# New Construction and Mortgage Default *

Tom Mayock
University of North Carolina at Charlotte
tmayock@uncc.edu

Konstantinos Tzioumis
ALBA Business School
American College of Greece
ktzioumis@alba.acg.edu

October 15, 2018

**Abstract**

In this paper we argue that because of non-linear depreciation schedules, appraisal complications, and homebuilders' significant bargaining power, loans collateralized by new construction are more likely to go into default relative to purchase loans for existing homes. Using loan-level mortgage records for more than 3 million loans originated between 2004 and 2009, we provide strong empirical evidence in support of this hypothesis. The unconditional default rate for mortgages used to purchase new construction was 5.6 percentage points higher than the default rates for other purchase loans in our sample. In our richest models that include extensive controls for borrower and loan characteristics as well as Census-tract-origination-year fixed effects, we find that loans for new homes were roughly 1.8 percentage points more likely to default.

*Keywords:* Mortgage Default, New Houses, Collateral Value.

*JEL Classification:* G01; G21; R14; R52.

---

*The opinions expressed in this paper are those of the authors alone, and do not necessarily reflect the views of the Office of the Comptroller of the Currency or the U.S. Department of the Treasury.

# 1 Introduction

The Great Recession spurred an enormous amount of research on the causes of mortgage default. None of the existing work on mortgage delinquency, however, has investigated whether loans that are collateralized by new construction are more likely to default than loans backed by existing homes. There are several reasons why we might expect loans for new construction to have elevated default rates.

First, previous work (Shilling, Sirmans and Dombrow, 1991; Harding, Rosenthal and Sirmans, 2007; Coulson, Morris and Neill, 2016) has demonstrated that home values exhibit non-linear depreciation schedules with sharp value declines early in the home's life akin to those that have been documented in the market for automobiles (Lacetera, Pope and Sydnor, 2012).[1] Second, new construction is typically more difficult to appraise than existing homes. Specifically, appraisers often use the 'cost approach' for valuing a new home rather than the typical 'market approach,' which is based on comparable sales. The cost approach is meant to better account for the fact that new houses are different in terms of housing characteristics, but this approach arguably yields noisier valuations because of limited independent information on price discovery.[2] Because of these complexities, appraisals for newly constructed housing are expected to be more susceptible to bias.

Third, previous work has provided evidence that bargaining power plays an important role in determining the transaction price for *existing* homes. For example, Harding, Rosenthal and Sirmans (2003) find that parents with school-aged children that face binding constraints on their search horizon pay more for housing services, a phenomenon that helps

[1] For additional background on the literature on housing depreciation, see Malpezzi, Ozanne and Thibodeau (1987), Clapp and Giaccotto (1998), and Harding, Rosenthal and Sirmans (2007).

[2] For instance, homes built during the 2000s have more square footage, more bedrooms and bathrooms, and are more likely to have a slab foundation, air-conditioning, and a garage or carport. For details see the 2015 report by the U.S. Department of Commerce titled "Characteristics of new housing" (`https://www.census.gov/construction/chars/pdf/c25ann2015.pdf`).

explains the well-known seasonality in home values. While we are unaware of any work that studies bargaining power in the context of *new* homes, given the nature of new home purchases, we would expect bargaining power to play an important role in these transactions as well. Homebuilders are real estate professionals that have significant experience bargaining in real estate transactions and extensive knowledge of the local market. The typical homebuyer, in contrast, buys and sells homes very infrequently. Furthermore, while the owner of a single home loses only one dollar for a one dollar reduction in sales price, a homebuilder with financial interests in many other properties in the local market will consider how price reductions on one home will affect the prices that he or she will receive on other homes.[3] Both of these factors imply that buyers will be in a weaker bargaining position – and thus are expected to pay more – when buying a new home from a builder relative to buying an existing home.

The three mechanisms discussed above all predict a sharp decline in a home's value as it transitions from new to existing home status. There are two primary schools of thought regarding the determinants of mortgage default: the 'strategic' model and the 'double trigger' model.[4] Both of these theories predict that the decline in value associated with the new-existing transition will elevate default risk. In the purely strategic model of default, the probability that a loan experiences delinquency between origination ($t = 0$) and some period $t = s$ – an event that we use to define the dependent variable in our empirical models – is simply the likelihood that a mortgage's loan-to-value ratio (LTV) exceeds some critical threshold between $t = 0$ and $t = s$ . It is clearly the case that all else equal, if the values of new homes fall more quickly (or appreciate less rapidly) than the values of existing homes, purchase loans for new construction are more likely

[3] This mechanism is conceptually similar to the mechanism discussed in Levitt and Syverson (2008), who find that homes owned by real estate agents are marketed longer and receive higher sales prices than other homes.
[4] For a detailed discussion of the strategic default framework, see Vandell (1995). For more on the double trigger framework, see Gerardi et al. (2017).

than purchase loans for existing homes to have LTVs that eventually exceed this LTV threshold, elevating default risk.

The impact of new home status on loan performance in the double trigger model can be analyzed in a similar fashion. In this framework, borrowers default on a mortgage when they have negative equity and they receive a negative income shock. If a new home's value falls sharply as it transitions from new to existing construction, then borrowers that purchased a new home are more likely than borrowers that purchased an existing home to experience negative equity over any fixed horizon. The higher likelihood of experiencing negative equity increases the probability of default for new home buyers in the double trigger model by increasing the risk that the borrower receives a negative income shock while simultaneously having negative equity in the property.

Drawing from these observations, we use loan-level mortgage data to test the hypothesis that mortgages used for purchasing newly-built houses are associated with a higher likelihood of delinquency. In this analysis we find that loans for new homes are more likely to go into default, even after accounting for the neighborhood in which the property is located, the year in which the loan was originated, and a multitude of loan and borrower characteristics. In a year-by-year analysis, we find that this new home effect was present during boom and bust periods. Our results thus do not appear to solely be an artifact of the early-2000s housing boom nor of the Great Recession. Furthermore, our results are robust to several variations in methodology, including models that allow for the potential endogeneity of the new home variable.

The remainder of the paper is organized as follows. In Section 2 we discuss the construction of the database that we use to conduct our empirical analysis. In Section 3 we discuss our empirical methodology and present our primary results. Section 4 describes the results of a series of robustness checks, and Section 5 concludes.

## 2 Data

To conduct our analysis, we combine a number of sources to create a unique dataset that contains information on a variety of property characteristics and loan performance. We utilize DataQuick's standardized assessment and transaction databases (hereafter, *DQ*) to identify arm's length sales of residential parcels that were purchased using a mortgage. In addition to fields that characterize the nature of a property transaction (e.g., sales price, mortgage amount), the DQ data also contains detailed information on financing activity and the structural characteristics of the property. We use the DQ financing data to construct combined-loan-to-value (CLTV) ratios by summing all of the mortgage debt that was originated at the time a property was purchased and dividing this sum by the property's sale price. During the housing boom, borrowers frequently took out multiple mortgages on a property simultaneously. The junior lien mortgage balances in such transactions – colloquially referred to as a 'piggyback loans' – are frequently not reported in servicers' CLTV fields, an omission that can result in a serious underestimate of a borrower's leverage. Our use of the DQ data to construct the CLTV variables thus allows us to create a far more accurate measure of a borrower's equity position in the property at the time of origination than measures based on servicing data alone.

We also use information in the DQ data to construct a measure of a property's age at origination, which we define as the difference between the year in which the mortgage is originated and the year in which the property was constructed.[5] We drop sales where that difference is above 110 years as such observations are likely either very unique historical homes or data entry errors. We also drop sales with a negative difference since these reflect pre-sales.[6] We classify a transaction as a new home sale if the age of the house is

[5] Across states, only about 2 percent of the sales have missing information on either the transfer date or the year built, resulting in a missing 'age at time of sale' variable. The only exception is California, in which these fields are missing for roughly 10 percent of records in the state.

[6] Except for Nevada, pre-sales are only a trivial part of the total home sales in our sample, about 0.7

either zero or one when the sale occurs; properties that were at least 2 years old at the time of the sale thus serve as the reference category in our regression analysis.[7]

We obtain loan performance information by combining an extended version of the OCC Mortgage Metrics database (hereafter, *OCCMM*) and loan-level records for mortgages securitized into private-label mortgage-backed securities (hereafter, *PLS*).[8] [9] Given the different focus of the loan performance datasets, mostly prime loans for the former and mostly subprime loans for the latter, their combination ensures that our sample contains loans from across the full credit spectrum. We monitor loan performance for four years following origination to create our measure of default. Specifically, we create an indicator variable ($Delinq_{ijt}$) that equals 1 if a loan was at least 90 days days past due or entered the foreclosure process within four years of origination. Because the market share of the OCCMM data declines significantly prior to 2004, we limit our sample to loans originated between 2004 and 2009. Additionally, we limit our analysis to those states where our final database contains at least 10 percent of the home-purchase originations for 1-4 family dwellings listed for that state in the Home Mortgage Disclosure Act (HMDA) loan application register data for each of the years in the sample. Lastly, for loans within a county in a given year to be included in the data, two conditions must be met. First, there must have been at least 30 new-home mortgages in that county-year in our data. This restriction will exclude counties that have particularly strict land use regulations or have little

percent.

[7] We consider as 'new' those houses with age equal to one because the date variable for construction completion in DQ contains only the year. In this way, the age of a house that was completed at the end of year $T$ and sold in the middle of year $T+1$ would be recorded as one. Notably, only a third of the new houses in our dataset have age equal to one.

[8] The initial version of the OCC Mortgage Metrics database reported loan performance for OCC-supervised mortgage servicers starting in 2008. In 2016, historical data from these servicers was added to the Mortgage Metrics files to create an 'extended' version of the mortgage data that includes performance and origination information for loans from before 2008.

[9] Because servicers in the OCCMM data service loans that were included in private-label mortgage-backed securities, there is some overlap in the coverage of the OCCMM and PLS databases. When a loan was found in both databases, we retained only one of the loan records to avoid duplication. More detailed information on the construction of this database can be found in Appendix A of Li and Mayock (2017).

developable land, thus limiting the potential for such counties to have a disproportionate impact on our results. Also, given the recent empirical evidence that strict supply constraints affect house price volatility (e.g., Glaeser, Gyourko and Saiz (2008) and Paciorek (2013)), limiting the data to counties with a non-trivial amount of new construction will help avoid any bias associated with including counties with severe housing supply issues. Second, new home sales had to comprise less than 90 percent of the loans in a county in a given year for those observations to be included in the data.[10]

We present in Table 1 the means of key borrower and loan characteristics for the overall sample as well as by default status and new home status. 18 percent of all loans in our sample were at least 90 days past due or entered the foreclosure process within 48 months of origination. The default rate for loans backed by new construction (23 percent) was significantly higher than the default rate for mortgages on existing homes (17 percent). Relative to borrowers purchasing existing homes, new home buyers in our sample were more likely to have loans with interest only periods and were less likely to have FHA loans or mortgages underwritten based on full documentation. These differences aside, the characteristics of the mortgages used to purchase new and existing construction were quite similar.

As expected, high-risk characteristics were overrepresented in the population of loans that experienced default. For example, defaulted loans had much lower FICO scores on average and were significantly more likely to have high-risk contract characteristics such as interest-only and option ARM periods, balloon payments, piggyback loans, prepayment penalties, terms in excess of 30 years, and zero down payments. Importantly for our analysis, mortgages for new construction accounted for 20 percent of all defaults in our sample but only 15 percent of the sample of mortgages that did not default.

[10] The motivation for this filter was identifying counties where the age of the property might not be reported correctly. This filter removed less than 1 percent of the county-year observations in our data.

Table 1: Average Loan Characteristics by Default and New Home Status

| | Loan Types | | | | |
|---|---|---|---|---|---|
| | Default[1] | No Default | New Construction | Existing Home | Total |
| Observations | 613,382 | 2,816,447 | 547,593 | 2,882,236 | 3,429,829 |
| Default Rate | 1.00 | 0.00 | 0.23 | 0.17 | 0.18 |
| New Construction[2] | 0.20 | 0.15 | 1.00 | 0.00 | 0.16 |
| Balloon Payment | 0.07 | 0.02 | 0.03 | 0.03 | 0.03 |
| Full Documentation | 0.50 | 0.61 | 0.52 | 0.60 | 0.59 |
| Interest Only | 0.30 | 0.15 | 0.22 | 0.17 | 0.18 |
| Option ARM | 0.07 | 0.04 | 0.05 | 0.04 | 0.05 |
| Owner Occupied | 0.88 | 0.85 | 0.87 | 0.85 | 0.86 |
| Piggyback Loan | 0.45 | 0.26 | 0.33 | 0.29 | 0.29 |
| Prepayment Penalty | 0.33 | 0.09 | 0.14 | 0.13 | 0.13 |
| FHA | 0.16 | 0.16 | 0.13 | 0.17 | 0.16 |
| VA | 0.02 | 0.04 | 0.05 | 0.04 | 0.04 |
| Term>30 Years | 0.12 | 0.02 | 0.04 | 0.03 | 0.04 |
| CLTV[3] | 94.30 | 86.38 | 87.58 | 87.84 | 87.80 |
| CLTV> 99[4] | 0.46 | 0.21 | 0.27 | 0.25 | 0.26 |
| FICO | 668.47 | 718.20 | 707.94 | 709.57 | 709.31 |
| Back-end DTI[5] | 40.05 | 36.51 | 37.89 | 36.97 | 37.12 |

[1] A loan is classified as defaulting if the loan becomes at least 90 days past due or enters the foreclosure process within 48 months of origination.
[2] A home is classified as new if it was zero or one year of age at the time of origination.
[3] CLTV is the total debt, including "piggyback" loans, secured by the property divided by the property value at the time of origination.
[4] CLTV>99 is equal to one if CLTV was greater than 99 at the time of origination.
[5] Back-end DTI is the sum of all required monthly debt payments,such as mortgage debt, automotive debt, and student loans - divided by monthly gross income.

# 3 Methodology and Results

We test the hypothesis that mortgages for new construction are more likely to default by estimating a series of progressively richer linear probability models in which the outcome variable is a mortgage default indicator and the key independent variable is the new construction indicator.[11] In the first of these models, we simply regress an indicator for mortgage default on an indicator for whether the home was new at the time that the mortgage was originated. In our second specification we regress the default indicator on origination year fixed effects. Our third specification includes a rich set of borrower and loan descriptors. In our fourth and fifth specifications we regress the default indicator on the same set of controls as in the third specification but include county-origination-year and Census-tract-origination-year fixed effects, respectively. In all of our specifications, standard errors are clustered at the county level.

More formally, the preferred models in which the specification includes geography-by-year fixed effects can be expressed as

$$Delinq_{ijt} = \alpha_{0jt} + \alpha_1 New_i + \boldsymbol{\kappa}'\mathbf{X}_i + \epsilon_{ijt} \tag{1}$$

where $Delinq_{ijt}$ is a dummy variable indicating whether the loan was ever least 90 days

[11] Our strategy for identifying the impact of new home status on mortgage performance requires that we condition on all factors that affect the likelihood that a loan defaults that could potentially be correlated with the new home indicator variable. Controlling for the post-origination path in housing values – a key determinant of mortgage default – is thus of critical importance for our identification strategy. Because neighborhood-level house price dynamics are notoriously hard to quantify, we opted to use fixed effects to control for post-origination house price dynamics in lieu of trying to quantify these dynamics directly. Given the large number of fixed effects that must be included in our model, we chose to utilize linear probability models in lieu of non-linear models such as Logit or Probit to avoid the incidental parameters problem (Greene, 2004). To investigate the sensitivity of our results to alternative modeling approaches, in an unreported robustness check we estimated Logit models that were identical to the specifications reported in Table 3 and Table 4 with one exception: in lieu of including Census tract-year fixed effects, the Logit models included state-origination-year dummy variables. The average marginal effects from this analysis implied that mortgages for new construction were roughly 3 percentage points more likely to default compared with loans to purchase existing homes.

past due (90+DPD) or worse or entered the foreclosure process within the first 48 months after origination, $i$ indexes the purchase loans in our sample, $j$ indexes the geography (county or Census tract) used to define the fixed effects and $t$ denotes the year of origination. The indicator variable $New_i$ denotes a new home sale, namely sales where the age of the house is either zero or one. $\alpha_{0jt}$ denotes a geography-by-origination-year fixed effect. We alternatively define our fixed effects at the county-origination-year and Census-tract-origination-year level. The vector $\mathbf{X}_i$ contains a large number of borrower and loan characteristics that previous studies have found to drive mortgage default risk, such as indicators for credit score ranges and CLTV ranges, a borrower's back-end debt-to-income ratio, the loan's amortization term, and indicator variables for exotic loan characteristics. Table 2 presents a complete list of these control variables and their respective definitions.

The primary complication associated with identifying the impact of new home status on mortgage performance is that new construction is not randomly located within a local housing market and households that choose to purchase new construction may differ from households that purchase existing homes along unobservable dimensions that also impact mortgage performance. If new homes are primarily located in neighborhoods that are more susceptible to downturns in the housing market, for example, then a simple regression of mortgage performance on new home status will reflect not just the impact of new home status on default but also the systematic concentration of new construction in riskier neighborhoods. Likewise if the buyers of new construction are more financially secure than buyers of existing homes along unobservable dimensions such as non-housing wealth, then the estimated new home effect in a simple regression will be biased downwards because of the positive correlation between new home status and wealth.

Our fixed effects models allow for the new home indicator variable to be correlated with unobservable factors – such as the post-origination path of housing values and other economic shocks – that are common to all loans originated in a given geography in a

Table 2: Definitions of Variables Used in Estimation

| **Variable** | **Definition** |
|---|---|
| Delinquent (0/1) | 1 if the loan became 90 days due or entered the foreclosure process in the four years following origination, 0 otherwise. |
| New (0/1) | 1 if loan is for a home whose age at the same of the sale is either zero or one, 0 otherwise. |
| CLTV 71-80 (0/1) | 1 if the CLTV ratio is between 70.01% and 80.0%, 0 otherwise. |
| CLTV 81-90 (0/1) | 1 if the CLTV ratio is between 80.01% and 90.0%, 0 otherwise. |
| CLTV 91-99 (0/1) | 1 if the CLTV ratio is between 90.01% and 99.0%, 0 otherwise. |
| CLTV 99plus (0/1) | 1 if the CLTV ratio is equal to or more than 99.01%, 0 otherwise. |
| FICO 620-659 (0/1) | 1 if the FICO score is between 620 and 659, 0 otherwise. |
| FICO 660-719 (0/1) | 1 if the FICO score is between 660 and 719, 0 otherwise. |
| FICO 720-769 (0/1) | 1 if the FICO score is between 720 and 769, 0 otherwise. |
| FICO 770plus (0/1) | 1 if the FICO score is equal to or more than 770, 0 otherwise. |
| DTI | Back-end DTI ratio (in percentage form). |
| Missing DTI (0/1) | 1 if the back-end DTI ratio is missing, 0 otherwise. |
| FHA (0/1) | 1 if the loan is insured by the Federal Housing Administration (FHA), 0 otherwise. |
| VA (0/1) | 1 if the loan is guaranteed by the Veterans Administration (VA), 0 otherwise. |
| Long term (0/1) | 1 if the loan's amortization period is more than 30 years, 0 otherwise. |
| Piggyback (0/1) | 1 if the loan is part of a piggyback loan, 0 otherwise. |
| Full doc (0/1) | 1 if there is full documentation for income, 0 otherwise. |
| Interest only (0/1) | 1 if loan has an interest-only period, 0 otherwise. |
| Fixed rate (0/1) | 1 if loan has a fixed rate, 0 if it has an adjustable rate. |
| Owner occupied (0/1) | 1 if property is not owner-occupied as a principal dwelling, 0 otherwise. |
| Balloon (0/1) | 1 if the loan does not fully amortize over its term, 0 otherwise. |
| Option ARM (0/1) | 1 if the loan is 'option ARM', 0 otherwise. |
| Prepay penalty (0/1) | 1 if the loan has a prepayment penalty, 0 otherwise. |

Indicator variables are denoted with '(0/1)'. The abbreviations CLTV and DTI stand for Combined Loan-to-Value and Debt-to-Income, respectively. We do not include in the sample any loans with extreme CLTV values, namely below 10 percent or above 110 percent. FICO scores below 620 and CLTV ratios below 70 percent serve as omitted categories in the respective indicators. Finally, because a non-trivial portion of the data (about 37 percent) had missing DTI information, we opted to include the DTI in the specification using a dummy variable adjustment in which we set missing cases to a constant and added in the specification an indicator flagging any missing DTI values.

particular year. If the potential confounders mentioned above are constant for purchase loans in a given origination cohort in a given geography (county or Census tract), then our fixed effects models can successfully identify the independent impact of new home status on mortgage performance. If, however, the new home indicator is correlated with the idiosyncratic factors that impact mortgage performance that vary *within* the cohort-geography combination that we have used to define the fixed effects, then our models will not identify the impact of new home status on mortgage default. We revisit this possibility in Section 4.

We report the results from our primary regression models in Table 3 and Table 4. Across all 5 specifications, the coefficient on the new construction indicator is positive and statistically significant at the 1-percent level. Our findings thus provide strong support for the hypothesis that mortgages backed by new construction are more likely to default. As expected, the estimated magnitude of the new construction parameter declines as the specifications become progressively richer. In the simplest model (Column 1) – which is effectively just a difference in means – we find that the purchase mortgages for new construction were 5.55 percentage points more likely to default than mortgages used to purchase existing homes. This estimate is roughly halved to 2.56 percentage points when the full set of controls and county-year fixed effects are added to the model (Column 4). As discussed above, there was little difference in the control set between borrowers purchasing new and existing homes. That said, the reduction in the estimated size of the new construction parameter between Column 1 and Column 4 is evidence that a significant fraction of the higher default rate for new construction mortgages presented in Table 1 can be explained by when and where new construction occurs.

While our specification with county-year fixed effects is more robust than the models that do not control for a property's geography or the timing of loan origination, this model is still susceptible to bias if intra-county-year unobservables are correlated with new home status. For instance, if new construction was concentrated in neighborhoods that experienced negative price or employment shocks that were systematically worse than the shocks experienced by the county in general, then our models will overstate the impact of new construction on default. To investigate the possibility of such bias, we re-estimated the default models with the full set of controls but, in lieu of county-year fixed effects, we included Census-tract-year fixed effects. In this specification, we are, in essence, identifying the impact of new construction status on default by comparing loans that were made to observationally identical borrowers with observationally identical mortgage contracts in the same neighborhood in the same year.

We report the results for this specification in Column 5 of Table 3 and Table 4. While the estimated new construction coefficient does shrink relative to that associated with the county-year fixed effects model, it remains statistically significant at the 1-percent level. The magnitude of the new construction effect also remains economically significant. The default rate for existing homes in our sample is 17 percent, and the results in Column 5 imply that, all else equal, mortgages for new construction are 1.75 percentage points more likely to go into default. Our findings thus suggest that relative to the baseline for existing homes, default risk is more than 10 percent higher for loans used to purchase new construction. As the change in the coefficient magnitude between Column 4 and Column 5 is consistent with the new home indicator being correlated with unobservable risk factors that vary within the county, for the remainder of our analysis we will focus on models that include tract-year fixed effects.

In the specifications with control variables, the estimated coefficients on the controls were consistent with expectations and previous research. For example, we find that regardless of how we model unobserved heterogeneity, default risk increases monotonically with CLTV, declines monotonically with FICO score, and increases with a borrower's back-end DTI. We also find that non-standard mortgage features such as periods that allow for payments that do not pay down principal, prepayment penalties, and terms in excess of 30 years all elevate default risk.

In the early years of our sample, credit was expanding, housing values were increasing rapidly, and many parts of the U.S. were in the midst of a residential construction boom. By the end of our sample, mortgage credit had contracted significantly, housing values were falling throughout the country, and residential building activity had ground to a halt. To get a sense of whether the new home effects that we reported in Table 3 were solely driven by mortgage repayment behavior in the boom or bust periods, we re-estimated our empirical models with the sample restricted to each origination year in our sample (2004 through 2009). As the results on the pooled sample suggested that it is important to control for intra-county heterogeneity, all of these models include Census tract fixed effects as well as the full set of controls. We report the results of this exercise in Table 5. To conserve space, we do not report the estimated coefficients for the control variables.

The coefficient on the new construction term is positive and statistically significant in 5 of the 6 years in our sample, with 2005 the lone exception. The estimated magnitude of the new home effect varied significantly over our sample period, ranging from a low of -0.14 percent in 2005 to 3.49 percentage points in 2008. Interestingly, the intertemporal variation in the magnitude of the new construction parameter did not rise and fall with the overall default rate for the loan cohort used to estimate the model, which is reported in the penultimate row of the table. While the default rate peaked at more than 30 percent for the 2006 origination cohort, the estimated new construction parameter for that subsample of the data was 2.14 percentage points. The new home effect was largest for the 2008 cohort, while the default rate for 2008 originations was less than 14 percent. The fact that the estimated parameter on the new construction term is statistically and economically significant throughout our sample is evidence that the impact of new home status on loan performance is not a phenomenon limited to boom or bust periods in the

Table 3: Mortgage Default Regressions: Purchase Mortgages 2004-2009

| | Dependent Variable: Loan Defaults in 48 Months Following Origination[1] *Specification* | | | | |
|---|---|---|---|---|---|
| Coefficient | (1) | (2) | (3) | (4) | (5) |
| New Construction[2] | 0.0555*** | 0.0386*** | 0.0424*** | 0.0256*** | 0.0175*** |
| | (0.00786) | (0.00702) | (0.00543) | (0.00202) | (0.00180) |
| Term>30 Years | | | 0.305*** | 0.205*** | 0.186*** |
| | | | (0.00660) | (0.00456) | (0.00519) |
| Piggyback Loan | | | -0.0137*** | -0.0138*** | -0.00737*** |
| | | | (0.00329) | (0.00159) | (0.00153) |
| Prepayment Penalty | | | 0.129*** | 0.109*** | 0.0988*** |
| | | | (0.00498) | (0.00376) | (0.00344) |
| Option ARM | | | 0.0555*** | 0.0181*** | 0.0197*** |
| | | | (0.00629) | (0.00450) | (0.00448) |
| Interest Only | | | 0.108*** | 0.0516*** | 0.0542*** |
| | | | (0.00779) | (0.00329) | (0.00298) |
| Full Documentation | | | -0.0614*** | -0.0447*** | -0.0419*** |
| | | | (0.00376) | (0.00187) | (0.00181) |
| Fixed Rate | | | 0.0102** | -0.0169*** | -0.0173*** |
| | | | (0.00449) | (0.00224) | (0.00191) |
| Owner Occupied | | | -0.0260*** | -0.0219*** | -0.0115*** |
| | | | (0.00325) | (0.00283) | (0.00201) |
| Balloon Payment | | | 0.0297*** | 0.0127*** | 0.00859*** |
| | | | (0.00486) | (0.00273) | (0.00280) |
| | | | (0.00460) | (0.00446) | (0.00389) |
| FHA | | | -0.0319*** | -0.0319*** | -0.0283*** |
| | | | (0.00396) | (0.00314) | (0.00255) |
| VA | | | -0.144*** | -0.132*** | -0.119*** |
| | | | (0.00709) | (0.00534) | (0.00451) |
| *Back-end DTI (DTI)* | | | 0.00183*** | 0.000902*** | 0.000841*** |
| | | | (8.18e-05) | (4.62e-05) | (4.90e-05) |
| *DTI Missing* | | | 0.0133*** | 0.00929** | 0.00672 |
| | | | (0.00428) | (0.00400) | (0.00423) |
| Model Includes Year Fixed Effects? | No | Yes | No | No | No |
| Model Includes County-Year Fixed Effects? | No | No | No | Yes | No |
| Model Includes Tract-Year Fixed Effects? | No | No | No | No | Yes |
| Observations | 3,429,829 | 3,429,829 | 3,429,829 | 3,429,829 | 3,429,829 |

[1] A loan is classified as defaulting if the loan becomes at least 90 days past due or enters the foreclosure process within 48 months of origination.
[2] A home is classified as new if it was zero or one year of age at the time of origination.
Standard errors clustered at the county level are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 4: Mortgage Default Regressions: Purchase Mortgages 2004-2009

| | Dependent Variable: Loan Defaults in 48 Months Following Origination[1] *Specification* | | | | |
|---|---|---|---|---|---|
| Coefficient | (1) | (2) | (3) | (4) | (5) |
| *CLTV Buckets* | | | | | |
| $70 < CLTV \leq 80$ | | | 0.0216*** | 0.0290*** | 0.0263*** |
| | | | (0.00199) | (0.00252) | (0.00255) |
| $80 < CLTV \leq 90$ | | | 0.0809*** | 0.0869*** | 0.0765*** |
| | | | (0.00459) | (0.00507) | (0.00504) |
| $90 < CLTV \leq 99$ | | | 0.109*** | 0.123*** | 0.110*** |
| | | | (0.00579) | (0.00635) | (0.00588) |
| $CLTV > 99$ | | | 0.191*** | 0.198*** | 0.175*** |
| | | | (0.00955) | (0.00927) | (0.00796) |
| *FICO Buckets* | | | | | |
| $620 < FICO \leq 659$ | | | -0.0988*** | -0.106*** | -0.100*** |
| | | | (0.00502) | (0.00467) | (0.00430) |
| $659 < FICO \leq 719$ | | | -0.187*** | -0.193*** | -0.183*** |
| | | | (0.00647) | (0.00626) | (0.00576) |
| $719 < FICO \leq 769$ | | | -0.251*** | -0.253*** | -0.238*** |
| | | | (0.00564) | (0.00567) | (0.00510) |
| $FICO > 769$ | | | -0.274*** | -0.279*** | -0.260*** |
| | | | (0.00460) | (0.00446) | (0.00389) |
| FHA | | | -0.0319*** | -0.0319*** | -0.0283*** |
| | | | (0.00396) | (0.00314) | (0.00255) |
| VA | | | -0.144*** | -0.132*** | -0.119*** |
| | | | (0.00709) | (0.00534) | (0.00451) |
| *Back-end DTI (DTI)* | | | 0.00183*** | 0.000902*** | 0.000841*** |
| | | | (8.18e-05) | (4.62e-05) | (4.90e-05) |
| *DTI Missing* | | | 0.0133*** | 0.00929** | 0.00672 |
| | | | (0.00428) | (0.00400) | (0.00423) |
| Model Includes Year Fixed Effects? | No | Yes | No | No | No |
| Model Includes County-Year Fixed Effects? | No | No | No | Yes | No |
| Model Includes Tract-Year Fixed Effects? | No | No | No | No | Yes |
| Observations | 3,429,829 | 3,429,829 | 3,429,829 | 3,429,829 | 3,429,829 |

[1] A loan is classified as defaulting if the loan becomes at least 90 days past due or enters the foreclosure process within 48 months of origination.
[2] A home is classified as new if it was zero or one year of age at the time of origination.
Standard errors clustered at the county level are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

housing and credit markets.

Table 5: Mortgage Default Regressions by Origination Year: Purchase Mortgages 2004-2009

| | Dependent Variable: Loan Defaults in 48 Months Following Origination[1] *Origination Year* | | | | | |
|---|---|---|---|---|---|---|
| Coefficient | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 |
| New Construction[2] | 0.00427** | -0.00144 | 0.0214*** | 0.0316*** | 0.0349*** | 0.0182*** |
| | (0.00168) | (0.00298) | (0.00300) | (0.00270) | (0.00321) | (0.00275) |
| Model Includes Tract-Year Fixed Effects? | Yes | Yes | Yes | Yes | Yes | Yes |
| Model Includes Control Variables? | Yes | Yes | Yes | Yes | Yes | Yes |
| Default Rate | 0.0672 | 0.1710 | 0.3020 | 0.2430 | 0.1370 | 0.0687 |
| Observations | 472,864 | 776,257 | 726,067 | 558,590 | 465,302 | 430,749 |

[1] A loan is classified as defaulting if the loan becomes at least 90 days past due or enters the foreclosure process within 48 months of origination.
[2] A home is classified as new if it was zero or one year of age at the time of origination.
Standard errors clustered at the county level are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

# 4 Robustness Checks

While our fixed effects models allow for arbitrary correlation between new home status and unobservables that are constant within the cohort-geography combination that we use to define our fixed effects, if unobservables that vary within the fixed effect level are correlated with the new home indicator, then our fixed effects models are biased. To investigate this possibility, we estimated a sequence of instrumental variable models in which new home status was treated as an endogenous variable. The instrument that we use to estimate these models is of the "shift-share" variety popularized by Bartik (1991). We constructed this instrument as follows. First, we used home sale records to count the number of new and existing home sales that occurred within a given Census tract – indexed by $j$ – in a given year ($t$). Next, we constructed similar counts of new and existing home sales at the state ($s$) level. On a tract-by-tract basis, we then subtracted the tract-level new and existing sales from the state-level new and existing sales; the remaining state-level sales were then used to create a variable ($PerNew_{sjt}$) that measures the fraction of new home sales relative to total sales in state $s$ in year $t$ *net of the sales in Census tract $j$ in year $t$.*

As is common in the literature, we define a base year for the shift-share instrument that predates the data used to estimate the model. Let $PerNew_j^{Base}$ denote the fraction of sales of new homes in Census tract $j$ in this base period.[12] We then use the change in the $PerNew_{sjt}$ terms to move the fraction of new home sales forward over time. For example, if we let $PerNew_{j,Base+1}^{IV}$ denote the value of our instrument for Census tract $j$ in the year following the base year, the value of the instrumental variable in tract $j$ in year $Base+1$ is constructed as follows

$$PerNew_{j,Base+1}^{IV} = \left(\frac{PerNew_{s,j,Base+1} - PerNew_{s,j,Base}}{PerNew_{s,j,Base}}\right) PerNew_j^{Base}$$

The value for the instrument in the second year after the base period is then defined as

$$PerNew_{j,Base+2}^{IV} = \left(\frac{PerNew_{s,j,Base+2} - PerNew_{s,j,Base+1}}{PerNew_{s,j,Base+1}}\right) PerNew_{j,Base+1}^{IV}$$

and so on.

For this exercise, we estimated 5 different models. In the first model, we simply regress the default indicator on the new construction indicator; this regression gives us the unconditional difference in default rates between mortgages used to purchase new and existing construction. The second model that we estimate is a simple instrumental variables (IV) regression of the default indicator on the new construction indicator where the shift-share variable serves as the instrument. The third and fourth models are IV regressions in which

[12] Ideally, we would have been able to use 2003 as the base year so that the temporal extent of the data used to estimate the instrumental variables (IV) models would align with that of the data used to estimate our fixed effects models. Because the reporting of transactions for many of the counties in our sample did not begin until 2004, however, taking this approach would result in a significant reduction in the geographic scope of our data. That said, we opted to use 2004 as our base year and use the loan originations from 2005 to 2009 to estimate the model.

we include origination-year fixed effects and origination-year fixed effects and all of the control variables used in Section 3, respectively. Our final and arguably most robust IV model includes the full set of controls as well as county-origination-year fixed effects.[13]

The conditions under which the fixed effects instrumental variables (FEIV) model is identified warrants discussion. In the context of our application, the FEIV model allows for the new home indicator to be correlated with loan-specific unobservables that affect mortgage performance. For this model to be identified, however, the instrument ($PerNew^{IV}_{j,t}$) must be strictly exogenous *conditional on the fixed effects*. That is, the FEIV estimator allows for arbitrary correlation between the instrument and the fixed effects (Wooldridge, 2010, p. 354).

The fixed effects in our richest specification are defined using the cross product of counties and origination years in our data.[14] In this specification, the identification assumptions for the FEIV model allow for $PerNew^{IV}_{j,t}$ to be correlated with factors that are common to loans originated in a given county in a given year, but the strict exogeneity condition requires that $PerNew^{IV}_{j,t}$ is orthogonal to all of the error terms within a county-origination-year cohort. Goldsmith-Pinkham, Sorkin and Swift (2018) demonstrate that the use of a shift-share instrument is numerically equivalent to using weighted values of the shares in the base period as instruments. That said, identifying our FEIV model in essence requires that, conditional on the county-year fixed effects, the fraction of new home sales in a Census tract in the base period is uncorrelated with idiosyncratic factors that affect loan performance.

We report the results of our IV models in Table 6. We have suppressed the estimated

[13] Because of the manner in which our instrument is constructed, we cannot include Census tract-year fixed effects as there is no within-tract-year variation in the instrument.

[14] For example, all loans originated in Miami-Dade County, Florida in 2005 would be assigned one fixed effect. Defining the fixed effects at the county-year level instead of simply at the county level allows us to control for unobserved heterogeneity in default risk that varies over time in a given county.

parameters and standard errors on control variables to conserve space. In an unconditional sense, loans backed by new construction in this estimation sample were more than 6 percentage points more likely to default than loans used to purchase existing homes. Consistent with the results we presented in Section 3, the coefficient on the new construction term is statistically significant in all of our IV specifications, and the first-stage F statistics suggest that our shift-share instrument is quite strong. The estimates are also economically significant. For example, in the FEIV model with county-year fixed effects – the specification in which the identification assumption is most likely to hold – we find that loans used to purchase new construction were 4.6 percentage points more likely to default. The magnitude of the estimates from the IV models are all larger than the estimates from the fixed effects models reported in Table 3. As the endogeneity test provides some evidence suggesting that the new construction variable is endogenous, the results of this robustness exercise suggest that the estimated new home effect reported in Table 3 is likely underestimated.[15]

# 5 Conclusion

Relative to an existing home, the purchase of new construction is unique for several different reasons. For example, homes have been shown to exhibit non-linear depreciation schedules, new homes are more difficult to appraise than existing construction, and buyers of new construction are likely at a significant bargaining disadvantage when purchasing homes from builders that have extensive experience in with real estate transactions. All of these factors increase the likelihood that, following origination, buyers of new homes will experience negative equity, a condition that the existing literature has linked strongly

[15] In addition to estimating models in which the new construction variable was treated as an endogenous regressor, we conducted several other robustness checks to see if our findings were driven by specification issues and whether there was significant heterogeneity in the new home effect. We provide a high-level summary of these analyses in Appendix A. The full results are available upon request.

Table 6: Mortgage Default Regressions: Purchase Mortgages 2005-2009

| | Dependent Variable: Loan Defaults in 48 Months Following Origination[1] *Specification* | | | | |
|---|---|---|---|---|---|
| Coefficient | (1) | (2) | (3) | (4) | (5) |
| New Construction[2] | 0.0606*** | 0.1418*** | 0.0812*** | 0.0855*** | 0.0463*** |
| | (0.0090) | (0.0327) | (0.0313) | (0.0221) | (0.0091) |
| New Construction Treated as Endogenous?[3] | No | Yes | Yes | Yes | Yes |
| First-Stage F Statistic | - | 597.62 | 578.14 | 597.52 | 556.75 |
| Endogeneity Test Statistic | - | 7.06*** | 2.02 | 4.94** | 3.83** |
| Model Includes Controls? | No | No | No | Yes | Yes |
| Model Includes Origination Year Fixed Effects? | No | No | Yes | Yes | No |
| Model Includes County-Year Fixed Effects? | No | No | No | No | Yes |
| Observations | 2,941,014 | 2,941,014 | 2,941,014 | 2,941,014 | 2,941,014 |

[1] A loan is classified as defaulting if the loan becomes at least 90 days past due or enters the foreclosure process within 48 months of origination.
[2] A home is classified as new if it was zero or one year of age at the time of origination.
[3] In models where the new construction variable is treatead as endogenous, the instrument used to identify the model is the shift-share variable that is described in the main text. The base year used to construct the instrument is 2004, and the loans that are used to estimate the model are restricted to mortgages originated between 2005 and 2009.
Standard errors clustered at the county level are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

to mortgage default.

In spite of the widespread interest in the determinants of mortgage default, to our knowledge the impact of new home status on mortgage delinquency has gone completely unstudied. This paper attempts to close this gap in the literature. Using a large database of mortgages from across the entire credit spectrum, we provide evidence that mortgages used to purchase new properties are more likely to experience delinquency than mortgages used to purchase existing homes. The differences in mortgage performance that we identify are also economically significant. The results from the richest econometric model in our baseline analysis – which is identified using within-Census-tract-origination-year variation in mortgage performance – imply that all else equal, mortgages collateralized by new construction were roughly 1.8 percentage points more likely to default than loans used to purchase existing homes. This differential represents a 10 percent increase over the baseline default rate of 18 percentage points in our sample. When we estimate these same fixed effects models on a year-by-year basis, we find evidence of the 'new home effect' for loans originated in years defined by rapidly rising housing prices and expanding credit markets as well as for loans originated in the opening years of the Great Recession. Our results thus provide evidence that the new home effect is a key determinant of mortgage performance in times of economic expansion as well as economic contraction.

Our results survive a battery of robustness checks. Perhaps most importantly, even when we allow for new home status to be correlated with unobservable mortgage risk factors, we find strong evidence that all else equal, homes for new construction were more likely to default. In our most robust instrumental variables model, for example, we find that loans used to purchase new homes were 4.6 percentage points more likely to default. As the estimates of the 'new home effect' from the instrumental variables models were significantly larger than those from our fixed effects models, the robustness checks suggest that the estimated new-existing performance differential in our baseline fixed effects models may be understated.

The primary contribution of our work is strong evidence that the uniqueness of new home purchases elevates default risk; to our knowledge the performance differential between loans for existing and new construction has not been explored previously in the mortgage default literature. We also view our findings as contributing to the ongoing debate on the causes of the mortgage crisis, and particularly to an emerging literature that challenges the conventional explanation that lays most of the blame on subprime lending (e.g., Ferreira and Gyourko (2015) and Adelino, Schoar and Severino (2016)). In the context of this debate, our findings provide one potential explanation for elevated default rates among prime borrowers in housing markets that experienced a significant increase in new construction.

# References

Adelino, M., A. Schoar and F. Severino. 2016. "Loan Originations and Defaults in the Mortgage Crisis: The Role of the Middle Class." *The Review of Financial Studies* 29(7):1635–1670.

Agarwal, S., G. Amromin, C. Gartenberg, A. Paulson and S. Villupuram. 2014. "Homebuilders, Affiliated Financing Arms, and the Mortgage Crisis." *Federal Reserve Bank of Chicago Economic Perspectives* 38(2):38–51.

Bartik, T. 1991. "Who Benefits from State and Local Economic Development Policies?" *WE Upjohn Institute for Employment Research* .

Clapp, J. and C. Giaccotto. 1998. "Residential Hedonic Models: A Rational Expectations Approach to Age Effects." *Journal of Urban Economics* 44(3):415–437.

Coulson, N., A. C Morris and H. Neill. 2016. "Are New Homes Special?" *Real Estate Economics* .

Ferreira, F. and J. Gyourko. 2015. "A New Look at the U.S. Foreclosure Crisis: Panel Data Evidence of Prime and Subprime Borrowers from 1997 to 2012." *National Bureau of Economic Research Working Paper* .

Gartenberg, C. 2014. "Do Parents Matter? Effects of Lender Affiliation through the Mortgage Boom and Bust." *Management Science* 60(11):2776–2793.

Gerardi, K., K. Herkenhoff, L. Ohanian and P. Willen. 2017. "Can't Pay or Won't Pay? Unemployment, Negative Equity, and Strategic Default." *The Review of Financial Studies* 31(3):1098–1131.

Glaeser, E., J. Gyourko and A. Saiz. 2008. "Housing Supply and Housing Bubbles." *Journal of Urban Economics* 64:198–217.

Goldsmith-Pinkham, P., I. Sorkin and H. Swift. 2018. "Bartik Instruments: What, When, Why, and How." *National Bureau of Economic Research Working Paper* .

Greene, W. 2004. "The Behaviour of the Maximum Likelihood Estimator of Limited Dependent Variable Models in the Presence of Fixed Effects." *The Econometrics Journal* 7(1):98–119.

Harding, J., S. Rosenthal and C.F. Sirmans. 2003. "Estimating Bargaining Power in the Market for Existing Homes." *Review of Economics and Statistics* 85(1):178–188.

Harding, J., S. Rosenthal and C.F. Sirmans. 2007. "Depreciation of Housing Capital, Maintenance, and House Price Inflation: Estimates from a Repeat Sales Model." *Journal of Urban Economics* 61(2):193–217.

Lacetera, N., D. Pope and J. Sydnor. 2012. "Heuristic Thinking and Limited Attention in the Car Market." *American Economic Review* 102(5):2206–36.

Levitt, S. and C. Syverson. 2008. "Market Distortions when Agents are Better Informed: The Value of Information in Real Estate Transactions." *The Review of Economics and Statistics* 90(4):599–611.

Malpezzi, S., L. Ozanne and T. Thibodeau. 1987. "Microeconomic Estimates of Housing Depreciation." *Land Economics* 63(4):372–385.

Paciorek, A. 2013. "Supply Constraints and Housing Market Dynamics." *Journal of Urban Economics* 77:11–26.

Shilling, J., C.F. Sirmans and J. Dombrow. 1991. "Measuring Depreciation in Single-Family Rental and Owner-Occupied Housing." *Journal of Housing Economics* 1(4):368–383.

Stroebel, J. 2016. "Asymmetric Information about Collateral Values." *The Journal of Finance* 71(3):1071–1112.

Vandell, K. 1995. "How Ruthless is Mortgage Default? A Review and Synthesis of the Evidence." *Journal of Housing Research* 6(2):245–264.

Wooldridge, J. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA.

# A Robustness Check Appendix

In addition to estimating models in which the new construction variable was treated as an endogenous regressor, we conducted several other robustness checks to see if our findings were driven by specification issues and whether there was significant heterogeneity in the new home effect. We provide below a high-level summary of these analyses. The full results are available upon request.

In the first such robustness check, we estimated Logit models that were identical to the specifications reported in Table 3 and Table 4 with one exception: in lieu of including Census tract-year fixed effects, the Logit models included state-origination-year dummy variables. The coefficient on the new construction variable in this model was statistically significant at the 1-percent level, and the average marginal effects from this analysis implied that mortgages for new construction were roughly 3 percentage points more likely to default compared with loans to purchase existing homes.

In the second portion of our robustness analysis, we investigated whether the new home effect varied between homes that were built as a part of a large-scale development and homes that were not part of large developments. This distinction is important since large-scale residential developments could take place in the periphery of cities, in contrast to infill development in the urban core. Also, a large-scale development could be part of a Planned Unit Development (PUD), a popular form of suburban development associated with additional legal and economic considerations and where lenders are obliged to follow stricter underwriting standards, partly as a result of requirements set by the secondary mortgage purchasers.

We conducted this analysis using loans that were originated in Arizona, California, Florida, and Nevada between 2004 and 2009. We used the transaction history data in these states to determine whether a home was a part of a large-scale development, which we defined as a residential subdivision where at least one seller sold 20 or more newly constructed homes.[16] We then re-estimated the linear probability models separately for each of the four states. These econometric models included Census tract-year fixed effects and a full set of controls as well as indicator variables indicating that the mortgage was for a new home that was located in a large-scale development ($NewLD_i$) and an indicator for mortgages backed by new homes that were not in a large-scale development ($NewNonLD_i$). The results of this exercise were mixed. The coefficient estimate for the $NewNonLD_i$ indicator was positive and statistically significant at the 1-percent level in Arizona and Florida, a finding which implies that new homes outside of large developments are more likely to default in these states. The estimated coefficient for the $NewLD_i$ indicator was positive and statistically significant at the 1-percent level in Florida, but negative and statistically significant in California. Notably, when comparing the coefficient estimates for these two key independent variables within each state, we found their difference to be statistically significant at the 5-percent level in Arizona and California. Overall, these results indicate that the positive 'new house' effect is present across both types of housing development and that, in two of the four states, mortgages associated with new homes which are not part of a large-scale residential development have an even higher delinquency rate.

Motivated by the work of Agarwal et al. (2014), Gartenberg (2014), and Stroebel (2016), we also used the data in Arizona, California, Florida, and Nevada to test for whether there were performance differentials between loans used to purchase new construction that were originated by homebuilder-associated lenders (hereafter, *HALs*) and purchase loans for new construction originated by other types of lenders. The models used to study

[16] Constructing the large-development indicator required querying the subdivision fields in the transaction records. Because the subdivision field was not formatted uniformly across markets, these queries required extensive manual review. Because of the labor-intensive nature of this process, we opted to focus on a subset of states that experienced a large number of defaults during the Great Recession.

the impact of HALs on loan performance were very similar to the linear probability measures used elsewhere in our analysis. These models contained a full set of borrower and loan controls, Census-tract-by-origination-year fixed effects, and indicators for loans originated by a subprime lender (regardless of new home status), loans for new construction originated by a subprime lender, loans for new construction originated by HALs, and the new home indicator. The results of this analysis did not provide any evidence that the performance of loans originated by HALs differed from that of loans originated by other lenders.

In our final robustness check, we studied the difference in the performance of mortgages used for 'pre-sales' – which we defined as loans for which the age of the property was negative at the time of the sale – relative to loans used to purchase new and existing construction.[17] From a collateral perspective, the valuation of homes prior to completion is even more complex than that of newly completed homes because the lender's appraisal is performed while the home is under construction or even shortly after the start of the construction (i.e. so-called 'plans and specs' appraisal). As a result, the appraisal follows the cost approach rather than the usual market approach. In other words, rather than using a selection of comparable sales to estimate a home's value, the appraiser considers value estimates from the information included in the plat (e.g., history of subdivision, access to public services), the design aspects reflected in the building plans, the specifications sheet, and the builder's break-down of projected expenses. The appraisal also considers the value from a settlement statement if the prospective buyer has already come to an agreement on the price with the builder. Also, as in the case of new homes, the appraisal value would not reflect monetary incentives offered by the developer to motivate the deal (e.g., closing costs paid by the developer).

[17] For the purposes of this analysis, we removed all loans for which the reported age of the property at the time of origination was less than negative 2 or exceed 110.

Given the aforementioned complications associated with valuing pre-sales, we are interested to see whether mortgages associated with such purchases performed differently from mortgages associated with other new home sales. Our analysis of the performance of loans used to purchase pre-sales was limited to loans that were originated in Clark County, Nevada between 2004 and 2009. We focused this analysis on Clark County for two reasons. First, Nevada's real estate market during the 2000s was an outlier in terms of pre-sales. For instance, using our dataset we find that in 2005 about 7 percent of home sales in Nevada were pre-sales, while nationally such sales corresponded to less than 1 percent of all transactions.[18] Second, due to its terrain, about 70 percent of the state's population is concentrated in Clark County, which contains only 7 percent of the state's total land area. Because of this highly concentrated population, the economic shocks affecting Clark County have an outsized impact on the Nevada economy.

We tested the impact of pre-sale status on loan performance using linear probability models with Census-tract-year fixed effects and the full set of controls that we used in all of our other specifications. The key difference in the specification of the pre-sales models is that the model contained dummy variables for homes that were between 0 and 1 years of age at the time of origination and an indicator for homes that had a negative aged ("pre-sales") at the time of origination. We estimated these models on a pooled sample of all originations between 2004 and 2009 as well as on a year-by-year basis.

In the pooled sample, the coefficient on the pre-sale variable was statistically significant at the 1-percent level and implied that pre-sales were more than 5 percentage points more likely to default than loans backed by existing homes or non-pre-sale new construction. The results from this analysis also suggest that the new home effect on mortgage performance in Clark County was driven entirely by the behavior of pre-sales, as the coef-

[18] Notably, this portion of pre-sales in Nevada is consistent with Figure 2 in Coulson, Morris and Neill (2016) which uses sales data from the Clark County Assessor's Office.

ficient on the new construction indicator was not statistically significant. Turning to the year-by-year results, the coefficient on the pre-sale indicator was positive and statistically significant at the 10% level between 2004 and 2006. While the estimated pre-sale effect was not statistically significant in 2007 or 2009, we found that in 2008 pre-sales were significantly less likely to default than existing homes. This 2008 result is consistent with lenders screening more carefully on borrower unobservables during period in which credit was contracting sharply.