Reconciling theory and empirics on the role of unemployment in mortgage default

Joseph Gyourko a, Joseph Tracy b,⇑

a The Wharton School, University of Pennsylvania & NBER, United States
b Federal Reserve Bank of New York, New York, United States

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 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 more than 100. In addition, our analysis indicates 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. Hence, aggregate portfolio-level risk estimates for mortgage guarantors such as FHA also are not improved.

These views represent those of the authors and not necessarily those of the Federal Reserve Bank of New York or the Federal Reserve System. This is a revised version of a paper that previously circulated under the title "Unemployment and Unobserved Credit Risk in the FHA Single Family Mortgage Insurance Fund (NBER Working Paper No. 18880). John Grigsby provided excellent research assistance. We appreciate the helpful comments of Andrew Haughwout, Wilbert van der Klaauw, the editor (Stuart Rosenthal) and referees, but remain responsible for any errors.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

There is a marked divergence between the important role played by unemployment in economic models of default and its modest empirical impact in loan level studies of default. In now standard theory, unemployment is one of the two key factors driving default according to the so-called double trigger hypothesis that became widely discussed during the recent housing bust (e.g., Gerardi et al. (2007), Foote et al. (2008), and Foote et al. (2010)). In this approach, a sufficiently severe income loss (typically from becoming unemployed) virtually guarantees default if the borrower also suffers from negative equity. Once an unemployed borrower in negative equity runs out of liquid financial resources, mortgage payments simply cannot be made and a default is inevitable, as not even a quick sale can pay off the outstanding balance.

However, this straightforward theoretical implication does not seem to be borne out in empirical work. One prominent housing agency, the Federal Housing Administration (FHA) did not even include a control for unemployment in its default specifications for several years. From 2008 until the most recent annual evaluation, FHA's outside actuarial reviewer discontinued use of a metropolitan area-wide unemployment rate in its default models on the grounds that it did not help in explaining default. 1 The most recent annual review of the FHA insurance program reversed this decision and included aggregate unemployment rates as controls in its underlying default (and prepayment) models. However, it found only a modest impact of unemployment on the probability of default, a finding that is consistent with other recent empirical evidence (e.g., Elul et al. (2010)) which also reports a much smaller influence on default for unemployment compared to other known risk factors such as borrower credit quality as reflected in FICO scores and loan-to-value (LTV) ratios.

1 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, p. A-13).
We bring new empirical analysis to bear that reconciles the discrepancy in favor of theory. The discrepancy arises because the default decision occurs at the household level, but the unemployment status of individual borrowers is not observed. Thus, a market-level unemployment rate is used to proxy for the unobserved unemployment status of the borrower in empirical models of individual default. This turns out to be a very noisy proxy.\(^2\) Even in a market with a high unemployment rate of 10%, the vast majority of borrowers still are employed, and an increase in that rate obviously does not mean that most borrowers lose their jobs and become income constrained. This leads to attenuation bias in the estimated effects of becoming unemployed on default risk.

We propose a new empirical strategy to measure this bias. It uses results from a simulation exercise involving transitions into and out of unemployment to help gauge the true impact of unemployment on default. We also construct new metropolitan area-level unemployment measures for homeowners with mortgages, in contrast to the standard government measures which include renters and owners without mortgages. Our results indicate that attenuation bias in the estimation of the impact of becoming unemployed on default is severe, with the standard approach underestimating the effect by a factor of more than 100. Correcting for attenuation bias indicates that borrower unemployment status is very important in determining default behavior, with the likely impact larger than that found for other well-known risk factors such as FICO scores.

This addresses the conundrum of why theory suggests that unemployment plays a critical role in individual borrower default decisions, but empirical default specifications do not find such a strong influence. However, this still leaves open the issue of whether including even a noisy proxy such as the official metropolitan area unemployment rate in regression models can help improve the accuracy of a portfolio-level default forecast by a large mortgage insurance guarantor like FHA. Our analysis shows that including such a noisy proxy for the borrower’s unemployment status does not improve the accuracy of the default forecast as measured by the mean square error. Overcoming this data limitation is important since the ability to merge borrower-specific unemployment information into default specifications would improve efforts to forecast and reserve for expected mortgage losses.

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 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. To better understand this issue, we turn next to how unemployment-related risk would affect a mortgage default. That is followed by a presentation and discussion of our key results in Section 4. This section also addresses the implication of our analysis for whether the FHA’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. There is a brief summary and conclusion.

2. Default and unemployment risk

2.1. Literature review

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 involves 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 study indicates that about 85% of households with a mortgage who are estimated to be in negative equity are current on their debt service payments.\(^7\) Empirical analysis in the academic literature also concludes that the decision to default is based on more than current and prospective negative equity.\(^4\) In particular, default has been shown to be associated with negative shocks to income, including that arising from becoming unemployed.\(^5\) 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.\(^2\) 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 mortgage payments. That will precipitate default because the borrower also cannot pay off the mortgage in full from sale proceeds.\(^7\)

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 of special servicers which tell it that income loss is the primary reason why the typical FHA borrower is no longer current on her mortgage payments.\(^8\)

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 default studies. If not, then theoretical models of mortgage default need to be reworked to emphasize risk factors other than unemployment. To better understand this issue, we turn next to how unemployment risk should be measured if it were to be controlled for in a loan level empirical investigation of mortgage default.

2.2. Measuring unemployment risk

To measure how unemployment-related risk would affect a portfolio of mortgages insured by a guarantor such as the FHA,

\(^2\) This point was raised by Haughwout et al. (2010, p. 17) footnote 22.

\(^3\) This particular estimate is from a study by CoreLogic in the second quarter of 2012. See the web article here (http://www.denesws.com/articles/borrowers-in-negative-equity-slowly-declining-as-home-values-gain-report-2012-09-12).

\(^4\) 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.

\(^5\) See Foster and van Order (1984) and Vandell (1995) for early discussions and presentation of data on this matter.

\(^6\) 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.

\(^7\) 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 Ferreras 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.

\(^8\) 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.
one would like to include a variable in a default specification that accurately reflects changes over time in the unemployment status of individual borrowers. That would allow researchers to estimate the extent to which becoming unemployed is positively correlated with default. The problem is the mortgage servicing data that are typically used to estimate default models do not track the borrower’s unemployment 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 \( t \). Further, let \( D_{ijt}^0 \) 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 \( x_{ijt} \), our simple default specification can be expressed as in Eq. (1):

\[
D_{ijt} = \beta_0 + \beta_1 x_{ijt} + \cdots + \beta_k x_{ijt-k} + \beta_{k+1} U_{jt} + \epsilon_{ijt},
\]

where \( \epsilon_{ijt} \) 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 Eq. (2):

\[
D_{ijt} = \tilde{\beta}_0 + \beta_1 x_{ijt} + \cdots + \beta_k x_{ijt-k} + \beta_k U_{jt} + \tilde{\epsilon}_{ijt}.
\]

Pinning down the relationship between \( \tilde{\beta}_0 \) and \( \beta_k \) obviously is key. To help do so, let the relationship between an individual’s employment status at time \( t \) and her market unemployment rate in the same period be described by the following simple linear equation\(^{10}\):

\[
I_{ijt} = \delta_0 + \delta_1 U_{jt} + \mu_{ijt}.
\]

Using these three equations, the coefficient on the market-level unemployment rate \( \beta_k \) in Eq. (2) will equal \( \beta_k \delta_1 \) (with \( \tilde{\beta}_0 = \beta_0 + \beta_k \delta_1 \) and \( \tilde{\epsilon}_{ijt} = \epsilon_{ijt} + \beta_k \mu_{ijt} \)). Thus, if the market unemployment rate is such a noisy proxy for individual unemployment status that \( \delta_1 \) in Eq. (3) is close to zero, then the regression results from estimating Eq. (2) would show that unemployment has little impact on default decisions because of attenuation bias even if individual borrower unemployment spells significantly increase the risk of default.

An estimate of \( \delta_1 \) is needed to recover the coefficient of interest, \( \beta_k \). To help gauge the potential magnitude of \( \delta_1 \) 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.\(^{11}\) These data are available only at the national level. Fig. 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 \( ee \). Similarly, the probability that the same individual becomes unemployed is given by \( uu \), while the probability that the individual leaves the labor market is \( nn \). The analogous permutations for individuals starting out either as unemployed or out of the labor market yield the nine possibilities depicted in Fig. 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. Fig. 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-insured mortgage data to be discussed below – from January 2005 to March 2012. As expected, the two series track each other very closely, with the small differences due to sampling variation.

This simulated data is then used to obtain an estimate for \( \delta_1 \). More specifically, we begin by randomly merging simulated employment histories into a sample of mortgages that we will 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

\(^{9}\) 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. 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.

\(^{10}\) For simplicity, this specification ignores that \( I_{ijt} \) is a limited dependent variable.

\(^{11}\) The labor flow data is produced by the Bureau of Labor Statistics. For details, see Frazis et al. (2005).
underwriting standards. In addition, we can directly compare our results to very recent findings reported by FHA's outside actuarial reviewer. The next section describes our underlying mortgage data and the key variables used in that analysis.

3. Data description

Our primary data source on loans that likely were insured by FHA is Lender Processing Services Inc. (LPS) Applied Analytics. The LPS data form a monthly panel of loans, which is derived from the portfolios of large U.S. mortgage servicers. 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, 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.

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. 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 primary unemployment data come from the BLS Local Area Unemployment Statistics (LAUS) database, which is comprised of non-seasonally adjusted county- and state-level unemployment rates. 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). We also experiment with a new measure more narrowly focused on owners with mortgages that is discussed more fully below. The current LTV ratio provided in the LPS data is calculated by dividing the balance remaining on the mortgage by the sale price

\[ R^2 = 0.0019. \]

This estimate of \( \delta_1 \) highlights how noisy a proxy the aggregate unemployment rate is for any given individual's true unemployment status. Eqs. (2) and (3) above indicate that the true impact of a borrower becoming unemployed on the probability of defaulting should be scaled up by \( 1/(\delta_1 \approx 1/0.00553, \text{as } \delta_k = \delta_0 \delta_1). \) Homeowners with mortgages are not a random sample of all adults, so the correlation based on actual borrower-specific unemployment would likely differ somewhat from this estimate. However, this simulation should capture the approximate magnitude of the attenuation bias.

To investigate this question further, we estimate a default specification like that in Eq. (1) on a sample of FHA-insured loans, and then use our imputed value of \( \delta_1 \) to adjust the regression coefficient on the market-level unemployment variable. Our results do not hinge on our focus on FHA mortgages and would apply to any sample of mortgages, although it is useful to focus on a single entity's pool because it helps minimize noise due to variation in

\[ \text{Unemployed} = 0.00042 + 0.00553 \times \text{Unemployment rate} \]

\[ (0.00052) (0.00006) \]

\[ R^2 = 0.0019. \]

12 Here we follow the convention of measuring the unemployment rate so that a unit change equals a one percentage point change.

13 This regression has 4.081 million simulated monthly observations.

15 Another advantage of focusing on FHA-insured mortgages is that we do not have to worry about biases arising from not being able to observe second liens (which is a potential problem with the LPS data base). Initial loan-to-value (LTV) ratios typically are so high on FHA-insured product that there are no other loans taken out. For other mortgage sectors, a downward bias results in the updated LTVs if we do not capture all subsequent loans. This, in turn, would lead to a less clean comparison of the relative impact of unemployment and LTV on default risk, and that is a trade-off we do not want to make. That said, a referee correctly noted that the correlation between the BLS metro area unemployment rates and average unemployment indicators for FHA borrowers well could differ from that of all mortgage borrowers. If so, one would expect some variation in the degree of attenuation across different categories of mortgages, so we caution against generalizing our precise results to the entire mortgage market.

16 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). We also experiment with a new measure more narrowly focused on owners with mortgages that is discussed more fully below.

14 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 to 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 to 2004.
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.\footnote{We use the Corelogic metro area overall house price indices. For information on them, see http://www.corelogic.com/products/corelogic-hpi.aspx.}

We also construct indicator 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.\footnote{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).} We measure the incentive for a borrower to refinance as the difference between the average 30-year fixed-rate mortgage rate in the origination month and the current month. Finally, we construct indicators 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 indicators 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.\footnote{Only 6.8\% of the monthly observations have durations of 49 months or longer.}

4. Estimation and results

4.1. Linear probability model results

We begin by estimating a simple linear probability model for the default indicator as described above in Eq. (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 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 to 680, and 680 + [left-out category]), indicators 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 indicators 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 to 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.\footnote{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 a 3.6 (column 1) to 6.3 (column 2) basis point increase in the monthly default risk. These effects are 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 Eqs. (1)-(3) that the point estimates reported in the first row of Table 1 are the product of $\beta_0$ 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 $\delta_3 = 0.00053$ implies that the impact of a borrower unemployment spell on default is likely very large, with an imputed

<table>
<thead>
<tr>
<th>Table 1</th>
<th>LPM default regressions.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>(1)</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>0.00036 $^*$</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
</tr>
<tr>
<td>Current loan-to-value (LTV)</td>
<td>0.0016 $^*$</td>
</tr>
<tr>
<td>80–99</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>100–119</td>
<td>0.0038 $^*$</td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
</tr>
<tr>
<td>120+</td>
<td>0.0105 $^*$</td>
</tr>
<tr>
<td></td>
<td>(0.0008)</td>
</tr>
<tr>
<td>Missing</td>
<td>0.0021 $^*$</td>
</tr>
<tr>
<td></td>
<td>(0.0005)</td>
</tr>
<tr>
<td>Credit score (FICO)</td>
<td></td>
</tr>
<tr>
<td>&lt;579</td>
<td>0.0120 $^*$</td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
</tr>
<tr>
<td>580–679</td>
<td>0.0049 $^*$</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
</tr>
<tr>
<td>Missing</td>
<td>0.0036 $^*$</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
</tr>
<tr>
<td>Not fully documented</td>
<td>−0.0008 $^*$</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
</tr>
<tr>
<td>Refinance mortgage</td>
<td>0.0015 $^*$</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
</tr>
<tr>
<td>Not 30-year term</td>
<td>−0.0016 $^*$</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
</tr>
<tr>
<td>Not single family residence</td>
<td>−0.0002 $^*$</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
</tr>
<tr>
<td>Recourse state</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
</tr>
<tr>
<td>Judicial foreclosure state</td>
<td>−0.0003 $^*$</td>
</tr>
<tr>
<td>Fixed effects included</td>
<td>(0.0003)</td>
</tr>
<tr>
<td>Vintage year of origination</td>
<td>Y</td>
</tr>
<tr>
<td>Year</td>
<td>Y</td>
</tr>
<tr>
<td>MSA</td>
<td>N</td>
</tr>
</tbody>
</table>

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. $^\dagger$ Significant at the 10\% level.
magnitude of 645 basis points. This is roughly 12 times the average monthly default risk.\(^{21}\)

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 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.

We performed the following robustness check on the extent of attenuation bias arising from using the unemployment rate as a proxy for the borrower specific unemployment indicator. First, we generated simulated defaults and prepayments for our sample of mortgages, with the monthly predicted default and prepayment probabilities 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). The prepayment coefficient for the unemployment indicator was set to set to a negative 651 basis points. These estimated monthly predicted default and prepayment probabilities are then used to simulate defaults and prepayments for each mortgage. More specifically, 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; the prepayment indicator is coded 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. The default data going forward are censored if the borrower prepaids. 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), which 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 associated with becoming unemployed is large.

Similar attenuation results when we simulate defaults and prepayments using a stylized double trigger model. Following the procedure described above for this analysis, the monthly default risk was increased by 1000 basis points for a borrower who becomes unemployed and is in negative equity. We assume no strategic default in this example so, there is no increase in default risk if an unemployed borrower has positive equity. Analogously, we increase the prepayment risk by 1000 basis points for a borrower who becomes unemployed and has positive equity, and assume 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.\(^{22}\)

The severe attenuation bias documented here arises from two sources. The first source reflects a scaling issue that affects the interpretation of the estimated impact. Because an individual borrower either is employed or unemployed, \(I^U_{jt}\) in Eq. (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 Eq. (4).

\[
D_{jt} = \beta_0 + \beta_1 X_{jt1} + \cdots + \beta_{k-1} X_{jt(k-1)} + \beta_k I^U_{jt} + \epsilon_{jt},
\]

where \(I^U_{jt}\) 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 \(I^U_{jt}\), in this regression ranges from zero to one so that \(\beta_k\), 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 change 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 Eq. (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 \(I^U_{jt}\) or \(I^E_{jt}\), 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, \(I^U_{jt}\), 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 \(I^U_{jt}\) in place of the BLS metro area unemployment rate yields an estimate for \(\beta_k\) 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, \(I^U_{jt}\), the coefficient equals one. That is, there is no attenuation bias from using \(I^U_{jt}\), as a proxy variable for \(I^U_{ijt}\). The main effect of switching from the borrower-specific unemployment-

\(^{21}\) Using column 1’s estimated coefficient implies that \(\beta_0 = 0.000357\). Rearranging leaves us that \(\beta_k = 0.000357/0.00006 = 6000\), which implies 654 basis points given the underlying units of measurement.

\(^{22}\) 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.
Table 2
Descriptive statistics for alternative metro area unemployment rates.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Renters</td>
<td>6.47</td>
<td>2.39</td>
<td>0.64</td>
<td>17.26</td>
</tr>
<tr>
<td>Homeowners</td>
<td>3.47</td>
<td>1.46</td>
<td>0.00</td>
<td>10.48</td>
</tr>
<tr>
<td>Homeowners w. mortgages</td>
<td>4.25</td>
<td>2.64</td>
<td>0.00</td>
<td>16.86</td>
</tr>
<tr>
<td>BLS rate</td>
<td>8.28</td>
<td>1.81</td>
<td>4.20</td>
<td>14.30</td>
</tr>
</tbody>
</table>

Notes: Number of observations 576. The calculations are restricted to the top-50 metro areas measured at the CBSA level. The unemployment rate for Renters, Homeowners and Homeowners with mortgages are calculated as survey weighted averages of individual level unemployment indicators taken from monthly Current Population Survey data. The BLS unemployment rate is taken from the BLS Local Area Unemployment Statistics (LAUS) and is based on all adults age 16 or older and is defined as the ratio of the number of unemployed to the sum of the number employed and the number unemployed. Individuals who are classified as “out-of-the-labor-force” are not counted.

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 microdata 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 \(I_{jt}\) is used instead. Measurement error arises because each variable is created from different samples of people. First, \(I_{jt}\) is averaged over homeowners with mortgages (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 \(I_{jt}\). A second factor is that the BLS unemployment rate excludes individuals who are classified as being out of the labor force.

If the attenuation bias associated with using \(U_{jt}\) as a proxy for \(I_{jt}\) could be estimated, then this would provide a check on our earlier estimate of the overall attenuation bias associated with \(U_{jt}\). We do so using the monthly Current Population Survey (CPS) to produce the reported unemployment rate. In addition to questions that determine an individual’s unemployment status, the basic monthly CPS survey asks if the household owns or rents. Starting in 2011, the March supplement to the CPS added a question on whether the household has a mortgage.

More specifically, the CPS interviews a residence for four consecutive months, rotates the residence out of the survey sample for eight months, and then re-interviews the residence for four additional months. Households and individuals within the household that remain in the same residence can be matched across the monthly surveys. Given the interview structure, the mortgage information from the March survey can be linked back to December of the prior year and forward to June of the current year.

Using the available 2011 and 2012 CPS data, we created a sample of households from the largest 50 metro areas. For each household head, we construct indicators for whether the individual is unemployed, is a homeowner or a renter, and for homeowners if a mortgage exists or not. Using the survey weights, for each metro area and month available, the average of the unemployment indicators for renters, homeowners and homeowners with mortgages is then constructed. Descriptive statistics on these various unemployment rates, in addition to the BLS unemployment rate are provided in Table 2. The data indicate that renters have higher unemployment rates than homeowners, and that homeowners with mortgages have higher unemployment rates than homeowners who own their homes without any debt. The BLS unemployment rates are magnified relative to the constructed unemployment rates due to exclusion from the denominator of individuals who are not in the labor force.

The attenuation from using the BLS unemployment rate as a proxy for the average unemployment for homeowners with mortgages now is estimated with the following results:

\[
\text{Unemployment rate, homeowner\_mortgage} = -0.083 + 0.523 \times \text{BLS metro area unemployment rate} \quad (0.483/0.057)
\]

\(R^2 = 0.128\).

The estimated coefficient (standard error in parentheses) on the BLS metro area unemployment rate implies an attenuation of 1.92 (\(=1/0.523\)). Combining this with the attenuation of 100 from the scaling, this gives an overall attenuation of 192. Even though we are working with a very short time period with these particular data, note that this closely matches our earlier overall attenuation of 180 derived from the simulated employment transition data. Moreover, we cannot tailor the CPS data to homeowners with FHA mortgages. That both approaches yield similar results increases our confidence in the magnitude of the attenuation bias being generated.

4.2. Discussion: individual borrower default risk

Table 1’s results imply that becoming unemployed is at least five times more influential than having a high versus a low FICO score: the 120 basis point impact of going from a 680+ FICO score to below a 580 FICO score is approximately one-half the 645 basis point impact implied after adjusting column 1’s coefficient on the metropolitan area unemployment rate variable. However, it is likely 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.

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, the year dummies would not be strongly correlated with unemployment status, the year dummies would not be statistically significantly correlated with default. In addition, Gyourko (2011) notes that time indicators 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.

We have also used a simple linear regression framework in this analysis rather than the more standard hazard model specification.

23 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 (Moulton (1990)).
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 \( t_1 \) 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 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 borrower unemployment risk is likely a very important factor in explaining mortgage default, at least on a par with other well-known risk factors such as borrower credit scores and leverage.26

4.3. Discussion: portfolio-level default risk

While use of the market-level proxy for individual borrower unemployment status clearly leads to attenuation bias in estimating the marginal effect of unemployment on the default behavior for individual 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 possibility. 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 improved and 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 \( \bar{D}_t \). 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.

\[
D_t = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_k x_{kt} + \beta_1 I_t + \epsilon_t
\]

If no proxy variable is used for \( I_t \) in the estimation of Eq. (5), then the predicted portfolio default rate would be given by Eq. (6).

\[
\bar{D}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1t} + \cdots + \hat{\beta}_k x_{kt}.
\]

Given our assumptions, the conditional expected value for the difference between the actual and predicted portfolio default rates is \( \beta_1 I_t \). Note further that adding time fixed effects in the form of year indicators to the specification that includes no proxy variable would absorb the unemployment risk. In that case, letting \( t = 0 \) denote the omitted time period, the conditional expected value for the estimated year effect associated with period \( t \) would be \( \beta_1 (I_t - I_0) \).

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.\(^{27}\) However, this estimation approach 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 \( \mu_k \) in (3). In this case, the portfolio default risk is given by Eq. (7).

\[
D_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1t} + \cdots + \hat{\beta}_k x_{kt} + \hat{\beta}_1 U_t + \tilde{\epsilon}_t.
\]

Given our assumptions, OLS will give unbiased coefficient estimates so that the conditional expected value of the difference between the actual and predicted yearly portfolio default rates will be \( -\beta_1 \mu_t \). In this case, year effects, if added to the specification would reflect the residual unemployment risk, \( \mu_t \). In addition to controlling for any left-out-variable bias on the coefficients of the other explanatory variables, adding the proxy variable may improve the ability to estimate the aggregate default risk in a portfolio.28

However, the reliability of that aggregate default risk forecast depends upon the strength of the relationship between the market unemployment rate proxy and individual unemployment status. More specifically, the forecast mean square error (MSE) actually 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 (Ohtani (1981)).29 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 simulated unemployment indicator, the BLS unemployment rate and dropping both. The square 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 generating 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 borrower’s unemployment status. A limitation of

26 This conclusion is corroborated in a recent working paper by Gerardi et al. (2013) that uses micro panel data to investigate the role of individual borrower unemployment on mortgage default. They use data from the 2009 Panel Study of Income Dynamics Supplement on Housing, Mortgage Distress, and Wealth, which consisted of 12 sub-waves conducted over the course of the year. The borrower’s equity position is constructed from the self-reported house value and the mortgage balances; no credit score is available; and, default is defined as the borrower being 60-days or more delinquent. The authors find that the average logit marginal effect on default associated with being unemployed for at least 6-months is 0.092, while the average marginal effect associated with having a current LTV above 120 is 0.058 (relative to a current LTV below 90). These average marginal effects on the probability that a borrower is 60+ days delinquent are not directly comparable to our transition matrix given that the probability that a borrower is 60+ days delinquent are not directly comparable to our transition matrix given that the probability that a borrower is 60+ days delinquent are not directly comparable to our transition matrix.

27 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 borrower’s unemployment status. A limitation of

28 This is an improvement over the specification that omits the proxy variable but includes year effects in that it is more straightforward 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.

29 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.
Table A1
Top 50 MSA list.

1. Albuquerque, NM (10740) 26. Memphis, TN-MS-AK (32820)
5. Birmingham–HOOVER, AL (13820) 30. Ogden–Clearfield, UT (36260)
6. Buffalo–Niagara Falls, NY (15380) 31. Oklahoma City, OK (36420)
10. Cleveland–Elyria–Mentor, OH (17460) 35. Pittsburgh, PA (38300)
13. Columbus, OH (18140) 38. Richmond, VA (40060)
15. Denver–Aurora, CO (19540) 40. Rochester, NY (40380)
16. Fayetteville, NC (21180) 41. Sacramento–Arden–Arcade–Roseville, CA (40900)
17. Hartford–CT (25540) 42. Saint Louis, MO–IL (41180)
18. Houston–Sugar Land–Baytown, TX (26620) 43. Salt Lake City, UT (41620)
19. Indianapolis–Carmel, IN (26500) 44. San Antonio, TX (41700)
20. Jacksonville, FL (27260) 45. San Diego–Carlsbad–San Marcos, CA (41740)
22. Killeen–Temple–Fort Hood, TX (28660) 47. Tucson, AZ (46060)
23. Las Vegas–Paradise, NV (28920) 48. Tulsa, OK (46140)
25. Louisville–Jefferson County, KY–IN (31140) 50. Wichita, KS (48620)

Table A2
LPM Prepayment Regression.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployment rate</td>
<td>0.0004</td>
<td>(0.0002)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Credit score (FICO)</td>
<td>579</td>
<td>–0.0055</td>
<td>0.005</td>
<td>1.0</td>
</tr>
<tr>
<td>Not fully documented</td>
<td>0.0007</td>
<td>(0.0002)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Refinance mortgage</td>
<td>0.0022</td>
<td>(0.0002)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Not 30-year term</td>
<td>0.0010</td>
<td>(0.0004)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Refinance incentive</td>
<td>0.0089</td>
<td>(0.0009)</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

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.

Table A3
Summary statistics for regression covariates.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployment rate</td>
<td>0.003</td>
<td>0.0015</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Credit score (FICO)</td>
<td>579</td>
<td>0.211</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Not fully documented</td>
<td>0.521</td>
<td>0.500</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Refinance mortgage</td>
<td>0.393</td>
<td>0.488</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Not 30-year term</td>
<td>0.044</td>
<td>0.020</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Notes: Monthly mortgage level servicing data from LPS for a 5% random sample of GNMA mortgages originated from January 2005 to March 2012.
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 an estimate of 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 (i.e., measurement error), 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 Fannie Mae default risks are necessary.

Acknowledgement

Gyourko thanks the Research Sponsor Program of the Zell/Lurie Real Estate Center at Wharton. He also appreciates the support of the National Research Initiative of the American Enterprise Institute for general research on the Federal Housing Administration.

Appendix A

See Tables A1–A3.

References


