impact of a borrower unemployment spell on default is likely very large, with an imputed magnitude of 645 basis points. This is roughly 12 times the average monthly default risk.[23]

A second thing to note in Table 1 is that the marginal effect associated with the unemployment proxy variable increases by 75 percent when we control for MSA effects. By adding MSA controls, only the variation within each metro area's unemployment rate is used to estimate the associated default risk. Any persistent differences in area unemployment rates between metro areas are removed from the estimation. For our sample period and our top-50 metro areas, 19% of the total variation in the area unemployment rates is associated with the between variation in average unemployment rates across the metro areas. The results indicate that the factors which impact these average unemployment rate differences across metropolitan

[23] Using column 1's estimated coefficient implies that $\beta_k\delta_1$=0.000357. Rearranging leaves us that β_k=0.000357/δ_1=0.00036/0.0055309=0.0645, which implies 645 basis points given the underlying units of measurement.

areas are not as strongly associated with whether a FHA borrower is unemployed as the time-series variation in the unemployment rate within a metropolitan area around this average.

To further explore the attenuation bias from using the unemployment rate as a proxy for the borrower specific unemployment indicator, we generate simulated defaults and prepayments for our sample of mortgages. The monthly predicted default and prepayment probabilities are based on the results from specification (1) of Table 1 (with the prepayment coefficient estimates provided in Appendix Table A2). These predicted probabilities are determined using the borrower specific unemployment indicator in place of the unemployment rate with a default coefficient of 0.0645 (to reflect the 645 basis point impact just discussed).[24] We then employ these estimated monthly predicted default and prepayment probabilities to simulate defaults and prepayments for each mortgage.[25] Finally, the simulated data are used to re-estimate the default specification replacing the unemployment indicator with the unemployment rate. The estimated coefficient (standard error) on the unemployment rate from that regression is a very small 0.00042 (0.00011). This reiterates how severe attenuation bias is likely to be when one uses an aggregate unemployment rate to proxy for individual employment status and the true default risk of becoming unemployed is large.

We obtain similar results when we simulate defaults and prepayments using a stylized double trigger model. In that analysis, the monthly default risk was increased by 1,000 basis points for a borrower who becomes unemployed and is in negative equity. No increase in default risk from unemployment was presumed if the borrower has positive equity. Analogously, we

---

[24] The prepayment coefficient for the unemployment indicator was set to negative 651 basis points.

[25] We draw pairs of uniform (0,1) random variables and code the default indicator equal to one if the first uniform random variable takes a value less than the predicted default probability, and we code the prepayment indicator equal to one if the second random variable takes a value less than the predicted prepayment probability. In the event that this process leads both indicators to be set to one, we decide which indicator to set to one by drawing a third uniform (0,1) and set the default indicator to one (zero) and the prepayment indicator equal to zero (one) if this random variable falls below (above) one half. We censor the default data going forward if the borrower prepays.

increase the prepayment risk by 1,000 basis points for a borrower who becomes unemployed and has positive equity, and presume no increase in the prepayment risk from unemployment if the borrower is in negative equity. The model is then estimated using the aggregate unemployment rate and the interaction of the unemployment rate with an indicator for negative equity. The coefficient on the unemployment rate is small and not statistically different from zero, while the coefficient (standard error) on the interaction term is 0.00044 (0.00008). The estimated interaction effect is only 0.44 percent the size of the actual effect.[26]

The severe attenuation bias documented here arises from two sources. The first source reflects a scaling issue that affects interpretation of the estimated impact. Because an individual borrower either is employed or unemployed, I_{ijt}^{U} in equation (1) takes on values of either 0 or 1. A unit change represents moving a borrower from a zero chance of being unemployed in a given time period to a one hundred percent chance of being unemployed. Now, consider aggregating specification (1) to the metro area level by averaging all of the individual variables by metro area as given in equation (4).

(4) $$D_{\bullet jt} = \beta_0 + \beta_1 x_{\bullet jt1} + \ldots + \beta_{k-1} x_{\bullet jtk-1} + \beta_k I_{\bullet jt}^{U} + \varepsilon_{\bullet jt} \ ,$$

where $I_{\bullet jt}^{U}$ is the average of the unemployment indicators across borrowers in metro area *j* at time *t*, with the same notation applying for the other variables. Note that $I_{\bullet jt}^{U}$ in this regression still ranges from zero to one so that β_k , which captures the impact of unemployment on the individual borrower's default risk, still is associated with a unit change in the average of the borrower specific unemployment indicators in the metro area for each time period. In this case, a unit

[26] Note that this exercise assumed no measurement error in the negative equity indicator variable. In practice, measurement error would even further bias downward the estimated interaction effect.

change still reflects a one hundred percentage point increase in the likelihood of becoming unemployed.

Note that the impact researchers typically estimate when they use a market-level employment rate as a proxy in a specification like that in equation (2) is not that of a 100 percent change in the unemployment rate. As we have done in the regression underlying Table 1's results, standard practice is to scale the proxy variable such that a 5% unemployment rate is entered as a 5 (not 0.05), so that a one unit change reflects a one percentage point change (say from 5% to 6% unemployment), not a 100 percentage point change. Rescaling this variable so that a unit change in it reflects a 100 percentage point change, as with I^U_{ijt} or $I^U_{\bullet jt}$, effectively scales up the coefficients on the unemployment rate proxy reported in Table 1 by 100. In that case, the 0.00036 coefficient becomes 0.036, which implies a 360 basis point higher default probability, which is seven times the average monthly default rate and three times the impact of moving from a higher credit quality FICO score (>680) to a subprime credit quality score (<580).

That this reflects the impact of unemployment on an individual borrower's default risk can be confirmed as follows. First, replace the metro area unemployment rate with the metro area averages of the unemployment status indicator variable, $I^U_{\bullet jt}$, in the loan level default regression specification using the simulated default data. In generating the default and prepayment outcomes in this analysis, recall that the simulated default risk was increased by 645 basis points when a borrower becomes unemployed. Using $I^U_{\bullet jt}$ in place of the BLS metro area unemployment rate yields an estimate for β_k of 662 basis points (with a standard error of 56 basis points). Similarly, if we regress the individual borrower unemployment status indicators,

I^U_{ijt}, on the metro area aggregates, $I^U_{\bullet jt}$, the coefficient equals one. That is, there is no attenuation bias from using $I^U_{\bullet jt}$ as a proxy variable for I^U_{ijt}. The main effect of switching from the borrower-specific unemployment indicator, I^U_{ijt}, to the metro area average, $I^U_{\bullet jt}$, is to increase the standard error of the estimate of β_k.[27]

While this shows that an unbiased estimate of the impact of unemployment risk on the probability of default by an individual borrower could be obtained with aggregated micro data on individual unemployment status if available, the second source of the attenuation bias is that the metro area unemployment rate, U_{jt}, an imperfect proxy variable for $I^U_{\bullet jt}$ must be used instead. Measurement error arises because each variable is created from different samples of people. First, $I^U_{\bullet jt}$ is averaged over borrowers (of the specific type of mortgages under study), while U_{jt} is averaged over all prime age adults who are participating in the labor market – including both renters and owners without mortgages who are not included in $I^U_{\bullet jt}$. A second factor is that the unemployment rate excludes individuals who are classified as being out of the labor force. To illustrate the differences between the two measures, when we regress $I^U_{\bullet jt}$ rescaled by 100 on U_{jt} we get a coefficient of 0.23.

This last point suggests one possible strategy to improve the ability to recover the impact of unemployment on default risk. The monthly *Current Population Survey* (CPS) is used to produce the reported unemployment rate, with the survey asking if a household rents or owns their residence. Using the same CPS data, one could reestimate metro area unemployment rates

[27] Using our simulated default data, the standard error increased by a factor of 6.6 (increasing from 0.00085 for the borrower specific unemployment indicator to 0.0056 for the metro average of the indicators). A related issue is downward bias to the reported standard errors on group level variables that have been merged into individual level data if independence is assumed across observations in the same group. See Moulton (1990).

restricting attention to homeowners and including all respondents in the denominator. In addition, if demographic characteristics are available for the targeted set of borrowers under study, one could tailor the CPS sample to align with those demographic characteristics. The resulting alternative metro area unemployment rates will likely have a higher correlation with $I^{U}_{\bullet jt}$, thereby reducing the second source of attenuation bias.

b. Discussion: Individual Borrower Default Risk

While Table 1's results indicate that unemployment is an important risk factor, we do not mean to imply that it is as least five times more influential than having a high versus a low FICO score. That is literally what one would conclude if those results were taken at face value: the 120 basis point impact of going from a 680+ FICO score to below a 580 FICO score is approximately one-fifth the 645 basis point impact implied after adjusting column 1's coefficient on the metropolitan area unemployment rate variable.

There are a number of factors that lead one to question the precision of the 645 figure, but none that lead us to conclude that unemployment risk is not at least the equal of FICO score in importance. Correctly scaling the unemployment rate measure implies a marginal effect for borrower unemployment that is larger by a factor of 100. This result does not rely on using the simulated employment histories. However, sizing the remaining attenuation bias from using BLS metro area unemployment rates likely is sensitive to using simulated versus actual employment histories. The simulations do not fully account for likely differences in employment dynamics between owners and renters. Another is that the simulation itself is done using national data, which means that we must presume the relationship between individual employment status and the aggregate unemployment rate is the same at the metropolitan area level. That will not be the case for any given MSA, but there also is no reason to presume that our estimate of δ_1 is not informative at all for sub-national populations. Indeed, there is good reason to suspect significant measurement error at that level, too (and, thus, attenuation bias in the estimation of β_k at the local market level).

It is likely also the case that some of the traditional risk factors included in our estimation and in most other default models, including FICO score and LTV, are correlated with individual unemployment risk. If so, some of the impact of the imperfectly controlled for unemployment risk is being picked up by those other variables.[28] That said, there are very strong reasons to believe this countervailing force is limited. Among them is the fact that even noisy proxies for unemployment status are statistically significantly correlated with default. In addition, Gyourko (2011) notes that time dummies for recent years are very influential in predicting default in the model used by FHA's actuarial reviewer, which strongly suggests the presence of omitted risk factors such as unemployment.[29]

We have also used a simple linear regression framework in this analysis rather than the more standard hazard model specification. That was done solely for ease in explaining and adjusting for attenuation bias with our simulation exercise results. In hazard model results not reported here, we still found than a one percentage point increase in a metropolitan area's unemployment rate was associated with about a 6% higher propensity to default, all else constant. That still is far less of an impact that the change from a high to low FICO score, but it does indicate that the key issue is not one of functional form. Thus, we find similar impacts of

[28] If the unemployment rate is a valid proxy variable for individual borrower unemployment status, then including the unemployment rate will eliminate this left-out-variable bias. However, some of the explanatory variables that are correlated with I_{it}^U may also be correlated with the error term in equation (3), μ_{it}.

[29] FHA's outsider actuary explicitly controls for FICO and LTV (among many other variables), so if they did largely capture unemployment risk because they were strongly correlated with unemployment status, the year dummies would not be influential predictors of default unless there are other important left-out variables in addition to the borrower unemployment status that vary by year.

noisy proxies for individual employment status in both linear probability and hazard specifications. We also strongly suspect that we would find similar attenuation bias in both settings.

In sum, we believe that the prudent way to interpret our findings is that unemployment risk is likely a very important factor in explaining default, quite plausibly more influential than other well-known risk factors such as borrower FICO score and the loan-to-value ratio.

c. Discussion: Portfolio-Level Default Risk

While use of the market-level proxy for individual borrower unemployment status clearly leads to gross error in predicting default behavior for specific borrowers, it still could be useful in an important policy sense if it helps the FHA generate more reliable forecasts of future default risk in its overall portfolio. Attenuation bias does not preclude that. If the proxy is positively correlated with individual unemployment status, as we have shown above is the case, then including it can help produce an unbiased forecast of the mean future default risk under certain conditions.

To better understand this, consider forecasting the average default rate for a portfolio of active FHA mortgages for a given date *t*, where we denote this average default rate by $D_{\bullet\bullet t}$. To simplify the discussion further, assume that the borrower-specific unemployment status variable is uncorrelated with the other explanatory variables in the default model. In that case, averaging across individuals and metro areas the actual portfolio default rate would give the following.

(5) $$D_{\bullet\bullet t} = \beta_0 + \beta_1 x_{\bullet\bullet 1t} + \ldots + \beta_{k-1} x_{\bullet\bullet k-1t} + \beta_k I^U_{\bullet\bullet t} + \varepsilon_{\bullet\bullet t}$$

If no proxy variable is used for $I^U_{\bullet\bullet t}$ in the estimation of equation (5), then the predicted portfolio default rate would be given by equation (6).

$$\hat{D}_{\bullet\bullet t} = \hat{\beta}_0 + \hat{\beta}_1 x_{\bullet\bullet 1t} + \ldots + \hat{\beta}_{k-1} x_{\bullet\bullet k-1t} \tag{6}$$

Given our assumptions, the expected value for the difference between the actual and predicted portfolio default rates is $\beta_k I^U_{\bullet\bullet t} + \varepsilon_{\bullet\bullet t}$. Note further that adding time fixed effects in the form of year dummies to the specification that includes no proxy variable would absorb the uncontrolled for unemployment risk. In that case, letting *t*=0 denote the omitted time period, the expected value for the estimated year effect associated with period *t* would be $\beta_k (I^U_{\bullet\bullet t} - I^U_{\bullet\bullet 0}) + (\varepsilon_{\bullet\bullet t} - \varepsilon_{\bullet\bullet 0})$. If the labor market worsened between these two periods, this year effect would likely be positive. Thus, adding year effects allows the model to capture within sample the differences in the mean default rates across years for the portfolio, given our assumptions.[30] However, this presents a problem when forecasting future portfolio default rates in that an assumption must be made in terms of future year effects.

Now consider adding the unemployment rate as a proxy variable and re-estimating the portfolio default risk. Here, we make the assumption that the unemployment rate is a valid proxy variable so that the other explanatory variables in (1) are uncorrelated with μ_{ijt}. In this case, the portfolio default risk is given by equation (7).

$$D_{\bullet\bullet t} = \tilde{\beta}_0 + \beta_1 x_{\bullet\bullet 1t} + \ldots + \beta_{k-1} x_{\bullet\bullet k-1t} + \tilde{\beta}_k U_{\bullet t} + \varepsilon_{\bullet\bullet t} + \mu_{\bullet\bullet t} \ . \tag{7}$$

[30] If we relax the assumption that the borrower specific unemployment indicator is uncorrelated with the other control variables, then some of the impact from the average incidence of unemployment will be captured through the other variables.

Given our assumptions, OLS will give unbiased coefficient estimates so that the expected value of the difference between the actual and predicted yearly portfolio default rates will be $(\varepsilon_{\cdot\cdot t} - \varepsilon_{\cdot\cdot 0}) + (\mu_{\cdot\cdot t} - \mu_{\cdot\cdot 0})$. In this case, year effects, if added to the specification would reflect only the residual unemployment risk. In addition to controlling for any left-out-variable bias on the coefficients of the other explanatory variables, adding the proxy variable improves the ability to estimate the aggregate default risk in a portfolio.[31] Importantly, forecasts of future portfolio default rates can incorporate a projected value for the unemployment rate instead of assuming a value for a future year effect.

However, the reliability of that forecast depends upon the strength of the relationship between the market unemployment rate proxy and individual unemployment status, not just its existence. Ohtani (1981) shows that the forecast mean square error (MSE) can be higher when including the proxy, as compared to leaving it out of the prediction specification, when the partial correlation between the missing variable and its proxy is weak enough.[32] Unfortunately, our case appears to be one in which there may be no improvement in the reliability of the portfolio-level forecast from including a market-level rate to proxy for individual borrower unemployment status.

We explore this issue using our simulated default data. Consider first the specification that included both origination year vintage effects as well as year effects. We can estimate this model using the unemployment indicator, the unemployment rate and dropping both. The square

[31] This is an improvement over the specification that omits the proxy variable but includes year effects in that it is more straight forward to construct an out-of-sample forecast. A forecast using the specification with the proxy variable requires a forecast for the unemployment rate. However, it is less clear what to do with the time effects in the specification that omits the proxy variable.

[32] Specifically, the condition for the MSE when using the proxy variable to be lower than the MSE when the proxy variable is omitted is that the square of the t-statistic on the unobserved variable times the square of the partial correlation coefficient between the unobserved variable and the proxy variable given the other control variables must exceed one (Ohtani (1981), page 627 equation (8)).

root of the mean square error (root MSE) from the model with the unemployment indicator is 0.0758. For this specification, we find that including the proxy variable does not change the root MSE from the specification that omits the proxy – both generate a higher root MSE of 0.0769. Not having the borrower-specific unemployment indicator increases the root MSE by 11 basis points relative to the average monthly default risk of 50 basis points. This conclusion is only reinforced if the relationship between the aggregate unemployment rate and individual unemployment status changes over time. Recent declines in the labor force participation rate suggest this may be the case over the recent period.

The only solution to this problem is better data. An improved proxy is needed which will covary more strongly with the average of the individual borrowers unemployment status. One can imagine such variables being created—perhaps, as discussed earlier, an unemployment rate series for homeowners (or FHA borrowers) and at the metropolitan area level. Alternatively, borrower employment status could be merged from state unemployment insurance records to mortgage servicing records if confidentiality concerns could be addressed.

In language borrowed from an entirely different context, this means that the question of how much aggregate portfolio default risk truly is affected by borrower unemployment status will remain a 'known unknown' over any reasonable horizon. The likely correlation between the market-level unemployment rate and individual borrower unemployment status is so weak that including the proxy variable does not improve the reliability of the portfolio-level default forecast. Thus, while we can address the academically interesting puzzle of why theory and empirics diverge on the importance of borrower unemployment status in accounting for individual default behavior, that does not solve the policy need to be able to control for unemployment risk empirically in a way that helps policy makers pin down its likely influence on portfolio-level default risk.

V. Conclusions

Economic theory and empirical practice regarding the estimation of default vary greatly in their treatment of unemployment risk. While the former assigns a large role for income loss associated with unemployment and other negative labor market shocks through the so-called double trigger hypothesis, empirical models of default used by academics and FHA's outside actuarial reviewer find only a minor influence, at best, associated with unemployment rates. Pinning down the impact of unemployment on default risk is hard because a noisy proxy for individual borrower unemployment status must be used in any such estimation. Of course, that only implies that the results of such an exercise are a lower bound on the true influence of unemployment because of attenuation bias, not that unemployment risk is small or nonexistent. That argues for developing credible ways to try to recover the true impact of this factor, not simply taking at face value estimates suffering from severe attenuation bias.

We suggest a new empirical approach to better gauge the likely impact of unemployment risk on borrower default. A simulation exercise using national data on labor market transitions into and out of work yields two noteworthy results: (a) it closely matches the BLS's national unemployment rate over time; and (b) it confirms that the aggregate unemployment rate is a very noisy proxy for individual employment status. We use a statistical measure of just how noisy that relationship is to recover the 'true' impact of aggregate unemployment on borrower default. Those results imply that unemployment has a powerful influence on default behavior. A large component of the attenuation bias is due to a scaling issue so that even if one ignored the second source of attenuation bias, unemployment is a key default risk factor. While our analysis helps reconcile the difference between theory and empirics on the role of borrower unemployment status on individual default behavior, that insight does not translate into more reliable forecasts of default risk at the portfolio level for entities such as FHA. Under certain conditions, a proxy will help generate unbiased forecasts of portfolio risk, but the precision of those estimates may still be very low. Unfortunately, that is likely the case here, so we cannot be confident that including the market unemployment rate helps generate more reliable forecasts of aggregate default risk in the FHA guarantee portfolio. Solving the problem requires better data, which almost certainly involves significant effort should policy makers decide that more reliable forecasts of FHA default risks are necessary.

References

Aragon, Diego, Andrew Caplin, Sumit Chopra, John V. Leahy, Marco Scoffier, and Joseph Tracy. "Reassessing FHA Risk." Working Paper No. 15802. National Bureau of Economic Research, March, 2010. http://www.nber.org/papers/w15802.pdf.

Bradley, Michael G., Amy Crews Cutts, Biplab K. Gosh, and Wei Liu. "Uncovering the Risks of Strategic Default." Working Paper. CoreLogic, 2012.

------, Amy Crews Cutts, and Wei Liu. "Strategic Mortgage Default: The Effect of Neighborhood Factors." Working Paper. CoreLogic, January, 2013.

Caplin, Andrew, Anna Cororaton, and Joseph Tracy. "Is the FHA Creating Sustainable Homeownership?" NBER Working Paper No. 18190, June, 2012. http://www.nber.org/papers/w18190.

------, Charles Freeman, and Joseph Tracy. "Collateral Damage: Refinancing Constraints and Regional Recessions." *Journal of Money, Credit, and Banking* 29 (November 1997): 496-516.

Deng, Yongheng, John M. Quigley, and Robert van Order. "Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options." *Econometrica* 68 (March 2000): 275-307.

Elul, Ronel, Nicholas S. Souleles, Chormsisengphet Souphala, Dennis Glennon, and Robert Hunt. "What 'Triggers' Mortgage Default?" *American Economic Review* 100 (May 2010): 490-94.

Ferreira, Fernando, Joseph Gyourko, and Joseph Tracy. "Housing Busts and Household Mobility." *Journal of Urban Economics* 68 (July 2010): 34-45.

------, Joseph Gyourko, and Joseph Tracy. "Housing Busts and Household Mobility: An Update." *Economic Policy Review* 18 (November 2012): 1-15.

Foote, Chris, Kristopher Gerardi, Lorenz Goote, and Paul Willen. "Reducing Foreclosures: No Easy Answers." In *NBER Macroeconomics Annual, 2009*, 89-138. Chicago and London, University of Chicago Press, 2010.

Foote, Christopher L., Kristopher Gerardi, and Paul S. Willen. "Negative Equity and Foreclosure: Theory and Evidence." *Journal of Urban Economics* 6 (September 2008): 234-245.

Foster, Chester, and Robert van Order. "An Option-Based Model of Mortgage Default." *Housing Finance Review* 3 (October 1984): 351-372.

Frazis, Harley J., Edwin L. Robison, Thomas D. Evans, and Martha A. Duff. "Estimating Gross Flows Consistent With Stocks in the CPS." *Monthly Labor Review*, September 2005, 3-9.

Gerardi, Kristopher, Adam Hale Shapiro, and Paul S. Willen. "Subprime Outcomes: Risky Mortgages, Homeownership Experiences, and Foreclosures." Working Paper No. 07-15. Federal Reserve Bank of Boston, December, 2007.

Ghent, Andra C., and Marianna Kudlyak. "Recourse and Residential Mortgage Default: Theory and Evidence from U.S. States." Working Paper No. 09-10. Federal Reserve Bank of Richmond, July, 2009.

Gyourko, Joseph. *Is FHA the Next Housing Bailout?* A Report Prepared for the American Enterprise Institute, 2011.

Haughwout, Andrew, Ebiere Okah and Joseph Tracy. "Second Chances, Subprime Mortgage Modifications and Re-Default", Staff Report No. 417, Federal Reserve Bank of New York, August 2010.

Integrated Financial Engineering, Inc. *Actuarial Review of the Federal Housing Administration Mutual Mortgage Insurance Fund for Fiscal Year 2008*. Prepared for U.S. Department of Housing and Urban Development. Washington, DC, October 13, 2008.

------. *Actuarial Review of the Federal Housing Administration Mutual Mortgage Insurance Fund Forward Loans for Fiscal Year 2011*. Prepared for U.S. Department of Housing and Urban Development. Washington, DC, October 12, 2011.

____. *Actuarial Review of the Federal Housing Administration Mutual Mortgage Insurance Fund for Fiscal Year 2012*. Prepared for the U.S. Department of Housing and Urban Development. Washington, D.C., November 5, 2012.

Kau, James, Donald Keenan, and Taewon Kim. "Default Probabilities for Mortgages." *Journal of Urban Economics* 35 (May 1994): 278-296.

Moulton, Brent R., "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units." *Review of Economics and Statistics* 72, no. 2, (May 1990): 334-338.

Vandell, Kerry D. "How Ruthless Is Mortgage Default? A Review and Synthesis of the Evidence." *Journal of Housing Research* 6, no. 2 (1995): 245-264.

Figure 1: Labor Market Transition Rates

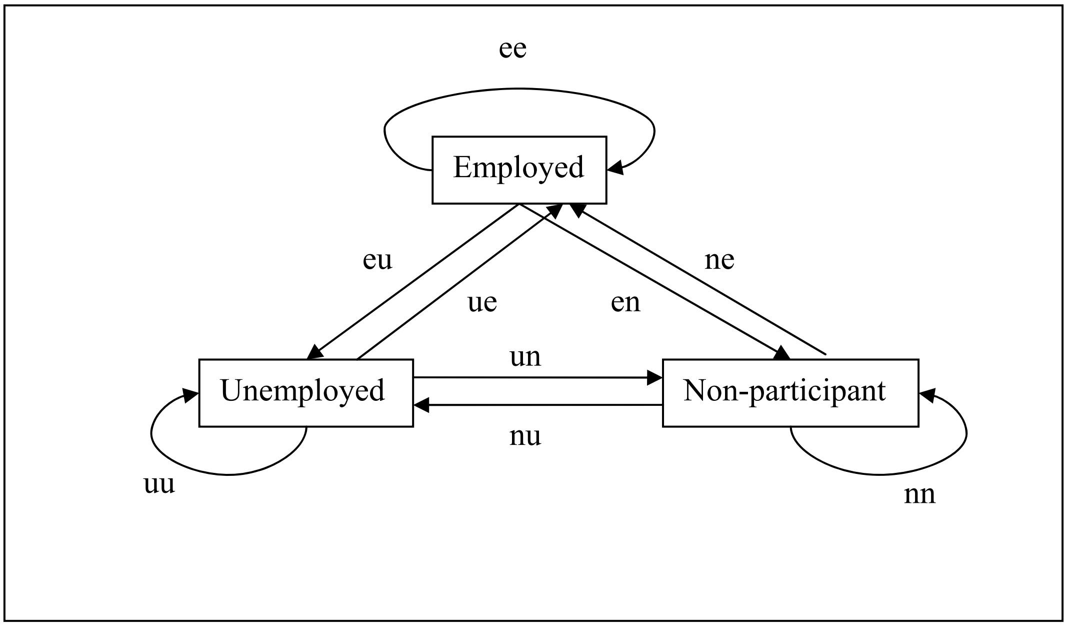

Figure 2: Actual and Simulated Unemployment Rates

Note: BLS non-seasonally adjusted national unemployment rate for individuals 16 year or older. Simulated unemployment rate based on BLS published monthly non-seasonally adjusted transition rates from February 1990 to October 2012 and 150,000 simulated individual histories.

Table 1. LPM Default Regressions

Variable	(1)	(2)
Unemployment rate	0.00036**	0.00063**
	(0.0001)	(0.0001)
Current loan-to-value (LTV):		
80 – 99	0.0016**	0.0016**
	(0.0002)	(0.0002)
100 -- 119	0.0038**	0.0036**
	(0.0003)	(0.0002)
120+	0.0105**	0.0100**
	(0.0008)	(0.0009)
Missing	0.0021**	0.0022**
	(0.0005)	(0.0005)
Credit score (FICO):		
<579	0.0120**	0.0119**
	(0.0004)	(0.0004)
580 – 679	0.0049**	0.0049**
	(0.0002)	(0.0002)
Missing	0.0036**	0.0036**
	(0.0002)	(0.0002)
Not fully documented	–0.0008**	–0.0007**
	(0.0001)	(0.0001)
Not purchase mortgage	0.0015**	0.0015**
	(0.0002)	(0.0002)
Not 30-year term	–0.0016**	–0.0015**
	(0.0002)	(0.0003)
Not single family residence	–0.0002*	–0.0003**
	(0.0001)	(0.0001)
Recourse state	0.0000	–0.0003
	(0.0004)	(0.0007)
Judicial foreclosure state	–0.0003**	–0.0006
	(0.0003)	(0.0002)
Fixed effects included:		
Vintage year of origination	Y	Y
Year	Y	Y
MSA	N	Y

Notes: Coefficients represent the average change in the monthly default risk for a unit change in the indicated variable. Standard errors are given in parenthesis and are clustered at the MSA level. Mean monthly default risk is 0.0054. Sample is restricted to the top-50 metropolitan areas. Sample size is 3,238,458.
** significant at the 5% level, * significant at the 10% level

Appendix Table A1. Top 50 MSA List

1. Albuquerque, NM (10740)	26. Memphis, TN-MS-AR (32820)
2. Atlanta-Sandy Springs-Marietta, GA (12060)	27. Milwaukee-Waukesha-West Allis, WI (33340)
3. Austin-Round Rock, TX (12420)	28. Minneapolis-St Paul-Bloomington, MN-WI (33460)
4. Baltimore-Towson, MD (12580)	29. Nashville-Davidson-Murfreesboro-Franklin, TN (34980)
5. Birmingham-Hoover, AL (13820)	30. Ogden-Clearfield, UT (36260)
6. Buffalo-Niagara Falls, NY (15380)	31. Oklahoma City, OK (36420)
7. Charlotte-Gastonia-Concord-Rock Hill, NC-SC (16740)	32. Omaha-Council Bluffs, NE-IA (36540)
8. Cincinnati-Middletown, OH-KY-IN (17140)	33. Orlando-Kissimmee, FL (36740)
9. Clarksville, TN-KY (17300)	34. Phoenix-Mesa-Scottsdale, AZ (38060)
10. Cleveland-Elyria-Mentor, OH (17460)	35. Pittsburgh, PA (38300)
11. Colorado Springs, CO (17820)	36. Portland-Vancouver-Beaverton, OR-WA (38900)
12. Columbia, SC (17900)	37. Raleigh-Cary, NC (39580)
13. Columbus, OH (18140)	38. Richmond, VA (40060)
14. Dayton, OH (19380)	39. Riverside-San Bernardino-Ontario, CA (40140)
15. Denver-Aurora, CO (19740)	40. Rochester, NY (40380)
16. Fayetteville, NC (22180)	41. Sacramento-Arden-Arcade-Roseville, CA (40900)
17. Hartford-CT (25540)	42. Saint Louis, MO-IL (41180)
18. Houston-Sugar Land-Baytown, TX (26420)	43. Salt Lake City, UT (41620)
19. Indianapolis-Carmel, IN (26900)	44. San Antonio, TX (41700)
20. Jacksonville, FL (27260)	45. San Diego-Carlsbad-San Marcos, CA (41740)
21. Kansas City, MO-KS (28140)	46. Tampa-St. Petersburg-Clearwater, FL (45300)
22. Killeen-Temple-Fort Hood, TX (28660)	47. Tucson, AZ (46060)
23. Las Vegas-Paradise, NV (29820)	48. Tulsa, OK (46140)
24. Little Rock, AR (30780)	49. Virginia Beach-Norfolk-Newport News, VA-NC (47260)
25. Louisville-Jefferson County, KY-IN (31140)	50. Wichita, KS (48620)

Appendix Table A2. LPM Prepayment Regression

Variable	(1)
Unemployment rate	−0.0004
	(0.0002)
Current loan-to-value (LTV):	
80 – 99	−0.0022**
	(0.0004)
100 -- 119	−0.0003
	(0.0005)
120+	0.0013
	(0.0011)
Missing	−0.0002
	(0.0006)
Credit score (FICO):	
<579	−0.0055**
	(0.0005)
580 – 679	−0.0030**
	(0.0001)
Missing	−0.0018**
	(0.0002)
Not fully documented	0.0007**
	(0.0002)
Not purchase mortgage	0.0021**
	(0.0002)
Not 30-year term	−0.0011**
	(0.0004)
Not single family residence	–0.0003
	(0.0004)
Recourse state	−0.0022*
	(0.0010)
Judicial foreclosure state	–0.0011
	(0.0010)
Refinance incentive	0.0089**
	(0.0009)
Fixed effects included:	
Vintage year of origination	Y
Year	Y
MSA	N

Notes: Coefficients represent the average change in the monthly prepayment risk for a unit change in the indicated variable. Standard errors are given in parenthesis and are clustered at the MSA level. Mean monthly prepayment risk is 0.0087. Sample is restricted to the top-50 metropolitan areas. Sample size is 3,238,458.
** significant at the 5% level, * significant at the 10% level

Appendix Table A3. Summary Statistics for regression covariates

Variable	Mean	Standard Deviation	Minimum	Maximum
Unemployment Rate	8.073	2.396	2.2	15
Current loan-to-value (LTV):				
80 – 99	0.554	0.497	0	1
100 – 119	0.331	0.471	0	1
120+	0.034	0.182	0	1
Missing	0.0162	0.126	0	1
Credit Score (FICO):				
< 579	0.047	0.211	0	1
580 – 679	0.410	0.492	0	1
Missing	0.132	0.339	0	1
Not full documented	0.521	0.500	0	1
Not purchase loan	0.393	0.488	0	1
Not 30-year term	0.044	0.205	0	1
Not single family residence	0.235	0.424	0	1
Refinance incentive	0.592	0.599	0	3.9