Fools or Crooks: Testing for Fraud in the Residential Mortgage Market

Paul Carrillo

Department of Economics

George Washington University

First version: June 2009

This version: December 2009

Abstract

Current explanations for the high rate of default and foreclosure in the U.S. emphasize house price fluctuations and lax lending criteria. An alternative explanation for default and foreclosure, that has generally been neglected, is fraud. One impediment to identifying and measuring fraud is the lack of a statistical test capable of detecting it. This paper proposes and implements the first test for fraud in mortgage lending. In a completely serendipitous turn of events, subsequent to the writing of this paper, indictments for fraud were announced in the area under study. Tests reveal that the proposed test successfully identified the fraudulent activity. The test proposed here is important for at least three reasons. First it can document the role of fraud in the mortgage foreclosure crisis. Second, it can serve as part of a forensic effort designed to detect and deter mortgage fraud. Third, the model developed here demonstrates that mortgage fraud distorts repeat sales house price indexes because it artificially elevates house prices during the period of fraud followed by a subsequent collapse due to the foreclosure sales. Accordingly, fraud can give the false impression that foreclosure lowers area house prices when it actually artificially inflates them. This suggests an alternative interpretation for the recent empirical literature on externalities from foreclosure.

Keywords: Foreclosure, Mortgage Fraud, Home Prices, Hedonic Model.

JEL Codes: D11, D12, G21, R20.

* Acknowledgements: The author is grateful to Anthony Yezer for his comments and encouragement to complete this project. He would also like to thank Bryan Boulier, Andrew Cohen, Ed Coulson, Mike Frantantoni, Donald Parsons, Christina Remhoff, Mark Schroder and Tara Sinclair for useful comments and discussions. The author is solely responsible for any errors.
1) Introduction

Current explanations for the high rate of default and foreclosure in the U.S. emphasize house price fluctuations (even bubbles), and lax lending criteria. Either of these explanations is consistent with the notion that many homebuyers were naive or made foolish choices. Another alternative explanation for default and foreclosure, that has generally been neglected, is fraud. This explanation implies that some homebuyers were knaves rather than fools.

This paper proposes and implements a new test to identify fraudulent transactions in the housing market. Specifically, we first assess if homeowners who defaulted on their mortgage soon after buying a residential property (early-payment-defaults) paid a price premium for their homes. Using housing sales data from an affluent suburban area of a large U.S. city, we provide robust evidence about the existence of such a premium, the early-default premium, which, on average, accounts for at least 1.6 percent of the home sale price. Then, we develop a battery of empirical tests which demonstrate that this premium can be attributed, at least in part, to the prevalence of mortgage fraud. These tests identify transactions that, in our view, are likely to be fraudulent. Soon after the first draft of this paper was finished, a criminal prosecution for mortgage fraud was announced in the area under study. Statistical analysis of the relation between the location of properties involved in the prosecution and the location of fraudulent transactions

---

1 See, for example, Gerardi, Shapiro, and Willen (2007), Danis and Pennington-Cross (2008), Glaeser, Gyourko and Saiz (2008), Haughwout, Peach, and Tracy (2008), Mayer, Pence, and Sherlund (2008), Demyanyk and van Hemert (forthcoming), Gerardi, Lehnert, and Sherlund and Willen (forthcoming).

2 The Federal Bureau of Investigation (FBI) reports on its web site that “losses attributed to mortgage fraud will most likely reach $4.2 billion for 2006. This figure does not take into account another estimated $1.2 billion spent on fraud prevention tools.” Despite these concerns, we are aware of only one study that discusses mortgage fraud (Shroder, 2008).

3 We are not aware of any other test in the literature that attempts to identify fraudulent transactions in the housing market.
identified by the statistical method demonstrates that the method successfully isolates areas and properties involved in the criminal prosecution.

If markets are competitive, it is not clear why homeowners would pay a price premium for their homes. If search and transaction costs are relevant and buyers face a distribution of sellers’ asking prices, however, it is likely that there is a distribution of prices paid by observationally identical buyers. Presumably, those buyers/borrowers who pay higher than average prices will be also more likely to default in the future. To illustrate these points, we build a simple model that depicts the home buyer search process and her future default option.\textsuperscript{4} The model shows that buyers’ reservation values are directly related to their default options in the future and provides a clear conceptual framework to analyze the relationship between average transaction prices and future default rates.

The model provides valuable insights about the early-default premium. Even if borrowers are well-informed, such a premium would arise, because a group of unlucky buyers would pay higher than average prices that perhaps are very close to or even equal to their reservation values. Because these transactions occur at prices higher than the mean, the chances of default in future periods compared to the average borrower are also greater. The premium can also be explained if a group of uninformed foolish borrowers have unrealistic presumptions about the distribution of prices in the future and overestimate home price appreciation rates. Presumably, these buyers would be willing to pay more for their units and, when future prices are revealed to them, default more often. Finally, we depart from our basic model setup and pose an alternative hypothesis that may also explain the relationship between price premiums and loan defaults. If sellers are

\textsuperscript{4} Related literature is discussed in the second section of the paper.
able to make side payments to buyers, it may be optimal for dishonest buyers to agree to pay a significantly larger price premium for their property and, soon after the transaction, take the default option. Of course, the side payment must be large enough to compensate the forgone down payment and other default costs borrowers may incur. Clearly, this practice is not legal.5

We have posited three potential explanations that are consistent with the early-default premium: a) price dispersion in seller’s asking prices, b) the behavior of foolish buyers, and c) the existence of fraudulent transactions. In the empirical section of this paper, we concentrate our efforts to measure the average premium and to implement a test for the possibility that early payment defaults reflect fraudulent transactions.

To achieve our empirical goals, we use residential real estate sales data from a large suburban county of a metropolitan area in the U.S.6 It is an area with relatively affluent and well-educated individuals where mortgage fraud is not generally expected. Home sales data are drawn primarily from the local Multiple Listing Service (MLS) and supplemented with data from the county assessor’s office and the U.S. Decennial Census. The data include 11,700 completed residential real estate transactions that were listed on the MLS between January and December 2006 and sold during the same period and contain detailed information about transaction prices and a comprehensive set of characteristics of the units and their neighborhoods.

MLS data are used to identify early-payment defaults. The MLS system allows sellers to provide remarks and comments about the characteristics of their units. It is common for home sellers, who have defaulted on their mortgage obligations, to include

---

5 This practice is an example of mortgage fraud as defined by the Federal Bureau of Investigation (http://www.fbi.gov/publications/fraud/mortgage_fraud06.htm).

6 For confidentiality reasons, the name of the county will remain anonymous.
remarks on their listings that disclose this information to interested buyers. For example, some sellers disclose in their MLS listings that their advertised unit is at or shortly before the foreclosure stage. We use this information to spot early defaults. Specifically, we identify an early-payment-default (EPD) if a) a property was sold during 2006 in our sample, b) it is listed again as a for-sale unit on the MLS during 2007 or 2008, and c) the new listing has remarks that contain the words “short-sale,” “pre-foreclosure,” or “foreclosure.”

Results from a standard hedonic model provide robust evidence of a large and positive EPD premium that, on average, is at least 1.6 percent (about $9,000) of the home value. The premium is considerably larger for apartments (3 percent) compared to detached and semi-detached housing units.

We then conduct three separate tests to identify the possible presence of fraudulent transactions. The approach we take here is similar in spirit to other methods in the literature used to detect corruption and fraud (Duggan and Levitt, 2002, and Jacob and Levitt, 2003, for example). In our first test, we check if the EPD premium is associated with time-to-default, the time between the date when the home was purchased and the EPD. Presumably, borrowers involved in fraudulent transactions would have no incentive to make any loan payments and will likely default shortly after the transaction is made. On the other hand, foolish borrowers would not default until they have learned the true distribution of home values. Because this process takes time, it is likely that several loan payments are made before foolish borrowers become delinquent. Because dishonest buyers would also be willing to pay a considerably higher price premium, a negative correlation between the EPD premium and time-to-default may suggest the
existence of some fraudulent transactions. Using several empirical specifications, our hedonic model provides robust evidence of such a negative relation.

The second test computes the correlation between the EPD premium and the loan-to-value ratio (LTV). Presumably, buyers involved in dishonest transactions would like to take a loan with the highest possible LTV. On the other hand, foolish buyers may have different incentives, because the cost of the mortgage generally increases with LTV. Because the benefits from mortgage fraud increase with the transaction price, a positive correlation between the early-default premium and LTV would be consistent with the existence of fraud. Our empirical results provide evidence of, precisely, a positive, large, and statistically significant correlation. For instance, the early-default premium of units with LTV greater than 0.99 is about 4 percent ($20,000) while that of their counterparts with lower LTVs is virtually non-existent.

Our final test searches for regularities in the data to assess if those early-defaults that paid higher than average premiums can be associated with particular actors. Unlucky and foolish buyers should be randomly distributed across all actors and, thus, no systematic differences in the early-default premium across them should be expected. However, if the premium is significantly larger and concentrated among certain groups, this raises the likelihood that fraudulent transactions may have occurred. Results from our empirical model show that EPDs that were listed by particular real estate brokers’ offices sold at a significantly higher price (more than $50,000) than comparable non-early-defaults listed by the same offices. On the other hand, the premium of EPDs listed by many other brokers’ offices is not significantly different than the average.

---

7 Actors of real estate transactions include loan originators, appraisals and real estate agents, for example. Data constraints force us to focus on the latter group.
In our view, the combined empirical findings show that the EPD premium can be explained, at least in part, by the existence of mortgage fraud, and that these models can be used to spot transactions where mortgage fraud may have occurred. For instance, we isolate 42 transactions in our sample that default very early (within one year), the LTV was unusually high (above 99 percent) and the transaction was facilitated by one of the broker offices with higher than average EPD premium (more than $10,000). Indeed, subsequent developments, including indictments for specific instances of fraudulent behavior subsequent to the completion of this research, identify the same areas and even the same properties that were identified in the tests. The methods developed in this paper could have a forensic application by lenders and regulators to the identification and prompt prosecution of fraudulent lending practices.

The rest of the paper is organized as follows. The next section develops the conceptual framework that guides the empirical tests. Section three provides details about the data and variables. The fourth section discusses the results. Finally, we conclude discussing the implications of the early-default premium for studies that attempt to measure the effects of foreclosure on home values.

2) Conceptual framework and empirical model

This paper tests if homeowners who defaulted on their mortgage soon after buying their property had paid a price premium for their homes. In this section, a simple model is developed in search for alternative hypothesis that rationalize this premium. The empirical models at the end of this section test for the possibility that early payment defaults reflect fraudulent transactions.
Model

Many models in the literature describe the default option as a put option of the borrower (Foster and Van Order 1984, 1985, and Epperson, Kau, Kennon and Muller, 1985). When there are no transaction costs and markets are competitive, well-informed borrowers would exercise this option (default) whenever they can increase their wealth; that is, they should default ruthlessly when the value of the home falls below the value of the loan balance. The default decision, however, also depends upon transaction costs (Stanton, 1995), the option to refinance the loan (Kau, Keenan, Muller, and Epperson, 1992, and Deng, Quigley, and Van Order, 2000) and cash-flow constraints (Foote, Geraldi and Willen, 2008). Default transaction costs, including negative stigma effects and changes in credit ratings, decrease the likelihood of default since rational borrowers would exercise this option only when negative equity is substantial. Similarly, even if equity is negative, it may be not optimal to exercise the default option since, by defaulting, the borrower would give up the chance to refinance in the future.8

Most of the previous models implicitly assume that (identical) borrowers should have paid the same purchase price for their units, $P_0$, at some previous time $t=0$. In fact, from a theoretical point of view, it is not immediately clear, particularly in competitive markets, why homeowners who default on their mortgages would pay a price differential for their homes. Since competition in the housing market is hindered by search and transaction costs, however, it is likely that the “law of one price” does not fully apply and that even identical buyers end up paying differentiated prices for their housing units. The

---

8 In a recent paper, Foote, Geraldi and Willen (2008) find that negative equity is a necessary but not a sufficient condition for delinquency (foreclosure). Besides negative equity, borrowers should face cash-flow constraints that make monthly mortgage payments unaffordable.
price that home buyers end up paying for their homes depends upon their optimal search strategies. For instance, when potential buyers search for housing units, presumably, they choose their reservation values (maximum price they are willing to pay) based upon the current distribution of sellers’ asking prices, their search costs and, also, upon the expected benefits from the default option. Of course, optimal default rules are also affected by the transaction price $P_0$, since it is a basic determinant of equity levels. We illustrate these points using a simple search model that depicts the home buyer search process and her future default option.

Let us first define $V_t(P_0)$ as the borrower’s value of the option to default in any period $t$. Here, $P_0$ represents the amount the home buyer paid for her home at time $t=0$. To simplify matters and without loss of generality we assume that interest rates are constant, the loan amount $M$ is a fixed proportion of the initial home value and it does not change over time. Specifically, we let $M = \alpha P_0$ where, $0 < \alpha < 1$ is the initial loan-to-value ratio, and assume that the periodic loan payments $m = r \alpha P_0$ only cover the corresponding interest rate $r$. From the point of view of the borrowers, current and future home values $P_t$ are random and their distribution $F_t$ can change over time. Interest rates and loan payments are deterministic.

Borrowers obtain utility from their accumulated equity $P_t - M$ but dislike making periodic payments; that is, we assume that the utility from honoring the periodic payments and “keeping” the home $u(P_t - M, m)$ is increasing in equity but decreasing in $m$. If a borrower chooses to default, she gives up $b$, the lifetime benefits she receives from

---

9 For simplicity and tractability, we also ignore the buyer’s option to refinance. Notice that, given the focus on EPD, refinancing is irrelevant.
10 For the EPD period, the assumption of no amortization is approximately correct even for a self-amortizing loan.
consuming the asset (homeownership) and pays other default costs $D$ that may include negative stigma effects and credit ratings. The value of the option to default in any given period can then be defined by

\[
V_t = \max\{u(P_t - M, m) + \beta E_t[V_{t+1}], -(D + b)\},
\]

where $E_t[V_{t+1}]$ is the borrower’s expected value of having the option to default next period and $\beta$ is the discount factor. Given the current home value $P_0$, rational borrowers would default when the costs from defaulting are smaller than the utility from keeping this option; that is, when $u(P_t - \alpha P_0, r \alpha P_0) + \beta E_t[V_{t+1}(P_0)] < -(D+b)$.

So far, we have assumed that (identical) borrowers pay the same price $P_0$ for their units. Since there are important search and transaction costs in the housing market, however, it is likely that even identical borrowers may pay different prices. A simple sequential search model is used to illustrate this point.\(^\text{11}\) For simplicity, we assume that all units are identical and that sellers’ asking prices are take-it-or-leave-it offers to potential buyers. Define $F_0$ as the distribution of seller’s asking prices. Sellers’ asking prices are heterogeneous due to differences in seller’s motivation to trade or idiosyncratic preferences.\(^\text{12}\) Buyers visit sellers sequentially, obtain a realization of an asking price $P_0$ and decide whether to buy the unit or continue the search process. If trade occurs, buyers pay a fraction of the price $(1-\alpha)P_0$ as a down payment and immediately obtain the

---

\(^{11}\) Search models have been widely used in the literature to analyze the housing market. Applications include Yinger (1981), Yavas (1992), Horowitz (1992), Haurin (1998), Arnold (1999) and Albrecht et al. (2007), among many others.

\(^{12}\) Glower, Haurin and Hendershott (1988) show that differences in seller’s motivation to trade influence their reservation values and, ultimately, transaction prices. Albrecht et al. (2007) assume that home buyers’ and sellers’ motivation to trade is heterogeneous (some are “relaxed” while others are “desperate”); this assumption is crucial to generate price dispersion in equilibrium.
benefits from homeownership $b$ in addition to the expected benefits from the future
default option $E_0 [V_1(P_0)]$. If trade does not occur, buyers pay a transaction cost $c$ and
continue the search process.\footnote{Notice that it is assumed that all the search process is fully completed during the initial period.} With these considerations, the buyer’s value of having an
opportunity to search, $W$, can be defined by

$$W = \max \left\{ b - (1 - \alpha)P_0 + E_0 [V_1(P_0)] , E[W] \right\} - c .$$

Assuming that the left hand side of the maximum operator in equation (2) is decreasing in
$P_0$, optimal buyer behavior should be characterized by a reservation strategy: if a seller’s
asking price $P_0$ is below a threshold $P_0^*$, trade should occur and vice versa. This is the
standard setup of a home buyer’s search problem except that we have included the
expected default option as part of the value borrowers get when purchasing a home. Thus,
buyer’s reservation values and actual transaction prices $P_0$ should also depend upon the
value and features of the default option. For example, changes in default costs $D$ and/or
down payment requirements could affect both default probabilities and transaction prices.

Rather than characterizing the solution of this general model, we illustrate the
case of a two period model. In period 0, buyers buy the home and, in period 1, they
choose between paying the loan and default. Assuming that the loan is repaid in full at
the end of period one and that the utility function is linear, the expected value of the
default option is defined by

$$E_0 [V_1(P_0)] = \int_0^\infty \max \left\{ P_1 - (1 + r)\alpha P_0 - (D + b) \right\} dF_i(P_1) ,$$
where, $F_1$ is the distribution of home prices in period 1. Given the transaction price $P_0$, a rational borrower would only choose to default if the realization of the home value in period 1 does not exceed $(1+r)\alpha P_0 - D - b$.

Now, let us turn our attention to the buyer’s search problem defined in equation (2). After some algebra and manipulation, we show in the appendix that the buyer’s optimal reservation price $P_0^*$ is defined by the unique solution to

$$
\mathbb{E}\left[ F_0(P_0)[1 - \alpha + \alpha r F_1(\alpha P_0 + r \alpha P_0 - D - b)]dP_0 \right] = c.
$$

This equation states that the optimal buyer’s reservation value must be such that the expected benefits from searching are equal to its costs. More importantly, it shows that buyers’ reservation values are directly related to borrower’s default options in the future and provides a clear instrument to analyze the relationship between average transaction prices $E[ P_0 | P_0 < P_0^* ]$ and default rates.

**Price premium and default rates**

Are home buyers who paid above average prices for their homes more likely to default on their loans? Our simple model suggests this is the case. In our model, rational well-informed borrowers engage in trade as long the asking price is below $P_0^*$. This means that some unlucky buyers would still pay higher than average prices that perhaps are very close to or even equal to their reservation values. Since these transaction prices are higher than the mean, the chances of default in the next period compared to the average borrower are also greater. The previous arguments suggest that there exists a positive correlation between price premiums and the likelihood of default in the future.
So far, we have assumed that borrowers are well informed; that is, they know the distribution of current and future home prices. We depart from perfect information and assume now that a group of foolish homebuyers rush into the market without taking the time or effort to learn about the distribution of prices. Foolish homebuyers know the distribution of current asking prices $F_0$ but, during period 0, have unrealistic presumptions about the distribution of home values in the future. Let $F^*_1$ be the incorrect distribution of prices in period 1 that the group of uninformed (optimistic) buyers expect and assume that $F^*_1(P)$ stochastically dominates $F_1(P)$ (that is, $F^*_1(P) < F_1(P)$ for all $P$). It is clear from equation (4) that these types of buyers will set higher reservation values and pay, on average, higher prices for their units. In period 1, borrowers realize the true distribution of prices $F_1$ and should be more likely to default given their lower equity levels (recall that they were willing to pay higher prices at $t=0$). Therefore, the behavior of a group of foolish buyers may also explain why borrowers who overpaid for their homes default more often.\textsuperscript{14}

Finally, we depart from our basic model and pose an alternative hypothesis that may also explain the relationship between early-default price premiums and loan defaults. Let us assume that sellers are able to make side payments to buyers who a) are willing to pay a significant price premium for their property and, b) will take the default option in the future. As long as the side payment is large enough to compensate for the forgone down payment and other default costs, buyers may agree to raise their reservation values

\textsuperscript{14} Due to language barriers and lower education levels, certain demographic groups may be less aware of the distribution of home prices and, in general, of the risks and benefits of homeownership. For instance, using housing sales data from two census tracts from the Washington D.C. region, Shroder (2008) finds that Hispanics paid a 5 percent premium for their homes. Since subprime origination was higher among this demographic group and default rates of subprime mortgages are higher than primes, a positive correlation between home prices and loan defaults could be expected.
and engage in trade. This practice is not legal. However, several reports in the press and in nonacademic journals have raised concerns about the prevalence of such fraudulent transactions in the housing market.\textsuperscript{15} Despite these concerns, there is little evidence in the literature about the prevalence of this practice.\textsuperscript{16} A positive early-default premium could also reflect the existence of a few of such fraudulent transactions.

\textit{Empirical model}

To investigate if homeowners who default on their mortgage payment requirements shortly after buying their home, early-payment-defaults, (EPDs) pay more/less for their housing units than owners of comparable units who do not default on their credit obligations, a standard hedonic model is used

\begin{equation}
\ln P_{i}^{t} = \alpha D_{i}^{t+1} + \beta X_{i}^{t} + \delta Z_{i}^{t} + \gamma M_{i}^{t} + u_{i}^{t},
\end{equation}

where, $P_{i}^{t}$ is the sale price of housing unit $i$ in period $t$, vectors $X$, $Z$ and $M$ include variables that describe the unit, the neighborhood, and market conditions, respectively; $\beta$, $\delta$ and $\gamma$ are vectors of parameters, and $u$ is an unobserved error term. The variable $D_{i}^{t+1}$ is an indicator for an EPD, it equals one if the owner of property $i$ defaults on his/her mortgage shortly after the transaction was made (in period $t+1$) and zero otherwise. We

\textsuperscript{15} Specifically, buyers and a sellers, with the complicity of appraisals and loan officers, would agree to trade a housing unit for a higher than average price. Buyers would take high cost mortgages that require small down payments and default soon after the sale. Typically, very few or even no loan payments would be made. The “surplus” from such fraudulent transaction would be shared by the buyer, seller and the other participants, while lenders bear all the costs. Details about mortgage fraud are provided at the FBI’s website (http://www.fbi.gov/publications/fraud/mortgage_fraud06.htm) and at the Mortgage Bankers Association mortgage fraud case reports (these reports are known as MARI’s reports and can be found online at http://www.marisolutions.com/resources-news/reports.asp).

\textsuperscript{16} We are aware of only one study that discusses this issue (Shroder 2008).
focus our attention on scalar \( \alpha \) which measures the conditional mean log price differential between home purchases that result in EPDs in the future and comparable sales during the same time period which did not result in an EPD. Given our discussion in the preceding section, a positive \( \alpha \) is expected.

We have posited three potential explanations that are consistent with such a price premium: a) price dispersion in asking prices, b) the behavior of foolish buyers, and c) the existence of fraudulent transactions. In what follows, we concentrate our efforts to test if the latter hypothesis can be ruled out. To conduct this test, we take three different approaches. First, we test if the early-default price premium (\( \alpha \)) is associated with time-to-default, the time between the date when the home was purchased and the loan default. Presumably, borrowers involved in fraudulent operations have no incentive to make any loan payments and will likely default shortly after the transaction is made. Meanwhile, foolish borrowers would not default until they have learned the true distribution of home values. Because this process takes time, it is likely that foolish buyers, on average, default later than dishonest ones. Presumably, fraudulent buyers are in collusion with sellers and should be actively trying to pay the highest price consistent with current appraisals. Thus, a negative correlation between the early-default premium and time-to-default may suggest the existence of some fraudulent transactions. To conduct this test, we estimate several variations of the following model

\[
\ln P_t = \alpha D_{t+1} + \alpha_2 D_{t+2} + \beta X_t + \delta Z_t + \gamma M_t + u_t,
\]

\(^{17}\) The methods we propose here are similar in spirit to other algorithms used in the literature to detect corruption and fraud (Duggan and Levitt, 2002, and Jacob and Levitt, 2003).
where, $D_{it^{+2}}$ indicates if the owner of property $i$ defaults on the loan 2 (or more) periods after the transaction was made.

The second test computes the correlation between the early-default premium and loan-to-value ratio (LTV). Presumably, buyers involved in fraudulent operations would like to take a loan with the highest possible LTV. Because the cost of the mortgage generally increases with LTV, foolish buyers may have different incentives. Thus, a positive correlation between the early-default premium and LTV would be consistent with the existence of fraud. To test this hypothesis, variations of the following model are estimated

$$\ln P_i = \alpha D_{it} + \delta LTV + \gamma M_{it} + \beta X_{it} + \zeta Z_{it} + u_i.$$ 

Our third and final approach searches for regularities in the data to assess if those EPDs that paid higher than average premiums are associated with particular actors (for example, originators, appraisals or real estate broker’s offices). Unlucky and foolish buyers should be randomly distributed across all actors and, thus, no systematic differences in the early-default premium across these groups should be expected. However, if the early-default premium is significantly larger and concentrated among certain groups, this would raise up the likelihood that mortgage fraud occurred. This method could be particularly useful to lenders and regulators to spot and further investigate actors that could have potentially been involved in fraudulent lending practices. To conduct this test, the following equation is estimated
\[
\ln P_{ij} = \alpha D_{i}^{t+1} + \gamma_j A_j * D_{i}^{t+1} + \eta_j A_j + \beta X_i^t + \delta Z_i^t + \gamma M_i^t + u_{ij}^t .
\]

Here, \( A_j \) equals one if the home sale was facilitated by a particular actor \( j \), and \( \gamma_j \) and \( \eta_j \) are scalar coefficients. We will focus our attention on \( \gamma_j \).

### 3) Data and variables

To estimate the empirical models we use residential real estate sales data from a large suburban county within a metropolitan area of the U.S. For confidentiality reasons, we cannot disclose the name of the county. It is noted, however, that it hosts more than 500,000 residents and more than 200,000 housing units.

Home sales data are drawn primarily from the local Multiple Listing Service (MLS), the primary outlet used by real estate agents to list home sales across the U.S.\(^{18}\) MLS data include detailed information about real estate transactions and units’ characteristics. For example, in addition to the transaction price and loan amounts, MLS data provide information about the home’s number of bedrooms, bathrooms, age and location (address). We gathered such data for all completed residential real estate transactions in the county that were listed on the MLS between January and December 2006 and sold during the same period.

The MLS data are supplemented with information from other sources. While sellers are required to include certain information on the MLS listing, other characteristics, such as lot size, residence size and school assignment, are optional and, thus, not always available on this database. To gather the missing information on housing

---

\(^{18}\) The author thanks the local association of real estate agents for sharing these data with him.
unit characteristics, we match the MLS records with county Assessor’s data. In addition, with the help from the county’s staff, we identified the elementary school assigned to each housing unit in our sample. Finally, data on neighborhood characteristics such as income and age distribution were collected from the U.S. Census Bureau’s 2000 Decennial Census.

To identify EPDs, ideally, we would like to use loan level records to find, among all residences that were sold during 2006 in our sample, those properties whose owners defaulted on their mortgage payments in later years (2007 and 2008). Unfortunately, we do not have access to such information. We take an alternative approach and identify EPDs using the MLS data.

The MLS system allows sellers to provide remarks and comments about the characteristics of their units. Generally, sellers (or their agents) include remarks to highlight the positive features of their homes and attract buyers. It is common for home sellers, who have defaulted on their mortgage obligations, to include remarks on their listings that disclose this information to interested buyers. At this point, it is useful to provide some background about how sales of foreclosed properties are generally handled in the county being studied.

Because foreclosure laws vary from state to state (Capone, 1996, and Pence, 2003) and counties can set specific regulations about how a foreclosed unit can be sold, the exact process by which lenders recover the amount owed on a delinquent loan varies from place to place. The process, however, has many similarities. Once a mortgage

---

19 A comprehensive review of this process is found in Pennington-Cross (2006).
20 Once a mortgage has become delinquent, the foreclosure process begins when lenders file a public default notice. This process ends when a) the borrower repays the default amount during a pre-foreclosure grace period; b) the owner sells the property to a third party during the pre-foreclosure period which allows
has become delinquent, lenders can take ownership of the collateral at virtually any stage of the foreclosure process and both borrowers and lenders have many options. In some cases, lenders allow owners to sell their properties to a third party during a pre-foreclosure grace period. Even if the loan balance exceeds the transaction price, lenders may be willing to accept the transaction to avoid the costs of foreclosure. These types of transactions are generally known as “short-sales.” Sales of real estate property that occur at the pre-foreclosure or foreclosure stage are also commonly known as “pre-foreclosure” and “foreclosure” sales, respectively.

Many borrowers and lenders use the MLS to sell the collateral of delinquent mortgages. More importantly, certain listings include remarks that inform if the listing is a “short,” “pre-foreclosure,” or “foreclosure” sale. We use this information to spot early defaults. Specifically, we identify an EPD if a) a property was sold during 2006 in our sample, b) it is listed again as a for-sale unit on the MLS during 2007 or 2008,21 and c) the new listing has remarks that contain the words “short-sale,” “pre-foreclosure,” or “foreclosure.”22 We are also able to compute time-to-default, the time between the home sale in 2006 and the date when the EPD listing appeared for the first time on the MLS during 2007-2008.23 We recognize that using MLS data to identify early-defaults introduces measurement error and, at the end of the results section, we discuss the limitations of our approach.

---

21 This includes all listings regardless if they were ultimately sold or withdrawn from the market.
22 Careful data mining was needed to identify cases where the remarks indicated that the listing was not a foreclosure or not a short-sale.
23 Similarly, data cleaning was needed to identify the time when listings appeared in 2007 and 2008 for the first time. In particular, a few properties that were withdrawn and later re-entered the market appear in more than one listing during 2007 and 2008.
Our combined database consists of 11,700 residential real estate transaction records that took place during 2006. The data include information about the sales price, the loan amounts, the housing unit, general market conditions, the neighborhood, and early-defaults. The complete list of variables, their descriptions and descriptive statistics, is listed in Table 1. The MLS data include information about the first and second loan amounts which are used to compute loan-to-value ratios. The property characteristics describe the unit’s square footage, acreage, number of bathrooms and bedrooms, age, number of fireplaces and if the unit is a detached residence, a townhome or an apartment. To measure market conditions, we create indicators for each month of the year when the property was sold. We select neighborhood characteristics that could affect the desirability of housing units. For example, since low income neighborhoods may offer fewer amenities, a housing unit located in a high income area presumably has a higher value than an identical counterpart in a lower income neighborhood. For this reason, measures of income and employment status are collected, namely, the median household income and unemployment rate in the census block group (CBG).

The hedonic price of housing services may vary with the racial and ethnic composition of the area. Thus, the fraction of the CBG African-American and Hispanic population are included. To account for the net effect of the unobserved neighborhood amenities that attract people to particular locations and the disamenities of high population density, the CBG population density is included. In addition, we add other demographic characteristics of the neighborhood’s residents that describe their age.

---

24 Excluding missing observations and a few cases with implausible high LTVs, the number of observations with valid LTV records decreases to 11,432.
25 To identify whether the housing unit is a detached home, a townhome, or an apartment, we use information from the field “type” on the MLS listing.
(proportion of the population older than 65) and school attainment (share of high school dropouts) distribution. Finally, we create indicators for each elementary school in the county to capture unobserved differences in school quality between locations.26

The average transaction price was $528,700 with a minimum of $125,000 and a maximum of $1,995,000.27 The mean of buyers’ loan-to-value ratios was close to 78 percent although a few transactions report very high or very low LTVs. A typical home in our sample is about 26 years old, has 1,694 square feet, two full bathrooms, and 0.2 acres of land. There is significant dispersion in the characteristics of the neighborhoods. For example, while there are many areas in our sample with virtually no Blacks or Hispanics living in them, there are several CBGs that are populated by these groups only.28

EPDs in this county were significantly lower than in many other counties in the U.S. For instance, out of the 11,700 properties that were sold in 2006, 214 (1.8 percent) and 513 (4.4 percent) were listed again in the MLS as a “foreclosure” or “short” sale during 2007 and 2008, respectively.29 More importantly, this is not an area where there have been general allegations of fraud.

Table 2 shows the distribution of early-default’s time-to-default and LTV, the variables used in our empirical tests. The median number of days between the home sale in 2006 and early-defaults in 2007 and 2008 was 418 and 694 days. Thus, it is likely that many of the 2007 early-defaults did not make any attempt to repay their loans, while most of the 2008 early-defaults presumably made at least a few mortgage payments. The

26 To maintain the anonymity of the county, the number of elementary schools is not disclosed. We report, however, that this number is above 100.
27 We exclude from the sample the upper 0.5 percentile of home sales (109 observations).
28 Again, to keep the anonymity of the county, we do not report descriptive statistics of the census block groups.
29 Since home sellers are not required to disclose on the MLS listing if they have defaulted on their loans, our estimates of early-defaults are likely underestimated.
median LTV of 2007 early-defaults is, not surprisingly, very close to one. It is surprising, however, that 40 and 30 percent of early-defaults had LTVs no greater than 0.95 and 0.80, respectively.

4) Results

Basic specification

We use the combined dataset to estimate equation (5) and show results from three different specifications in Table 3. In all specifications, the dependent variable is the log of the transaction price. The first column shows estimates of the most basic model where the only covariate added is the early-default ($D_{t+1}$) variable. Its coefficient ($\alpha$) is negative and statistically significant suggesting that the (unconditional) mean sale price of early-defaults was about 11 percent lower than the rest of units. Notice, however, that failure to control for other covariates may introduce important negative omitted variable biases. Mortgage defaults are concentrated in less affluent neighborhoods, where units are generally smaller, and have lower quality. Since units’ size/quality and neighborhood income are positively correlated with home values, omitting these variables from equation (1) produces a spurious negative correlation between home prices and EPDs.

The specification shown in the second column of Table 3 adds property characteristics and market conditions (month fixed effects) but omits any neighborhood control. Adding property characteristics notably raises our coefficient of interest ($\alpha$). However, it is still negative. When covariates that describe the neighborhood as well as elementary school fixed effects are included, the estimate of $\alpha$ becomes positive and is statistically significant. Results from this full specification are shown in the last column
of Table 3 and suggest that, all other things equal, homeowners who bought their homes in 2006 and later defaulted on their mortgage payments (during 2007) paid about 1.6 percent more for their housing units than their counterparts who did not default on their credit obligations.

Despite the large set of covariates included in the model, it is likely that we are still not picking up other unobserved characteristics of the unit and neighborhood that are correlated with both home prices and EPDs. Thus, the estimate of \( \alpha \) (1.6 percent) discussed above is a lower bound of the true coefficient. To illustrate this point, we divide our sample into three groups: detached units, townhomes and apartments. Apartments are likely to be less heterogeneous than detached units or townhomes. Thus, omitted variable biases should be less relevant and the estimated EPD premium should be larger. To test this hypothesis, we estimate \( \alpha \) for each subsample using the same set of covariates as those used in the model with the full set of controls (last column of Table 3) and display results in Table 4. As expected, the price premium associated with early-defaults is much larger for the sample of apartments (3 percent) than for the other two groups.

In sum, results from this basic specification suggest that homeowners who bought their homes in 2006 and defaulted on their mortgage payments within the next year paid an average price premium of at least 1.6 percent.

Fools or crooks?

In this section, we search for empirical evidence to rule out the existence of fraudulent transactions. First, we analyze the relationship between the early-default price premium
discussed above and time-to-default. To explore this relationship, several versions of equation (6) are estimated and results are shown in Table 5. The first column shows our baseline results that correspond to those displayed in the third column of Table 3. To evaluate price differentials of EPDs that occurred after two years from their sale, in the second column, we add an indicator that equals one if the early-default took place in 2008. The coefficient of this variable is very small and statistically not different than zero suggesting that homeowners who bought their homes in 2006 and defaulted on their mortgage in 2008 did not pay any price premium for their homes. It is reassuring that the average price gap of 2007 early-defaults (the estimate of $\alpha$) barely changes and remains positive and statistically significant. Taken together, these results evidence a negative correlation between the early-default price premium and time-to-default.

To provide additional evidence about this relationship, two alternative models are estimated. In the third column of Table 5, we include two binary variables that equal one if the early-default occurred in the first or in the last six months of 2008, respectively. Results confirm the negative relationship between the price-premium and time-to-default. For instance, our estimates suggest that properties whose owners defaulted during 2007 paid an average 1.7 percent price premium; this premium decreases to 1 percent for early-defaults in the first six months of 2008 and virtually disappears for those early-defaults in the last six months of this year. Finally, we estimate a model that includes an indicator for early-defaults (either in 2007 or 2008) and an interaction term between this variable and time-to-default. Results are shown in the fourth column of Table 5. The coefficient of our variable of interest (-0.014) is negative and statistically significant, providing additional evidence about the negative relationship between the EPD premium and time-to-default.
For instance, these estimates suggest that homeowners who default on their mortgages after six, twelve and eighteen months of buying their property paid a price premium of about 2.5, 1.8, and 1 percent, respectively.

To measure the correlation between the early-default premium and LTV, several variations of equation (7) are estimated and results are shown in Table 6. The first column contains our baseline specification. In the second column, an indicator that equals one if LTV is higher than 0.99 is added. Notice that the estimate of the EPD premium remains virtually unchanged.30 We then interact this covariate with the early-default variable. Results are shown on the third column of Table 6 and suggest a clear positive correlation between LTV and the early-default premium. For instance, the premium of units with LTV no less than 0.99 is about 3.8 percent ($20,000) while that of their lower LTV’s counterparts is virtually non-existent. In the fourth and fifth columns we estimate equation (7) using the level of the loan-to-value ratio instead of a binary dummy variable. These alternative specifications also highlight a statistically significant positive correlation between LTV and the EPD premium.

In our final test, we estimate equation (8) to assess if those EPDs that paid higher than average premiums are associated with particular actors. Results are shown in Table 7 and its first column displays our baseline results. To illustrate this method, we first select two real estate agent broker’s offices that represented sellers in our data. For confidentiality reasons, we do not disclose their names and only inform that each of these offices sold more than 50 but fewer than 200 housing units during our sample period.31

30 Notice that the coefficient on LTV is negative and statistically significant suggesting that buyers with high LTV pay on average lower prices (1.2 percent less) for their homes. Genesove and Mayer (1997), on the other hand, find that sellers with high LTVs stay longer on the market and obtain higher prices.

31 Selected broker’s offices had the same number of early-defaults.
To assess if home sales of early-defaults that involved Office “A” sold at a different premium than those involving Office “B”, we interact binary broker office dummy variables with our early-default covariate. Parameter estimates are shown in the second column of Table 7. These results suggest that EPDs that were listed by Office A sold at an average 9.2 percent (0.015 + 0.077) higher price (about $49,000) than comparable non-EPDs. On the other hand, the price premium of EPDs listed by Broker B is not significantly different than the average 1.5 percent premium (about $8,000). As it is shown in the third column of this table, these differences remain even after broker office fixed effects are added.

Brokers A and B have been chosen to illustrate that differences in the premium across actors can be large. In a more general specification, we create binary variables for each of the 18 offices that had 3 or more EPDs in our sample and interact them with the early-default variable. Results are shown in the last column of Table 7. The coefficient on the early-default variable (0.007) suggests that the EPD premium for offices with one or two early-defaults is virtually non-existent. To analyze the premium across brokers with three or more EPDs, we plot in Figure 1 histograms of our 18 coefficients ($\gamma$) of interest. If the premium is randomly distributed across actors, a symmetric plot is expected. A concentration at the right tail of the distribution, however, would suggest that the EPD premium is being generated by a small group of brokers. The evidence shown in Figure 1 supports the latter case. For instance, in a few broker offices, the early-default premium can be as large as 10 percent or $50,000. These differences are surprisingly

---

32 Early-defaults occurred in 132 broker offices. We focus on offices with at least 3 transactions. The maximum number of EPDs in a broker office is 10.
large and puzzling. Why would (even foolish) buyers agree to pay a $50,000 premium for a home when they could choose a comparable broker office and pay none?

Testing for fraud

Our previous results provide robust empirical evidence of a larger early-default premium for a) borrowers who default very early and, presumably, made few or no loan payments, b) borrowers with high LTV, and c) transactions associated with particular broker offices. In our view, this circumstantial evidence points out transactions where mortgage fraud may have occurred. In particular, there are 214 properties in our sample that default very early (within one year). Among these 214 EPDs, we identify 42 sales where the LTV was unusually high (above 99 percent) and the transaction was facilitated by one of the broker offices with higher than average EPD premium (more than $10,000). We believe that the likelihood of finding fraudulent transactions among these 42 home sales is high and that lenders and regulators could focus their audit efforts among them.

Shortly after the first draft of the paper was finished, several individuals were indicted for mortgage fraud in the area under study. Details of these cases cannot be revealed without disclosing the name of the county. The indictment covers mortgage fraud among other offenses of a criminal enterprise.33 Addresses of properties were revealed in the indictment and eight of them are part of our sample.34 The spatial association between these eight properties and the locations of the forty-two transactions

33 According to the indictment, the scheme involved real estate agents, loan officers, and individuals recruited to serve as buyers just as hypothesized in this paper.
34 Some properties named were traded in other periods (either before or after 2006) and would not have been part of the sample.
identified in the test proposed here is very close. The addresses of 3 of the named properties exactly match addresses of one of the 42 transactions, 4 are within 0.25 miles, and the final one is 0.5 miles from the nearest transaction that we identify. The null hypothesis of no spatial association between the 42 sales we identify and the properties involved in the prosecution can be rejected at virtually any significance level. In our view, this evidence demonstrates that the method suggested in this research successfully isolates areas and properties where mortgage fraud may have occurred.

Limitations

Using MLS data to assess the price effects of EPDs has some limitations. Our first concern is related to measurement error. Since home sellers self-report on the MLS if they have defaulted on their loans, our estimates of EPDs may be inaccurate. Furthermore, the actual date when the owner of a home defaults on his/her a mortgage payment is not known. Instead, we observe the date when the listing is initially posted on the MLS and, presumably, the decision to sell and advertise a residence in the MLS occurs several months after the owner’s default decision. It is reassuring that, despite the attenuation biases introduced by measurement error, our coefficients of interest are statistically significant. If loan level data are used to accurately compute the early-default and time-to-default variables, we expect that the EPD premium would rise.

Our second concern is related to sample selection issues. Because the contents of the MLS are self reported (by the real estate agent) and sellers are not required to disclose

---

35 To conduct the tests we estimate the distribution of several statistics under the null hypothesis that the 42 sales identified by our method are a random draw from the sample. The statistics include a) the average minimum distance between the indicted units and the sample of 42 properties, and b) the share of the indicted residences that are located within 0.1, 0.2 and 0.5 miles from one of the 42 selected addresses.
this information, one could worry that unobserved characteristics of a delinquent owner influence both a) the price she paid for her home in 2006 and b) the decision to disclose this information on the real estate listing. This should not be a serious concern, however, to the extent that the sample selection we introduce by using MLS data to identify foreclosures in 2007 and 2008 is uncorrelated with home prices in the past (2006).

Other implications
A large and growing number of studies analyze the relationship between foreclosures and home prices. Generally, these studies conclude that foreclosed units sell at a large price discount (as large as 20 percent) and provide evidence of large negative externalities associated with foreclosures in the neighborhoods where these are concentrated (Immergluck and Smith, 2006, Lin et al. 2009, and Rogers and Winter, forthcoming). Most of these studies use individual home sales data to estimate the price gap between foreclosure sales and sales of otherwise identical housing units. Several statistical concerns, such as omitted variable biases and simultaneity, have been raised and efforts have been made to account for these problems (Dubin, 2008, and Clauretie and Daneshvary, 2009, for example). The early-default premium found in this paper points out a new source of bias.

To assess the effects of foreclosures on home prices, all of the studies cited above use home sales data which, presumably, include records of properties that were sold more than once (repeat-sales). If the second transaction of a repeat-sale in their sample is a foreclosure and the first sale occurred less than two years before the foreclosure, our

---

36 Earlier studies that include a binary variable for foreclosure status in a hedonic model include Shilling, Benjamin and Sirmans (1990), Forgey, Rutherford and Wolverton (1996), Springer (1996), Carroll, Clauretie and Neill (1997).
previous findings suggest that the first sale occurred at an average premium. Even if foreclosure had no effect on prices, estimates from a hedonic model that includes a “foreclosure” covariate would suggest that foreclosures sell at price discount. Furthermore, it is likely that locations selected for fraudulent mortgage activity by criminal enterprises are not chosen at random from the entire housing stock. There may well be factors, not easily observed, that distinguish these neighborhoods that could result in classical omitted variable bias in estimating the neighborhood effects of concentrated foreclosure. Thus, the negative effects of foreclosure on prices may be overestimated.

5) Conclusions

This paper documents the existence of a substantial housing price premium paid by homeowners who quickly default on their mortgages. Specifically, it is found that homeowners who bought their homes during 2006 and defaulted on their loan payments shortly after the sale (within one year) paid an average premium of at least 1.6 percent of the home value. Several competing hypothesis are consistent with such price differential. However, a battery of empirical tests point out to mortgage fraud as the main source of the EPD premium. In our view, the empirical tests can be used as a general method for detecting fraud in housing market sales. Finally, the early-default premium highlights a new source of bias for studies that analyze the effects of foreclosure on home values.

Our results evidenced a substantial EPD premium in a county where mortgage defaults and foreclosures rates have been significantly smaller and there have been less concerns about mortgage fraud compared to other parts of the country. We anticipate, however, that similar patterns will emerge in other areas with higher default and
mortgage fraud rates. More importantly, the methods illustrated in this study could be useful for lenders and regulators to focus their audit efforts among those actors where the likelihood of finding fraudulent transactions is the highest.
References

Economics*, (27), 453-481.
Discount Corrected for Spatial Price Interdependence and Endogeneity of
Demyanyk, Y. and O. van Hemert, forthcoming, “Understanding the subprime mortgage
Heterogeneity and the Exercise of Mortgage Options,” *Econometrica*, 68(2), 275-
307.
Working paper.
in Mortgages,” *Journal of The American Real Estate and Urban Economics
Association*, 13, 261-272.


Appendix

Let us start solving the buyer’s search problem in period 0

\[ E[W] = \int_{0}^{\infty} \max\{g(P_0), E[W]\}dF_0(P_0) - c, \]

where it is assumed (for now) that \( g \) is a continuous decreasing function. As it is well established in these type of models, the solution to the buyer’s problem is a reservation strategy. The buyer should buy the home if the price is below a reservation value \( P_0^* \); otherwise, she should keep searching and get another price draw. Since \( g \) is continuous and decreasing, it must be the case that \( g(P_0^*) = E[W] \). With these considerations, we can manipulate the previous equation as follows

\[
E[W] = \int_{P_0: g(P_0) > W} g(P_0)dF_0(P_0) + \int_{P_0: g(P_0) \leq W} E[W]dF_0(P_0) - c
\]

\[
g(P_0^*) = \int_{P_0: g(P_0) > g(P_0^*)} g(P_0)dF_0(P_0) + \int_{P_0: g(P_0) \leq g(P_0^*)} g(P_0^*)dF_0(P_0) - c
\]

\[
g(P_0) = \int_{0}^{P_0^*} g(P_0)dF_0(P_0) + g(P_0^*)(1 - F_0(P_0^*)) - c. \quad (A1)
\]

We then use integration by parts to compute the value of the integral

\[
\int_{0}^{P_0^*} g(P_0)dF_0(P_0) = [g(P_0)f_0(P_0)]_{0}^{P_0^*} - \int_{0}^{P_0^*} F_0(P_0)g'(P_0)dP
\]

\[
= g(P_0^*)F_0(P_0^*) - \int_{0}^{P_0^*} F_0(P_0)g'(P_0)dP_0,
\]

where, \( g' \) is the partial derivative of \( g \) with respect to \( P_0 \). We substitute this result in equation (A1) and find that the optimal buyer’s reservation value is defined as the solution to

\[
- \int_{0}^{P_0^*} F_0(P_0)g'(P_0)dP_0 = c. \quad (A2)
\]

We now need to show that \( g \) is decreasing and compute its derivative \( g' \). Recall from equation (2) in the text that

\[
g(P_0) = b - (1 - \alpha)P_0 + E_0[V_1(P_0)].
\]
We use equation (3) to first manipulate and rearrange

\[ E_0[V_1(P_0)] = \int_0^\infty \max\{P - (1 + r)\alpha P_0, -(D + b)\} dF_1(P) \]

\[ = \int_0^\infty \{P - (1 + r)\alpha P_0\} dF_1(P) - \int_0^\infty (D + b) dF_1(P) \]

\[ = \int_0^\infty P dF_1(P) - (D + b) F_1[(1 + r)\alpha P_0 - (D + b)] - (1 + r)\alpha P_0 (1 - F_1[(1 + r)\alpha P_0 - (D + b)]). \]

We may then use Liebnitz rule to show that the expected value of the default option is a decreasing function with respect to the transaction price; that is

\[ E_0[V_1(P_0)] = \int_0^\infty \max\{P - (1 + r)\alpha P_0, -(D + b)\} dF_1(P) \]

\[ = - (1 + r)\alpha [(1 + r)\alpha P_0 - (D + b)] f_1[(1 + r)\alpha P_0 - (D + b)] \]

\[ - (1 + r)\alpha (D + b) f_1[(1 + r)\alpha P_0 - (D + b)] \]

\[ - (1 + r)\alpha (1 - F_1[(1 + r)\alpha P_0 - (D + b)]) \]

\[ + [(1 + r)\alpha]^2 P_0 [(1 + r)\alpha P_0 - (D + b)] \]

\[ = -(1 + r)\alpha (1 - F_1[(1 + r)\alpha P_0 - (D + b)]) \]

\[ \leq 0. \]

The previous results allow us to compute

\[ g'(P_0) = -(1 - \alpha) - (1 + r)\alpha (1 - F_1[(1 + r)\alpha P_0 - (D + b)]) < 0. \]

We may replace this last result in (A2), rearrange and find that the optimal buyer’s reservation value solves

\[ \int_0^{P_0^*} F_0(P_0)[1 - \alpha (1 - r F_1[(1 + r)\alpha P_0 - D - b)])] dP_0 = c. \quad \text{(A3)} \]

It is easy to see that the left hand side of equation (A3) is increasing in \( P_0^* \) and that it is equal to zero when \( P_0^* = 0. \) Since the right hand side is positive and constant, a unique solution exists.
Table 1: List of variables and descriptive statistics
Residential properties listed and sold during 2006
Number of observations: 11,700

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Transaction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price</td>
<td>Transaction price (in thousands $) of property listed and sold during 2006</td>
<td>528.7</td>
<td>241.2</td>
<td>125.0</td>
<td>1,995.0</td>
</tr>
<tr>
<td>Early-Default 2007</td>
<td>Indicator if property (sold during 2006) is listed again on the MLS in 2007 as a “pre-foreclosure”, “foreclosure” or “short” sale</td>
<td>0.018</td>
<td>0.134</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Early-Default 2008</td>
<td>Indicator if property (sold during 2006) is listed again on the MLS in 2008 as a “pre-foreclosure”, “foreclosure” or “short” sale</td>
<td>0.044</td>
<td>0.205</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Loan-to-value ratio</td>
<td>Ratio of buyer's first loan + second loan to transaction price</td>
<td>0.779</td>
<td>0.267</td>
<td>0</td>
<td>1.010</td>
</tr>
<tr>
<td><strong>Unit</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sqft</td>
<td>Living area square feet</td>
<td>1,694.0</td>
<td>824.4</td>
<td>426</td>
<td>9,590</td>
</tr>
<tr>
<td>Acreage</td>
<td>Lot acreage</td>
<td>0.20</td>
<td>0.42</td>
<td>0</td>
<td>8.6</td>
</tr>
<tr>
<td>Bedrooms</td>
<td>Number of bedrooms</td>
<td>3.29</td>
<td>1.07</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Full Bathrooms</td>
<td>Number of full bathrooms</td>
<td>2.28</td>
<td>0.83</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Half Bathrooms</td>
<td>Number of half bathrooms</td>
<td>0.77</td>
<td>0.65</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Basement</td>
<td>Indicator if unit has a basement</td>
<td>0.69</td>
<td>0.46</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Central</td>
<td>Indicator if unit has central heating</td>
<td>0.94</td>
<td>0.24</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Fireplace</td>
<td>Number of fireplaces</td>
<td>0.89</td>
<td>0.70</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>New</td>
<td>Indicator if unit is new</td>
<td>0.03</td>
<td>0.16</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Age</td>
<td>Age of the unit (in years)</td>
<td>25.8</td>
<td>15.4</td>
<td>0</td>
<td>136</td>
</tr>
<tr>
<td>HOA</td>
<td>Indicator if property has a home ownership</td>
<td>0.61</td>
<td>0.49</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Detached</td>
<td>Indicator if unit is a detached single family home</td>
<td>0.44</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Townhome</td>
<td>Indicator if unit is a townhome</td>
<td>0.39</td>
<td>0.49</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Apartment</td>
<td>Indicator if unit is an apartment</td>
<td>0.17</td>
<td>0.37</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Neighborhood</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td>Population density in Census Block Group (CBG)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>Proportion of Blacks in CBG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>Proportion of Hispanics in CBG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greater than 65</td>
<td>Proportion of population older than 65 in CBG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS dropouts</td>
<td>Proportion of high school dropouts in CBG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployment</td>
<td>Unemployment rate in CBG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>Median household income in CBG (in 1999 thousands $)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:

- The number of observations with LTV information is 11,432.
- Characteristics of the neighborhoods are not reported.
Table 2: Distribution of time-to-default and loan-to-value ratio

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Number of days between home sale in 2006 and early-default during 2007</th>
<th>Number of days between home sale in 2008</th>
<th>Loan-to-value ratio of units sold during 2006 and early-default during 2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>244 (N = 214)</td>
<td>511 (N = 513)</td>
<td>0.799</td>
</tr>
<tr>
<td>20</td>
<td>305</td>
<td>569</td>
<td>0.800</td>
</tr>
<tr>
<td>30</td>
<td>343</td>
<td>616</td>
<td>0.800</td>
</tr>
<tr>
<td>40</td>
<td>371</td>
<td>649</td>
<td>0.950</td>
</tr>
<tr>
<td>50</td>
<td>418</td>
<td>694</td>
<td>0.996</td>
</tr>
<tr>
<td>60</td>
<td>451</td>
<td>724</td>
<td>0.998</td>
</tr>
<tr>
<td>70</td>
<td>489</td>
<td>724</td>
<td>1.000</td>
</tr>
<tr>
<td>80</td>
<td>528</td>
<td>816</td>
<td>1.000</td>
</tr>
<tr>
<td>90</td>
<td>579</td>
<td>872</td>
<td>1.000</td>
</tr>
</tbody>
</table>
### Table 3: Home prices and early defaults

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early-Default 2007</td>
<td>-0.111 (0.017) ***</td>
<td>-0.019 (0.008) **</td>
<td>0.016 (0.006) **</td>
</tr>
<tr>
<td>Log Sqft</td>
<td>0.399 (0.008) ***</td>
<td>0.341 (0.007) ***</td>
<td></td>
</tr>
<tr>
<td>Acreage</td>
<td>0.084 (0.008) ***</td>
<td>0.082 (0.008) ***</td>
<td></td>
</tr>
<tr>
<td>Bedrooms</td>
<td>0.014 (0.003) **</td>
<td>0.029 (0.002) ***</td>
<td></td>
</tr>
<tr>
<td>Full Bathrooms</td>
<td>0.066 (0.003) ***</td>
<td>0.043 (0.003) ***</td>
<td></td>
</tr>
<tr>
<td>Half Bathrooms</td>
<td>0.021 (0.004) ***</td>
<td>0.012 (0.003) ***</td>
<td></td>
</tr>
<tr>
<td>Basement</td>
<td>0.082 (0.005) ***</td>
<td>0.065 (0.004) ***</td>
<td></td>
</tr>
<tr>
<td>Central</td>
<td>0.010 (0.005) *</td>
<td>0.001 (0.004)</td>
<td></td>
</tr>
<tr>
<td>Fireplace = 1</td>
<td>0.038 (0.003) ***</td>
<td>0.023 (0.003) ***</td>
<td></td>
</tr>
<tr>
<td>Fireplace ≥ 2</td>
<td>0.138 (0.006) ***</td>
<td>0.067 (0.005) ***</td>
<td></td>
</tr>
<tr>
<td>New</td>
<td>0.017 (0.014)</td>
<td>0.014 (0.011)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-0.007 (0.001) ***</td>
<td>-0.009 (0.001) ***</td>
<td></td>
</tr>
<tr>
<td>Age square</td>
<td>0.0001 (0.000) ***</td>
<td>0.0001 (0.000) ***</td>
<td></td>
</tr>
<tr>
<td>HOA</td>
<td>-0.040 (0.005) ***</td>
<td>0.003 (0.003)</td>
<td></td>
</tr>
<tr>
<td>Detached</td>
<td>0.293 (0.009) ***</td>
<td>0.354 (0.007) ***</td>
<td></td>
</tr>
<tr>
<td>Townhome</td>
<td>0.079 (0.006) ***</td>
<td>0.150 (0.005) ***</td>
<td></td>
</tr>
<tr>
<td>Log Density</td>
<td>-0.0001 (0.002)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>-0.158 (0.023) ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.033 (0.026)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greater than 65</td>
<td>0.260 (0.031) ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS dropouts</td>
<td>0.020 (0.030)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployment</td>
<td>0.100 (0.079)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>6.187 (0.004) ***</td>
<td>2.921 (0.062) ***</td>
<td>2.425 (0.095) ***</td>
</tr>
<tr>
<td>Month fixed effects (11)</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Elementary school fixed</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R square</td>
<td>0.001</td>
<td>0.877</td>
<td>0.935</td>
</tr>
<tr>
<td>Number of observations</td>
<td>11,700</td>
<td>11,700</td>
<td>11,700</td>
</tr>
</tbody>
</table>

Notes: Table shows results from a linear OLS model. Dependent variable is the log of transaction price of residential units sold during 2006. The early-default variable equals one if a property was sold during 2006 and is listed again on the MLS in 2007 as a “pre-foreclosure”, “foreclosure” or “short” sale. Robust standard errors in parenthesis. *, **, *** denote significance at the 10, 5, and 1 percent level, respectively.
Table 4: Home prices and early-defaults by unit type

<table>
<thead>
<tr>
<th></th>
<th>(1) All</th>
<th>(2) Detached</th>
<th>(3) Townhome</th>
<th>(4) Apartment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early-Default 2007</td>
<td>0.016 **</td>
<td>0.021 *</td>
<td>0.014 **</td>
<td>0.030 **</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.011)</td>
<td>(0.006)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>R square</td>
<td>0.935</td>
<td>0.875</td>
<td>0.906</td>
<td>0.867</td>
</tr>
<tr>
<td>Number of observations</td>
<td>11,700</td>
<td>5,206</td>
<td>4,534</td>
<td>1,960</td>
</tr>
</tbody>
</table>

Note: Table shows results from a linear OLS model. Dependent variable is the log of transaction price of residential units sold during 2006. The early-default variable equals one if a property was sold during 2006 and is listed again on the MLS in 2007 as a “pre-foreclosure”, “foreclosure” or “short” sale. In addition to the early-default variable shown above, independent variables include housing characteristics, neighborhood characteristics, month fixed effects and elementary school fixed effects; that is, the same variables used in column (3) of Table 3. Robust standard errors in parenthesis. *, **, *** denote significance at the 10, 5, and 1 percent level, respectively.
Table 5: Early-default premium and time-to-default

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early-Default 2007</td>
<td>0.016 **</td>
<td>0.017 ***</td>
<td>0.017 ***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td></td>
</tr>
<tr>
<td>Early-Default 2008</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early-Default January - June 2008</td>
<td>0.010 *</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early-Default July - December 2008</td>
<td>-0.002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early-Default 2007 or 2008</td>
<td></td>
<td></td>
<td>0.032 ***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.013)</td>
<td></td>
</tr>
<tr>
<td>Time (in years) from sale in 2006 until Early-Default in 2007 or 2008 (zero if Early-Default does not occur)</td>
<td></td>
<td></td>
<td>-0.014 **</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.007)</td>
<td></td>
</tr>
<tr>
<td>R square</td>
<td>0.935</td>
<td>0.935</td>
<td>0.935</td>
<td>0.935</td>
</tr>
<tr>
<td>Number of observations</td>
<td>11,700</td>
<td>11,700</td>
<td>11,700</td>
<td>11,700</td>
</tr>
</tbody>
</table>

Note: Table shows results from a linear OLS model. Dependent variable is the log of transaction price of residential units sold during 2006. The early-default variables equal one if a property was sold during 2006 and is listed again on the MLS as a “pre-foreclosure”, “foreclosure” or “short” sale in later corresponding periods. In addition to the early-default variable shown above, independent variables include housing characteristics, neighborhood characteristics, month fixed effects and elementary school fixed effects; that is, the same variables used in column (3) of Table 3. Robust standard errors in parenthesis. *, **, *** denote significance at the 10, 5, and 1 percent level, respectively.
### Table 6: Early-default premium and loan-to-value ratio (LTV)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early-Default 2007</td>
<td>0.016 **</td>
<td>0.017 ***</td>
<td>-0.003</td>
<td>0.017 **</td>
<td>-0.044 *</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.009)</td>
<td>(0.007)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>1 ( LTV &gt; 0.99 ) * Early-Default 2007</td>
<td></td>
<td></td>
<td>0.036 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.013)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LTV * Early-Default 2007</td>
<td>0.068 ***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 ( LTV &gt; 0.99 )</td>
<td>-0.009 ***</td>
<td>-0.010 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LTV</td>
<td>-0.023 ***</td>
<td>-0.024 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R square</td>
<td>0.935</td>
<td>0.935</td>
<td>0.935</td>
<td>0.935</td>
<td>0.935</td>
</tr>
<tr>
<td>Number of observations</td>
<td>11,700</td>
<td>11,432</td>
<td>11,432</td>
<td>11,432</td>
<td>11,432</td>
</tr>
</tbody>
</table>

Note: Table shows results from a linear OLS model. Dependent variable is the log of transaction price of residential units sold during 2006. The early-default variable equal one if a property was sold during 2006 and is listed again on the MLS as a “pre-foreclosure”, “foreclosure” or “short” sale in 2007. In addition to the early-default variable shown above, independent variables include housing characteristics, neighborhood characteristics, month fixed effects and elementary school fixed effects; that is, the same variables used in column (3) of Table 3. Robust standard errors in parenthesis. *, **, *** denote significance at the 10, 5, and 1 percent level, respectively.
Table 7: Early-default premium and real estate broker offices

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early-Default 2007</td>
<td>0.016 **</td>
<td>0.015 **</td>
<td>0.015 **</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Office A * Early-Default 2007</td>
<td>0.077 ***</td>
<td>0.077 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.021)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Office B * Early-Default 2007</td>
<td>0.003</td>
<td>0.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.038)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Office A</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Office B</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early-Default 2007 * Broker office with 3 or more early-defaults fixed effects (18)</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Broker office fixed effects (include indicators for all offices)</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>R square</td>
<td>0.935</td>
<td>0.935</td>
<td>0.935</td>
<td>0.941</td>
</tr>
<tr>
<td>Number of observations</td>
<td>11,700</td>
<td>11,700</td>
<td>11,700</td>
<td>11,694</td>
</tr>
</tbody>
</table>

Note: Table shows results from a linear OLS model. Dependent variable is the log of transaction price of residential units sold during 2006. The early-default variable equal one if a property was sold during 2006 and is listed again on the MLS as a “pre-foreclosure”, “foreclosure” or “short” sale in 2007. In addition to the early-default variable shown above, independent variables include housing characteristics, neighborhood characteristics, month fixed effects and elementary school fixed effects; that is, the same variables used in column (3) of Table 3. Robust standard errors in parenthesis. *, **, *** denote significance at the 10, 5, and 1 percent level, respectively.
Figure 1: Early-default premium across brokers

a) Three-bin histogram

b) Five-bin histogram