Estimating Loan-to-Value and Foreclosure Behavior

PRELIMINARY AND INCOMPLETE

Arthur Korteweg

Morten Sorensen

July 2011

Abstract

We develop and estimate a unified model of house prices, loan-to-value ratios ("LTVs"), and trade- and foreclosure behavior. House prices are only observed for traded properties, and trades are endogenous, creating sample-selection problems for traditional estimators. We use a Bayesian filtering procedure to recover the price path for each property in the data and produce selection-corrected estimates of historical LTVs and foreclosure behavior, both showing large unprecedented changes since 2007.

Stanford Graduate School of Business (korteweg@stanford.edu) and Columbia Business School, NBER and SIFR (ms3814@columbia.edu). We are grateful to Chris Mayer and the Paul Milstein Center for Real Estate at Columbia Business School for generous access to data and research assistance. We thank Chris Downing, Monika Piazzesi, Stijn van Nieuwerburgh, Chester Spatt, Richard Stanton, Nancy Wallace, along with seminar participants at the 2011 Summer Real Estate Symposium, Berkeley Capital Markets Real Estate seminar, Columbia Business School, and Stanford Graduate School of Business for helpful comments and feedback.

Price indices, loan-to-value ratios ("LTVs"), and trade- and foreclosure behavior are important measures of economic activity in real-estate markets. Yet, common estimators of these measures have severe limitations. Particularly, estimation of LTVs is a nontrivial inference problem, and to our knowledge our empirical approach provides the first analysis in the literature of this problem. Additionally, our analysis sheds new light on the losely related problems of estimating price indices and trade- and foreclosure behavior.

LTVs are mostly estimated by survey, which raises two immediate concerns: Home owners may not be fully informed and objective when hypothetically valuing their home for a survey; and the estimated LTVs are restricted to the geographical scope and intermittent timing of the survey, obviously. This difficulty of estimating LTVs makes it difficult to estimate trade- and foreclosure behavior as well. Natural empirical specifications are probit and hazard models of trade- or foreclosure decisions as functions of current LTVs and other explanatory variables. Without LTV data, however, these models are difficult to estimate. Additionally, when trade decisions depend on the current price, a concern about sample selection arises, because prices are only observed for the traded properties. When appreciating properties are more likely to trade, as documented by Case, Pollakowski, and Wachter (1997), the price dynamics of the properties with observed prices (the traded ones) are no longer representative of the price dynamics in the overall population. This selection potentially biases standard indices of property prices (see Haurin and Hendershott, 1991).[1] Our empirical approach addresses these concerns.

Our model extends the standard repeat-sales model (Bailey, Muth, and Nourse (1963) and Case and Shiller (1987)) in two ways: First, we explicitly model and recover the entire price path of each property in our data, even when it is not traded. Second, we add a trade

[1]This problem is aptly described by Calnea, who calculates the UK's national land registry house price index using a repeat-sales regression (RSR): "The RSR index is naturally more reflective of properties that transact more frequently. In so far as a differential in price appreciation exists between properties based on the relative frequency of transactions, the RSR measure will be naturally weighted towards the more frequently transacting subset of properties. There are a variety of reasons why the holding duration of properties might be unevenly distributed. The increase in transaction costs for more expensive properties due to stamp duty may result in a decreased turnover of more expensive homes. "Life-cycle" theories on property holding period posit that less expensive properties are traded more frequently - when people move up the property "ladder" they tend to move home less often. In addition the Buy-to-Let market is more active in the lower price brackets. Policy-makers need to be aware of the price appreciation differentials between submarkets, especially when there is systematic variation in the frequency of transactions between these submarkets" (Calnea Analytics (2010), p. 14).

process that specifies the decision to trade for each property at each point in time.

Simultaneously estimating prices, LTVs, and trade- and foreclosure behavior brings several advantages: First, econometrically, the trade process is a dynamic extension of the selection equation in the standard sample-selection model (Heckman (1979, 1990)). Simultaneously estimating the price and trade process corrects for selection bias. (Below, we sometimes refer to the trade process as the *selection process* or *trade-selection process* to emphasize this role.) Second, the estimated parameters in the trade process reflect trading behavior. When we separate regular sales and foreclosure sales and allow the coefficients to vary over time, we find large unprecedented changes in recent foreclosure behavior. Third, the approach estimates the full cross-sectional and time-series distribution of LTVs. Estimating time series of LTVs from transaction data is difficult, because it requires modeling the price behavior of untraded properties. We present the first historical estimates of LTVs calculated from transaction prices. Finally, our procedure consistently incorporates price information revealed by untraded properties. To illustrate, when appreciating properties are more likely to trade, a period with few trades suggests declining prices. Consequently, volume is informative about prices and should be included in the estimation. This represents a fundamental difference between our dynamic extension and the standard Heckman (1979) cross-sectional sample-selection model. In the standard model, observations with unobserved outcomes are only informative about the first stage, not the second one.[2] In our dynamic extension, property prices (the unobserved outcomes) are serially correlated. Hence, the fact that a property does not trade in a current period is informative about its current price but also about its previous and subsequent prices, due to serial correlation, and it cannot be ignored in the second stage.[3]

[2]In the standard selection model unobserved outcomes are independent and can be integrated out of the likelihood function for the second stage. With serial correlation, unobserved outcomes are no longer independent and cannot be integrated out. To incorporate the information in nontraded properties, our estimator must track the entire price path of each individual property.

[3]Some studies construct corrected real-estate indices by estimating a standard Heckman selection model (e.g., Jud and Seaks (1994), Hwang and Quigley (2004), and Munneke and Slade (2000)). These studies suffer from two problems. First, the standard Heckman model does not allow for serial correlations in the unobserved outcomes, and the price information revealed by untraded properties is not incorporated, raising concerns about these estimators being misspecified and inconsistent. Second, this approach cannot include the (unobserved) house price in the first-stage selection equation, and the studies estimate this equation with only observable house characteristics, such as size, location, and prevailing mortgage rates, meaning that these studies only corrects for selection to the extent that the particular price dynamics of the traded properties is fully explained by these observable characteristics. They do not capture selection directly on price. By

Bayesian estimation brings two important advantages:[4] It is efficient, with robust convergence properties. Modeling price dynamics for individual properties is computationally intensive, leaving standard ML estimation numerically infeasible. Moreover, it is straightforward to construct posterior distributions for nonlinear transformations of the estimated parameters, such as calculations of price indices, LTVs, and foreclosure- and trade intensities. Standard statistical inference, such as calculating standard errors, is difficult for nonlinear functions of the estimated parameters, particularly functions of parameters that are not asymptotically normal distributed, such variance parameters.

Finally, note that our model extends the empirical finance literature about estimating the risk and return of illiquid assets with unobserved prices, such as venture capital investments in privately held companies (Cochrane (2005) and Korteweg and Sorensen (2010)). The empirical issues are similar, and our approach may be useful for understanding prices and trading behavior of other illiquid or asynchronously-traded assets, such as corporate bonds, small-cap stock, or index arbitrage. More generally, our empirical approach exploits recent advances in Bayesian computational methods — specifically, MCMC, Gibbs Sampling, and FFBS — that permit estimating new models of behavior at the level of individual agents, which may be useful in other applications.

In section I, we present the empirical model and discuss the estimation procedure and identification. Section II presents the data. Section III compares price indices estimated with and without including the trade process to correct for sample selection. Section IV presents estimates of the LTV distribution. Section V presents estimates of trade- and foreclosure behavior. Section VI concludes. More details about the estimation procedure are in the appendix.

I Model of Prices and Trades

To fix ideas and notation, it is useful to derive the discrete-time price dynamics from continuous-time fundamentals. Let $\mu(t)$ be a common exogenous determinant of price

filtering out the unobserved prices, our approach allows the trade-selection process to depend directly on each property's contemporaneous price.

[4] Our procedure is substantially different from the Bayesian estimators of the repeat-sales model described and compared by Goetzmann (1992), which do not exploit these advantages to the same extent.

appreciation (a price index), and let the price of property i follow the Brownian motion

$$\frac{dP_i(t)}{P_i(t)} = \mu(t)dt + \sigma_i dB_i(t). \tag{1}$$

Define the log-price $p_i(t) = \ln(P_i(t))$ and let $\delta_i(t) = \int_{t-1}^{t} \mu(\tau)d\tau - \frac{1}{2}\sigma_i^2$. Using Ito's lemma, the change in the log-price from time t to t' is

$$p_i(t') = p_i(t) + \left[\sum_{t+1}^{t'} \delta_i(\tau)\right] + \varepsilon_i(t,t') \tag{2}$$

with $\varepsilon_i(t,t') \sim N\left(0,(t'-t)\sigma_i^2\right)$.

The standard repeat-sales regression ("RSR") is estimated from this equation; using properties that trade (at least) twice, the change in the observed (log-)prices is regressed on indicator variables that equal one for the intermediate periods between the two trades, represented by the $\delta_i(\tau)$ terms in equation (2). To correct for heteroscedasticity, the RSR is often implemented as a GLS estimator by weighing each observation by the inverse of the square root of the time between trades. With sufficiently many partially overlapping trades (assuming no heterogeneity in σ_i), all $\delta(t)$ coefficients are identified. Moreover, the estimated $\delta(t)$ coefficients are consistent when ε is independent of the indicator variables. This independence fails, however, when the decision to trade a property is not independent of its price appreciation. When appreciating properties trade faster, a higher error term is correlated with more indicators equaling zero, creating a sample-selection problem and potentially biasing the estimated coefficients.

Setting $t = t' - 1$, the one-period transition equation is

$$p_i(t) = p_i(t-1) + \delta_i(t) + \varepsilon_i(t) \tag{3}$$

with $\varepsilon_i(t) \sim N(0,\sigma_i^2)$. Since most prices are unobserved, this equation cannot be estimated directly. Treating the price as an unobserved state variable, however, the estimation procedure uses this equation to filter out the unobserved intermediate prices between trades.

An index can be constructed from the estimated $\delta(t)$ coefficients. Normalizing by the price level at time t_0, it is natural to define an index as the average in the population of the current price relative to the time t_0 price (in actual prices, not log-prices) as

$$I(t) = E\left[\frac{P_i(t)}{P_i(t_0)}\right] = E\left[\exp\left(\left[\sum_{t_0+1}^{t}\delta(\tau)\right] + \varepsilon_i(t_0,t)\right)\right] = \prod_{t_0+1}^{t}\exp\left(\delta(\tau) + \tfrac{1}{2}\sigma^2\right), \quad (4)$$

with the one-period change int the index given by $I(t)/I(t-1) = \exp\left(\delta(t) + \frac{1}{2}\sigma^2\right)$. Although, typical real-estate indices, such as those from S&P/Case-Shiller, CoreLogic, and FHFA (formerly OFHEO), are defined without the $\frac{1}{2}\sigma^2$ adjustment. Below, we compare indices with and without this adjustment. Goetzmann (1992) discusses this adjustment in more detail, and denote indices with and without it *arithmetic* and *geometric* indices, respectively.

A Trade- and Foreclosure Processes

To model home sales, it is convenient to define the latent discrete-time process $w_i(t)$ such that property i trades between time $t-1$ and t, and hence $p_i(t)$ is observed, when

$$w_i(t) \geq 0. \quad (5)$$

This trade process is parametrized as

$$w_i(t) = W_i'(t)\alpha_0 + p_i(t)\alpha_p + \eta_i(t), \quad (6)$$

where $\eta_i(t)$ is *i.i.d.*[5] The log price is $p_i(t)$. Property characteristics and prevailing mortgage rates (and a constant term) are in $W_i(t)$, along with the log-loan amount with a coefficient fixed to $-\alpha_p$, implying that α_p is the negative of the coefficient on log-LTV.

The resulting model is a dynamic extension of the standard cross-sectional sample-selection model from Heckman (1979, 1990) with the trade process as the selection process. Under standard conditions, jointly estimating the these processes corrects for the selection bias that arises when the price dynamics of the properties with observed prices are not representative of the price dynamics in the overall population. Note that the trade process

[5] This model of trades is equivalent to a binary probit model, and the parameters are only identified up to scale. Without loss of generality, the scale is normalized by fixing the error term's variance to one.

depends on the contemporaneous LTV ratio. This is a natural specification, yet it has been difficult to implement in previous studies, since LTVs are unobserved. Our model solves this problem, by estimating the entire price process jointly with the trade process.

Our data distinguish normal sales from foreclosure sales. One might suspect, as we find empirically, that these sales are governed by different processes. Appreciating properties may be more likely to trade in normal sales, whereas depreciating properties may be more likely to wind up in foreclosure. Hence, we estimate some specifications with separate trade and foreclosure processes. In these specifications, we include a foreclosure process such that a foreclosure sale (but not necessarily a normal sale) occurs between time $t-1$ and t when

$$z_i(t) \geq 0, \tag{7}$$

and this process is parametrized as

$$z_i(t) = Z_i'(t)\gamma_0 + p_i(t)\gamma_p + \xi_i(t). \tag{8}$$

As above, $Z_i(t)$ contains observed variables that affect the probability of foreclosure sales, and ξ is *i.i.d.* $N(0,1)$. Again, $Z_i(t)$ contains the log-LTV, property characteristics, and prevailing mortgage rates.

B Empirical Implementation

To summarize, the baseline model specifies the price- and trade processes:

$$p_i(t) = p_i(t-1) + \delta(t) + \varepsilon_i(t) \tag{9}$$

$$w_i(t) = W_i'(t)\alpha_0 + p_i(t)\alpha_p + \eta_i(t) \tag{10}$$

$$w_i(t) \geq 0 \Leftrightarrow p_i(t) \text{ is observed} \tag{11}$$

The price process, $p_i(t)$, is mostly unobserved, except when $w_i(t) \geq 0$. The trade (or selection) process, $w_i(t)$, is entirely unobserved. The vector $W_i(t)$ is observed data. The error terms are *i.i.d* with $\varepsilon_i(t) \sim N(0,\sigma^2)$ and $\eta_i(t) \sim N(0,1)$. The estimated parameters of interest are α, σ^2, and $\delta(t)$.[6]

This model defines a likelihood function. ML estimation is complicated, however, by the large number of latent variables, since evaluating the likelihood requires jointly integrating over them. Specifying the model at the quarterly level over a twenty-year period results in 160 ($= 2 \times 4 \times 20$) latent variables per property. With about 70,000 properties in our data, evaluation of the likelihood requires numerically evaluating 70,000 160-dimensional integrals, rendering ML estimation intractable. Alternatively, we use a Bayesian procedure based on a Markov Chain Monte Carlo (MCMC) method known as Gibbs sampling to dramatically reduce the computational burden. We provide an overview of this procedure below, and more details are in the appendix (see also Korteweg and Sorensen (2010)).

The model is constructed such that its variables can be divided into three blocks: The first block contains the parameters, α, σ^2, and $\delta(t)$; the second one contains the variables in the trade processes, $w_i(t)$; and the third block contains the price processes, $p_i(t)$. The Gibbs sampler simulates the (augmented) posterior distribution by iteratively drawing the variables in each block conditional on the previous draws of the variables in the other blocks (see Geman and Geman (1984), Tanner and Wong (1987), Gelfand and Smith (1990), and Johannes and Polson (2006)).

In the first block, conditional on the previous draws of the price and trade processes, the parameters α, σ^2, and $\delta(t)$ are given by the linear equations 9 and 10, and they are drawn from a standard Bayesian linear regression. Drawing the trade process in the second block is similarly straightforward. Conditional on the parameters and prices, the distribution of $w_i(t)$ follows a truncated normal distribution that is constrained to be negative when there is no trade and positive when a trade is observed, as specified by equation 11. These first two blocks are analogous to Bayesian estimation of probit models (Albert and Chib (1993)). The key to the model is the third block where the entire path of unobserved prices is drawn using the Forward Filtering Backwards Sampling (FFBS) procedure (Carter and

[6] Although not pursued here, these assumptions can be relaxed somewhat. It is possible to estimate specifications with property-specific coefficients using hierarchical priors, it is possible to account for observation errors in the observed prices, and the normality assumptions generalize to mixtures of normals without losing tractability of the Bayesian procedure. The log-price must enter the trade process linearly, however, to maintain a linear Kalman filter, and generalizing this specification add substantial numerical complexity. See Korteweg and Sorensen (2010) for details.

Kohn (1994) and Fruhwirth-Schnatter (1994)). Conditional on the parameters and trade process, the price process can be viewed as being defined by a linear state space called a Kalman Filter. Under this view, $p_i(t)$ is the unobserved state variable; the transition rule is the one-period price equation (9); the index $\delta(t)$ is an "observed" control acting on the state; and, conditional on $w_i(t)$, the trade process is an observation equation, providing noisy "observations" of the state. Given this setup, the FFBS procedure efficiently draws from the conditional posterior distribution of the entire price path, as required in the third block of the Gibbs sampler.[7]

This Gibbs sampling procedure iteratively draws from the joint posterior distribution of the parameters and the individual price paths for the properties. Using these draws, the posterior distributions of the price index and LTV distributions are straightforward to construct. For the price index, fixing t, in each iteration, calculate $I(t) = \prod_{t_0+1}^{t} \exp\left(\delta(\tau) + \frac{1}{2}\sigma^2\right)$ using the current draws of $\delta(t)$ and σ^2. Across iterationrs, the resulting distribution is the posterior distribution of $I(t)$.[8] Repeating this calculation for all t constructs the posterior distribution of the entire index.

For the LTV distribution, the construction is slightly more complicated conceptually, since the object of interest is the LTV distribution, and the posterior becomes a distribution over distributions (or, more precisely, the time series of this distribution of distributions). In each iteration, calculate the $LTV_i(t) = Principal_i(t)/P_i(t)$ using the calculated outstanding principals (described below) and the current draws of the price processes. Collecting these values across properties produces one draw from the posterior distribution of the cross-sectional distribution of LTVs. Collecting these cross-sectional distributions across iterations and time produces a time series of their posterior distribution. From this distribution it is straightforward to calculate, for example, the time series of the estiamated fraction of "underwater" properties, i.e., properties with LTVs exceeding one.

[7]While it is convenient to describe the blocks in this order, the actual sampling procedure has better numerical properties when starting with the prices in the the third block. By setting the initial value of $\alpha_p = 0$, the first iteration can draw the prices independently of the initial values of the trade process, which speeds up convergence.

[8]Note that it would be difficult to perform standard classical asymptotic inference on this index, such as calculating its standard error, since $I(t)$ is a nonlinear function of the estimated parameters and the asymptotic distribution of σ^2 is not normal.

C Identification

Heckman (1990) shows that semi-parametric identification of the standard selection model requires exogenous variation in the selection equation. We include the time since the previous trade (*Time*) as exogenous variation in the selection process. There are two requirements for *Time* to be valid. First, the exclusion restriction requires that *Time* is independent of the error term in the price process. When the price process follows a martingale, which is commonly thought to be a reasonable assumption for price processes, the exclusion restriction holds mechanically. Second, *Time* must be directly related to the probability of a sale. Due to transaction costs, it is reasonable to think that new owners do not intend to resell properties immediately after buying them. Immediately after a trade, the trade intensity declines and then it gradually increases over time. This behavior is consistent with the well-known phenomenon of "seasoning" of mortgage-backed securities, where new loans prepay slower than older loans. The empirical results confirm this pattern as well. When the two requirements are satisfied, *Time* provides valid exogenous variation for the identification of the model. Note, however, that the empirical results are largely similar for specifications with and without *Time*, suggesting that the parametric identification from the distributional assumptions for the error terms is not unreasonable.

II Data

We use data with transactions of single-family residences in Alameda County, California. Alameda county is located in the San Francisco Bay Area and includes the cities of Oakland and Fremont. The data are from the Real Estate Center at Columbia Business School and contain all transactions in the twenty-year period from 1Q1988 to 3Q2008, including sales dates and prices, mortgage amounts, and refinancing information obtained from the deeds records. In addition, tax records contain information about the property's characteristics, such as size, number of bedrooms, single- or multi-family residence, etc.

We restrict the sample to properties that satisfy the following criteria: The property is a single-family residence; it trades at least twice during our sample period, and both sales are full sales (no partial sales); it is owner-occupied, and the owner is not a corporation;

the tax records have no missing property characteristics; the property's characteristics have not changed between the two sales; it has no more than 10 bedrooms, no more than 5 bathrooms, no more than 3 stories, is located on 5 or fewer acres, and has a living space of no more than 10,000 square feet. The resulting sample contains 164,824 transactions of 68,700 properties. Table 1 presents summary statistics, and Figure 1 shows the time-series of normal trades and foreclosure sales of the properties in our sample.

While the data contain initial mortgage amounts, they contain no information about amortization schedules. We construct an amortization schedule assuming a thirty-year fixed rate at the prevailing mortgage rate at the time of each sale. Whenever possible we update the outstanding mortgage amounts using refinancing information. Based on this calculation of outstanding loans, we define LTV as the loan-to-value ratio, where the value is given by the price process specified in our model. We include the log of this LTV as an explanatory variable in our selection process.

This calculation fails for unmortgaged properties. These properties have LTVs equal to zero, leaving the log-LTV undefined. For these transactions, we set log-LTV to -3, corresponding to a LTV ratio of 5%. The particular number has a negligible effect on our estimates, as does excluding these properties altogether. We do not truncate the upper tail of the LTV distribution, but in unreported estimates, we find that doing so has a negligible impact on the estimates as well. The top plot in Figure 5 shows the distribution of the buyers' LTV ratios at the time of the transactions. The bottom plot shows the sellers' distribution (note that both are observed at the time of sale but not in between sales).

Mortgage rates are collected from FRED (a database provided by the Federal Reserve in St. Louis).

III Price Dynamics

A natural starting point for the empirical analysis is to investigate the estimated price dynamics. Ignoring the trade process and selection correction for the moment, Figure 2 compares indices estimated using the standard GLS procedure to indices from our MCMC procedure (without selection, i.e., fixing $\alpha_p = 0$), along with the S&P/Case-Shiller home price index for the San Francisco metro area.[9] Arithmetic indices include the $\frac{1}{2}\sigma^2$ adjustment (the annualized GLS estimate of σ^2 is 0.2811^2); Geometric indices are calculated without the adjustment. All indices are normalized to 100 in 1Q2000.

Apart from the large differences between arithmetic and geometric indices, the indices are very similar. The S&P/Case-Shiller index is very similar to our geometric indices, although it is about ten points lower at the peak. This ten-point difference probably arises because the S&P/Case-Shiller index includes transactions from the entire San Francisco MSA, comprising Alameda, Contra Costa, Marin, San Francisco, and San Mateo counties; our data cover only Alameda, which is around one-third of the MSA by population. Additionally, unlike our indices, the S&P/Case-Shiller index is smoothed and value-weighted.

In Figure 2, the indices from the GLS and MCMC estimators are virtually identical. This is reassuring. Statistically, the GLS estimator is defined by moment conditions and the assumption that the variance of the error term increases linearly with the time between trades. The MCMC estimator imposes additional distributional assumption on the error terms. The similarity between the resulting indices suggests that, at least absent the trade process and selection correction, the MCMC estimator's additional distributional assumptions are not unreasonable.

Figure 3 presents indices corrected for selection by including the trade process in the MCMC estimator. Models A to F refer to different specifications of the selection process given in Table 2. The baseline index is from the MCMC estimates without selection correction from Figure 2. The patterns are broadly similar. Unlike the corrected indices, the uncorrected one shows a small increase by the end of the sample period. At this point, the uncorrected index is at the level of early 2005 whereas the corrected ones is at the level of early 2004. Whether this difference is substantial is somewhat a matter of interpretation. The broad similarities, however, suggest that selection bias arising from changes in the composition of transacted properties are modest, at least as captured by our model, and that these composition changes do not substantially distort the standard repeat-sales index. This is reassuring as well. It suggests that the price processes generated by our procedure, including the trade process, are not unreasonable.

[9] Downloaded from the S&P/Case-Shiller website on February 25, 2010.

IV LTV Distributions

The LTV distribution is an important indicator for the state of the real-estate market and household finance more generally. For example, Lamont and Stein (1999) find evidence that higher LTVs exacerbate the effects of income shocks on house prices. And, there is some debate whether high LTVs impede labor mobility (Ferreira, Gyourko, and Tracy (2010) and Schulhofer-Wohl (2011)). Bajari, Chu, and Park (2010) and Melzer (2011) argue that LTVs are important determinants of household expenditures and defaults.

LTV distributions are difficult to measure, however. One common approach is by survey, and one typical example is the American Housing Survey (AHS), which was started in 1974 and visits included cities at irregular four- to nine-year intervals. Another example is the Consumer Expenditure Interview Survey (CE) by the Bureau of Labor Statistics. There are two problems with survey-based LTV estimates: First, they depend on the home owners' personal assessments of property values. These assessments may be inaccurate, particularly during periods with few or atypical transactions (such as the recent surge in foreclosure sales). According to Melzer (2011), "homeowners might be optimistic about their home's value or reluctant to acknowledge that they owe more than their home is worth." Moreover, Melzer (2011) find substantial differences between the fractions of underwater properties reported by the AHS and CE surveys. Second, since they only cover the cities and time periods where surveys were actually performed, these LTV estimates have limited scope. In light of these problems, it seems desirable to have an empirical procedure to estimate LTVs for individual properties from transaction prices. Transaction data exist for vastly longer time series and geographical areas than surveys, and transaction prices are arguably more objective and comparable across places and periods.

The data contain the LTV ratio at the time of each trade. Panel A in Figure 5 plots the LTV ratio of the new owner's mortgage. Panel B, plots the selling owner's mortgage, constructed from amortizing the individual mortgages as described. Figure 6, Panel A, plots the median seller's LTV over 20 years, together with the fraction of the housing stock that is sold in each quarter, showing an increase in median LTVs towards the end of the sample paired with a drop in trading volume. The increase in the median LTV appears fairly innocuous, essentially restoring the median LTV to its level before the housing boom during the 2000s. In contrast, Figure 6, Panel B, shows a dramatic increase in the fraction of sales that are foreclosure sales, and Panel C plots the fraction of properties sold with a LTV ratio above one, i.e., "underwater" properties. This increase is even more dramatic, suggesting that not only is the median LTV increasing, but the cross-sectional distribution of LTVs is becoming more dispersed, with an increasing fraction of properties in the tail with very negative equity. These plots are all made from the LTVs at the time of sales.

Figures 7 to 9 present the historical LTV distributions estimated by the model. Figure 7 shows the percentiles of the LTV distribution resulting from the various specifications. Figure 8 and 9 show the fractions of properties with LTVs greater than 1, 1.25, and 1.5. In all cases, the plots are sobering. In all cases, the LTV distribution appears to deteriorate substantially during the last years of our sample.

Another, albeit less common, approach to estimating LTVs is to starts from property values and mortgage principals observed when properties trade and update these values over time using a local house price index. The LTV distribution is then constructed from the updated values and the amortized principals (e.g., Bajari, Chu and Park, 2011). There are two problems with this *index approach*: First, most property price indices are geometric, not arithmetic, meaning that they reflect the relative changes in log-prices, not actual prices. For the calculation to be consistent, the $\frac{1}{2}\sigma^2$ adjustment should be included. Second and more importantly, the index approach underestimates the cross-sectional dispersion in prices and LTVs. To illustrate, if the index shows an average appreciation of 5%, it makes a big difference for the LTV distribution whether this appreciation means exactly 5% for all properties, or whether it means, say, a 15% increase for half the properties and a 5% decline for the other half. By assuming the former, the index approach mechanically underestimates dispersion in the LTV distribution.

We compare LTV estimates from the index approach to those from our model. In Figures 8 and 9, the plots denoted RSR represent LTV estimates using the index approach. The RSR plots are substantially below both the baseline plot, calculated without selection correction, and the two plots correcting for selection (using models A and F). Not surprisingly, our model shows substantially greater dispersion in LTVs, and consequently a much higher fraction of underwater properties, than the index approach.

Figures 8 and 9 also shows a large effect of the selection on the dispersion LTVs. In other words, moving from the index method to the baseline model shows a substantial increase in dispersion of LTVs, which is expected. But, moving from the baseline model to Models A and F also leads to substantial differences in the LTV distribution, which may more surprising. The intuition is the following. Properties with lower LTVs are more likely to trade and LTV ratios are reset when properties are refinanced following a trade. In contrast, properties with higher LTVs are less likely to trade and have their LTV ratio reset. Hence, even when the average price remain flat, properties below average with declining prices have worsening LTVs. Properties above average, with appreciating prices, have improving LTVs, but are also more likely to trade and reset their improved LTV back to the starting point. Overall, this selection leads means that average LTVs can deteriorate even when house prices remain flat, on average. This effect explains the increase in the number of underwater properties from the baseline to Model A. Model F further allows the LTV coefficient to vary by sub period, and it changes sign in the last one, reversing the direction of selection and resulting in Model F estimating fewer underwater properties than the baseline specification. Overall, this evidence suggests that the cross-sectional selection of property trades and refinancings, depending on LTVs, is important for the distribution of LTVs in general and the fraction of underwater properties specifically. In all cases, however, the index method underestimates this fraction.

V Trade- and Foreclosure Behavior

The estimates of the different specifications of the trade process are in Table 2. Log-LTV is a primary determinant of trades and is included in all models. Models B to F include additional explanatory variables, and Model F allows some coefficients to vary by sub period to reveal changes in trade- and foreclosure behavior.

In model A only log-LTV is included and the estimated coefficient is negative and statistically highly significant.[10] Since a lower LTV corresponds to a higher price, the negative coefficient suggests that properties with higher prices trade at higher intensities. This is

[10]Although a slight abuse of standard terminology for Bayesian statistics, we term a coefficient statistically significant at a given level when the corresponding symmetric credible interval does not contain zero.

consistent with Case, Pollakowski, and Wachter (1997) who finds that appreciating properties trade faster. Moreover, in Figure 3, we see that the correction attenuates the size of the price bubble relative to the baseline. With a negative coefficient, this attenuation is intuitive. When prices appreciate generally, properties with more rapid price appreciation are more frequently traded, their prices are more frequently observed, and this causes standard indices to exaggerate the price appreciation. Correcting for this selection, shows that the overall population of properties had more attenuated price dynamics.

Model B adds the time (in years) since the previous sale. The estimated coefficients are significantly positive for *Time* and negative for *Time-Squared*. This is reasonable. It shows that the trade intensity follows an inverse-U shape. When a property is just traded the probability of another trade immediately drops and then gradually increases, peaking after about nine years[11] holding LTV constant. As discussed previously, including *Time* improves the statistical identification of the model, although Table 2 and Figure 3 suggest it has this improvement has only a modest effect on the estimated parameters and indices.

Model C further includes the mortgage rate, which comes out with a negative and significant coefficient, showing that periods with higher mortgage rates have fewer transactions. Including this variable has little effect on the LTV coefficient, however, and Figure 3 shows almost no change in the estimated index.

Model D further adds the size and age of the property, finding that larger and older houses trade less frequently. In Model E these house characteristics are also interacted with the log-LTV ratio. The positive significant coefficient on the interaction of size and log-LTV means for larger houses the trade intensity is less sensitive to LTVs than for smaller houses. The interaction of age and log-LTV shows a very small effect.

Finally, in Model F the sample is divided into four sub periods with separate intercepts and coefficients on log-LTV. This specification is motivated by the sharp increase in foreclosures toward the end of our sample, which raises concerns about a structural break in trade- and foreclosure behavior. Interestingly, the coefficient on LTV increases monotonically over the sample (its absolute magnitude declines). The other coefficients in Model F are largely unchanged compared to the other models. The increase in the LTV coefficients

[11] In Model B the maximum intensity occurs after 9.1 years ($-0.0381/2 \times -0.0021 = 9.1$).

reflect a gradual shift in the market where sales are becoming less sensitive to LTV ratios. In the early period (1988-93) more valuable properties had relatively higher trade intensities, but these higher intensities diminished until the last sub period where the coefficient changes sign and the effect reverses. Now, less valuable properties are more likely to trade. This change is consistent with the large increase in foreclosure sales during the last period, since foreclosure sales are predominantly depreciating properties with high LTV ratios.[12] Indeed, Figure 3 shows that the home price index created from the estimates of Model F declines about 10-15 index points less during 2007 and 2008 than the other indices. Moreover, since the index from Model F peaks at a higher level, the proportional decline is even smaller.

Up to this point, the sample of trades only includes properties that trade at least twice to make our analysis to the standard repeat-sales analysis. One concern is that properties that trade less than twice are different, and that our sample is selected. Given the long time period covered by our data, this concern is probably a smaller concern. But, unlike the tradition repeat-sales analysis, our model can be estimated on all properties in the data, including those that trade only once. Obviously, these properties are not informative about the price appreciation directly, but including them may improve the estimates of the trade process. This roughly doubles the sample to over 140,000 propertis. Table 3 compares the coefficient on LTV in the trade equation across samples. In the larger sample, this coefficient is closer to zero, but still statistically highly significant. Another concern is that foreclosure sales may not be arms-length sales and that the resulting prices may not reflect the true market values. To investigate this concern, the reestimate our model after dropping foreclosures from the sample altogether. Again, Table 3 shows that the coefficient on LTV in the trade equation remains largely unchanged compared to the repeat-sales sample.

A Equivalent Constant Intensity

To interpret the economic magnitudes of the estimation variation in the trade-selection equation, it is useful to calculate the equivalent constant trade intensities. Assume that

[12] This effect may be confounded by the enactment of the 2005 Bankruptcy Abuse Prevention and Consumer Protection Act (BAPCPA), but it is unclear in which direction the coefficient would be affected.

trades follow a Poisson arrival process with constant intensity λ_i. The number of trades of property i up to time t is denoted $d_i(t)$ and is distributed

$$P[d_i(t) - d_i(t-1) = n] = \frac{\lambda_i^n \exp[-\lambda_i]}{n!} \text{ for } n \geq 0. \tag{12}$$

Under the Poisson specification, the probability of no trades from $t-1$ to t is

$$P[d_i(t) - d_i(t-1) = 0] = \exp[-\lambda_i]. \tag{13}$$

From the estimated coefficients in the trade process, the probability of no trade is

$$P[w_i(t) < 0] = \Phi[-W_i'(t)\alpha_0 - p_i(t)\alpha_p]. \tag{14}$$

Equating these two expressions of no-trade probabilities yields

$$\exp[-\lambda_i] = \Phi[-W_i'(t)\alpha_0 - p_i(t)\alpha_p], \tag{15}$$

or

$$\lambda_i = -\ln[\Phi(-W_i'(t)\alpha_0 - p_i(t)\alpha_p)]. \tag{16}$$

We denote λ_i the *constant equivalent intensity*. Its posterior distribution is straightforward to calculate from the posterior distribution of α, and as an intensity, it is more readily interpretable than the coefficients in the selection equation alone. Note that λ_i is only used for interpreting the economic magnitude of the estimated coefficients in the trade process. It is not a structural element of the model. Explicitly modeling the trade intensity as a continuous-time process depending on the LTV (or price) turns it into a *doubly stochastic Poisson process*, or Cox process (see Cox (1955)), which is beyond the present scope.

Figure 10 graphs the time series of λ_i evaluated at the median LTV across properties for Models A and F. The top panels present estimates from the specifications that combine normal trades and foreclosure sales. The bottom plots are from specifications that treat foreclosures separately. Not surprisingly, the top plots show slightly higher intensities, because they plot the intensities of either normal trades or foreclosure sales. Moving from the Model A in the left panels to Model F on the right ones, the intensity process becomes much more volatile. Model F allows the LTV coefficient to vary across sub periods, and it includes a larger number of explanatory variables, including mortgage rates and quarterly indicators to capture seasonality in trades, overall capturing more time variation in the process. Specifically, the seasonal adjustments results in “spiky” movements in the intensity.

Finally, to illustrate the economic magnitude of the estimated LTV coefficients, in addition to the trade intensity evaluated at the median LTV, the thin gray lines plot the intensities evaluated at the bottom and top quartiles of the LTV distribution. Moving from Model A to F, we see that the effect of LTV increases substantially. In some periods, moving from the top to the bottom quartile of the LTV distribution increases the trade intensity by a factor of two to three, which is a substantial economic effect.

B Foreclosure Behavior

Finally, Table 4, Panel B, presents the estimated coefficients for the foreclosure process for specifications with separate processes for foreclosure sales and normal trades. The corresponding coefficients for the trade process are in Panel A. Note that the coefficients in Panel A are largely similar to those from Table 2, which combined normal trades and foreclosure sales in the estimation. This follows from the relatively low number of foreclosures over most of the sample. Hence, normal trades dominate the sample of traded properties and estimates using just the normal trades are very similar to those that combine them with (relatively few) foreclosures sales.

Comparing Panels A and B in Table 4, the greatest difference between the trade and foreclosure processes is the LTV coefficient for the last sub period in Model F. In Panel A, Model F the coefficient on LTV in the last sub period is smaller compared to Table 2. Much of the increase in this coefficient from Table 2 appears to be caused by foreclosure sales among high LTV properties. This is confirmed in Table 4, Panel B where Model F ha a large positive coefficient on LTVs in the last period, confirming that those foreclosure sales were predominantly driven by properties with high LTVs. The negative coefficients during the previous periods are less intuitive, but these were periods with fewer foreclosure sales, so those may result from more atypical transactions.

Comparing Panels A and B in Table 4 more broadly, the two processes appear largely similar. In addition to the differences for log-LTV, the two processes appear to respond slightly differently to the property age, with older properties being more likely to wind up in foreclosure and less likely to trade in normal trades. Comparing Figures 3 and 4 shows that the indices are largely unchanged after including these processes separately, despite their differences.

VI Conclusion

The facts that property prices are only observed when the property trades and that appreciating properties trade more frequently create problems for traditional measures of house prices, loan-to-value (LTV) ratios, and trade- and foreclosure behavior.

First, since trades are endogenous, the sample of observed house prices is selected, potentially biasing standard price indices and measures of LTV distributions. Second, trade- and foreclosure behavior depend directly on current prices or LTVs, even though these variables are unobserved in the data. These econometric problems have not previously been formally addressed.

To address these problems, we present and estimate a new econometric model. It filters out the entire price path of each property in our data, even when the property is nontraded and its price is unobserved. This filtering is numerically intensive, but exploiting recent advances in computational Bayesian estimation — specifically, MCMC, Gibbs Sampling, and FFBS — produces a tractable estimation procedure with robust convergence properties. The model estimates prices jointly with the trade and foreclosure processes, which allows those decisions to depend directly on price or LTV, and which uses the trade- foreclosure processes as dynamic extensions of the standard Heckman sample selection process to correct the estimates for selection arising when the price dynamics for the traded properties for which prices are observed are not representative of the overall population.

This estimation procedure produces selection corrected estimates of price indices, where we find that the effect of selection is relatively minor. It produces estimates of the LTV distribution — to our knowledge, this is the first procedure to estimate LTVs from transaction prices — where we find that the effect of selection is substantial. The deterioration in the LTV distribution toward the end of our sample is large and unprecedented. Finally, our estimates of the coefficients for the trade- and selection processes suggest large changes in behavior towards the end of our sample.

We present results from transactions in Alameda county, CA. To confirm the robustness of our findings, we reestimate (but do not report) our model with data for Maricopa county, AZ (Phoenix and Scottsdale metro areas) and Clark county, NV (Las Vegas). We find qualitatively similar results, specifically the dramatic deterioration of the LTV distributions and the increase in the coefficients on LTVs toward the end of our sample appear to be robust across geographical locations.

Our sample ends in 2008, unfortunately, but the deterioration of the real estate markets have continued since then, making the problems with traditional estimators of prices, LTVs, and behavior even more apparent. In April, 2010, Fiserv, Inc. (which produces the S&P/Case-Shiller index) issued a press release stating that "the housing market has experienced significant turmoil and the last two-to-three years have seen large increases in foreclosures as well as other market dislocations." This statement was followed by a recommendation against using previous adjustments for seasonality, because they were no longer found to be working correctly. The most recent (June 2011) figures from Las Vegas show that volume of sales were up 8 percent since last year and up 16.7% of the previous month. But almost 70% of these sales were distress sales (47.2% of all sales were bank-owned properties and another 21.6% were short sales). Around 50% of all homes sold were purchased with cash. Clearly, both the press release and the recent Las Vegas figures show the need for ways to analyze distressed markets for illiquid assets.

Appendix: Estimation Procedure

For each property i the natural logarithm of price, $p_i(t)$, follows the process:

$$p_i(t) = p_i(t-1) + \delta(t) + \varepsilon_i(t), \tag{17}$$

with $\varepsilon_i(t) \sim N(0, \sigma^2)$ and i.i.d. across firms, $i = 1 \ldots N$, and across time, $t = 1 \ldots T$. Note that $\delta(t)$ is the index return between time $t-1$ and t, so $\delta(t)$ ranges from $\delta(2)$ to $\delta(T)$. The price is observed whenever $w_i(t) \geq 0$, where $w_i(t)$ is given by the selection equation

$$w_i(t) = W_i'(t)\alpha_0 + p_i(t)\alpha_p + \eta_i(t). \tag{18}$$

The vector of covariates $W_i(t)$ is observed.[13] In some models we separate out foreclosure sales by including an additional selection equation in which a foreclosure occurs when $z_i(t) \geq 0$, where $z(t)$ is

$$z_i(t) = Z_i'(t)\gamma_0 + p_i(t)\gamma_p + \xi_i(t). \tag{19}$$

The error terms η and ξ are distributed *i.i.d.* normal with variance equal to one, and are uncorrelated with each other at all leads and lags (including contemporaneously).

For some of the calculations in this appendix it is convenient to stack the selection equations into a 2x1 vector

$$s_i(t) = S_i'(t)\phi_0 + p_i(t)\phi_p + \zeta_i(t), \tag{20}$$

where $s_i(t) = \begin{bmatrix} w_i(t) \\ z_i(t) \end{bmatrix}$, $S_i'(t) = \begin{bmatrix} W_i'(t) & 0 \\ 0 & Z_i'(t) \end{bmatrix}$, $\phi_0 = \begin{bmatrix} \alpha_0 \\ \gamma_0 \end{bmatrix}$, $\phi_p = \begin{bmatrix} \alpha_p \\ \gamma_p \end{bmatrix}$, and $\zeta_i(t) = \begin{bmatrix} \eta_i(t) \\ \xi_i(t) \end{bmatrix}$. The covariance matrix of $\zeta_i(t)$ is the identity matrix. For the models in which we only use one selection equation, $s_i(t) = w_i(t)$ and equation (20) is identical to (18).

The set of parameters to be estimated is $\theta = \left(\delta(2) \ldots \delta(T), \sigma^2, \alpha_0, \alpha_p, \gamma_0, \gamma_p\right)$. We augment the parameter set with the latent variables and use a Bayesian estimation algorithm

[13] Explain how to get LTV into the selection equations: include in W and force its coefficient to equal α_p

that simulates the posterior distribution $f(\theta, \{p_i(t), s_i(t)\} | data)$ using a Gibbs sampler (Gelfand and Smith, 1990). By the Clifford-Hammersley theorem, we can break up the posterior into three *complete conditionals*:

1. Latent prices: $f(\{p_i(t)\} | \{s_i(t)\}, \theta, data)$

2. Selection variables: $f(\{s_i(t)\} | \{p_i(t)\}, \theta, data)$

3. Parameters: $f(\theta | \{p_i(t), s_i(t)\}, data)$

We sample from each distribution 1-3 in turn, after which we return back to step 1 and repeat. The resulting sequence of parameter draws forms a Markov chain, the stationary distribution of which is exactly the posterior distribution. Given a sample of draws of the posterior distribution it is then straightforward to numerically integrate out the latent variables and obtain the marginal posterior of parameters, $f(\theta | data)$, or the unobserved prices, $f(\{p_i(t)\} | data)$, for example. In the remainder of this appendix we discuss how to draw from each conditional distribution.

A1 Latent prices

We draw latent prices in the period between property sales using the Forward-Filter-Backward-Sample (FFBS) algorithm (Carter and Kohn (1994) and Fruhwirth-Schnatter (1994)), which provides an efficient way to sample a path of state variables defined by a linear state space model. Since the error terms are assumed i.i.d. across firms, we can sample $p_i(t)$ separately for each firm. For expositional simplicity we describe the algorithm for a particular firm, suppressing the dependence on i.

Interpreting the econometric model as a linear state space model, $p(t)$ is the state variable, and equation (17) is the transition rule. Conditional on the parameters, $\delta(t)$ is an "observed" control acting on the state, and conditional on $s(t)$, the selection equations (20) are noisy observations equation for the state. This setup allows us to calculate the filtered distribution of $p(1) \dots p(T)$, using the Kalman filter.

The Kalman filter produces the distribution of $p(t)$ conditional on $s(1) \dots s(t)$, for any time t. However, $p(t)$ needs to be sampled conditional on the entire time series $s(1) \dots s(T)$.

This is achieved by a backward smoother, which effectively runs a Kalman filter backwards, starting at time T. The conditional distribution of the state vector of latent valuations is given by the identity (Lemma 2.1 in Carter and Kohn (1994))

$$f\left(p(1)\dots p(T)|s^T\right) = f\left(p(T)|s^T\right)\prod_{t=1}^{T-1} f\left(p(t)|s^t, p(t+1)\right), \tag{21}$$

where $s^t = \{s(1)\dots s(t)\}$ contains the selection variables up to time t. Next we describe the forward filtering and backward sampling steps in detail.

Define $m(t|j) = E\left[p(t)|s^j\right]$ and $v(t|j) = Var\left[p(t)|s^j\right]$ as the mean and variance of $p(t)$ conditional on the selection variables up to time j. Note that all conditional distributions are Normal and hence fully characterized by their means and variances (see Kalman (1960) and Anderson and Moore (1979)).

For the forward filtering step, for $t = 1\dots T$, we calculate $m(t|t)$ and $v(t|t)$ by iterating on the forward filter, through a forecasting and an updating part. The forecasting part involves the two equations

$$m(t|t-1) = m(t-1|t-1) + \delta(t), \tag{22}$$

and

$$v(t|t-1) = v(t-1|t-1) + \sigma^2. \tag{23}$$

For the updating part, as long as $p(t)$ remains unobserved, we update

$$m(t|t) = m(t|t-1) + K' \cdot \left[s(t) - S'(t)\phi_0 - m(t|t-1)\phi_p\right], \tag{24}$$

where the Kalman gain K is given by

$$K = \left[I + \phi_p v(t|t-1)\phi_p'\right]^{-1} \cdot v(t|t-1)\phi_p, \tag{25}$$

and I is the 2x2 identity matrix (in the case of one selection equation this is simply a scalar unity). When K is large, more weight is placed on the information from the selection equation. This happens when either ϕ_p or $v(t|t-1)$ is large, i.e. when either the selection equations are more informative about the valuations or when the valuations are more uncertain. Further,

$$v(t|t) = v(t|t-1) \cdot \left(1 - K' \cdot \phi_p\right). \tag{26}$$

To estimate the model without correcting for selection, we force $\phi_p = 0$. Then $m(t|t) = m(t|t-1)$ and $v(t|t) = v(t|t-1)$, and no information is used in periods where $p(t)$ is unobserved. In periods where $p(t)$ is observed, $m(t|t) = p_t^{OBS}$ and $v(t|t) = 0$.

For the backward sampling part, $p(T)$ is first simulated from the Normal distribution with mean $m(T|T)$ and variance $v(T|T)$, as given by the Kalman filter. For $t = T-1\ldots 1$, we draw $p(t)$ from the conditional distribution $p(t)|s^t, p(t+1)$. This distribution can be derived from a filtering problem where the draw of $p(t+1)$ provides an additional observation of $p(t)$. The distribution is

$$p(t)|s^t, p(t+1) \sim \mathcal{N}(r, q), \tag{27}$$

where

$$r = m(t|t) + G \cdot [p(t+1) - m(t+1|t)], \tag{28}$$

$$q = v(t|t) \cdot (1 - G), \tag{29}$$

with

$$G = \frac{v(t|t)}{v(t|t) + \sigma^2}. \tag{30}$$

From equation (30), G can be interpreted as a Kalman gain similar to K in equation (25). As such, the backwards sampler weighs the information from the filtered distribution $p(t)|s^t$ and the information in $p(t+1)|s^T$ to obtain a draw of $p(t)|s^T$, with the weight depending on the relative variance of the filtered estimate, $v(t|t)$, and the variance of a one-period price change. If the filtered estimate $m(t|t)$ is very precise relative to the variance of the valuation change from one period to the next, then G is close to zero, and most weight is put on the distribution of $p(t)$ from the Kalman filter. The more imprecise the Kalman filter distribution relative to how much the valuation can possibly change (as captured by sigma), the more weight is put on the "observed" $p(t+1)$.

A2 Selection variables

The selection variables are sampled conditional on the valuations, parameters, and whether the valuation is observed or not. Simulating this block is similar to simulating the (augmented) posterior distribution of a probit model (Albert and Chib (1993)). Under the assumption that η and ξ are independent we may draw the selection variables, w and z, separately. In addition, by the i.i.d. assumption we may draw each property-quarter variable separately.

When homes are sold, the home price is observed and the posterior distribution of the first selection variable, $w_i(t)$, is

$$w_i(t)|\{p_i(t)\},\theta,data \sim \mathcal{N}_L\left(W_i'(t)\alpha_0 + p_i(t)\alpha_p, 1\right). \tag{31}$$

When home price is unobserved, it is

$$w_i(t)|\{p_i(t)\},\theta,data \sim \mathcal{N}_U\left(W_i'(t)\alpha_0 + p_i(t)\alpha_p, 1\right). \tag{32}$$

Here, $\mathcal{N}_L\left(\mu,\sigma^2\right)$ denotes a lower-truncated Normal distribution with mean μ and variance σ^2, truncated at zero from below. Similarly, $\mathcal{N}_U\left(\mu,\sigma^2\right)$ is the upper-truncated distribution, truncated at zero from above.

Drawing $z_i(t)$ is analogous but using foreclosures instead of regular home sales for observations.

A3 Parameters

Conditional on $\{p_i(t)\}$, $\{W_i(t)\}$ and $\{Z_i(t)\}$ the distributions of α, γ, $\{\delta(t)\}$ and σ^2 are given by the three Bayesian linear regressions (17), (18) and (19). Since ε, η and ξ are independent by assumption, we may estimate the three equations separately.

In the valuation equation, $\delta = [\delta(2)\ldots\delta(T)]'$, and σ^2 are defined by the regression of Y_p on X_p, where the vector Y_p stacks the one-period returns, $p_i(t) - p_i(t-1)$, across all properties and time periods. Let $N(t)$ be the number of companies for which $p(t) - p(t-1)$ exists, so Y_p is a $\sum_{t=2}^{T} N(t)$ by 1 vector. The matrix X_p is a $\sum_{t=2}^{T} N(t)$ by $T-1$

matrix of zeros and ones. Each row of X_p contains $T-2$ zeros and a one in column $t-1$, corresponding to the timing of the return in Y_p (such that a one in the first column of X_p indicates a return from time 1 to time 2).

The standard conjugate Normal-Inverse Gamma prior with prior parameters a_0,b_0, μ_0, and Σ_0 is

$$\sigma^2 \sim IG(a_0, b_0), \tag{33}$$
$$\delta|\sigma^2 \sim \mathcal{N}\left(\mu_0, \sigma^2\Sigma_0^{-1}\right). \tag{34}$$

The posterior distributions for the parameters in the valuation equation are then (e.g. Rossi, Allenby, and McCulloch (2005)):

$$\sigma^2|Y_p, X_p \sim IG(a, b), \tag{35}$$
$$\delta|\sigma^2 \sim \mathcal{N}\left(\mu, \sigma^2\right), \tag{36}$$

with parameters

$$a = a_0 + \sum_{t=2}^{T} N(t), \tag{37}$$
$$b = b_0 + e'e + (\mu - \mu_0)' \Sigma_0 (\mu - \mu_0), \tag{38}$$
$$\Sigma = \Sigma_0 + X_p' X_p, \tag{39}$$
$$\mu = \Sigma^{-1}\left(\Sigma_0\mu_0 + X_p' Y_p\right). \tag{40}$$

The vector $e = Y_p - X_p\mu$ contains the stacked error terms.

The above regression pools the entire panel data set and quickly leads to memory issues and slow computation speeds. For example, in a data set with 100,000 properties observed for 80 quarters, the X_p matrix is 8 million by 79. The solution to these problems is to exploit the unique structure of X_p. In particular, $X_p'X_p$ is a diagonal $T-1$ by $T-1$ matrix that contains the number of trades on the diagonal. With a diagonal prior Σ_0 the inverse $\Sigma^{-1}$is then simply a diagonal matrix with the inverse of each element of Σon the diagonal. The $T-1$ by 1 vector $X_p'Y_p$ contains the sum of returns for each period. These quantities can be efficiently computed and plugged into (37)-(40).

The selection equations are considerably simpler. To obtain a draw of $\alpha = \begin{bmatrix} \alpha_0 \\ \alpha_p \end{bmatrix}$, we regress Y_w on X_w, where Y_w is a vector that stacks $w_i(t)$ over all properties and time periods. Similarly, X_w stacks $\begin{bmatrix} W_i'(t) & p_i(t) \end{bmatrix}$ over all properties and periods. Recall that we normalize the variance of the error term to one in order to identify the scale of the parameters, and we can consequently treat the inference problem as a standard Bayesian regression with known variance. The prior distribution is

$$\alpha \sim \mathcal{N}\left(\theta_0, \Omega_0^{-1}\right), \tag{41}$$

and the posterior becomes

$$\alpha | Y_w, X_w \sim \mathcal{N}\left(\theta, \Omega^{-1}\right), \tag{42}$$

with

$$\Omega = \Omega_0 + X_w' X_w, \tag{43}$$

and

$$\theta = \Omega^{-1}\left(\Omega_0 \theta_0 + X_w' Y_w\right). \tag{44}$$

Drawing γ works analogously to drawing α.

References

Albert, James and Siddhartha Chib, 1993, Bayesian Analysis of Binary and Polychotomous Response Data, *Journal of the American Statistical Association* 88, 669-679.

Bailey, Martin J., Richard F. Muth, and Hugh O. Nourse, 1963, A Regression Method for Real Estate Price Index Construction, *Journal of the American Statistical Association* 58, 933-942.

Bajari, Patrick, Sean Chu, and Minjung Park, 2010, An Empirical Model of Subprime Mortgage Default from 2000 to 2007, *Working Paper*, University of Minnesota.

Bhutta, Neil, Jane Dokko, and Hui Shan, 2010, The Depth of Negative Equity and Mortgage Default Decisions , *Working Paper*, Federal Reserve Board.

Calnea Analytics, 2010, Land Registry House Price Index Methodology: Discussion and Description of Methodology, http://www.calnea.net/LandRegistryIndex.pdf, downloaded March 3, 2010.

Carter, Chris K. and Robert J. Kohn, 1994, On Gibbs Sampling for State Space Models, *Biometrika* 81, 541-553.

Case, Karl E., Henry O. Pollakowski, and Susan M. Wachter, 1997, Frequency of Transaction and House Price Modeling, *Journal of Real Estate Finance and Economics* 14, 173-187.

Case, Karl E. and Robert J. Shiller, 1987, Prices of Single Family Homes since 1970: New Indexes for Four Cities, *New England Economic Review* Sept/Oct, 46-56.

Case, Karl E. and Robert J. Shiller, 1989, The Efficiency of the Market for Single-Family Homes, *American Economic Review* 79, 125-137.

Cochrane, John, 2005, The risk and return of venture capital , *Journal of Financial Economics* 75, 3-52.

Cox, John C., Jonathan E. Ingersoll and Stephen A. Ross, 1985, A Theory of the Term Structure of Interest Rates, *Econometrica* 53, 385-407.

Cox, D. R., 1955, Some statistical methods connected with series of events, *Journal of the Royal Statistical Society, Series B* 17, 129-164.

Downing, Chris, Richard Stanton and Nancy Wallace, 2005, An Empirical Test of a Two-Factor Mortgage Valuation Model: How Much do House Prices Matter?, *Real Estate*

Economics 33, 681-710.

Ferreira, Fernando, Joseph Gyourko and Joseph Tracy, 2010, Housing Buses and Household Mobility, *Journal of Urban Economics* 68, 34-45.

Fruhwirth-Schnatter, Sylvia, 1994, Data Augmentation and Dynamic Linear Models, *Journal of Time Series Analysis* 15, 183-202.

Gatzlaff, Dean H. and Donald R. Haurin, 1997, Sample Selection Bias and Repeat-Sales Index Estimates, *Journal of Real Estate Finance and Economics* 14, 33-50.

Gatzlaff, Dean H. and Donald R. Haurin, 1998, Sample Selection and Biases in Local House Value Indices, *Journal of Urban Economics* 43, 199-222.

Gelfand, Alan and Adrian Smith, 1990, Sampling Based Approaches to Calculating Marginal Densities, *Journal of the American Statistical Association* 85, 398-409.

Geman, Stuart and Donald Geman, 1984, Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 721-741.

Genovese, David and Christopher Mayer, 2001, Loss Aversion and Seller Behavior: Evidence from the Housing Market, *Quarterly Journal of Economics* 116, 1233-1260.

Goetzmann, William N., 1992, The Accuracy of Real Estate Indices: Repeat Sale Estimators, *Journal of Real Estate Finance and Economics* 5, 5-53.

Goetzmann, William N., 1993, Accounting for Taste: Art and the Financial Markets over Three Centuries, *American Economic Review* 83, 1370-1376.

Goetzmann, William N. and Liang Peng, 2006, Estimating House Price Indexes in the Presence of Seller Reservation Prices, *Review of Economics and Statistics* 88, 100-112.

Goetzmann, William N. and Matthew Spiegel, 1995, Non-Temporal Components of Residential Real Estate Appreciation, *The Review of Economics and Statistics* 77, 199-206.

Guiso, Luigi, Paola Sapienza, Luigi Zingales, 2009, Moral and Social Constraints to Strategic Default on Mortgages, *Working Paper*, University of Chicago.

Haurin, Donald R. and Patric H. Hendershott, 1991, House Price Indexes: Issues and Results, *AREUEA Journal* 19, 259-269.

Heckman, James, 1976, The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models, *Annals of Economic and Social Measurement* 5, 475-492.

Heckman, James, 1979, Sample Selection Bias as a Specification Error, *Econometrica* 47, 153-162.

Heckman, James, 1990, Varieties of Selection Bias, *American Economic Review* 80, 313-318.

Himmelberg, Charles, Chris Mayer, Todd Sinai, 2005, Assessing High Hourse Prices: Bubbles, Fundamentals, and Misperceptions, *Journal of Economic Perspectives* 19, 67-92.

Hwang, Min and John M. Quigley, 2004, Selectivity, Quality Adjustment and Mean Reversion in the Measurement of House Values, *Journal of Real Estate Finance and Economics* 28, 161-178.

Johannes, Michael and Nick Polson, 2006, MCMC Methods for Financial Econometrics, in Yacine Ait-Sahalia and Lars P. Hansen, eds.: Handbook of Financial Econometrics.

Jud, G. Donald, and Terry G. Seaks, 1994, Sample Selection Bias in Estimating Housing Sales Prices, *The Journal of Real Estate Research* 9, 289-298.

Kain, John F. and John M. Quigley, 1970, Measuring the Value of Housing Quality, *Journal of the American Statistical Association* 65, 532-548.

Korteweg, Arthur G., and Morten Sørensen, 2010, Risk and return characteristics of venture capital-backed entrepreneurial companies, *Review of Financial Studies* , Forthcoming.

Laffont, Jean-Jacques, Herve Ossard, Quang Vuong, 1995, Econometrics of First-Price Auctions, *Econometrica* 63, 953-980.

Landvoigt, Tim, Monika Piazzesi, Martin Schneider, 2010, The Housing Market(s) of San Diego, *Working Paper*, Stanford University.

Lin, Zhenguo, and Kerry D. Vandell, 2007, Illiquidity and Pricing Biases in the Real Estate Market, *Real Estate Economics* 35, 291-330.

Meese, Richard A. and Nancy E. Wallace, 1997, The Construction of Residential Housing Price Indices: A Comparison of Repeat-Sales, Hedonic-Regression, and Hybrid Approaches, *Journal of Real Estate Finance and Economics* 14, 51-73.

Melzer, Brian, 2011, Mortgage Debt Overhang: Reduced Investment by Homeowners with Negative Equity, *Working Paper*, Northwestern University.

Munneke, Henry J. and Barrett A. Slade, 2000, An Empirical Study of Sample-Selection Bias in Indices of Commercial Real Estate, *Journal of Real Estate Finance and Economics* 21, 45-64.

Munneke, Henry J. and Barrett A. Slade, 2001, A Metropolitan Transaction-Based Commercial Price Index: A Time-Varying Parameter Approach, *Real Estate Economics* 29, 55-84.

Peng, Liang, 2001, Building a Venture Capital Index, *Working Paper*, University of Boulder, Colorado.

Rossi, Peter E., Greg M. Allenby and Robert McCullough, 2005, Bayesian Statistics and Marketing, John Wiley and Sons, Chichester, UK.

Schulhofer-Wohl, Sam, 2011, Negative Equity Does Not Reduce Homeowners' Mobility, *Working Paper*, Federal Reserve Bank of Minneapolis.

Tanner, Martin and Wing Wong, 1987, The Calculation of Posterior Distributions by Data Augmentation, *Journal of the American Statistical Association* 82, 528-549.

Webb, Cary, 1988, A Probabilistic Model for Price Levels in Discontinuous Markets, In Wolfgang Eichhorn, ed., Measurement in Economics: Theory and Applications of Economic Indices, Springer-Verlag, New York.

Figure 1: Descriptive Statistics of Home Sales. The figure presents descriptive statistics of our data for Alameda county, CA. Panel A contains the total number of home sales, separating regular transactions and foreclosure sales. Panel B presents foreclosures as a fraction of all sales. Panel C and D present regular transactions and foreclosure sales as a fraction of the total number of properties the data.

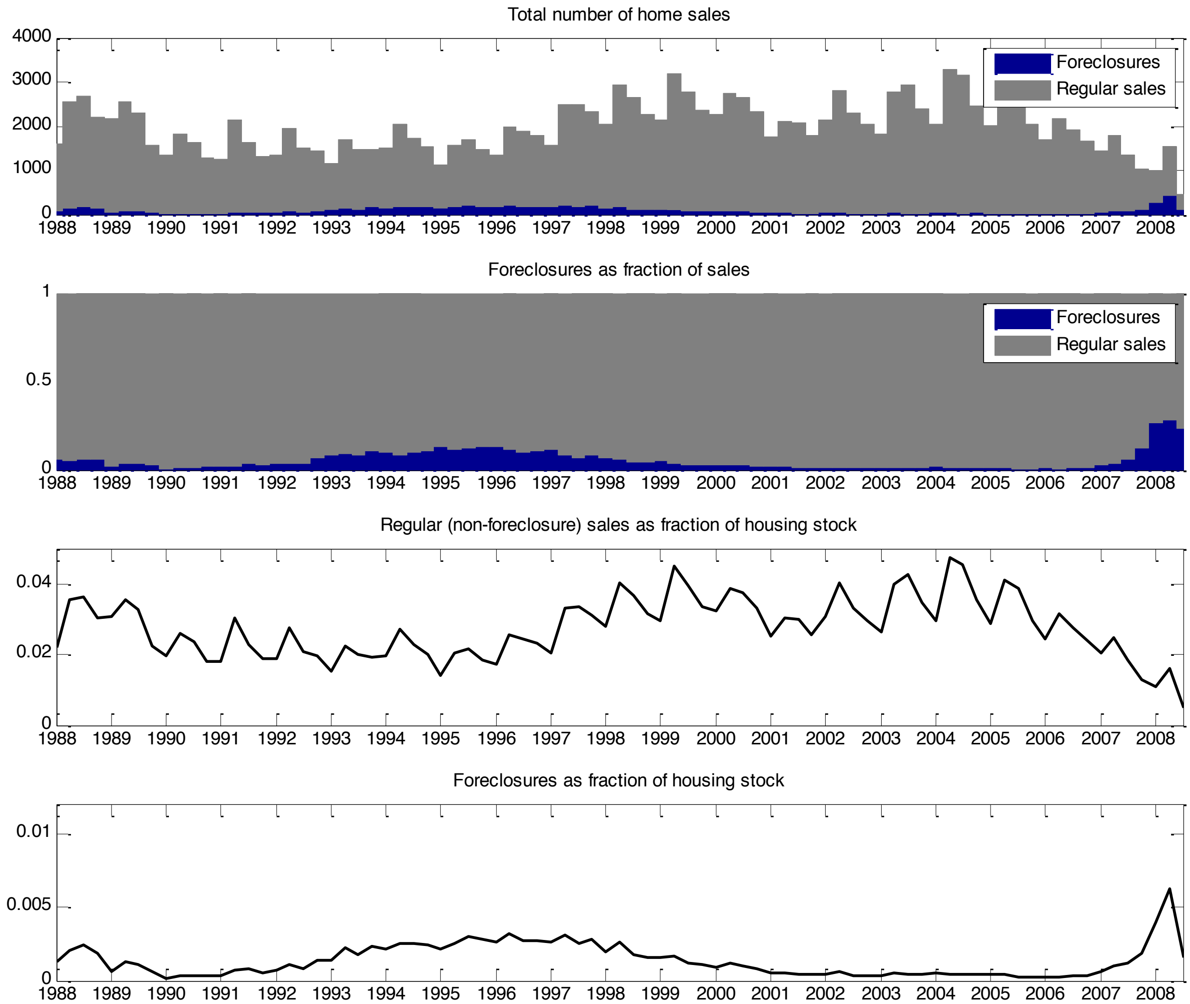


Figure 2: Price Indices without Trade-Selection Process. The figure shows house price indices for Alameda county, CA, without adjustment for sample selection. The indices are normalized to equal 100 in the first quarter of 2000. GLS indices are estimated using repeat-sales regressions using generalized least squares with weights proportional to the square root of the time between sales. MCMC indices are estimated using the Bayesian procedure described in the paper (without including the sales or foreclosure processes). Arithmetic indices are calculated with the 1/2 σ^2 adjustment discussed in the text. The other indices are calculated without this adjustment.

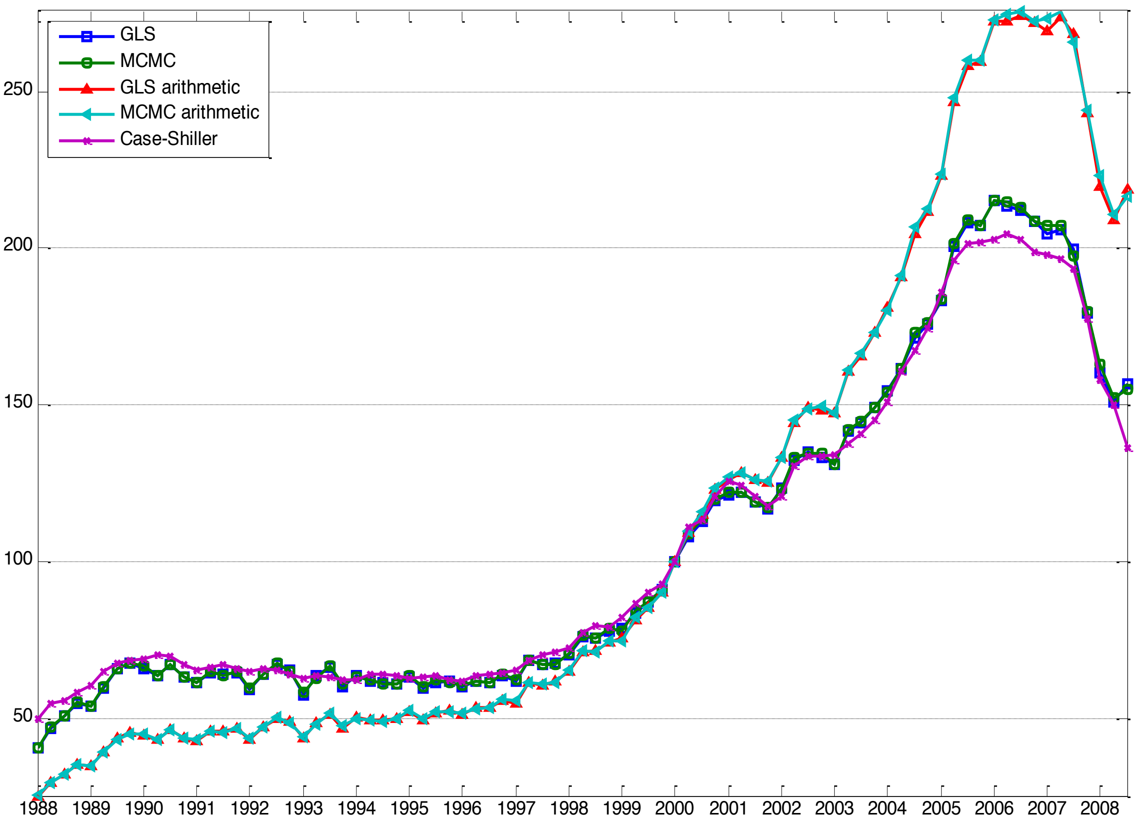
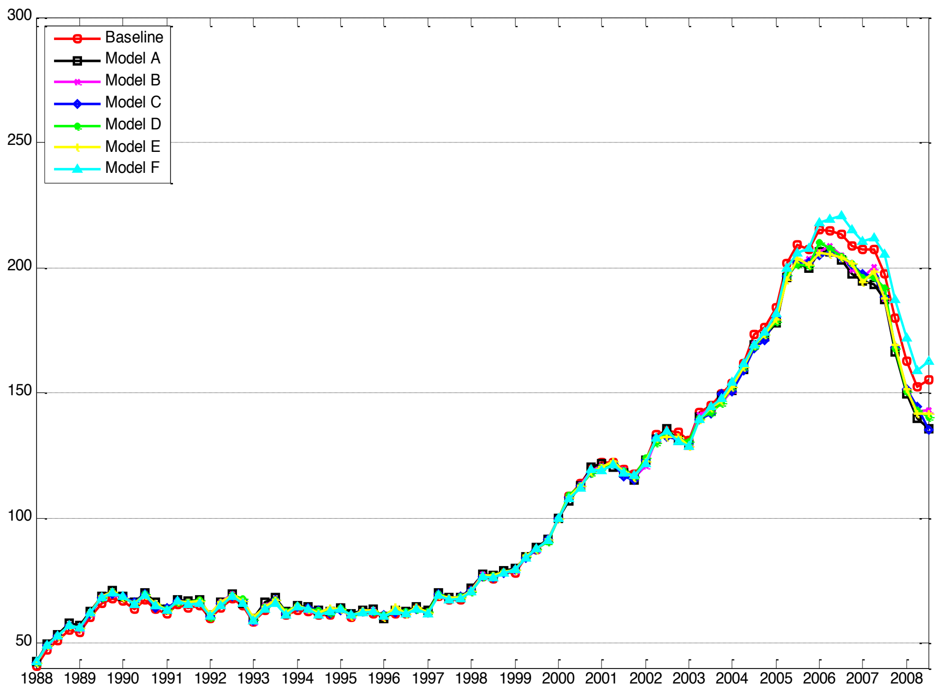


Figure 3: Price Indices with Trade-Selection Process. The figure shows house price indices for Alameda County, CA, after adjusting for sample selection. The indices are normalized to equal 100 in the first quarter of 2000. All indices are estimated using the Bayesian procedure described in the paper. The baseline index is calculated without adjusting for sample selection. The remaining indices are calculated using four different specifications of the trade process. The bottom plot presents arithmetic indices, calculated with the $1/2\ \sigma^2$ adjustment discussed in the text. The top graph presents indices without this adjustment.



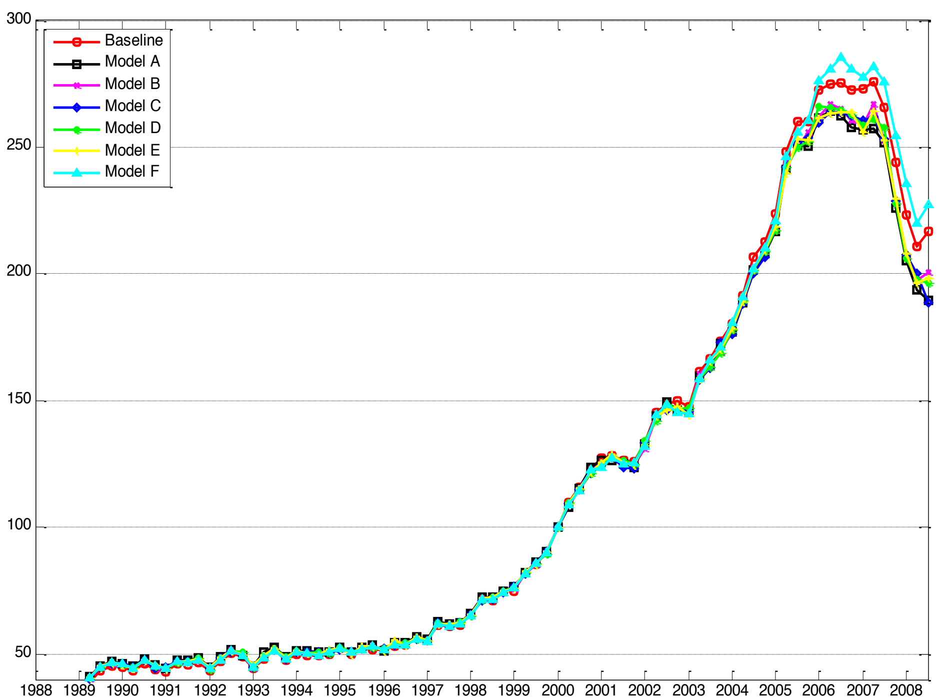


Figure 4: Price Indices with Separate Trade- and Foreclosure-Selection Processes. The figure shows house price indices for Alameda County, CA, after adjusting for sample selection. The indices are normalized to equal 100 in the first quarter of 2000. All indices are estimated using the Bayesian procedure described in the paper. The baseline index is calculated without adjusting for sample selection. The other indices are calculated using four different specifications of the trade process. The bottom plot presents arithmetic indices, calculated with the 1/2 σ^2 adjustment discussed in the text. The top graph presents indices without this adjustment.

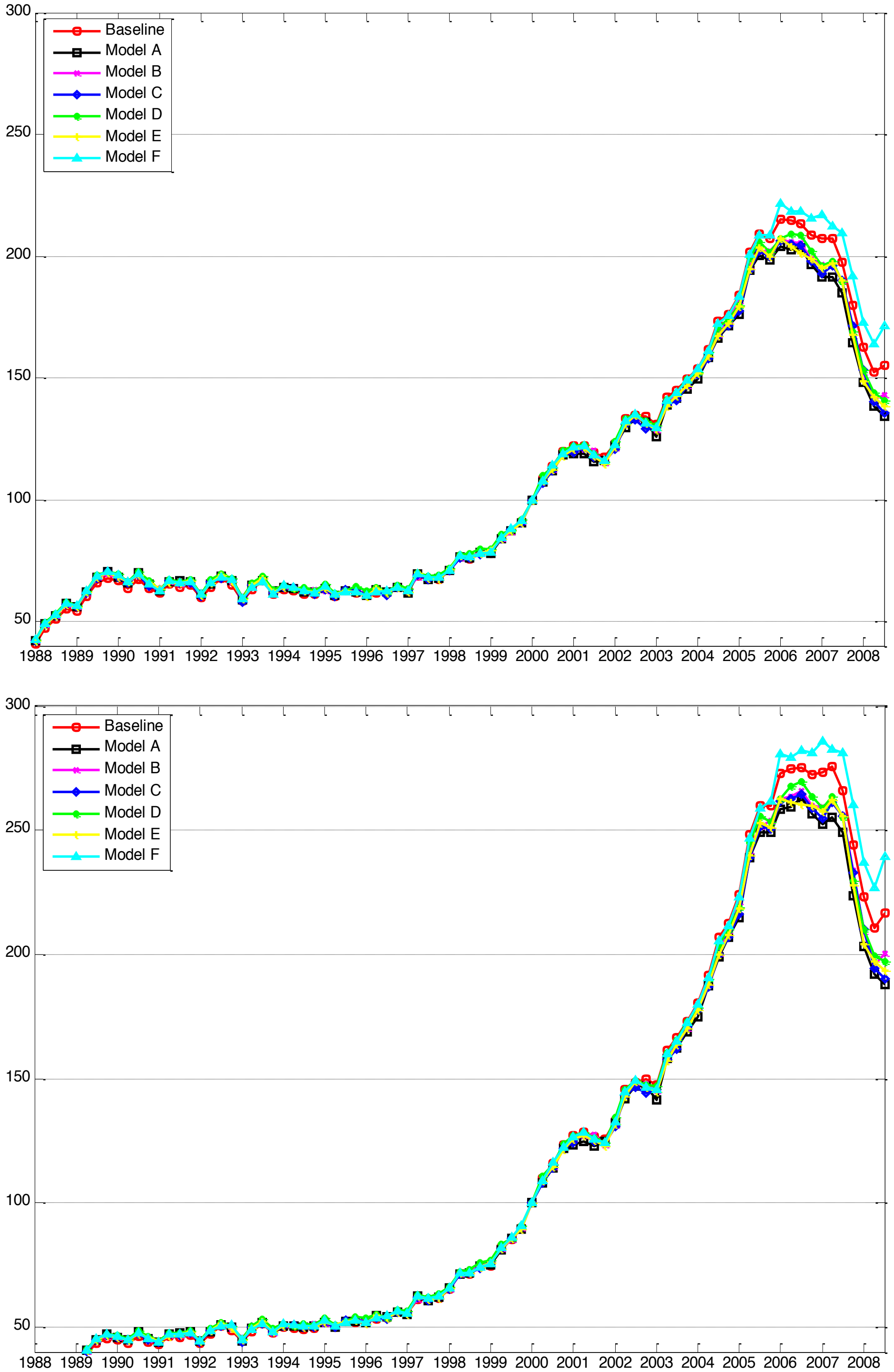


Figure 5: Distribution of Loan-to-value (LTV) at sale. The figure contains descriptive statistics of the buyers and sellers' LTVs as reported at the date of a property sales. The top plot shows the histogram of LTVs where the loan amount represents the buyer's mortgage amount. The bottom plot uses the seller's remaining mortgage balance (computed as described in the text) to calculate LTVs.

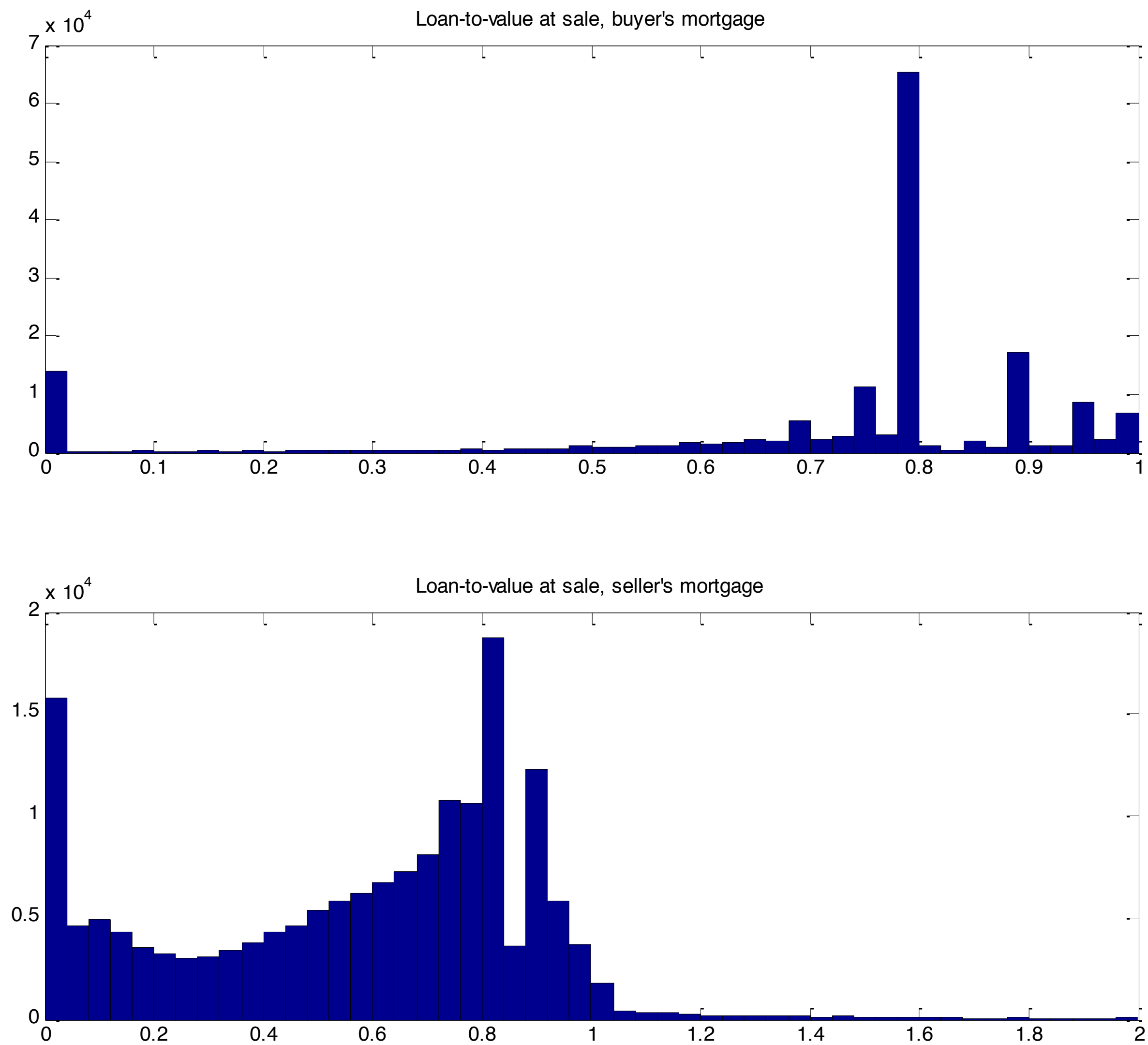


Figure 6: Historical LTVs, Foreclosures, and Fractions of Underwater Properties. The figures report the time series of median observed LTVs (left-hand side scale), and the fraction of the housing stock that sold (top plot, right-hand scale) foreclosures as a fraction of total sales (middle plot, right-hand scale), and the fraction of sold properties that are underwater at the time of sale (bottom plot, right-hand scale).

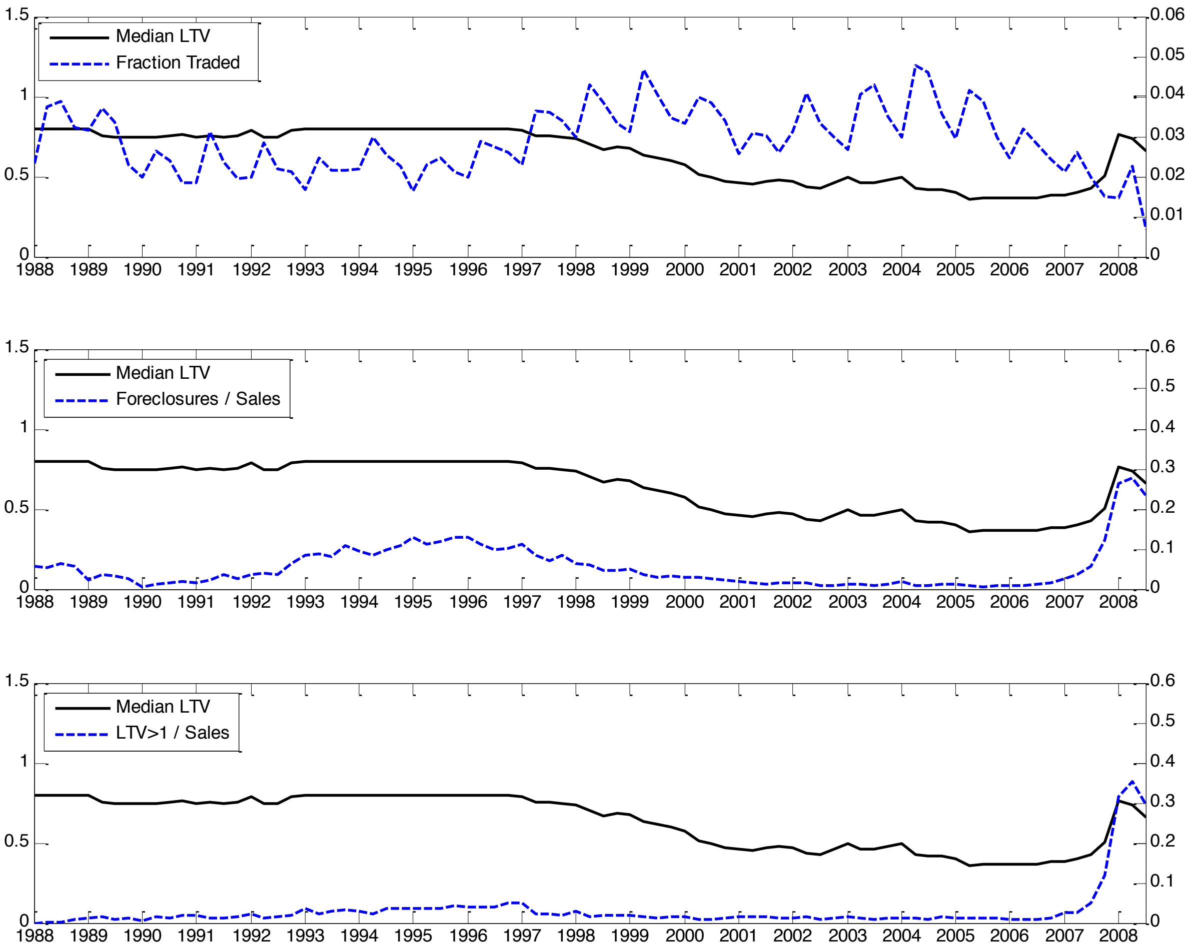
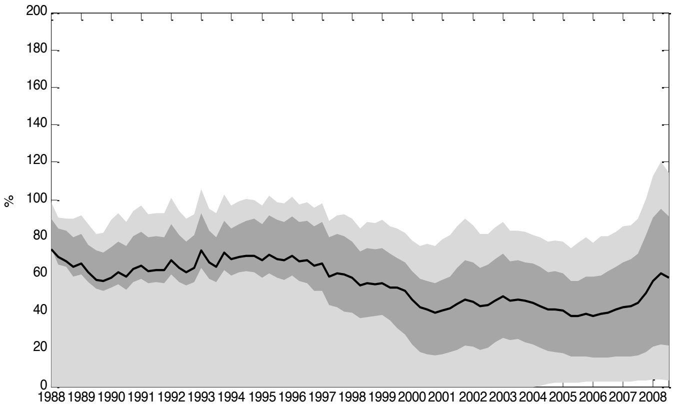
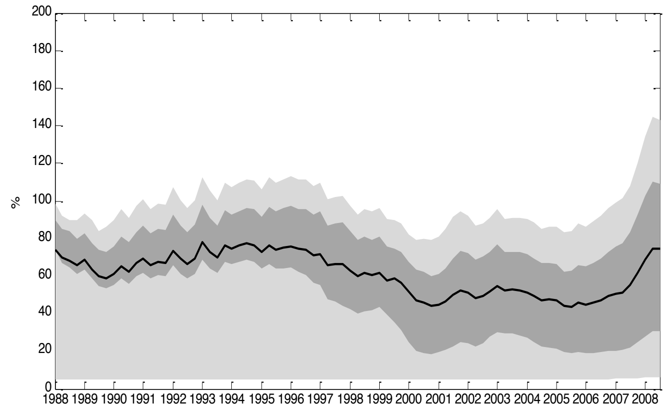


Figure 7: LTV Percentiles with Trade-Selection Correction.
Time-series plots of the LTV ratio for the standard repeat-sales (RSR, top-left), the baseline (top-right), and various selection models (See Tables 2 and 4 for model specifications). The light shaded band shows the (5%, 95%) LTV bound, the dark shaded band show the (25%,75%) LTV bound.

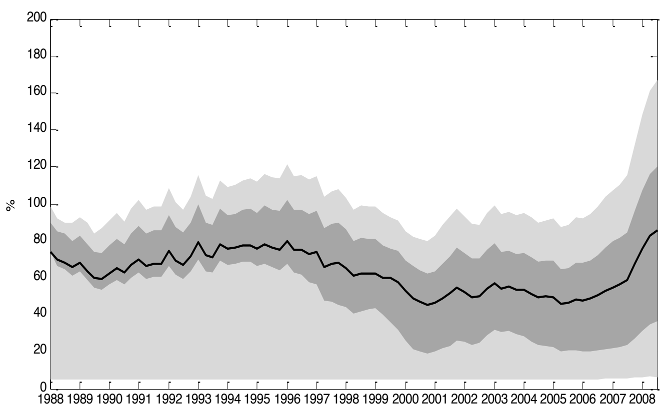
RSR:



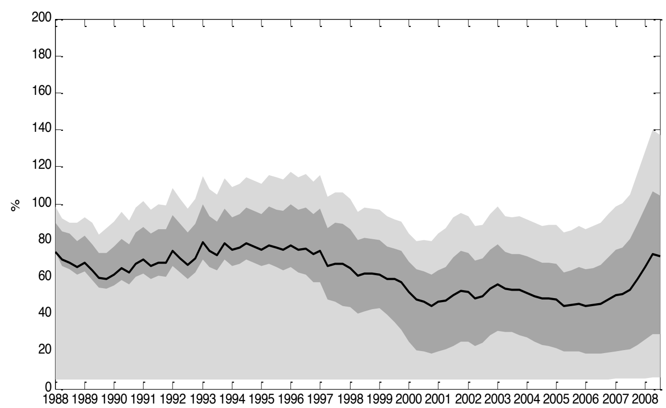
Baseline:



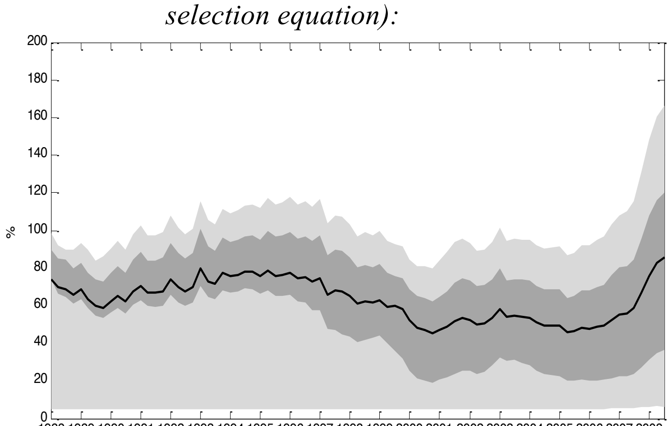
Model A (trade-selection equation):



Model F (trade-selection equation):



Model A (trade and foreclosure selection equation):



Model F (trade and foreclosure selection equation):

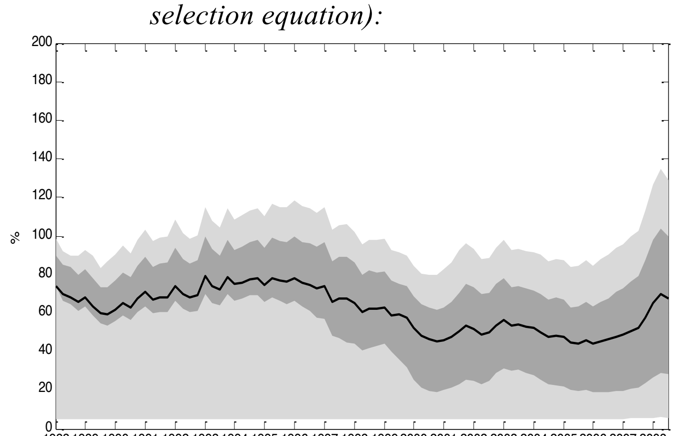


Figure 8: Distribution of Underwater Properties with Trade-Selection Process.
Time-series plots of the percentage of properties (vertical axis) that are underwater, as estimated by standard repeat-sales (RSR), our baseline model and two selection-corrected models (as described in Table 2). We show the distribution for several degrees of "underwaterness": LTV > 100% (top plot), LTV > 125% (middle plot), and LTV > 150% (bottom plot).

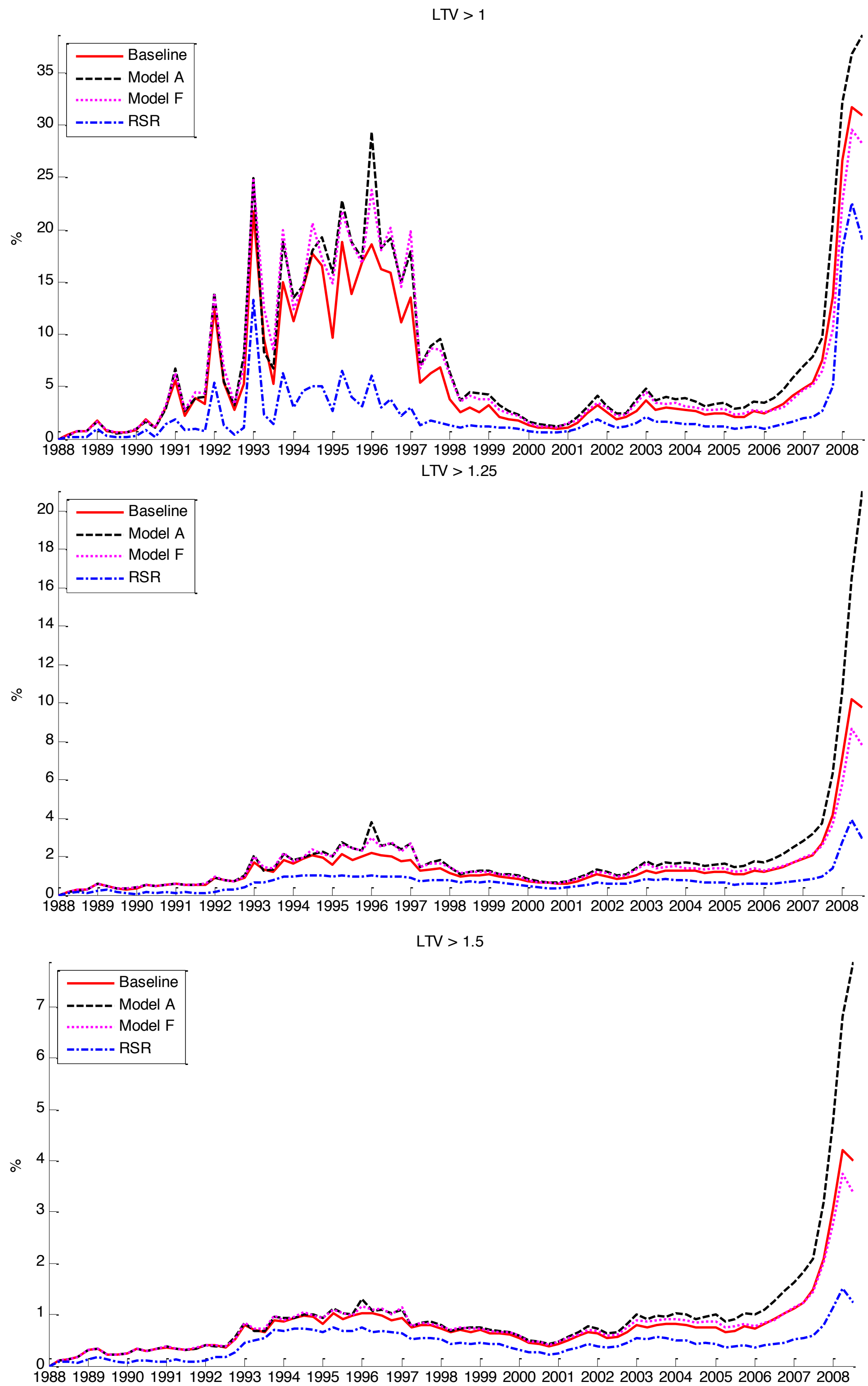


Figure 9: Distribution of Underwater Properties with Separate Trade- and Foreclosure Selection Processes.

Time-series plots of the percentage of properties (vertical axis) that are underwater, as estimated by standard repeat-sales (RSR), our baseline model and two selection-corrected models (as described in Table 4). We show the distribution for several degrees of “underwaterness”: LTV > 100% (top plot), LTV > 125% (middle plot), and LTV > 150% (bottom plot).

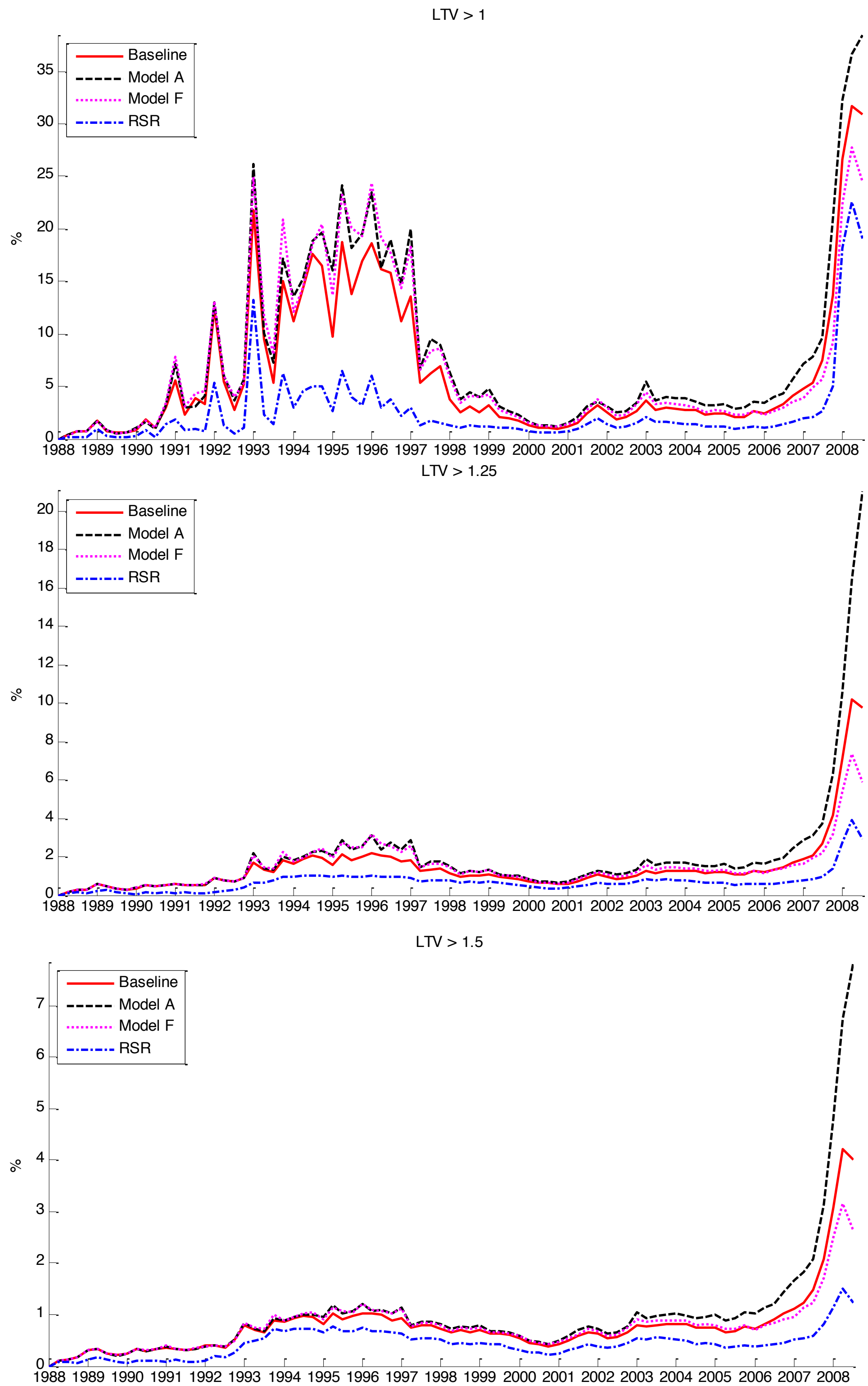
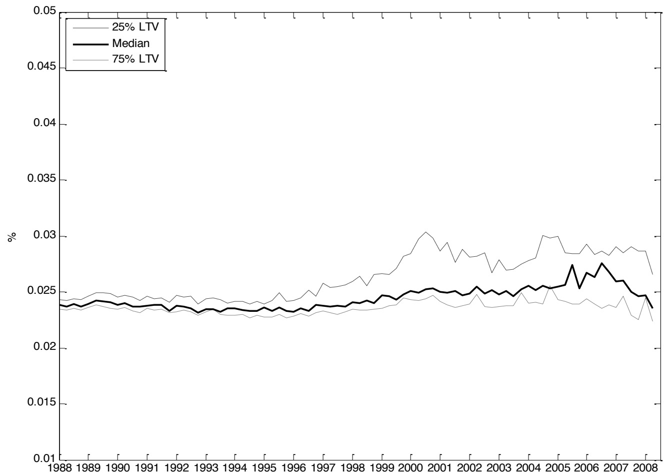
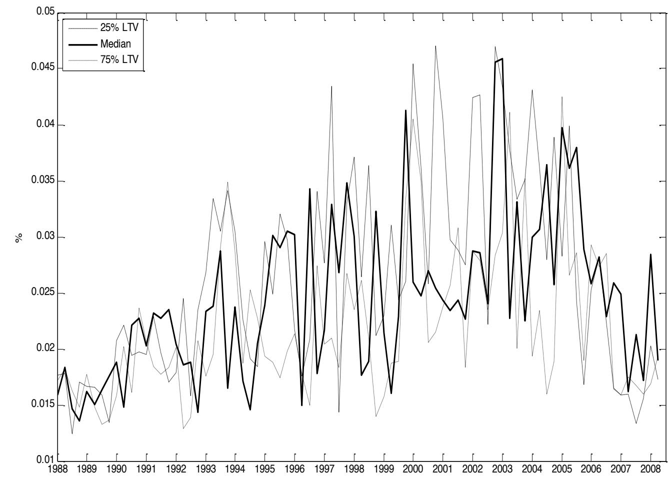


Figure 10: Equivalent Constant Trade Intensities
The time-series of trade intensities (computed as described in the text) at the median LTV ratio in each quarter, as well as at the 25th and 75th percentiles. The two top plots refer to models from in Table 2 with one trade-selection process. The bottom two plots refer to models from Table 4 with separate trade- and a foreclosure processes.
Artie,

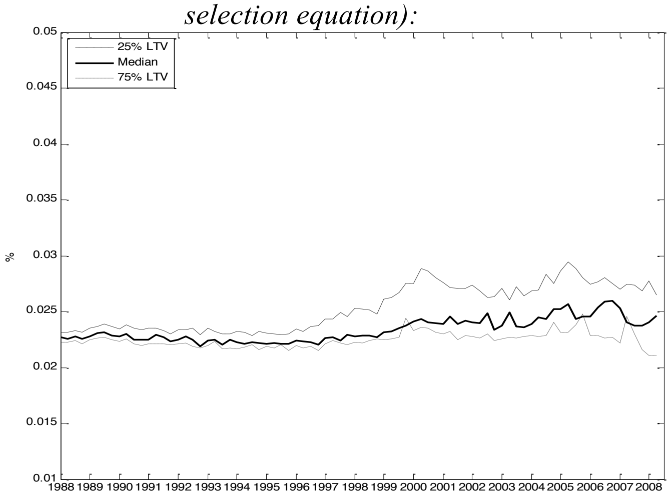
Model A (trade-selection equation):



Model F (trade-selection equation):



Model A (trade and foreclosure selection equation):



Model F (trade and foreclosure selection equation):

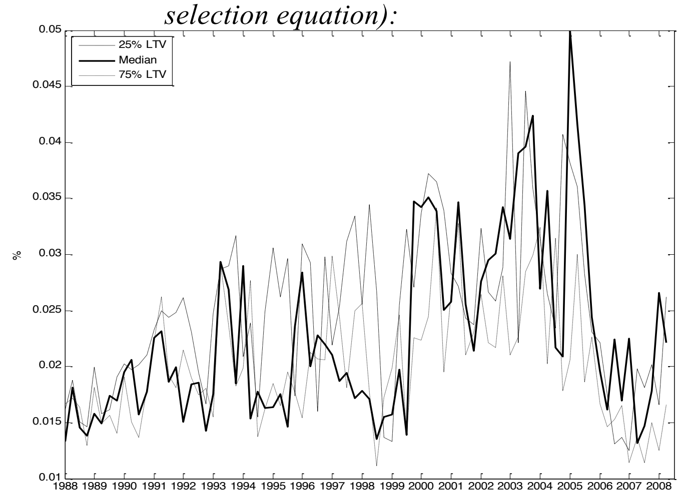


Table 1: Summary Statistics. The repeat-sales sample contains single-family residences in Alameda County, CA, over the period 1988Q1- 2008Q3. Foreclosure rate is the number of foreclosures as a percentage of all properties in the sample.

Panel A: Trade Data

	Mean	St.Dev	Min	Q10	Q50	Q90	Max
Number of prop.	68,700						
Number of trades	164,824						
per property	2.40	0.85	1	2	2	3	10
Loan-to-value ratio							
buyer	0.71	0.26	0.00	0.24	0.80	0.95	1.00
seller	0.58	0.48	0.00	0.05	0.65	0.90	68.79
Sale-to-sale return (%)							
arithmetic	64.44	79.34	-86.40	-4.90	42.26	167.52	500.00
log	0.40	0.45	-2.00	-0.05	0.35	0.98	1.79
Time between sales (years)	5.15	3.71	0.25	1.00	4.25	10.50	20.50
Foreclosure rate (%)	0.13						

Panel B: Property Characteristics

	Mean	St.Dev	Min	Q10	Q50	Q90	Max
Acres	1.34	0.68	0.03	0.62	1.24	2.12	5.00
Space (000s sqft)	1.66	0.64	0.34	1.00	1.52	2.52	7.53
Bedrooms (#)	3.15	0.88	1	2	3	4	9
Bathrooms (#)	2.17	0.85	1	1	2	3	5
Total rooms (#)	6.50	1.58	2	5	6	9	15
Basement	0.03						
Garage	0.90						
Fireplace	0.33						
Pool	0.06						
Stories	1.31	0.39	1	1	1	2	3
Construction year	1962.95	25.89	1901	1924	1965	1996	2006

Table 2: Parameters of Trade-Selection Process, Alameda Repeat-Sales Sample. LTV is the natural logarithm of the loan-to-value ratio. Time indicates time since the previous sale (in years). Square footage is the property's size in thousands of square feet. Sigma is the annualized standard deviation of the error term in the observation (price index) equation.

	A	B	C	D	E	F
LTV	-0.0598***	-0.0401***	-0.0448***	-0.0426***	-0.0406***	
	(0.0012)	(0.0017)	(0.0015)	(0.0014)	(0.0051)	
LTV '88-'93						-0.0789***
						(0.0041)
LTV '94-'00						-0.0504***
						(0.0026)
LTV '01-'06						-0.0457***
						(0.0020)
LTV '07-'08						0.0389***
						(0.0033)
LTV *					0.0130***	
Square Footage					(0.0024)	
LTV *					-0.0006***	
Age (years)					(0.0001)	
Time (years)		0.0505***	0.0381***	0.0511***	0.0519***	0.0413***
		(0.0010)	(0.0013)	(0.0011)	(0.0011)	(0.0012)
Time Squared		-0.0022***	-0.0021***	-0.0022***	-0.0023***	-0.0020***
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Mortgage rate			-5.4128***			-3.6849***
			(0.1463)			(0.1939)
Square footage				-0.0902***	-0.0671***	-0.0809***
				(0.0040)	(0.0061)	(0.0047)
Age (years)				-0.0008***	-0.0013***	-0.0006***
				(0.0001)	(0.0001)	(0.0001)
Intercept	-2.0073***	-2.1199***	-2.1132***	-2.0573***	-2.0471***	
	(0.0017)	(0.0034)	(0.0046)	(0.0055)	(0.0062)	
Intercept '88-'93						-2.1061***
						(0.0075)
Intercept '94-'00						-2.0671***
						(0.0065)
Intercept '01-'06						-2.0036***
						(0.0059)
Intercept '07-'08						-2.0664***
						(0.0079)
Seasonal Dummies	No	Yes	Yes	Yes	Yes	Yes
Sigma	0.2814***	0.2816***	0.2813***	0.2812***	0.2815***	0.2810***
	(0.0004)	(0.0005)	(0.0004)	(0.0005)	(0.0005)	(0.0005)

Table 3: Parameters of Trade-Selection Process Comparing Samples. The Repeat-Sales sample is the sample used in the all Tables. The Single Sales sample includes all properties that had at least one trade during the sample period. The No Foreclosures sample is the Repeat Sales sample dropping properties that experienced a foreclosure during the sample period. LTV is the natural logarithm of the loan-to-value ratio. Time indicates time since the previous sale (in years). Sigma is the annualized standard deviation of the error term in the observation (price index) equation.

	Repeat Sales	Single Sales	No Foreclosures
LTV	-0.0598***	-0.0203***	-0.0620***
	(0.0012)	(0.0014)	(0.0017)
Intercept	-2.0073***	-2.2122***	-2.0125***
	(0.0017)	(0.0017)	(0.0021)
Seasonal Dummies	No	No	No
Sigma	0.2814***	0.2812***	0.2695***
	(0.0004)	(0.0005)	(0.0005)
# Properties	68,700	142,794	41,983

Table 4: Parameters of Trade- and Foreclosure-Selection Processes, Alameda Repeat-Sales Sample

Panel A: Trade-Selection Process

	A	B	C	D	E	F
LTV	-0.0656***	-0.0455***	-0.0513***	-0.0486***	-0.0443***	
	(0.0017)	(0.0018)	(0.0017)	(0.0019)	(0.0053)	
LTV ’88-‘93						-0.0810***
						(0.0043)
LTV ’94-‘00						-0.0500***
						(0.0024)
LTV ’01-‘06						-0.0425***
						(0.0027)
LTV ’07-‘08						0.0122***
						(0.0041)
LTV *					0.0148***	
Square Footage					(0.0026)	
LTV *					-0.0007***	
Age (years)					(0.0001)	
Time (years)		0.0482***	0.0358***	0.0487***	0.0493***	0.0402***
		(0.0012)	(0.0011)	(0.0012)	(0.0012)	(0.0012)
Time Squared		-0.0020***	-0.0020***	-0.0020***	-0.0020***	-0.0018***
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Mortgage rate			-5.3842***			-3.1686***
			(0.1537)			(0.1919)
Square footage				-0.0835***	-0.0558***	-0.0706***
				(0.0047)	(0.0068)	(0.0046)
Age (years)				-0.0010***	-0.0016***	-0.0008***
				(0.0001)	(0.0001)	(0.0001)
Intercept	-2.0298***	-2.1325***	-2.1278***	-2.0666***	-2.0522***	
	(0.0025)	(0.0043)	(0.0040)	(0.0059)	(0.0067)	
Intercept ’88-‘93						-2.1075***
						(0.0073)
Intercept ’94-‘00						-2.0839***
						(0.0061)
Intercept ’01-‘06						-1.9943***
						(0.0062)
Intercept ’07-‘08						-2.1233***
						(0.0081)
Seasonal Dummies	No	Yes	Yes	Yes	Yes	Yes
Sigma	0.2815***	0.2816***	0.2816***	0.2813***	0.2815***	0.2806***
	(0.0004)	(0.0004)	(0.0004)	(0.0007)	(0.0004)	(0.0004)

Panel B: Foreclosure-Selection Process

	A	B	C	D	E	F
LTV	0.0728***	0.0800***	0.0788***	0.0714***	0.0176	
	(0.0038)	(0.0040)	(0.0038)	(0.0031)	(0.0145)	
LTV ’88-‘93						0.0019
						(0.0150)
LTV ’94-‘00						-0.0283***
						(0.0067)
LTV ’01-‘06						-0.1292***
						(0.0041)
LTV ’07-‘08						0.4018***
						(0.0084)
LTV *					0.0155*	
Square Footage					(0.0065)	
LTV *					0.0008***	
Age (years)					(0.0002)	
Time (years)		0.0877***	0.0750***	0.0877***	0.0844***	0.0516***
		(0.0038)	(0.0038)	(0.0038)	(0.0064)	(0.0026)
Time Squared		-0.0061***	-0.0057***	-0.0063***	-0.0061***	-0.0050***
		(0.0003)	(0.0002)	(0.0003)	(0.0004)	(0.0002)
Mortgage rate			-3.5461***			-8.7778***
			(0.3114)			(0.5034)
Square footage				-0.1863***	-0.1841***	-0.1805***
				(0.0131)	(0.0181)	(0.0120)
Age (years)				0.0029***	0.0036***	0.0028***
				(0.0002)	(0.0002)	(0.0001)
Intercept	-3.0504***	-3.3177***	-3.2843***	-3.3598***	-3.3867***	
	(0.0055)	(0.0140)	(0.0162)	(0.0232)	(0.0193)	
Intercept ’88-‘93						-3.4580***
						(0.0219)
Intercept ’94-‘00						-3.2079***
						(0.0094)
Intercept ’01-‘06						-3.8252***
						(0.0112)
Intercept ’07-‘08						-2.9094***
						(0.0155)
Seasonal Dummies	No	Yes	Yes	Yes	Yes	Yes