A More Timely House Price Index *

Elliot Anenberg
Federal Reserve Board

Steven Laufer
Federal Reserve Board

October 14, 2014

Abstract

We construct a new "list-price index" that uses the repeat-sales approach to measure house prices but for recent months uses *listings* data instead of *transactions* data. Because listings data describe the current offering price and are available essentially in real time, our index is more timely than existing house price indices in two ways. First, our index describes house values at the contract date when the price is determined rather than at the closing date when the property is transferred. As a result of this difference in timing, our index displays a stronger correlation with stock prices and less short term serial correlation than a standard repeat-sales index. Second, our index accurately reveals trends in house prices several months before existing sales price indices like Case-Shiller become available. In a sample of nine large MSAs over the years 2008-2012, our index (i) accurately forecasts the Case-Shiller index several months in advance, (ii) outperforms forecasting models that do not use listings data,

*elliot.anenberg@frb.gov. steven.m.laufer@frb.gov. We thank participants at the 2013 Urban Economics Association meetings, Federal Reserve System Applied Micro Conference, and DC Urban Day as well as seminar participants at the Kansas City Fed, University of Toronto and Wharton. We also thank Pat Bayer, Moshe Buchinsky, Ingrid Ellen, Joe Gyourko, Ed Kung, Karen Pence, Shane Sherlund, Todd Sinai and Maisy Wong for their helpful comments. Jessica Hayes provided excellent research assistance. All remaining errors are our own. An earlier version of this paper was circulated under the title "Using Data on Seller Behavior to Forecast Short-run House Price Changes." The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors.

and (iii) outperforms the market's expectation as inferred from prices on Case-Shiller futures contracts.

1 Introduction

Changes in house prices have important consequences for the real economy as they affect both households' wealth and their ability to borrow. However, the measurement of house prices differs from the measurement of the prices of other important assets, such as stocks and bonds, in two significant ways. First, they differ in the date to which the transaction price is attached. For financial assets, prices describe the value of the assets at the "trade date," the date on which the buyer and seller establish a price and enter into a contract, rather than the "settlement date" when the legal transfer of the asset occurs. By associating prices with the time at which market participants actually agree to them, analysts are able to study the relationships between these price movements and other macroeconomic events. In contrast, for housing transactions, prices are typically associated with the settlement (or "closing") date rather than the trade (or "contract") date. Measuring the relationship between house prices and other high-frequency economic variables is made more difficult because it is not possible to deduce from the recorded closing date exactly when a particular transaction was negotiated and therefore what the macroeconomic conditions were at the time the price was determined.

Second, housing differs from stocks and bonds in how quickly price measures are available. While prices in financial markets are available almost instantaneously, house price indices are reported with lags of several months. This delay is a significant information friction with measurable effects on important economic variables. We show, for example, that a release of the Case-Shiller house price index has an immediate effect on the stock prices of homebuilding companies, despite the fact that this release contains information about housing market conditions from a few months earlier.[1] If the stock market is unable to

[1] The Case-Shiller index, developed by Bailey et al. [1963] and modified by Case and Shiller [1987] and Case and Shiller [1989], is the most widely followed measure of U.S. house price trends. The index is calculated using repeated transactions of the same house so as not to be distorted by changes in the mix of homes sold over time.

overcome the reporting delays associated with house prices, it seems likely that individual homeowners, policy makers, lenders, etc. aren't either, suggesting that this information friction may have much broader effects on financial markets and real economic activity.

This delay in house price reporting emerges because once a buyer and seller have found each other and agreed on a sale price, there is little incentive for either party to publicize the negotiated price or other details of the transaction. Even once the sale price is disclosed (by law) at the closing, which is typically a couple of months following the sale agreement, there is another delay of a couple additional months before the public record becomes available.[2] Further, while local jurisdictions require that these documents contain the closing date, the contract date is not recorded. In contrast, before a contract is signed, the seller has a strong incentive to broadcast the current offering price, both as an advertisement that the house is for sale as well as a signal to potential buyers of the likely price at which the house can be purchased.[3] Thus, information on listing prices is disseminated on internet platforms such as Multiple Listing Services (MLS) in essentially real time. On such forums, when a sale agreement is reached, the listing is removed immediately to indicate to potential buyers that the property is no longer available. By using information on the list prices of homes that are delisted, we can potentially learn about the level of sale prices well in advance of what is currently possible. In addition, by observing the date on which they are delisted, we can place the transactions into the context of the economic conditions that prevailed at the time those prices were actually determined, potentially allowing us to better measure how house prices respond to high-frequency economic variables.

In this paper, we develop a new house price index that exploits the informational content of listings data to overcome these difficulties in the measurement of house prices. A key aspect of our methodology is that we associate each delisting with the most recent *prior* sale of that property. This creates a pair of observations analogous to a pair of repeat sales in the construction of the Case-Shiller index and other conventional repeat-sales indices. This is important because it allows us to provide a more timely index of house price trends without

[2]For example, the Case-Shiller house price index summarizing sales prices that close in month t is not released until the end of month $t+2$.

[3]See Chen and Rosenthal [1996] for a discussion of the role of listing prices as a commitment device for sellers.

sacrificing the most attractive feature of the repeat-sales index: its ability to control for changes in the mix of homes sold over time by partialing out a house-specific fixed effect from each price. The differences between our "list-price index" and a standard repeat-sales index are that (i) the price of the second sale in each pair is not an observed transaction price but rather an estimate based on the final list price observed before delisting and (ii) the price of that second sale is associated with the delisting date rather than the eventual closing date.[4]

Our approach is complicated by the facts that the sale-to-list price ratio (i.e. the ratio of the actual sale price to the price at which the seller had listed the house) varies, both in the cross section and across time, and that many delistings do not ultimately result in transactions. We build a simple model of the home-selling problem that shows how some of the variation in sale-to-list price ratios and the propensity to transact can be explained by other observable information on seller behavior, such as the time on market (TOM) and the history of list price changes. We show that the model's predictions are consistent with the data, and we use this additional information to adjust the final list price up or down and to weight delistings according to their predicted probabilities of becoming sales. These adjustments turn out to be quite helpful for performance, as about 80% of the time series variation in the aggregate sale-to-list price ratio and the share of delistings that transact is explained by observable information in our listings data.

We demonstrate the advantages of our house price measure relative to a repeat-sales index using a sample of 2 million listings drawn from nine major U.S. metro areas over the 2008-2012 time period. The first advantage is that the list-price index describes house values at the time that the transaction price is negotiated whereas measures of house prices based on closing dates describe sales negotiated over a range of previous dates, as the lag length between the agreement and closing date varies across transactions. As a result of this difference in timing, we show that our list-price index is positively correlated with contemporaneous movements in equity prices, whereas the repeat-sales index is not. The latter index is instead correlated with lagged equity prices and the strength of this correlation

[4]Like the Case-Shiller index, our index can also be constructed to allow for both heteroskedastic errors and value weighting, which we describe in more detail in Section 3.

is somewhat attenuated, consistent with the expected effect of the measurement error in timing just described. Additionally, we provide evidence that this timing convention for sales-based measures is the source of some of the positive short-term serial correlation in house prices. Intuitively, the timing of the repeat-sales index causes a one-time shock to house prices to be reflected in the prices of houses that close over a span of several months, leading to additional positive correlation between the growth rates over this time period. While the month-over-month growth rates in a standard repeat-sales index are positively autocorrelated, our list-price index exhibits a slightly negative autocorrelation, as one might expect from a price estimated with *i.i.d* sampling error.

A second advantage of our index is that because list prices reveal information about house prices sooner than transaction documents, we are able to use our list-price index to forecast more standard house price measures, which rely entirely on such records. We show that our index (i) accurately forecasts the Case-Shiller index several months in advance; (ii) outperforms forecasting models that do not use listings data; and (iii) for the seven MSAs in which data on futures contracts are available, outperforms the market's expectation as inferred from prices on Case-Shiller futures contracts. We find that correcting for variation in sale-to-list price ratios and propensity to sell (the "adjusted list-price index") reduces our forecasting errors by approximately 30% relative to a simple model in which we neither adjust list prices nor weight delistings differently (the "simple list-price index"). Nonetheless, the simple list-price index, which is less parametric than the adjusted list-price index, also performs quite well.

Our paper contributes to a large literature that has studied the time-series properties of house prices. On the correlation between house prices and stock prices, several papers, including Flavin and Yamashita [2002] and Goetzmann and Spiegel [2000] failed to find any correlation in quarterly data, while Favilukis et al. [2011] document a positive correlation at annual frequencies between the price-rent ratio for housing and the price-dividend ratio for equities. While we do find a positive correlation between house prices and stock prices, our data spans a rather narrow time period and so our contribution on this question is not to establish the strength of this correlation, but mainly to demonstrate how the timing convention inherent in the sales-based measurement of prices leads to a downward bias in its measurement. The positive autocorrelation in house prices was famously introduced by Case and Shiller [1989] and Cutler et al. [1991]. More recently, Glaeser et al. [2014], have argued that this persistence cannot be explained by fundamentals. In this paper, we provide a new mechanism that leads to additional positive serial correlation at higher frequencies. Perhaps most similar to the current paper in terms of its goal is Wallace [2011], who discusses measurement problems associated with several of the most popular US house price indices and their implications for the measurement of risk in residential and commercial mortgages. Our paper introduces an additional distortion related to the timing of standard repeat-sales indices and studies its effects on the measurement of house price dynamics.

Our paper is also related to the empirical and theoretical literature that studies various aspects of the home-selling process. Anenberg [2013], Carrillo [2012], and Merlo et al. [2013] estimate various extensions of the Chen and Rosenthal [1996] model of the home-selling problem using the type of micro data used in our paper. These empirical search models highlight how and why seller choice variables like the list price and marketing time relate to the sales price at a micro level. Genesove and Mayer [2001] and Bucchianeri and Minson [2013] study how behavioral factors such as loss aversion and anchoring influence seller behavior and ultimately sales prices. Hendel et al. [2009] and Levitt and Syverson [2008] focus on how the seller's decision to use a realtor affects the selling process and selling outcomes. In the current paper, we exploit the relationships between seller behavior and sales prices highlighted by these existing papers to forecast the final sales price.[5]

Finally, we also contribute to the literature on house price forecasting (Gallin [2008], Malpezzi [1999], Rapach and Strauss [2009], and Case and Shiller [1990], among others). The existing literature mostly focuses on the explanatory power of variables that measure macroeconomic conditions like rents, income, unemployment rates, mortgage rates, etc. An exception is a recent paper by Carrillo et al. [2012], who show that including aggregate listings variables like average TOM in standard time-series forecasting models improves forecasting performance. In our paper, listings data provide predictive power for an entirely different

[5]Other related papers that study the home-selling problem include Knight [1996], Genesove and Mayer [1997], Salant [1991], Anenberg [2011], Merlo and Ortalo-Magne [2004], Horowitz [1992], Han and Strange [2013], and Haurin [1988].

but complementary reason. That is, we exploit the *timeliness* of listings data relative to transaction data. We are also unique in that we use the micro data on listings, rather than aggregates, to tie each individual list price to a previous sale price, as discussed above.[6]

This paper proceeds as follows. Section 2 describes our data sources and the particular sample we use to test the performance of our new index. Section 3 reviews the Case-Shiller sale price index methodology. Section 4 introduces our basic methodology with the simple list-price index, and discusses its advantages and potential issues. Section 5 presents theory and evidence on how and why we should use other available information on seller behavior to augment the simple list-price index and outlines our methodology for this adjusted list-price index. Section 6 compares the behavior of both versions of the index to a standard repeat-sales index. Section 7 studies the ability of our indices to forecast the Case-Shiller index and Section 8 concludes the paper.

2 Data

In this section we describe the data sources and the particular sample we use in the analysis that follows.

Our list-price methodology uses two types of data. The fist data requirement is the type of micro data on housing transactions used to produce the Case-Shiller index. We obtained this data from Dataquick, who collects records from local governments throughout the U.S. on home transactions (which in most cases are required to be publicly disclosed by law). For each home sale, these data include the sales price, the closing date, the precise address of the home, home characteristics, and information about the lender, buyer, and seller. A point of emphasis for us is that these transaction data become available with a lag of several months because it takes time for a sale closing to be recorded in the public record.

[6]Recently, several companies have starting using listings data to forecast house prices. For example, the Trulia Price Monitor is a measure of trends in current (not necessarily final) asking prices, adjusted for changes in several observable hedonic characteristics. CoreLogic incorporates new listing prices into a time series regression model to construct a "Pending HPI" (CoreLogic, 2012). Based on our reading of their published materials, neither of these measures fully exploits the informational content of the listings data as we do in the present study.

Our second data requirement is micro data on home listings, which we purchased from Altos Research. For the universe of homes listed for sale on the Multiple Listing Service (MLS), the dominant platform through which homes for sale are advertised in the U.S., these data include the listing price of the home at a weekly frequency. This allows us to observe the week in which a property is delisted, which may occur when there is a sale agreement or when the seller decides to withdraw the home from the market. There is no variable that indicates why a property is delisted, and consequently, if it is delisted because of a sale agreement, we observe nothing from the listings data about the terms of the agreement such as the sales price. Using the date of initial listing and listing prices from previous weeks, we can infer the time-on-market (TOM) at the time of delisting and the full history of list price changes. In addition to the list price, the data include the precise address of the home and some house characteristics. Importantly, and in contrast to sales data, this listing data can be obtained in real time.

In order to construct our list-price index, we we are able to obtain sales and listings data for the following MSAs: Chicago, Denver, Las Vegas, Los Angeles, Phoenix, San Diego, San Francisco, Seattle and Washington DC.[7] Our Dataquick sample generally runs from 1988-2012 for each metro area; the Altos Research sample spans 2008-2013.[8] As we describe below, our list-price index requires linking each home in the listing data to its previous sales record in the transaction data. We do so using the address, which is common to both datasets. The construction of our adjusted list-price index also requires linking a sample of the delistings that result in sales to the associated transaction record, which we again do using the address. We associate a sale with a previous delisting whenever there is a lag of less than twelve months between the most recent delisting of the property and the closing date. To be consistent with the sample of home sales used in the Case-Shiller index (which is described in more detail in the next section), we drop (i) delistings that do not merge to a previous transaction, (ii) delistings where the length of time since the last transaction is

[7]The MSA definitions are the ones used by Case-Shiller. See http://us.spindices.com/documents/methodologies/methodology-sp-cs-home-price-indices.pdf.

[8]Data are available from Dataquick for years beyond 2012, but we were not able to obtain the most recent years for this project. Altos Research did not begin collecting listings data until 2008 and so listings data before 2008 are not available from this particular provider.

less than six months, and/or (iii) delistings that are not single-family. In the end, we are left with a large micro dataset of several million properties that includes the full history of list price changes for each listing, as well as the house's transaction history.

Figure 1 presents the Case-Shiller index for each of the MSAs over the time period in which our transactions data and listings data overlap (2008 - 2012). Most of the cities in our sample experienced significant declines in house prices during the beginning of the sample period, although the magnitude of the decline varied considerably across cities. Our sample also covers periods of price increases; prices rose by varying degrees in 2009 when the first-time home buyer tax credit was in effect and in 2012, when the most recent house price recovery started in many US cities.

Table 1 presents summary statistics of the 1.9 million single-family home listings that we can merge to a previous transaction record and that are delisted during our sample period. Appendix Table A1 presents the same statistics separately for each metro area in our sample. Table 1 shows that a majority of listings are delisted without a list price change and the median TOM is between one and two months. Of those properties that are delisted, we are able to identify approximately half of them with subsequent sales that appear in our transactions data. Many of delistings that do not result in closings are relisted soon after delisting, which may be due to sales agreements that fall through because mortgage contingency or home inspection fails. However, as mentioned above, our listings data do not provide the specific reason.

3 Case-Shiller Sales Price Index

We begin with a stylized presentation of the Case-Shiller repeat sales methodology.[9] Our list-price indices will build off of the equations and notation introduced in this section.

The Case-Shiller regression equation is

$$p_{it} = v_i + \delta_t + \varepsilon_{it} \tag{1}$$

where p_{it} is the log sales price of house i sold in month t, v_i is a house fixed effect, δ_t is

[9] For the full methodology, see Shiller [1991] and the Case-Shiller website.

a month effect that captures the citywide level of house prices at month t, and ε_{it} is the unexplained portion of the house price. Case and Shiller [1989] interpret ε_{it} as a noise term due to randomness in the search process, the behavior of the real estate agent, or other imperfections in the market for housing. Estimates of δ_t, which we denote δ_t^{CS}, are the basis for the Case-Shiller index. For example, $\delta_t^{CS} - \delta_{t'}^{CS}$ is interpreted as the percent change in house prices in the city between months t' and t.

To estimate equation (1), Case and Shiller employ a repeat sales approach. For each home sale, they use the previous home sale to difference out the house fixed effect, v_i. This gives

$$p_{it} - p_{it'} = \delta_t - \delta_{t'} + \varepsilon_{it} - \varepsilon_{it'} \tag{2}$$

where t' denotes the month of the previous sale of house i. The time effects can be estimated through weighted OLS on the pooled sample of sales pairs, where sales pairs with a longer interval between sales are downweighted to account for heteroskedasticity in $\varepsilon_{it} - \varepsilon_{it'}$ ("interval weighting"). Case and Shiller drop (i) homes that cannot be matched to previous sales (e.g. new construction) (ii) home pairs where the interval between sales is less than six months and (iii) all non single-family homes. In practice, Case and Shiller also effectively weight each sale pair by the level of the first sale price, p_{i0}, to ensure that the index tracks the aggregate value of the real estate market ("value weighting"). They also use a three-month moving average index to reduce month-to-month noise in $\varepsilon_{it} - \varepsilon_{it'}$. This is implemented by including a pair with a sale in month t as a pair in months t, $t+1$, and $t+2$.

It is important to emphasize that the time subscript in equation (1) reflects the month in which the sale officially closes. The closing date lags the date when the sale price was agreed upon by a month or two on average, as we have discussed above and as we will show below. Furthermore, Case-Shiller do not release their price index for month t until the last Tuesday of month $t+2$ because the sale prices become available with significant lags, as discussed above. In contrast, our list-price index, which we present next, reflects the value of housing at the time the sale price is negotiated and can be computed essentially in real time.

4 Simple List-Price Index

In this section we outline the methodology of the simple list-price index, which is the simplest way to use listings data to create a house price measure analogous to the repeat-sales index described above.[10] Then, we discuss the potential issues with the simple list-price index from a theoretical perspective, followed by an empirical investigation to determine which issues are important in practice. The empirical work will motivate the adjusted list-price index, which we present as our preferred index in the subsequent section.

4.1 Methodology

The simple list-price index is estimated using the same regression equation as Case-Shiller (equation (2)), except that for the second sale of each transaction pair, we substitute sales price with the final list price and we include each house in the month it is delisted rather than the month in which it closes.

Let p_{it}^L denote the log of the final list price of house i that is delisted at time t and define $\mu_{it} = p_{it} - p_{it}^L$ to be the log of the idiosyncratic sale-to-list price ratio.[11] For convenience, we further define $\overline{\mu} = E(\mu_{it})$, the expected sale-to-list price ratio, and $\tilde{\mu}_{it} = \mu_{it} - \overline{\mu}$ so that $E(\tilde{\mu}_{it}) = 0$. Then to obtain the month-t simple list-price index value δ_t^L, we substitute into equation (2) as follows

$$p_{it}^L - p_{it'} = \delta_t^L - \delta_{t'} + \varepsilon_{it} - \varepsilon_{it'} - \mu_{it} = \delta_t^L - \delta_{t'} + \overline{\mu} + \nu_{it} \tag{3}$$

where $\nu_{it} = \varepsilon_{it} - \varepsilon_{it'} - \tilde{\mu}_{it}$.

Rather than jointly estimating the previous house price level $\delta_{t'}$ along with δ_t^L, we use an estimate of $\delta_{t'}$ calculated from the transaction data alone using the Case-Shiller methodology. This means that when we estimate δ_t^L, we take $\delta_{t'}$ as given and move it to the left-hand side

[10] We tailor our list-price index methodology to track the Case-Shiller index specifically because the Case-Shiller index is currently the most widely followed measure of house price trends. Our approach could equally well be applied to track any other measure of house prices that is based on transactions data.

[11] Note that the time-subscript on p_{it} now refers to when the property is delisted rather than when the transaction closes. We return to this distinction when we do our repeat sales forecasts in Section 7.

of the equation.[12] Finally, because δ_t^L is an index and the absolute level of the index is arbitrary, we can drop $\overline{\mu}$ from the equation, effectively shifting the entire index (in logs) by a constant amount $\overline{\mu}$ relative to standard repeat-sales indices. This gives our estimating equation

$$p_{it}^L - p_{it'} + \delta_{t'} = \delta_t^L + \nu_{it}. \tag{4}$$

Our estimate of δ_t^L, which we denote $\hat{\delta}_t^L$, is the simple list-price index value for month t. In practice, when estimating equation (4), we reproduce the interval weighting done by Case-Shiller and other repeat-sale indices, as described above.

4.2 Discussion

The simple list-price index is attractive because it exploits the timely nature of listings data without compromising the key properties of the repeat-sales index. In particular, like the Case-Shiller repeat-sales index, the simple list-price index accounts for changes in the mix of homes sold over time and is estimated using a methodology that is transparent and simple to compute. This version of the list-price index, however, relies on several additional assumptions. In this section, we identify those assumptions and examine empirically the degree to which they actually hold in the data. In the following section, we will present an alternative list-price index where these assumptions are relaxed.

To start this discussion, we note that at the time of delisting, the researcher cannot observe which transactions will close and which will not. Our index therefore uses all delistings, some of which will not ultimately result in transactions. We introduce the random variable κ_{it} and say that the delisting of house i at time t results in a transaction if $\kappa_{it} > 0$, where the threshold 0 is chosen *wlog*. With this notation in hand, we examine the assumptions necessary to estimate δ_t from equation (4).

For the OLS estimator $\hat{\delta}_t^L$ to be consistent, it must be the case that

$$E(\delta_t^L \cdot \nu_{it}) = E(\delta_t^L \cdot (\varepsilon_{it} - \varepsilon_{i0} - \tilde{\mu}_{it})) = 0. \tag{5}$$

[12] Moving $\delta_{t'}$ to the left hand side is a convenience that we can take because the Case-Shiller methodology uses a chain weighting procedure in which the estimate of $\delta_{t'}$ is not affected by data after time t. If this were not the case, we could simply estimate $\delta_{t'}$ along with δ_t.

We can break up this expression into several terms:

$$\begin{aligned}E(\delta_t^L \cdot (\varepsilon_{it} - \varepsilon_{i0} - \tilde{\mu}_{it})) = 0 = E(\delta_t^L \cdot (\varepsilon_{it} - \varepsilon_{i0}) | \kappa_{it} > 0) \cdot Pr(\kappa_{it} > 0) \\ + E(\delta_t^L \cdot (\varepsilon_{it} - \varepsilon_{i0}) | \kappa_{it} < 0) \cdot Pr(\kappa_{it} < 0) - E(\delta_t^L \cdot \tilde{\mu}_{it}). \qquad (6)\end{aligned}$$

Equation (6) will hold if (but not only if) each of the three expressions on the right-most side of the equation equals zero. We consider each term separately. First,

$$E(\delta_t^L \cdot (\varepsilon_{it} - \varepsilon_{i0}) | \kappa_{it} > 0) = 0, \qquad (7)$$

says that among delistings that are in fact sales, the error terms cannot be correlated with the time effects. This condition was already necessary for the consistent estimation of the standard repeat sales model in equation (2).

The next term,

$$E(\delta_t^L \cdot (\varepsilon_{it} - \varepsilon_{i0}) | \kappa_{it} < 0) = 0, \qquad (8)$$

requires that the error terms of listings that are withdrawn and do *not* result in transactions satisfy the same exogeneity restrictions as the error terms for the observations of houses that *do* sell (equation (7)). If delisted houses that do not sell have list prices that imply higher or lower values for the level of house prices, then including these observations will bias our estimates.

The final piece of equation (6) is

$$E(\delta_t^L \cdot \tilde{\mu}_{it}) = 0, \qquad (9)$$

which says that that the sale-to-list price ratio cannot co-vary with the time effects. A sufficient but not necessary condition would be that the average sale-to-list price ratio be time invariant (i.e. $E_t(\tilde{\mu}_{it}) = 0$). The intuition behind this condition is that if variation in prices is caused by movements in the sale-to-list price ratio, we would not be able to identify this variation by looking only at list prices.

If these three conditions discussed above are satisfied, then δ_t^L can be consistently estimated from equation (4). Our list price model makes an additional assumption that we abstracted from in the discussion above, namely that all housing transactions first appear as delistings in the MLS. In fact, not all homes that sell are listed on the MLS and if homes that are not sold via the MLS are a *selected* group of transactions, then the simple list-price index may be biased.

4.3 Descriptive Evidence

We next examine the empirical relevance of each potential issue with the simple list-price index in turn.

We first examine trends in the sale-to-list price ratio. Figure 2 shows the 25th, 50th, and 75th percentiles of the distribution sale-to-list price ratios for our full sample. The sale-to-list price ratio fluctuates within a band of several percentage points, and the variation appears to be correlated with the house price cycle, in violation of the assumptions of our simple list-price index. Periods of rising prices tend to have high sale-to-list price ratios.

Another potential source of bias for the simple list-price index is the inclusion of all delistings rather than just those that lead to sales. Figure 3 shows that indeed, delistings that result in closings are a selected group of delistings that tend to have lower list prices relative to delistings that do not result in closings, and the magnitude of the list price difference is negatively correlated with the house price cycle.[13] Figure 4 presents the share of delistings that result in a sale by quarter. This share is also volatile over time, with hotter markets being associated with a higher probability of sale. Together, the data shown in Figures 3 and 4 suggest that including all delistings, rather than only the ones that result in sales, will bias the index due to selection.

Finally, we investigate the potential for selection bias arising from the types of homes that are listed on the MLS. Figure 5 shows that the sales that do not appear in our listings data represent only a small minority of total sales, which is consistent with reports from the National Association of Realtors.[14] This suggests that this type of selection should not have a large effect on the performance of the simple list-price index.[15]

[13] In comparing list prices across properties, we normalize each list price by $(p_{it'} - \delta_{t'})$ to control for differences in house quality.

[14] See for example the 2012 NAR Profile of Home Buyers and Sellers which reports that 88 percent of home sales are broker assisted.

[15] Sales may not merge to a delisting either because the home is sold without the assistance of a broker (e.g. in a foreclosure auction) or the address is coded with error, preventing a successful merge.

To summarize, the empirical evidence suggests two problems with the simple list-price index. First, the sale-to-list price ratio varies with the housing cycle so that the final list price is a good, but not unbiased, predictor of the final sales price. Second, since this price index uses all delistings rather than only the ones that result in closed transactions, it is susceptible to selection bias. We next discuss an alternative specification meant to address these issues.

5 Adjusted List-Price Index

In our simple list-price index, the only elements of the listings data we use are the date at which the property is delisted and the final list price. This section examines whether we can use other information available at the time of delisting – such as time on market and the list price history – to improve the performance of the simple list-price index.

5.1 Model

We first present a model of the home selling problem that generates variation in sale-to-list price ratios and the probability of sale conditional on delisting, which is precisely the variation that is an issue for the simple list-price index. The model delivers predictions for how these outcomes should vary with observable listings variables such as TOM and the list price history. This exercise therefore gives us a theoretical motivation for why such information should be useful in constructing an alternative list-price index meant to address the limitations of the simpler version.

The model is in the spirit of Chen and Rosenthal [1996] and describes the behavior of a homeowner trying to sell her house. The model generates variation in the outcomes of interest from two sources. The first is heterogeneity in the valuation that sellers place on not selling and staying in the home and the second is a finite selling horizon.[16] We keep the

[16]In practice, heterogeneity in the seller's value from remaining in the house may arise from factors such as employment opportunities and changes in the seller's familial or financial situation. A finite selling horizon may be a good approximation of reality if things like the start of a school season or the closing date on a trade-up home purchase impose limits on the date by which the owner must sell.

model simple enough so that we can analytically derive predictions that can be tested in the data.

The model contains two periods and in each period t, the seller sets a list price p_t and potential buyers arrive with a probability $\alpha_0 - \alpha_1 p_t$. We assume that $\alpha_1 > 0$ so that a higher list price discourages buyers from visiting the home.

We assume that all of the bargaining power rests with the seller so that when a potential buyer arrives, the negotiated price is equal to the buyer's reservation value. However, the list price functions as a commitment device so that if the buyer's reservation price is higher than the list price, the seller commits to selling the house at the list price, leaving the buyer with positive surplus. Thus, when setting the list price, the seller faces a trade-off: a high list price discourages buyers from visiting a home, but a high list price results in a higher sales price conditional on a buyer arriving. This result is consistent with the empirical evidence (e.g. Merlo and Ortalo-Magne [2004]).

There are two type of buyers in the market. A fraction β are high types with sufficiently high valuation that the seller's commitment always binds and the negotiated sale price, p^* equals the list price p_t. A fraction $(1-\beta)$ are low types with valuation v, which is sufficiently low that the commitment does not bind and the negotiated price equals v. If the seller is unable to negotiate a sale with a prospective buyer by the end of the second period, she remains in the house, an outcome to which she assigns a reservation value of $w_i \in [\underline{w}, \bar{w}]$. We assume $v > \bar{w}$ so that the negotiation with any buyer results in an acceptable sale price and the house goes unsold only if no buyer arrives.

5.2 Model Predictions and Evidence

The theoretical results from this simple model illustrate how variables such as TOM, the history of list price changes, and indicators of the seller's reservation value may provide information about the heterogeneity among sellers and could therefore help us better predict variation in the sale-to-list price ratio and the probability of sale. These results follow largely from the model's basic prediction that that sellers with higher reservation values tend to set higher list prices and therefore take longer to sell their homes. We leave a formal presentation of these results to the appendix, together with their derivations. Instead, we move on to a set of empirical results, all of which are consistent with the model's predictions.

Column 1 of Table 2 shows the results of a regression with the sale-to-list price ratio as the dependent variable. We find that homes that sell with shorter TOM have larger sale-to-list price ratios, as do homes where sellers have lowered their list prices. We also include dummy variables for whether a house is being sold by a bank that has foreclosed on the property and whether the final listing price is lower than the home's previous sales price, both of which may indicate that the seller has a lower reservation value.[17] In both cases, we find higher sale-to-list price ratios, consistent with the predictions of our model.

We also estimate the likelihood that a property is delisted because of a sale rather than because of a withdrawal by the seller for other reasons. Marginal effects from a probit model are shown in Column 2 of Table 2. Properties that are taken off the market soon after they are first listed are much more likely to reflect sales compared with properties with longer TOM. Sellers who have changed their list prices are more likely to delist their properties due to a sale, as are those who reduce prices by larger amounts relative to the initial list price. These results are again consistent with the model and the interpretation that sellers who make larger reductions in their list prices have lower reservation values. Foreclosure sales and sellers who list their properties for less than the previous sales price are also more likely to sell, again consistent with the idea that these sellers have lower reservation values.[18]

Ultimately, the effectiveness of using the listing history to augment the simple list-price index depends on the extent to which listing history can explain the *time-series* variation in the sale-to-list price ratio and sales rate. To examine this, Columns 3 and 4 of Table

[17]See Campbell et al. [2011] for evidence that banks are more motivated to sell than the typical non-bank seller. Genesove and Mayer [2001] argue that sellers are subject to loss aversion and are reluctant to re-sell their homes for less than they originally paid for it. Sellers who have posted list prices below the previous sales prices are essentially guaranteed to realize a nominal loss on the transaction. We might therefore expect that sellers would be less willing to do this unless they assigned a particularly low value to staying in the house.

[18]The regressions in Table 2 include several explanatory variables that do not have a direct mapping to the motivating model, but still provide explanatory power for the variation of interest. For example, we find that there is a discrete jump down in the probability of selling at a TOM of exactly six months, perhaps because many listing contracts with realtors expire after six months. The variables related to the digits of the list price level are motivated by the findings in Beracha and Seiler [2013].

2 present aggregate versions of the regressions in Columns 1 and 2 where each observation is a city-month combination and the dependent variable and regressors are averages over all the delistings in a given month-city. We find that our regressions can explain about 80 percent of the variation in both the sale-to-list price ratio and the sales rate over time. This suggests that incorporating information on the listing history may significantly improve the performance of the simple price index.

In addition to the variation that can be captured by changes in the variables in the listing data, some variation in the sale-to-list price ratio and probability of sale is attributable to macroeconomic factors, which are likely to be persistent. As a result, we would expect that the errors in our list-price index are likely to be positively correlated over time and indeed, we find this to be the case. Including lagged dependent variables in our regressions allows us to explain several additional percentage points of the variation of the sale-to-list price ratio and the propensity to sell (not reported).

5.3 Adjusted List-Price Index: Methodology

In this section, we outline the methodology of our preferred list-price index, which takes advantage of the additional information in the listings data in a way that is consistent with the model and evidence presented in Sections 5.1 and 5.2.

Step 1: Estimate Expected Sale Price and Probability of Sale

From the set of previous observations that are available at time τ, we see which delistings resulted in transactions and, for those that did lead to sales, the sale price. Based on these data, we estimate the empirical relationship between variables that are observable at the time of delisting, such as TOM and the list price history, and the two variables describing the subsequent sale of the property: whether the sale occurred and if it did, the sale-to-list price ratio. In what follows, i indexes the individual seller and t indexes the periods prior to τ.

1. For the sample of delistings that did sell, estimate the equation for the expected sale-

to-list price ratio

$$p_{it} - p_{it}^{L} = \alpha_t^p + \beta_\tau^p X_{it}^p + \varepsilon_{it}^p \tag{10}$$

using OLS, where X_{it}^p is the vector of observables that explain variation in the ratio and α_t^p is a time fixed effect that captures additional time-series variation in the sale-to-list price ratio.

2. For the entire sample of past delistings, estimate the probability that a delisting results in a sale

$$Sell_{it} = I(\beta_\tau^s X_{it}^s + \varepsilon_{it}^s > 0) \tag{11}$$

where $Sell_{it}$ is an indicator that equals one when a sale is observed, X_{it}^s is the vector of observables that explain variation in the propensity to sell and $\varepsilon_{it}^s \sim \mathcal{N}(0,1)$. The expected probability of sale conditional on X_{it}^s is then $\Phi(\beta_\tau^s X_{it}^s)$ where Φ is the standard normal c.d.f.

Step 2: Estimate Serial Correlation in Sale-to-List Price Ratios

In addition to the *cross-sectional* variation in the sale-to-list price ratio that can be predicted using our above estimates of $\beta_\tau^p X_{it}$, we also find evidence that there is predictable *time-series* variation in the time effects α_t^p, which capture variation in the average sale-to-list price ratio over time beyond what is implied by changes in the observable covariates. While the estimates of α_t^p reveal these time effects for past data, they do not directly tell us about what we expect the average sale-to-list price to be in the current period. In order to use these estimates to help predict current sale-to-list ratios, we assume these time effects have a simple serial correlation structure.

1. Estimate the serial correlation in the estimated time fixed-effects $\hat{\alpha}_t^p$ from the equation:

$$\hat{\alpha}_t^p = \rho_0 + \rho_1 \hat{\alpha}_{t-1}^p + e_t \tag{12}$$

using OLS where t denotes the month and ρ_1 measures the degree of serial correlation.

2. Let L denote the number of months since the most recent available sales data, which means that we have estimates $\hat{\alpha}_t^p$ for $t \leq \tau - L$.Then we can estimate

$$\hat{\alpha}_\tau^p = \hat{\rho}_0(1 + \hat{\rho}_1 + \hat{\rho}_1^2 + ... + \hat{\rho}_1^{L-1}) + \hat{\rho}_1^L \hat{\alpha}_{\tau-L}^p. \tag{13}$$

This expression results from iteratively substituting into the right hand side of equation (12) until we get back to the observable (as of time τ) estimate $\hat{\alpha}_{\tau-L}^p$. In this equation, $\hat{\rho}_0$ and $\hat{\rho}_1$ denote the OLS estimates from (12).

Step 3: Estimate Adjusted List-Price Index

Based on the above calculations, we use the information that is available at the time of delisting to generate estimates of the expected sale price and probability of sale for each delisting that occurs at time τ. We then use these estimates to calculate our price index.

1. From Step 2, the final estimate of the sale price for each delisting is given by

$$\hat{p}_{i\tau} = p_{i\tau}^L + \hat{\alpha}_\tau^p + \hat{\beta}_\tau^p X_{i\tau}^p \tag{14}$$

and the final estimate of the probability of sale is

$$\hat{\pi}_{i\tau} = \Phi(\hat{\beta}_\tau^s X_{i\tau}^s). \tag{15}$$

2. From these estimated transaction prices and probabilities, estimate price levels $\delta_\tau^{L_A}$ using the same estimating equation as we used for the simple list-price index. In this case, the equation takes the form

$$\hat{p}_{i\tau} - p_{it'} + \delta_{t'} = \delta_\tau^{L_A} + \eta_{i\tau}. \tag{16}$$

We estimate this equation using weighted least squares, where the weighting is proportional to $\hat{\pi}_{i\tau}$ and also depends on the elapsed time between the two transactions, as described in Section 3.

The full set of regressors X^p and X^s used to estimate equations (10) and (11) are those shown in Table 2. It is important to emphasize that all of the variables in X^p and X^s are computed using listings data or historical transaction data, so that all of the data inputs required to compute the adjusted list-price index are indeed available on a timely basis.

Because we have a relatively short time series of listings data, the adjusted list-price index methodology that we use to generate the results presented in the next two sections deviates from the methodology described above in two places. First, we estimate the parameters of equations (10) and (11) using the same, full sample of listings and transactions data for each time period and we estimate them separately for each MSA. Second, rather than including time fixed effects in equation (10) and adjusting for serial correlation, we include instead the two-month lag of the average sale-to-list price ratio.

6 Comparing List-Price and Sales-Price Indices

We next explore how our new measure of house prices compares to a more standard repeat-sales index. To perform this comparison, we construct each index at a weekly frequency throughout the period spanned by both the listings and sales data. For each methodology, we compute a single composite index based on listings and transactions pooled across all the cities in our data-set.[19] Figure 6 plots the adjusted list-price index and the sale-price index over our sample period. As expected, the list-price index appears to lead the sale price index and displays greater volatility. The former effect arises because delisting occurs before closing. The latter effect arises because the sale price index is effectively a moving average of the list-price index due to variation in the lag between delisting and closing.

The remainder of the section compares the properties of the different indices with the goal of understanding how the timing conventions associated with the repeat-sales methodology may influence our conclusions about how house prices behave. In order to fully isolate the effects of this timing, we introduce an additional index, which uses the sale price recorded in the transaction records but associates the sale with the earlier date at which the property was taken off the MLS database. This hybrid index has the same timing convention as the list-price index but the same price measure as the sales data. Therefore, we can be confident

[19] This differs somewhat from the construction of the 10 and 20-city composite Case-Shiller indices in which the index for each MSA is weighted by the aggregate value of the housing stock in that metro area.

that any difference between this index and the repeat-sales index reflects only the difference in the timing and is not affected by the difference in how we measure prices for each sale.[20]

6.1 Correlation with Movements in Stock Prices

First, we want to study how house prices co-move with other macroeconomic variables and in particular how the answer to this question depends on which measure of house prices we use. As an example, we consider the two-week growth in house prices and in the S&P 500 stock market index.[21] Results are shown in Table 3. As seen in columns 1 and 2, a one percent increase in stock prices is associated with a 0.15% change in the simple list-price index and a 0.24% change in the adjusted list-price index, controlling for seasonal effects and lagged house price growth. The fact that the correlation is slightly larger for the adjusted list-price index than for the simple list-price index suggests that the relationship between stock prices and house prices is *not* fully captured in the change in prices posted by sellers. The positive correlation between stock and house prices may arise for a number of reasons such as common shocks (e.g. monetary policy news, economic news) and/or wealth effects. As we discussed in Section 1, our goal in this paper is not to identify the specific mechanism that generates the correlation.

Not unexpectedly, the sales price index, which reflects prices determined several weeks earlier, is uncorrelated with contemporaneous changes in stock prices (column 3) but is positively correlated with lagged changes in stock prices (column 4). We find the strongest correlation using a six week lag, roughly consistent with the average time between the contract and closing dates. In addition to the sales index responding to lagged rather than contemporaneous movements in stock prices, we find that the measured correlation between house prices and equity prices is *stronger* when house prices are measured using our list-price

[20] For applications where one is more concerned with using the most accurate measurement of the sale price and less concerned with being able to measure prices in real time, this construction may in fact be more appropriate than our list-price index.

[21] Since our listings data are available at a weekly frequency, the list-price index for time t actually reflects delistings that occur at various point over the seven days subsequent to t. Therefore, for consistency, we take the stock price index for time t to be a simple average of the daily stock prices over the seven days subsequent to t.

index than when they are measured with a standard repeat-sales index.[22]

To understand the intuition for these results, it is useful to imagine an idealized one-time positive shock to the economy which causes an instantaneous rise in both stock prices and in negotiated house prices. From the listings data, we can identify with relatively little measurement error which prices were determined before the shock and which after the shock. However, if we consider the set of houses that close in a given period several weeks after the shock, these closings will represent a mix of purchases negotiated after the shock that close with short lag times and purchases that were negotiated before the shock but take longer to close. This mixture causes the measured size of the response in house prices to be biased downwards. We observe that our hybrid index, which uses prices from the sales data but ties the transaction to the delisting date, produces results that are more similar to the list-price index. This further supports our interpretation that the differences are coming from the timing conventions rather than other differences between the indices.

6.2 Serial Correlation

The estimates shown in Table 3 also provide some additional insights about one of the most striking features of house prices, the strong amount of short-run positive autocorrelation first studied in the seminal paper by Case and Shiller [1989]. We find that the list-price index displays significantly less positive serial correlation than the sales-price index.[23] Quantitatively,

[22] While each of these estimates are themselves statistically different from zero, the difference between the coefficients does not quite reach statistical significance. The p-value for the two sided test is 0.24. However, given the theoretical prediction (described below) that the coefficient should be smaller for the repeat-sales index than for the list-price index, one might think that a one-sided sided test for the difference in parameters is more appropriate. In this case, the p-value would be 0.12. Thus our result almost reaches a low level of statistical significance despite the very short time period spanned by our data and the high amount of volatility in the S&P 500 index. In conducting these hypothesis tests, we treat the pair of equations as seemingly unrelated regressions using feasible GLS in order to account for the correlations between the error terms in the two regressions.

[23] At the very high, two-week frequency that we consider, we actually find a negative autocorrelation in the sales-price index. But as we approach longer, more commonly studied frequencies, or if we combine the total effect from including several lags, we reproduce the well-documented positive autocorrelation in house prices.

a one percent increase in the list-price index over the previous two-week period predicts an increase in the following period that is 0.3 percentage points lower than does a corresponding change in the sales-price index.

The fact that each of these indices contains sampling or measurement error, at least some of which is purely transitory, will tend to create a negative short-run autocorrelation as unusually high or low readings of the index are unwound in the following observation period. This explains why in all of the house price indices shown, growth is more positively correlated with its second lag than with its first and the one-period autocorrelation is actually negative for all indices at very high frequencies. However, the fact that the one-period autocorrelation is larger for the sales index than for the list-price index suggests that the difference in timing convention generates additional persistence in measured house prices. The explanation is similar to the reasoning presented above for why the sales-price index is less strongly correlated with equity prices. An instantaneous shock to negotiated prices will be reflected in sale prices of houses that close over a span of several periods, so that, for example, a positive shock to prices will cause prices to rise in the transactions that close in the following month as well as in the subsequent month, leading to additional persistence in growth rates.

The results from this exercise suggest that much of the very short term persistence in house price growth can be explained by the measurement issues that are the focus of this paper. However, the literature has documented persistence even in one year house price changes, a frequency for which our mechanism has little to contribute because the lag between agreement and closing is rarely longer than several months. Therefore, we view our results here as complementing existing studies such as Anenberg [2013], Guren [2013], Head et al. [2014], which explore various mechanisms that can help to explain some of the puzzling short term persistence in house prices. More generally, we view the main contribution of our results in Section 6 as illustrating how our more timely measure of house prices allows for a more accurate measurement of high frequency movements in home prices and the correlation of those movements with other economic variables of interest.

7 Forecasting with the List-Price Index

Our list-price index is tightly connected to standard repeat-sales indices such as Case-Shiller by the simple fact that the delistings that underlie our index will ultimately become the transactions on which these standard indices are based. In order to test this connection, we next show how the information that goes into the construction of our list-price index can alternatively be used to generate a forecast of the Case-Shiller index several months ahead of its release. This exercise serves two purposes. First, the ability to better forecast house prices could help alleviate some of the information frictions we described in the introduction. Second, the exercise helps establish the validity of our index by demonstrating that our index captures essentially the same information about house prices that is contained in standard repeat-sales indices, while making that information available much sooner.

7.1 Motivating Empirical Exercise

We begin our forecasting discussion with a brief empirical exercise that highlights the economic significance of the information lag associated with house prices. The Case-Shiller index is released in the last week of each month, with a two-month delay to the release. From futures contracts traded on the Chicago Mercantile Exchange (CME), we can infer market expectations about the house price levels that will be reported in upcoming releases.[24] Based on these expectations, we can measure the surprise in the Case-Shiller index, which we calculate as the percent change in the actual index value relative to the market's expectation of the index value on the day prior to its release. Figure 7 shows the results of an event study relating surprises in the 10-city Case-Shiller index to changes in the stock price of six different home building companies.[25] For a sample of 25 Case-Shiller index release days for

[24] Futures prices are available for four releases each year from August 2006 through August 2013. Section 7.3.2 provides more details about these futures prices.

[25] Changes in stock prices are measured as the difference in the opening price on the day of a Case-Shiller index release relative to the closing price on the day before, which is the appropriate comparison because the index is always released before the market opens. We use the companies in the Google finance homebuilding sector. The stock tickers are TOL, RYL, BZH, PHM, DHI, KBH, WLH, HXM. We drop HXM from our analysis because it is a Mexican homebuilding company, although the result still holds if this company is

which data are available on futures prices, a one percent positive surprise is associated with a 0.35 percent increase in homebuilder stock prices and the effect is statistically significant. We cluster standard errors by release date.[26]

The key point of this exercise is that the Case-Shiller index release describes housing transactions that were negotiated up to four months earlier and the pricing information contained in these transactions appears to be important for valuing these companies. Yet during these intervening months, market participants were not fully able to incorporate this information. By using information from listings data that was available at the time of the contract negotiations, our list-price index can mitigate this information friction.

7.2 A Simulated Repeat-Sales Index

In order to move from the list-price index, which describes the value of houses at the time of contract, to a measure that resembles a repeat-sales index, we must associate a closing date with each transaction. Even if we could observe which delistings close, we do not know exactly when the closing dated will be, as the lag between the delisting date t_i^d and closing date t_i is idiosyncratic. To address this uncertainty, we assume that the lag between delisting and closing dates, $l_i = t_i - t_i^d$, is drawn randomly from a discrete distribution C, which we estimate non-parametricly using the empirical distribution of $t_i - t_i^d$. (The support of C is 0,...,365 days.) Then for each delisting, we simulate a range of closing dates by drawing from this distribution. Because the Case-Shiller index is a monthly index, each closing is assigned to the calendar month in which it falls. Once we have assigned dates to these simulated transactions, we use them to estimate a "simulated repeat-sales index," following the methodology of the Case-Shiller index as closely as possible.

Our approach assumes that the lag between the delisting date and closing date has a constant, time-invariant distribution. Figure 8 shows the percentiles of the distribution of *Closing date - Delisting Date* for delistings that result in sales over time. On average, there

included.

[26]For this result and in Figure 7, we difference off the overnight change in the S&P500 index from each homebuilder stock price change. When we do not difference off this change, the coefficient estimate rises to 0.43 and the robust t-statistic drops from 3.8 to 1.95.

is an average delay of about six weeks between delisting and closing. The figure shows that the distribution of delays does not vary much over time, so our assumption that the lag times follow a time-invariant distribution is likely a reasonable approximation.

The construction of our simulated repeat-sales index is designed to mimic the construction of the actual Case-Shiller index, but uses information that is available several months earlier. In this sense, the simulated repeat-sales index serves as a forecast for what the Case-Shiller index will look like when it is finally released. We construct two versions of this forecast, based on the two versions of the list-price index. The first forecast is based on the simple list-price index, where the estimated transaction price is simply the final listing price p_{it}^L plus an average sale-to-list price ratio $\overline{\mu}$. We construct this forecast as follows:

1. For each observed delisting i at time t_i^d, draw R random realizations of the time to closing: l_{itr} for $r = 1...R$, which gives a simulated transaction that closes at date $t_{ir} = t_i^d + l_{itr}$ with price $p_{it_{ir}} = p_{it}^L + \overline{\mu}$.

2. To mimic the smoothing approach of Case-Shiller, generate three copies of each simulated closing and each prior sale by adding 0, 1, and 2 months to the time subscript on the current and previous sales price, respectively.

3. Take as given the level of the Case-Shiller house price index at the time of the previous sale, $\delta_{t'}^{CS}$. Then, using the simulated transactions, estimate price levels δ_t from

$$p_{it_{ir}} - p_{it'} + \delta_{t'}^{CS} = \delta_{t_{ir}} + \eta_{itr} \tag{17}$$

   using weighted least squares. Again following the Case-Shiller methodology, the weighting depends on the elapsed time between the two transactions and also the initial sales price, as described in Section 3.

Constructing a forecast based on the adjusted list-price index rather than the simple list-price index follows the same steps with two adjustments:

1. Rather than using the marked-up final listing price $p_{it_{ir}} = p_{it}^L + \overline{\mu}$, use the expected sale price $p_{it_{ir}} = p_{it}^L + \hat{\alpha}_{t_{ir}}^p + \hat{\beta}_\tau^p X_{it}^p$ that we estimated in the construction of the adjusted list-price index

2. Multiply the weight on each simulated transaction by the probability of sale: $\hat{\pi}_{it_{ir}} = \Phi(\hat{\beta}^s_\tau X^s_{it})$

In examining these forecasts, it is important to bear in mind that the transactions that close in a particular month come from delistings that are observed over a fairly large time period. If the distribution of lag times between delisting and closing has support of $[0, 365]$ days, then sales that contribute to a particular month's price index may be delisted as early as 365 days before the start of the month and as late as the last day of the month. In practice, our forecast at time t of the price level in month t' is based on all delistings observable at time t that could possibly close in month t'.

7.3 Forecasting Performance

In this section, we report the performance of our two forecasts over the sample period. We examine our ability to forecast the Case-Shiller HPI at various horizons, which we define as the number of weeks from the date of the last observed listings data until the end of the month we are trying to forecast. For example, at a horizon of one week, we observe all listings information for the first three weeks of the month and we are trying to forecast the HPI based on all transactions that will close in that month. Given that closing dates lag agreement dates by several weeks, at this horizon we should observe close to the entire universe of delistings that would contribute to the Case-Shiller index for that month. At longer horizons, an increasing share of the the sales are from properties for which we have not yet observed delistings. However, even five months into the future, we find that our index still has significant predictive power. Our index is able to predict prices so far into the future because some transactions take a significant amount of time to close and also because the smoothing process (described in Step 2 of the simulated repeat-sales index) causes sales that close in a given month to affect the price index for the two subsequent months as well. Also, because the level of the Case-Shiller index is not released until almost two months after the end of that month, we can sensibly write down "forecasts" for the HPI of months that have already ended but for which price data have not yet been released. These forecasts will have negative forecast horizons.

Since the Case-Shiller index level itself has no meaning, we forecast the change in the index level relative to the latest available index value. Thus, a forecasting error of x means that our forecasts under/over estimates the percent change in sales prices by $100 \cdot x$ percentage points. Performance is based on a comparison of our simulated repeat-sales indices to a true repeat-sales index that we estimate using our transaction data, following the Case-Shiller methodology.[27]

7.3.1 Absolute Performance

Table 4 summarizes the absolute performance of both the simple list-price index and the adjusted list-price index at various horizons. The number of months ahead of the Case-Shiller release of that month's HPI is reported in the second column. The forecast based on the adjusted list-price index performs well, even at forecasting horizons of up to 13 weeks, which is six months in advance of the Case-Shiller release. The root mean square error (RMSE) associated with a forecasting horizon of 13 weeks is .041, the mean absolute error (MAE) is .032, and the adjusted list-price index explains 53 percent of the variation in the six month percent change in the Case-Shiller index. As expected, performance improves as more listings information about the month we are trying to estimate becomes available. Even the simple list-price index, despite its issues discussed in Section 4, performs well. When the forecasting horizon is five weeks, the RMSE is .035 and the MAE is .029. Relative to the simple list-price index, using the adjusted list-price index delivers improved performance by about 20-30 percent for shorter forecasting horizons.

Figures 9-10 show additional detail for select forecasting horizons using the adjusted list-price index. The figures show that the index performs well (i) in each MSA individually, (ii) over the entire sample period, and (iii) during turning points. For example when sales prices started to come out of their multi-year slump in mid 2012, our forecasts from the list-price index did so as well. In addition, when sales prices ticked up in 2009 due to the Obama administration's first time home buyer tax credit, this increase was captured by list-price index as well. However, the fit is not perfect: for example, when sales prices decreased

[27] In particular, we compute the index at a monthly frequency using the value-weighting and three-month smoothing procedure described in Section 3.

significantly in early 2012 in Chicago, the list-price index did not predict that prices would fall nearly as much.

We want to make clear that our index achieves excellent performance even though we are not doing any forecasting in the usual sense. In other words, we are not extrapolating any trends or projecting relationships forward. Rather, we are simply processing data on seller behavior in a novel way and exploiting the long lag between when seller behavior is observed and when the corresponding sales price index is released.

We should also emphasize that our sample period covers one of the most volatile time periods in U.S. housing market history, and some of the most volatile sub-markets (e.g. Phoenix). During such a period of heightened volatility, one might expect list prices to be the least informative about sales prices, as sellers may have difficulty assessing their home values when market conditions are changing so drastically.

7.3.2 Relative Performance

In this section we address two outstanding questions about performance. First, does listings information provide any *additional* explanatory power for short-run house price changes relative to a forecasting equation that does not use listings data? And a second, more challenging question: is the informational content of the listings data that we exploit already known to market participants?

To address the first question, we report the performance of an alternative short-run forecast calculated based on the following AR(3) specification:

$$\delta_{j,t}^{CS}-\delta_{j,t-L}^{CS} = \rho_{0,j}+\rho_{1,j}(\delta_{j,t-L}^{CS}-\delta_{j,t-2L}^{CS})+\rho_{2,j}(\delta_{j,t-2L}^{CS}-\delta_{j,t-3L}^{CS})+\rho_{3,j}(\delta_{j,t-3L}^{CS}-\delta_{j,t-4L}^{CS})+\beta_j X_{j,t-L}+\varepsilon_{j,t} \tag{18}$$

where L, as defined above, is the appropriate lag-length associated with the forecast horizon of interest, $\delta_{j,t}^{CS}$ is the Case-Shiller index for city j in month t, and $X_{j,t-L}$ is a vector of controls including seasonal dummies, national mortgage rates, and state level unemployment rates.[28]

[28] One might be concerned that we are omitting some key observable fundamental in equation (18). This seems unlikely because the fundamental would need to be available and to vary at a very high frequency given the short forecasting horizons we are considering. Furthermore, the explanatory power of any fundamental is weakened by the idiosyncratic lag between the agreement date and the closing date for a sale price as

We estimate equation (18) separately for each city, and so the parameters in (18) depend on j. The estimation sample is the full sample of index values available for each city (typically 1988-2013).

Table 5 shows that the gains in performance by forecasting with the adjusted list-price index are large and statistically significant.[29] Compared to the AR(3) model, the adjusted list-price index delivers 34 percent and 33 percent improved performance in terms of RMSE and MAE, respectively, for an estimate of Case-Shiller five months in advance.

To address the second question, we compare the performance of our index with the performance of the market's expectation as implied by the prices of futures contracts for the Case-Shiller index over our sample period. Futures contracts trade on the Chicago Mercantile Exchange for each individual city in the 10-city Case-Shiller composite index, as well as for the composite index as a whole. Contracts extending 18 months into the future are listed four times a year (February, May, August, November). Each of these contracts trades on a daily basis until the day preceding the release of the Case-Shiller index value for the contract month, at which point there is a cash settlement. We interpret the price of the contract (i.e. the midpoint of the bid-ask spread) on day t as the market's expectation of the house price index $S - t$ days into the future, where S denotes the settlement day (i.e. the day that the index value is released). This interpretation is supported by the motivating exercise depicted in Figure 7, which shows that surprises in the index level measured relative to these futures prices shift around stock prices in the expected way.

Seven of the nine cities in our sample are contained in the 10-city composite index and therefore have futures traded on the CME. We obtained daily price history for each of the 15 futures contracts for these cities that expired during our sample period. Table 6 shows that the RMSE of the futures prices decline over time as the expiration date approaches. This is to be expected if traders are incorporating new information that arrives over time into their expectations. Table 6 also summarizes the performance of the adjusted list-price index

illustrated in Section 6. Finally, we note that the literature has emphasized the role of search frictions and momentum in explaining house price dynamics, which may be best captured in the reduced form by the AR terms in equation (18).

[29] Our test statistic is a panel version of the Diebold-Mariano test statistic with a bartlett kernel (see Diebold and Mariano [2002]).

compared to the performance of the futures market over our sample period. The detail for a few select forecasting horizons is presented in Figures 11 and 12. At a forecasting horizon of five weeks, the RMSE from our adjusted list-price index represents a 50 percent improvement over the forecast implied by the CME futures. For all of the forecast horizons considered in Table 6, we can reject the null hypothesis of no improvement in favor of the alternative hypothesis that the performance of the adjusted list-price index is superior.[30] This suggests that the information we exploit in our index is novel and not already known to the market. However, because these futures contracts are thinly traded and bid-ask spreads are wide, we do not claim that we have necessarily identified a profitable trading strategy.

8 Conclusion

In this paper, we have presented a new "list-price index," which attempts to fully use the information contained in listings data in order to create a more timely measure of house prices in two respects. First, the listings data contain information about the contract date at which the buyer and seller negotiate the price and allow us to associate the measure of the house's value to this date. Second, the listings data are available several months before the records of the actual transactions, allowing us to construct a measure of house prices that is available with almost no delay. In working towards these goals, our methodology includes two novel aspects that let us fully use the listings data to measure house prices. First, we link each listing to its previous sale in a manner that is fully analogous to a standard repeat-sales index and accounts for the composition of houses that are sold each month. Second, we adjust for differences between the list prices and the expected transaction prices by exploiting other information in the listings data, such as time on market and the history of list-price changes. While the timing of our index is its primary advantage, the last two points are important because ultimately it is the transaction prices, not the list prices, that

[30] In evaluating statistical significance for a given forecasting horizon, here we ignore the possibility that forecast errors may be serially correlated and we thus test for significance using a differences in means test. We make this assumption because the futures contracts are spaced three months apart, and thus the data that contribute to the forecast of one observation are essentially orthogonal to the data that contribute to the forecast of another observation.

are the standard measure of house values.

Our list-price index uses listings information only for those properties that are removed from the MLS database and therefore more likely to become sales in the coming weeks. We do not use the data on the many other properties that are still listed for sale, but one insight from our study is that this larger pool of listings will ultimately influence measured house prices based on the likelihood that they lead to sales and the expected date and price at which those sales occur. Understanding the best way to use this wealth of data to better inform our understanding of housing market conditions remains an open question for future research.

A Model Derivations

Model Predictions

This appendix presents the theoretical results from this model of the seller's problem, which illustrate how variables such as TOM, the history of list price changes, and indicators of the seller's reservation value may provide information about the heterogeneity among sellers and could therefore help us better predict variation in the sale-to-list price ratio and the probability of sale.

We start with a presentation of the main results and the intuition behind them. We provide formal proofs of these statements in the following section.

Because sellers with higher reservation values set higher list prices, they are less likely to attract a buyer. This means that for sellers with higher reservation values, a greater share of the delistings will occur because the model has ended without the arrival of a buyer. When these sellers do find a buyer, the final sale price is expected to be lower when compared to that higher list price. This implies that:

1. The sale-to-list price ratio is decreasing in the reservation value of the seller, w_i.

2. The probability of sale conditional on delisting is decreasing in the reservation value of the seller, w_i.

Next we consider the effect of time on market. In the model, a longer TOM means we are considering a seller in the second period rather than the first. With regard to the sale-to-list price ratio, there are two changes in the second period relative to the first. The first change is that all sellers who are still in the market will lower their list prices. This increases the expected sale-to-list price ratio. The second change is a difference in composition. Sellers with higher reservation values will post higher prices and be less like to match with a buyer in the first period and will therefore make up a larger fraction of sellers in the second period. Because these sellers tend to have lower sale-to-list price ratios relative to sellers with lower valuations, this change in composition will have the opposite effect. Over-all, the effect of TOM is ambiguous. However, if we control for the size of the list-price change, differences in TOM should capture only this composition effect. In this case, the model predicts that, after controlling for the changes in list-price, sellers with greater TOM are more likely to have lower sale-to-list price ratios. This implies that:

3. The sale-to-list price ratio is decreasing in TOM, holding fixed the size of the list price change.

In the second period, some sellers are able to sell their homes and some withdraw, having not met a buyer. In the first period, sellers only delist their homes if there is a sale. This means that by construction, the probability that a delisting is a sale is higher in the first period. Therefore, the model predicts that:

4. The probability of sale conditional on delisting is decreasing in TOM.

Over time, sellers tend to adjust their list prices downward and the model makes predictions about how the size of this reduction in list price is related to the sale-to-list price ratio and the probability of sale. Because sellers with higher reservation values start out with higher list prices, this mechanically increases the measured change in the size of the list price over time. On the other hand, sellers with lower reservation values are more eager to attract a buyer and face additional pressure to lower their list prices if the homes remain unsold. Which of these mechanisms is stronger depends on the values of the model parameters and in particular on how the sellers' reservation values compare to the valuation of an expected buyer. There are several possible cases:

Case 1: If $\underline{w} > (1-\beta)v$, then the size of the reduction in the list price is decreasing in the reservation value of the seller, w_i. In this case:

5a. The sale-to-list price ratio is increasing in the size of the list price reduction, holding fixed TOM.

6a. The probability of sale conditional on delisting is increasing in the size of the list price reduction, holding fixed TOM.

Case 2: If $\bar{w} < (1-\beta)v$, then the size of the reduction in the list price is increasing in the reservation value of the seller, w_i. In this case:

5b. The sale-to-list price ratio is decreasing in the size of the list price reduction, holding fixed TOM.

6b. The probability of sale conditional on delisting is decreasing in the size of the list price reduction, holding fixed TOM.

Case 3: If $(1-\beta)v$ falls within the support of the distribution of w_i, then the size of the reduction in the list price is non-monotonic in the reservation value of the seller, w_i. In this case, the model does not predict a monotonic relationship between the size of the list price reduction and the sale-to-list price ratio or the probability of sale.

We provide additional intuition for this last set of results below, when we present the formal proofs.

Proofs

This section provides proofs of the model predictions outlined above. We start with a series of propositions.

Proposition 1 *In the second period, sellers with higher reservation values post higher prices than sellers with lower reservation values.*

Proof. Working backwards, we consider the the second period problem of a seller with valuation w_i. If she posts a list price p_2, a buyer arrives with probability $\alpha_0 - \alpha_1 p_2$. Of these buyers, a fraction β will be high types, resulting in a sale at price p_2 and a fraction $1-\beta$ will be low types, resulting in a sale at price v. With probability $1-(\alpha_0-\alpha_1 p_2)$, no buyer arrives and the seller is left with value w_i.

The seller's problem

$$V_2(w_i) = \max_{p_2}(1-\alpha_0+\alpha_1 p_2)w_i + (\alpha_0-\alpha_1 p_2)(\beta p_2 + (1-\beta)v)$$

is solved by

$$p_2(w_i) = \frac{1}{2\alpha_1\beta}\left(\alpha_1 w_i + \alpha_0\beta - \alpha_1(1-\beta)v\right)$$

The derivative with respect to w_i,

$$\frac{dp_2}{dw_i} = \frac{1}{2\beta} > 0$$

so that the posted list price p_2 is higher for sellers with larger values of w_i.

Proposition 2 *In the second period, sellers with higher valuations receive, on average, lower sale prices relative to their list prices compared to sellers with lower valuations.*

Proof. Conditional on a buyer arriving, the expected sale price is

$$Ep^* = \beta p_2 + (1-\beta)v$$

The ratio of the expected sale price to the list price is given by

$$\mu_2 = Ep^*/p_2 = \beta + (1-\beta)v/p_2,$$

which is a decreasing function of the listing price. This happens simply because when the list price is not a binding constraint, the sale price is determined by the buyer's valuation. If listing price is higher, it will be higher relative to that valuation and the sale will occur at a smaller fraction of the list price. The derivative of this ratio of expected sale price to list price with respect to the seller's valuation,

$$\frac{d\mu_2}{dw_i} = -(1-\beta)vp_2^2\frac{dp_2}{dw_i} = -(1-\beta)vp_2^2\frac{1}{2\beta} < 0.$$

Because they post higher list prices, sellers with higher valuations receive, on average, a lower sale price relative to that list price.

Proposition 3 *In the second period, sellers with higher valuations are less likely to sell their homes than sellers with lower valuations.*

Proof. The probability that the house is sold is equal to $\alpha_0 - \alpha_1 p_t$, which is a decreasing function of the listing price. Since $\frac{dp_2}{dw_i} < 0$, this implies that sellers with higher value of w_i are less likely to sell their homes.

Proposition 4 *The results in Propositions 1-3 hold in the first period as well.*

Proof. Moving backwards to the first period, the seller faces the same problem except that if a buyer does not arrive in this period, the seller enters the second period so that the value of not selling is $V_2(w_i)$ rather than w_i. All of the above equations continue to hold with the substitution $V_2(w_i)$ for w_i.[31] Given the above solution for $p_2(w_i)$, we can write

$$V_2(w_i) = \frac{1}{2\alpha_1\beta}\left(\alpha_1 w_i + \alpha_0\beta - \alpha_1(1-\beta)v\right)^2 + (1-\alpha_0)w_i + \alpha_0(1-\beta)v$$

[31]This requires a further assumption that $v > V_2(w_i)$, i.e. that it is still optimal to accept an offer from a low-type buyer rather than reject that offer in hopes of matching with a high-type buyer in the second period. This assumption will hold if w_i is sufficiently low that the risk of not matching in the second period outweighs the potential gain of meeting a buyer willing to pay the second-period asking price.

so that

$$\frac{dV_2(w_i)}{dw_i} = \frac{1}{\beta}\left(\alpha_1 w + \alpha_0\beta - \alpha_0(1-\beta)v\right) + 1 - \alpha_0 = \alpha_1 p_2(w_i) + (1 - (\alpha_0 - \alpha_1 p_2(w_i))) > 0,$$

which means that $V_2(w_i)$ is a strictly increasing function and the results we derived for the second period also hold in the first period. That is, sellers with higher values of w_i have higher list prices, lower sale-to-list price ratios, and lower probability of sale in the first period as well.

Proposition 5 *Sellers lower the list price in the second period, i.e.* $p_2 < p_1$*.*

Proof. Consider the behavior of a seller who fails to attract a buyer in the first period and must now set a new list price in the second period. The change in list price from period one to period two is

$$p_2(w_i) - p_1(w_i) = \frac{1}{2\beta}\left(w_i - V_2(w_i)\right).$$

In the second period, the seller receives value w_i if no buyer arrives but a strictly higher value if one does. This implies $w_i - V_2(w_i) < 0$ so that from the above equation, the new list price is always lower than the original.

Proposition 6 *If* $\bar{w} < (1-\beta)v$*, then the decline in list prices,* $p_1 - p_2$*, is larger for sellers with higher reservation values. If* $\underline{w} > (1-\beta)v$*, then it is the sellers with lower reservation values that decrease their list prices more.*

Proof. The derivative of the change in list prices with respect to the seller's reservation value is given by:

$$\frac{d}{dw_i}\left(p_1(w_i) - p_2(w_i)\right) = \frac{1}{2\beta}\frac{d}{dw_i}\left(V_2(w_i) - w_i\right) = \frac{\alpha_1}{2\beta}\left((1-\beta)v - w_i\right).$$

If $\bar{w} < (1-\beta)v$, then $(1-\beta)v > w_i$, the derivative is positive, and the decline in list prices is larger for sellers with higher values of w_i. Alternatively, if $\underline{w} > (1-\beta)v$, then $w_i > (1-\beta)v$, the derivative is negative, and it is the sellers with *lower* reservation values that decrease their listing prices more. The intuition for this result is as follows. The seller in the first period is forward looking. A seller with a higher valuation knows that in the second period, she will set a higher list price in order to capture the higher benefit of matching with a potential buyer willing to pay that list price. A consequence of this higher list price is that it becomes less likely that she will attract a buyer in the second period. In particular, there is a lower probability that she will attract a low-type buyer and a higher probability that she will instead receive her reservation value w_i. If w_i is high compared with the expected benefit of matching with a low-type buyer ($w_i > (1-\beta)v$), this increases the value of reaching the second period. This makes the seller with the higher valuation marginally raise her list price in the first period in order to increase the probability of reaching the second period. This higher price in the first period makes the size of the list price change larger. Conversely, if w_i is low compared with the expected benefit of matching with a low-type buyer ($w_i > (1-\beta)v$), then a lower probability of matching with a low-type buyer decreases the value of reaching the second period. In this case, the seller with the relatively higher valuation will set a slightly lower list price in the first period in order to *decrease* the chance of reaching the second period. In this case, sellers with higher valuations will have smaller decline in list prices between the two periods.

References

Elliot Anenberg. Loss aversion, equity constraints and seller behavior in the real estate market. *Regional Science and Urban Economics*, 41(1):67 – 76, 2011.

Elliot Anenberg. Information frictions and housing market dynamics. *mimeo Federal Reserve Board*, 2013.

Martin J. Bailey, Richard F. Muth, and Hugh O. Nourse. A regression method for real estate price index construction. *Journal of the American Statistical Association*, 58(304): 933–942, 1963.

Eli Beracha and Michael J Seiler. The effect of listing price strategy on transaction selling prices. *The Journal of Real Estate Finance and Economics*, pages 1–19, 2013.

Grace Bucchianeri and Julia Minson. A homeowner's dilemma: Anchoring in residential real estate transactions. Technical report, mimeo, University of Pennsylvania, 2013.

John Campbell, Stefano Giglio, and Parag Pathak. Forced sales and house prices. *American Economic Review*, 2011.

Paul Carrillo, Erik De Wit, and William Larson. Can tightness in the housing market help predict subsequent home price appreciation? evidence from the u.s. and the netherlands. *mimeo George Washington University*, 2012.

Paul E. Carrillo. An empirical stationary equilibrium search model of the housing market. *International Economic Review*, 53(1):203–234, 2012.

Karl. E. Case and Robert. J. Shiller. Prices of single-family homes since 1970: New indexes for four cities. *New Eng. Econ. Rev.*, Sep/Oct:45–56, 1987.

Karl E. Case and Robert J. Shiller. The efficiency of the market for single-family homes. *The American Economic Review*, 79(1):125–137, 1989.

Karl E. Case and Robert J. Shiller. Forecasting prices and excess returns in the housing market. *Real Estate Economics*, 18(3):253–273, 1990.

Yongmin Chen and Robert W. Rosenthal. Asking prices as commitment devices. *International Economic Review*, 37(1):129–155, 1996.

David M. Cutler, James M. Poterba, and Lawrence H. Summers. Speculative dynamics. *The Review of Economic Studies*, 58(3):pp. 529–546, 1991.

Francis X Diebold and Robert S Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 20(1):134–144, 2002.

Jack Favilukis, Sydney C. Ludvigson, and Stijn van Nieuwerburgh. "the macroeconomic effects of housing wealth, housing finance, and limited risk-sharing in general equilibrium". *NBER Working Paer*, w15988, 2011.

Marjorie Flavin and Takashi Yamashita. Owner-occupied housing and the consumption of the household portfolio. *American Economics Review*, 92(1):345–362, 2002.

Joshua Gallin. The long-run relationship between house prices and rents. *Real Estate Economics*, 36(4):635–658, 2008.

David Genesove and Christopher Mayer. Loss aversion and seller behavior: Evidence from the housing market. *The Quarterly Journal of Economics*, 116(4):1233–1260, 2001.

David Genesove and Christopher J. Mayer. Equity and time to sale in the real estate market. *The American Economic Review*, 87(3):255–269, 1997.

Edward L Glaeser, Joseph Gyourko, Eduardo Morales, and Charles G Nathanson. Housing dynamics: An urban approach. *Journal of Urban Economics*, 81:45–56, 2014.

W.N. Goetzmann and M. Spiegel. "the policy implications of portfolio choice in underserved mortgage markets". *Yale International Center for Finance*, Working Paper 00-18, 2000.

Adam Guren. The causes and consequences of house price momentum. *mimeo Harvard University*, 2013.

Lu Han and Will Strange. What is the role of the asking price for a house? Technical report, mimeo, University of Toronto, 2013.

Donald Haurin. The duration of marketing time of residential housing. *Real Estate Economics*, 16(4):396–410, 1988.

Allen Head, Huw Lloyd-Ellis, and Hongfei Sun. Search, liquidity, and the dynamics of house prices and construction. *The American Economic Review*, 104(4):1172–1210, 2014.

Igal Hendel, Aviv Nevo, and Franois Ortalo-Magne. The relative performance of real estate marketing platforms: Mls versus fsbomadison.com. *American Economic Review*, 99(5): 1878–98, 2009.

Joel L. Horowitz. The role of the list price in housing markets: Theory and an econometric model. *Journal of Applied Econometrics*, 7(2):115–129, 1992.

John R. Knight. Listing price, time on market, and ultimate selling price: Causes and effects of listing price changes. *Real Estate Economics*, 30(2):213–237, 1996.

Steven D. Levitt and Chad Syverson. Market distortions when agents are better informed: The value of information in real estate transactions. *Review of Economics and Statistics*, 90(4):599–611, 2008.

Stephen Malpezzi. A simple error correction model of house prices. *Journal of Housing Economics*, 8(1):27–62, 1999.

Antonio Merlo and Francois Ortalo-Magne. Bargaining over residential real estate: evidence from england. *Journal of Urban Economics*, 56(2):192–216, September 2004.

Antonio Merlo, Franois Ortalo-Magne, and John Rust. The home selling problem: Theory and evidence. Technical report, mimeo, University of Pennsylvania, 2013.

David E. Rapach and Jack K. Strauss. Differences in housing price forecastability across {US} states. *International Journal of Forecasting*, 25(2):351 – 372, 2009.

StephenW. Salant. For sale by owner: When to use a broker and how to price the house. *The Journal of Real Estate Finance and Economics*, 4(2):157–173, 1991.

Robert J. Shiller. Arithmetic repeat sales price estimators. *Journal of Housing Economics*, 1(1):110–126, 1991.

Nancy Wallace. "real estate price measurement and stability crises". *UC Berkeley: Fisher Center for Real Estate and Urban Economics. Working Paper*, 2011.

Table 1: Summary Statistics for Delistings

This table shows summary statistics for the 1.9 million houses that are delisted from the MLS databases between 2008-2012.

$I[\cdot]$ denotes the indicator function.

Percentile	Final List/ Initial List Price	Number of List Price Changes	List Price	Days on Market	I[House Relisted Within 1 Month]	I[House Relisted Within 2 to 6 Months]
10	0.86	0	110000	0	0	0
25	0.94	0	170000	14	0	0
50	1.00	0	282500	56	0	0
75	1.00	1	469000	119	0	0
90	1.00	3	775000	196	1	1

Table 2: Variation in Sale-to-List Price Ratio and Probability of Sale

This table shows regressions of the log sale-to-list price ratio and the probability of sale on variables in the listings data. The specification in column 1 is estimated on the full sample of delistings that sell. The specification in column 2 is estimated on the full sample of delistings. In columns 3 and 4, each observation is a MSA-month and all variables are averages over all of the delistings in the MSA-month. For example, I[Sell] in MSA j in month t is the share of all delistings that result in sales in MSA j in month t. Change List Price equals one if the seller adjusted the list price at least once before delisting. "Common" List Price Level includes list prices with ten thousandths and thousandths digits equal to 5,0; 9,9; and 9,5 respectively. A "Precise" list price digit is 1,2,3,6,7, or 8 following the definition in Beracha and Seiler [2013]. In columns 2 and 4, the dependent variable, "I[Sell]", is one if the delisting results in a closed transaction and zero otherwise and the estimation uses a probit specification. In order to avoid censoring issues, we exclude delistings in 2012 from the estimation sample.

VARIABLES	(1) Log (Sale Price) - Log (List Price)	(2) I[Sell]	(3) Log (Sale Price) - Log (List Price)	(4) I[Sell]
Months on Market	-0.0154***	-0.0080***	0.0044	-0.0137
	(0.0002)	(0.0010)	(0.0050)	(0.0550)
Months on Market Squared	0.0016***	-0.0049***	0.0003	-0.0069
	(0.0000)	(0.0002)	(0.0012)	(0.0153)
I[Months on Market>6]*Months on Market	0.0095***	-0.0169***	0.0066	-0.0077
	(0.0002)	(0.0010)	(0.0060)	(0.0687)
I[Months on Market>6]*Months on Market Squared	-0.0014***	0.0055***	-0.0004	0.0085
	(0.0000)	(0.0002)	(0.0012)	(0.0160)
(Final List Price/Initial List Price)*I[Change List Price=1]	-0.0242***	-0.2044***	0.5138***	0.2443
	(0.0027)	(0.0091)	(0.1146)	(0.9143)
I[Final List Price > Initial List Price]	0.0115***	0.0655***	-0.0369	3.0318***
	(0.0008)	(0.0036)	(0.0949)	(0.7645)
I[Change List Price=1]	0.0138***	0.2118***	-0.5572***	-0.1553
	(0.0025)	(0.0083)	(0.1039)	(0.8852)
I[Final List Price < Previous Sales Price]	0.0215***	0.0577***	-0.0175***	-0.0330
	(0.0002)	(0.0009)	(0.0067)	(0.0583)
Foreclosure Dummy	0.0037***	0.1510***	0.0346**	0.1707
	(0.0003)	(0.0013)	(0.0145)	(0.1356)
I[Property Delisted < 1 month ago]	-0.0181***	-0.1163***	-0.0554***	-0.4898***
	(0.0003)	(0.0011)	(0.0101)	(0.0668)
I[2 months ago < Property Delisted < 6 months ago]	-0.0082***	-0.1016***	0.0116	0.6197***
	(0.0003)	(0.0012)	(0.0174)	(0.1374)
I[List Price Level = "Common"]	-0.0082***	-0.0315***	-0.0799***	-0.3948
	(0.0003)	(0.0011)	(0.0307)	(0.2431)
I[Hundredths Digit of List Price Level > 0]	0.0052***	0.0723***	0.0437**	0.0098
	(0.0002)	(0.0010)	(0.0198)	(0.1757)
I[Ten Thousandths Digit of List Price Level > 0]	-0.0009**	0.0061***	-0.1346***	0.3888
	(0.0004)	(0.0016)	(0.0476)	(0.4667)
I[Thousandths Digit of List Price Level = "Precise"]	0.0054***	0.0469***	0.0642*	0.9841***
	(0.0003)	(0.0014)	(0.0358)	(0.2727)
I[Thousandths Digit of List Price Level = 4 or 9]	-0.0028***	0.0294***	-0.1583***	0.3665
	(0.0003)	(0.0011)	(0.0335)	(0.2721)
I[3 < Months on Market] * I[Change List Price]		0.0536***		-0.0224
		(0.0018)		(0.2238)
I[Days since last price change < 30]		-0.0537***		-0.4658***
		(0.0015)		(0.1587)
I[Days on Market = 180]		-0.0920***		-0.1992
		(0.0027)		(0.3394)
Seasonal Dummies	X	X	X	X
MSA Dummies	X	X	X	X
Observations	902975	1552736	507	426
R-squared	0.063	0.0519	0.809	0.787

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: The Response of House Prices to Changes in Stock Prices

This table shows regression results for the different house price indices regressed against their own lags and changes in the the S&P 500 stock price index. Each house price index is computed at a weekly frequency with no smoothing across weeks. $s\&p500_t$ denotes the simple average of the daily closing prices of the S&P 500 stock price index over the seven days subsequent to week t. We estimate the regressions as seemingly unrelated regressions using feasible GLS in order to account for the correlations between the error terms in the regressions.

Dependent Variable = Log(priceindex$_t$)-Log(priceindex$_{t-2}$)

	(1) Adjusted List Price	(2) Simple List Price	(3) Sale Price	(4) Sale Price	(5) Sale Price tied to Delisting Date
Log(s&p500$_t$)-Log(s&p500$_{t-2}$)	0.2444***	0.1497**	0.0255		0.2199***
	(0.0809)	(0.0642)	(0.0355)		(0.0709)
Log(s&p500$_{t-6}$)-Log(s&p500$_{t-8}$)				0.1420***	
				(0.0347)	
Log(priceindex$_{t-2}$)-Log(priceindex$_{t-4}$)	-0.5973***	-0.5791***	-0.3189***	-0.3523***	-0.5845***
	(0.0398)	(0.0423)	(0.0642)	(0.0629)	(0.0444)
Log(priceindex$_{t-4}$)-Log(priceindex$_{t-6}$)	-0.2401***	-0.1883***	0.3226***	0.2839***	-0.1869***
	(0.0455)	(0.0492)	(0.0622)	(0.0612)	(0.0500)
Log(priceindex$_{t-6}$)-Log(priceindex$_{t-8}$)	-0.1356***	-0.0972**	0.2046***	0.1958***	-0.0571
	(0.0397)	(0.0427)	(0.0634)	(0.0612)	(0.0437)
Seasonal Dummies	X	X	X	X	X
Observations	229	229	229	229	229
R-squared	0.501	0.497	0.573	0.603	0.494

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Forecasting Performance of Simple and Adjusted List-Price Indices

This table shows the forecasting performance of the two list-price indices at different forecast horizons. The forecast horizon in the first column is measured from the date of the last observed listings data until the end of the month we are trying to forecast. The second column shows the number of months until the release of the the Case-Shiller house price index for the month we are forecasting, which is released with a two-month delay. We forecast changes in the Case-Shiller index (i.e. changes in the log of the price level). RMSE abbreviates root mean square error; MAE abbreviates mean absolute error. Each observation is a MSA-month.

Forecast Horizon (Weeks)	# Months Ahead of Case Shiller	Adjusted Index			Simple Index			Adjusted Index/Simple Index		
		RMSE	MAE	R-squared	RMSE	MAE	R-squared	RMSE	MAE	R-quared
-3	2	0.014	0.011	0.637	0.021	0.017	0.233	0.688	0.676	2.729
1	3	0.020	0.015	0.677	0.029	0.023	0.296	0.678	0.658	2.286
5	4	0.025	0.019	0.685	0.035	0.029	0.340	0.691	0.664	2.014
9	5	0.031	0.025	0.634	0.043	0.035	0.323	0.736	0.709	1.963
13	6	0.041	0.032	0.531	0.050	0.041	0.282	0.809	0.778	1.879
17	7	0.049	0.039	0.433	0.057	0.046	0.260	0.875	0.861	1.668
21	8	0.057	0.044	0.355	0.062	0.048	0.241	0.921	0.916	1.477
25	9	0.063	0.048	0.224	0.066	0.050	0.148	0.954	0.957	1.515
29	10	0.067	0.050	0.109	0.069	0.051	0.072	0.980	0.978	1.511
33	11	0.070	0.051	0.019	0.070	0.051	0.020	1.000	0.990	0.973
37	12	0.074	0.052	-0.077	0.073	0.053	-0.060	1.008	0.991	1.284
41	13	0.078	0.055	-0.191	0.078	0.056	-0.200	0.996	0.976	0.955

Table 5: Forecasting Performance of Adjusted List-Price Index Relative to Alternative Forecasts

This table compares the forecasting performance of the adjusted list-price index to a forecast regression using an AR(3) in house prices. The alternative AR(3) forecast includes seasonal dummies, controls for changes in national mortgage rates, and controls for changes in state level unemployment rates estimated separately for each MSA on the entire history of Case-Shiller values. Each observation is a MSA-month. The first column shows the number of months until the release of the the Case-Shiller house price index for the month we are forecasting. The index is released with a two-month delay.

	Root Mean Square Error			Mean Absolute Error		
# Months in advance of Case-Shiller	Forecasting Regression	Adjusted Index	Percent Improvement	Forecasting Regression	Adjusted Index	Percent Improvement
2	0.016 ***	0.014	12%	0.013 **	0.011	10%
3	0.027 ***	0.020	26%	0.020 ***	0.015	24%
4	0.037 ***	0.025	33%	0.028 ***	0.019	32%
5	0.048 ***	0.031	34%	0.037 ***	0.025	33%
6	0.056 ***	0.041	28%	0.044 ***	0.032	27%
7	0.064	0.049	23%	0.049	0.039	21%

*, **, *** denotes that we can reject the null of forecast error equality in favor of the alternative that the forecast error of the adjusted list-price index is lower at the 1, 5, and 10 percent levels according to the Diebold-Mariano test.

Table 6: Forecasting Performance Relative to CME Futures

This table compares the forecasting performance of the adjusted list-price index to the forecast inferred from Chicago Mercantile Exchange (CME) futures prices. The forecast horizon shown in the first column is measured from the date of the last observed listings data until the end of the month we are trying to forecast. The second column shows the number of months until the release of the the Case-Shiller house price index for the month we are forecasting. The index is released with a two-month delay. Performance for both the adjusted index and the CME futures prices is for the full sample of MSAs excluding Phoenix and Seattle. Futures contracts extending 18 months into the future are listed four times a year. Each of these contracts trades on a daily basis until the day preceding the Case-Shiller release day for the contract month. We use the price of the futures contract relative to the realized index value to calculate performance. Only the months in which a CME contract exists are used to calculate the performance of the adjusted list-price index.

Forecast Horizon (Weeks)	# Months Ahead of Case Shiller	Root Mean Square Error			Mean Absolute Error		
		Adjusted Index	CME Futures	Percent Improvement	Adjusted Index	CME Futures	Percent Improvement
-3	2	0.013	0.027 ***	51%	0.011	0.020 ***	48%
1	3	0.019	0.041 ***	53%	0.015	0.032 ***	54%
5	4	0.026	0.051 ***	50%	0.020	0.039 ***	50%
9	5	0.035	0.063 ***	45%	0.027	0.049 ***	45%
13	6	0.045	0.072 ***	37%	0.036	0.056 ***	37%
17	7	0.053	0.075 ***	29%	0.042	0.060 ***	30%

Table A1: Summary Statistics for Delistings by MSA

This table shows summary statistics for the samples of delistings from each MSA.

	Percentile	Final List/ Initial List Price	Number of List Price Changes	List Price	Days on Market	I[House Relisted Within 1 Month]	I[House Relisted Within 2 to 6 Months]
Chicago	10	0.82	0	99000	14	0	0
(N=192717)	25	0.92	0	149900	42	0	0
	50	1.00	1	229000	92	0	0
	75	1.00	2	349900	175	0	0
	90	1.00	4	569900	287	1	1
DC	10	0.86	0	158100	7	0	0
(N=171987)	25	0.95	0	230000	15	0	0
	50	1.00	0	345648	49	0	0
	75	1.00	1	500000	105	1	0
	90	1.00	2	747000	182	1	1
Denver	10	0.89	0	129900	0	0	0
(N=135785)	25	0.95	0	180000	14	0	0
	50	1.00	0	254900	49	0	0
	75	1.00	1	374900	112	0	0
	90	1.00	3	550000	189	1	1
LA	10	0.89	0	189000	0	0	0
(N=450167)	25	0.97	0	275000	14	0	0
	50	1.00	0	399900	49	0	0
	75	1.00	1	649000	106	0	0
	90	1.00	2	1100000	182	1	1
Phoenix	10	0.84	0	79900	0	0	0
(N=382772)	25	0.93	0	114900	14	0	0
	50	1.00	0	169900	42	0	0
	75	1.00	1	274900	104	0	0
	90	1.00	3	450000	176	1	1
SanDiego	10	0.89	0	209900	0	0	0
(N=103708)	25	0.96	0	279900	7	0	0
	50	1.00	0	389900	35	0	0
	75	1.00	1	599000	98	0	0
	90	1.00	2	960000	175	1	1
SanFran	10	0.89	0	174900	0	0	0
(N=171249)	25	0.96	0	264900	7	0	0
	50	1.00	0	435000	28	0	0
	75	1.00	1	725000	77	0	0
	90	1.00	2	1150000	133	1	1
Seattle	10	0.83	0	155900	14	0	0
(N=158921)	25	0.92	0	219777	36	0	0
	50	0.98	1	315000	84	0	0
	75	1.00	2	475000	154	0	0
	90	1.00	3	699950	246	1	1
LasVegas	10	0.83	0	85000	14	0	0
(N=198130)	25	0.93	0	118000	42	0	0
	50	1.00	0	164900	77	0	0
	75	1.00	1	245000	147	0	0
	90	1.00	3	369900	231	1	1

Figure 1: Case-Shiller House Price Index

This figure shows the Case-Shiller House Price Index for Chicago, Washington DC, Phoenix, San Francisco, Las Vegas, Denver, Los Angeles, San Diego and Seattle over the time period in which our transactions data and listings data overlap (2008 - 2012). The index in each city is normalized to 1 in January 2008.

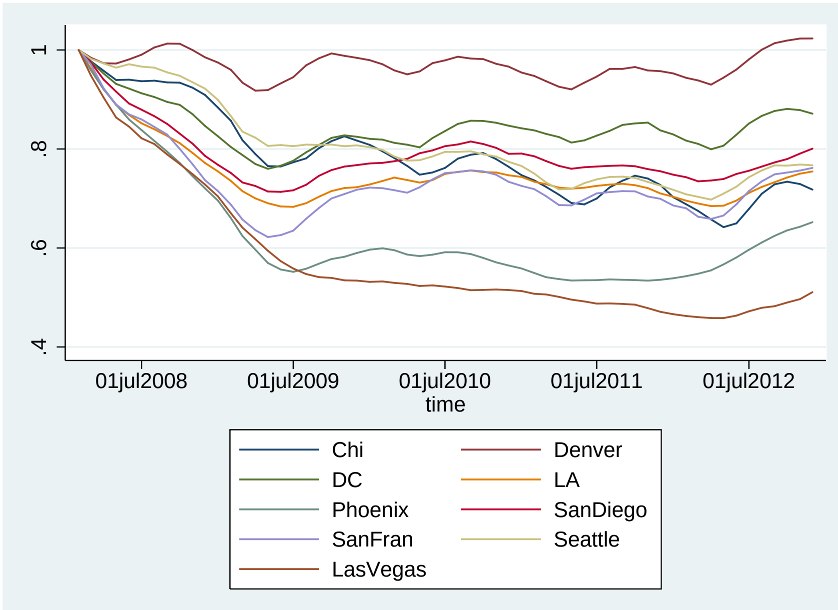

Figure 2: Median Sale-to-List Price Ratio

This figure shows the 25th, 50th, and 75th percentiles of the distribution of sale-to-list price ratios.

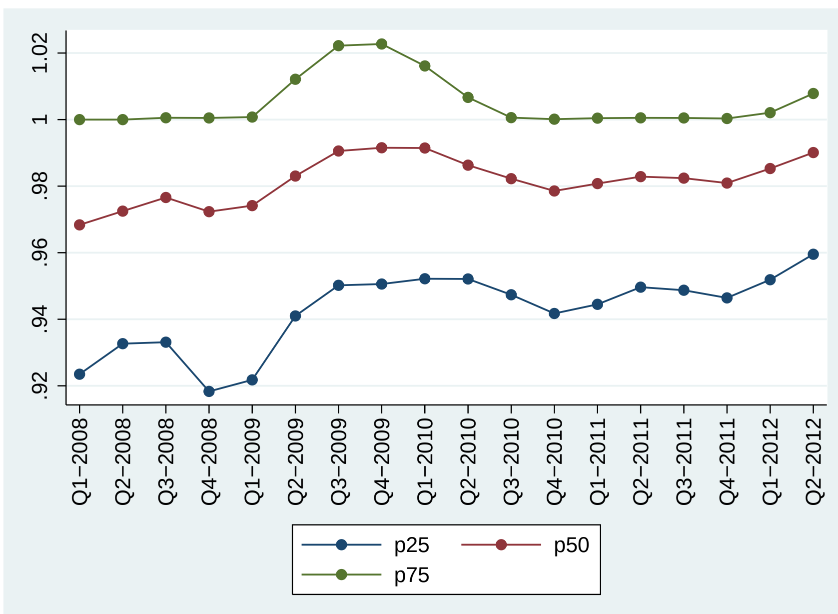

Figure 3: List Price of Withdrawals Relative to List Price of Sales

This figure shows the difference between the median log list price of houses that are withdrawn in a given quarter-year relative to the median log list price of homes that are sold. Withdrawals are defined as delistings that do not result in closed transactions, while sales are delistings that do. An estimate of time-invariant house quality is partialed out of list prices, as discussed in the main text.

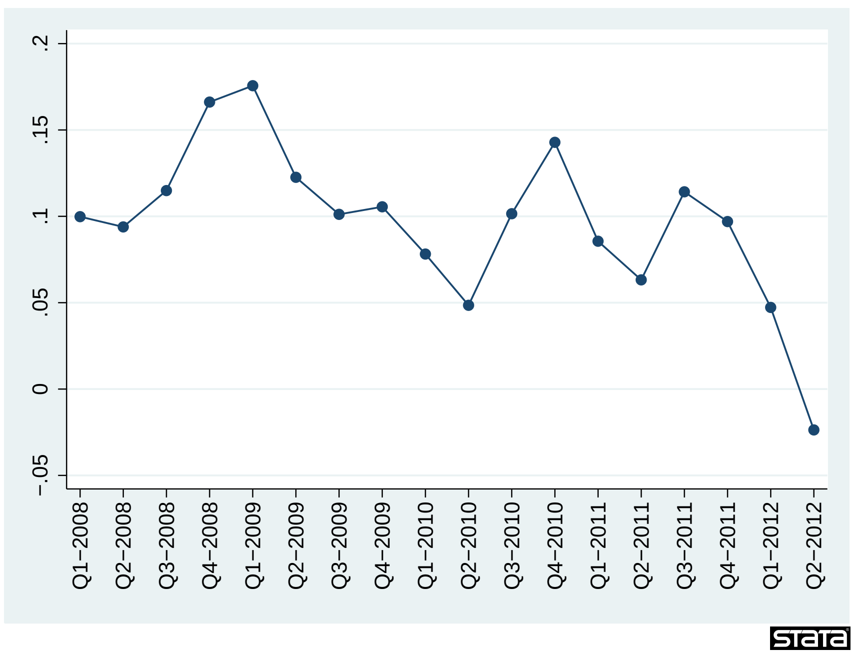

Figure 4: Share of Delistings that Result in Sales

This figure shows the share of delistings that are observed to lead to sales in each quarter.

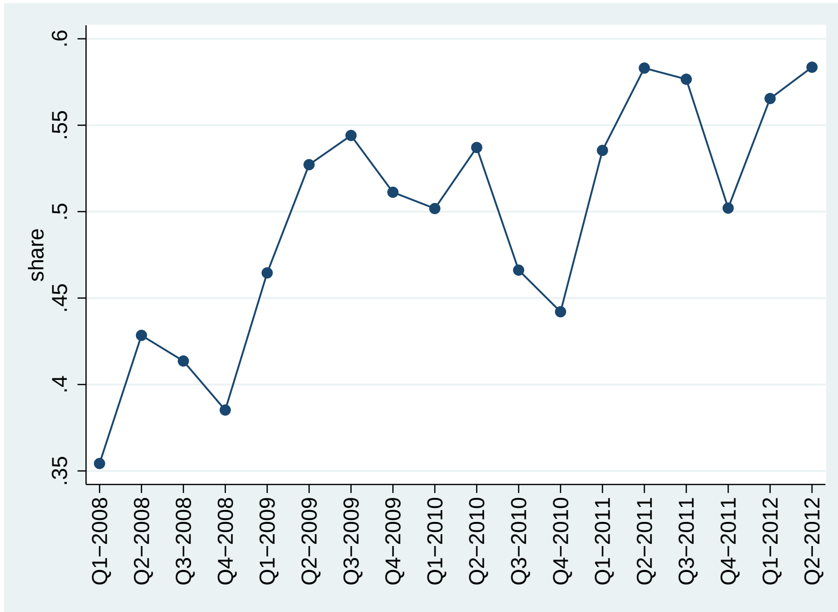

Figure 5: Share of Sales Transactions Appearing in the Listings Data

This figure shows the share of sales in each quarter and in each city that can be linked back to a listing in the MLS database.

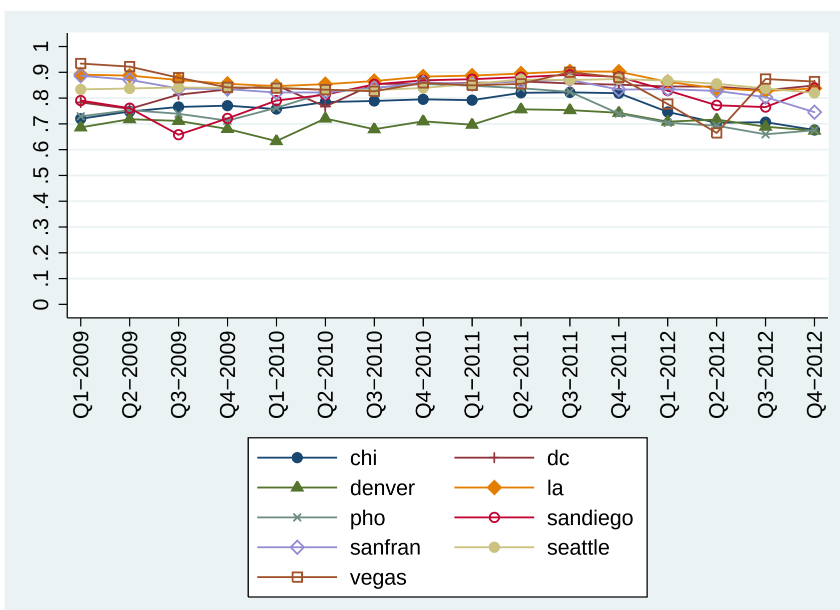

Figure 6: Adjusted List-Price Index and Sales-Price Index

This figure shows the adjusted list-price index and the sale-price index over our sample period spanned by the transactions and listings data. The two indices are computed at a weekly frequency with no smoothing across weeks. The Adjusted List Price-Index methodology is described in Section 5. The Sales Price Index is computed using the standard repeat sales approach described in Section 3.

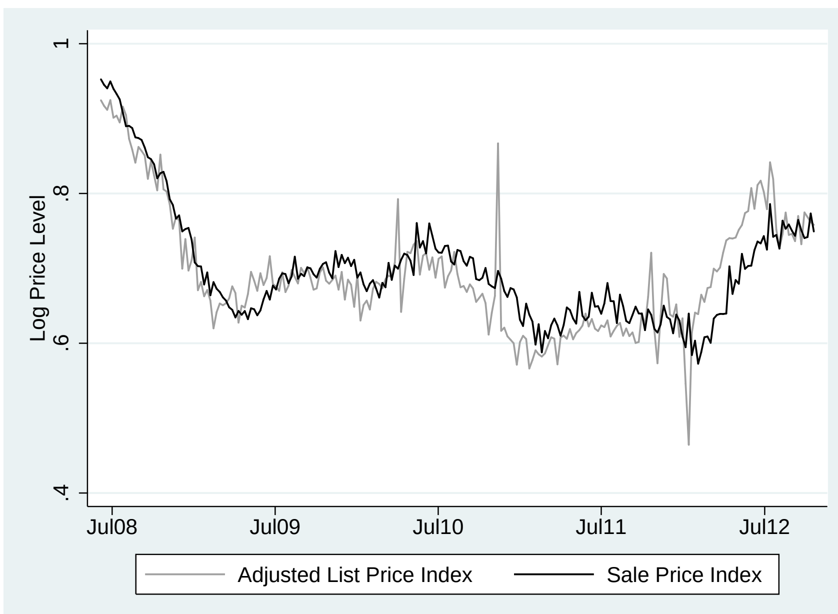

Figure 7: Stock Price Response to Case-Shiller Index Release

This figure shows the response of the stock prices of six different home-building companies to surprises in the Case-Shiller index upon its release. The surprise is measured as the difference between the released index value and market expectations based on futures contracts traded on the Chicago Mercantile Exchange. The figure shows a sample of 25 different Case-Shiller index release days for which data are available on futures prices. Changes in stock prices are measured as the opening price on the day of a Case-Shiller index release relative to the closing price on the day before. We difference off the overnight change in the S&P 500 index from each homebuilder stock price change.

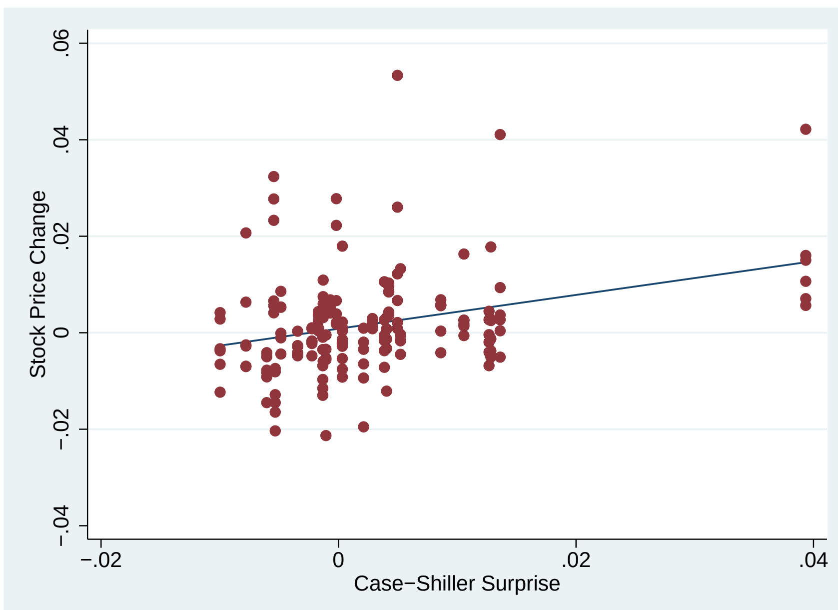

Figure 8: Lag Between Delisting and Closing Dates

This figure shows the 25th, 50th, and 75th percentiles of the distribution between closing dates and delisting dates for delistings that result in closed transactions. Closing date is when ownership of the house is transferred from the seller to the buyer, and the transaction is recorded in the public record.

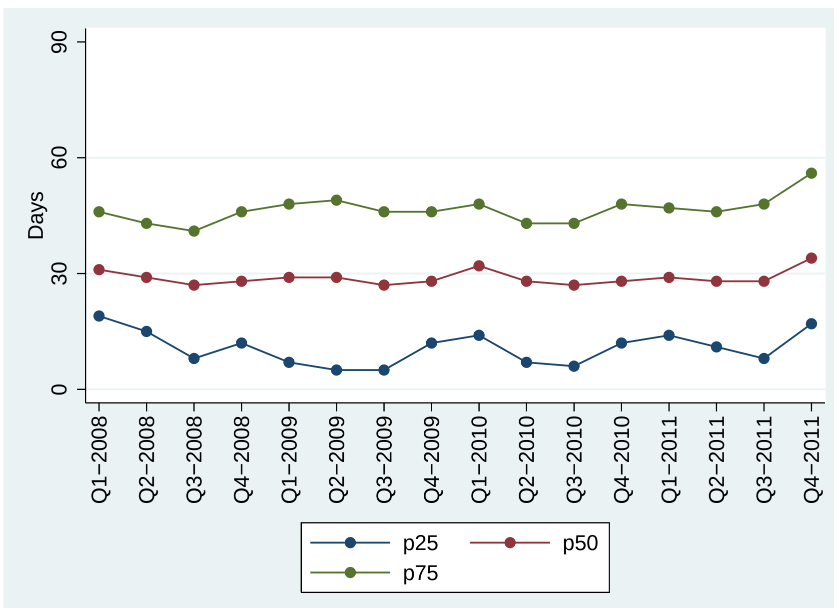

Figure 9: Minus Three-weeks-ahead Forecast of List-Price Index

The darker lines in the figure show the two-month change in house prices based on a repeat-sales index calculated following the Case-Shiller methodology. The lighter lines show the forecast of this two-month change based on adjusted list-price index at a forecasting horizon of negative three weeks, which is two months prior to the release of the Case-Shiller Index. Changes are calculated as the index value (which is the log of the price level) minus the index value two months before (which is the log of the price level from two months before).

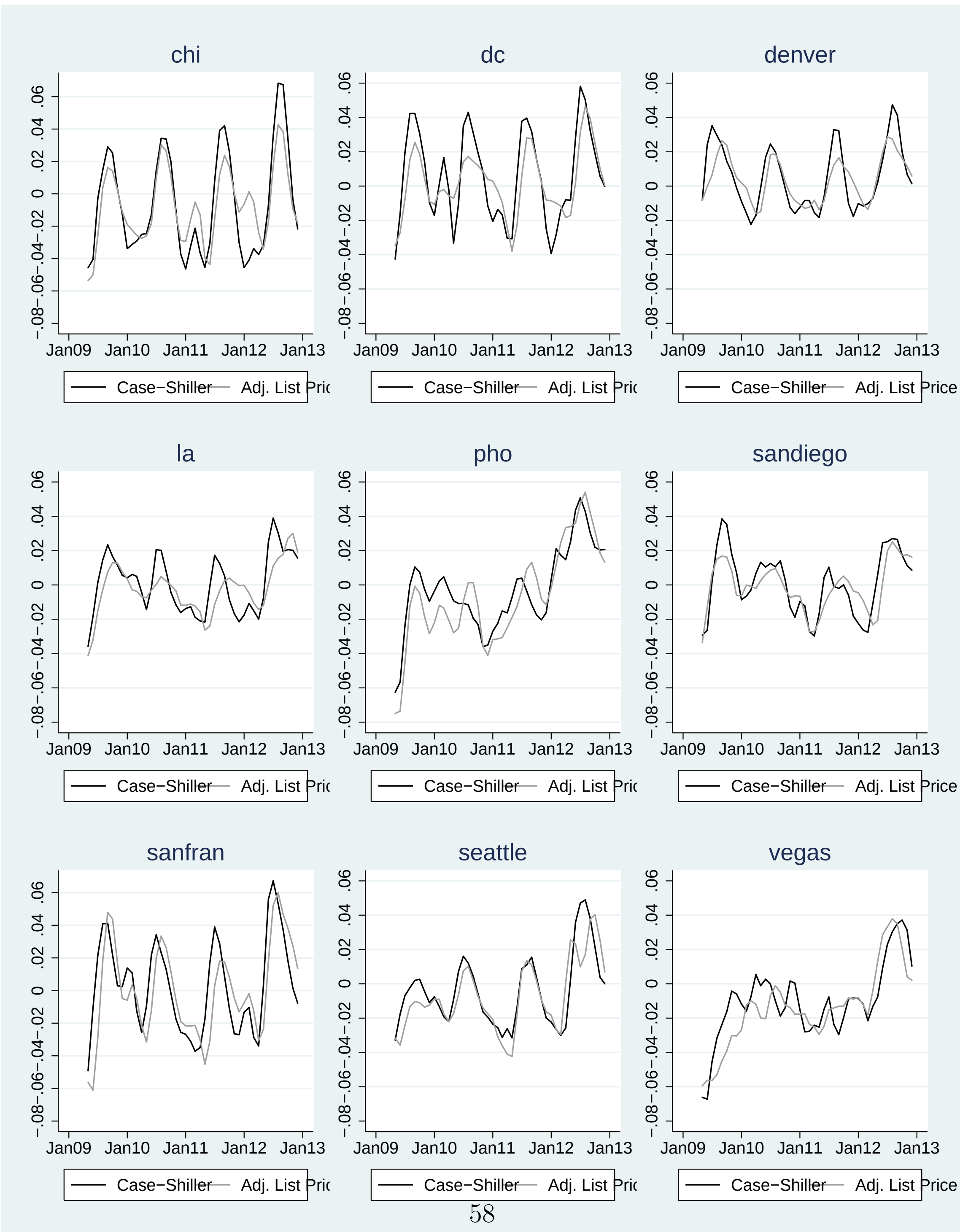

Figure 10: Five-weeks-ahead Forecast of List-Price Index

The darker lines in the figure shows the four-month change in house prices based on the repeat-sales index calculated following the Case-Shiller methodology. The lighter lines show the forecast of this four-month change based on adjusted list-price index at a forecasting horizon of five weeks. Changes are calculated as the index value (which is the log of the price level) minus the index value four months before (which is the log of the price level from four months before).

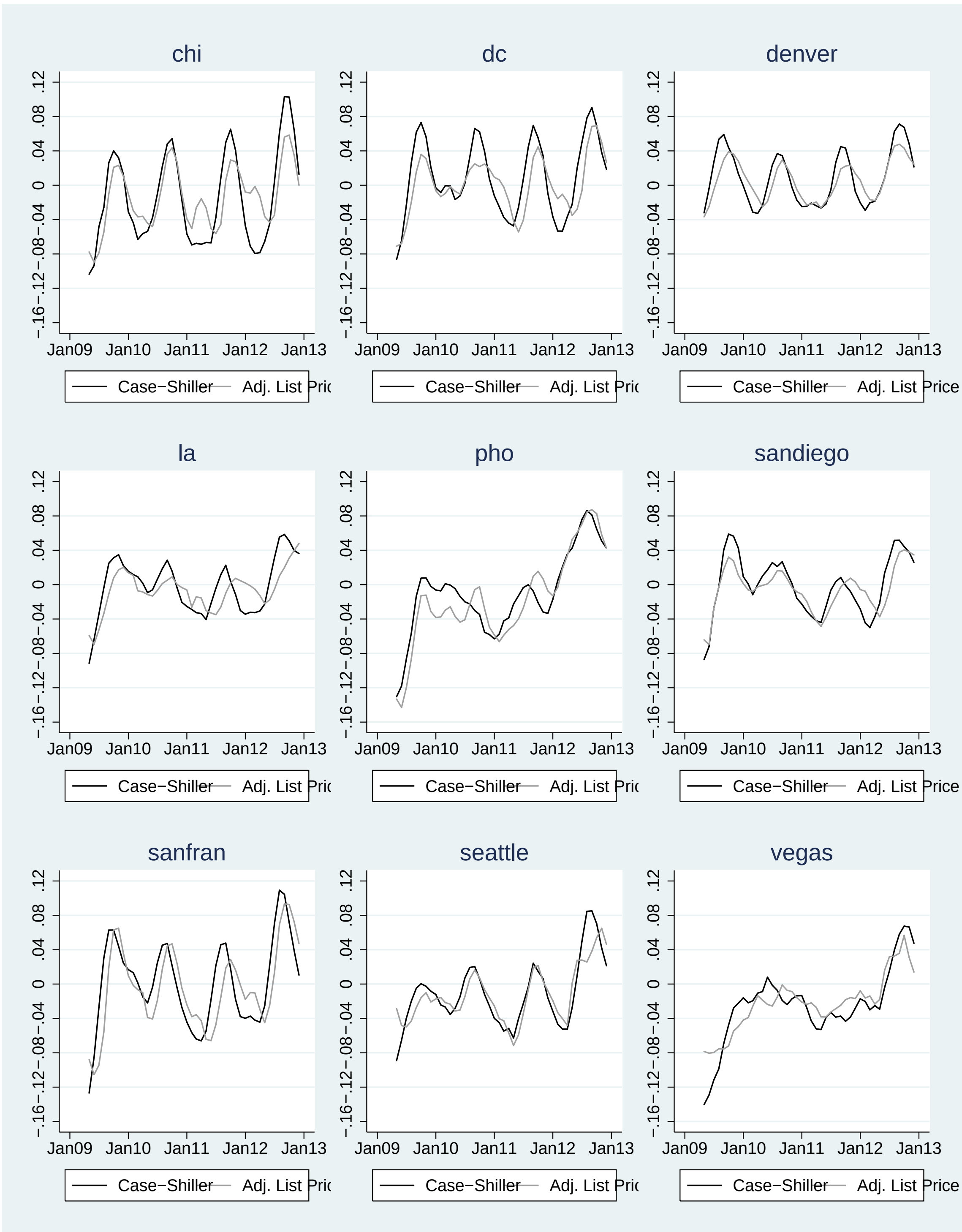

Figure 11: Forecast Errors of List-Price Index and CME Futures (6 Weeks Ahead of Case-Shiller Release)

The darker lines in the figure show the forecast error associated with futures prices on the Chicago Mercantile Exchange (CME) six weeks ahead of the Case-Shiller release. The lighter lines in the figure show the forecast error associated with the adjusted list-price index five weeks ahead of the Case-Shiller release. Forecast errors are calculated as the predicted index value (which is the predicted log of the price level) relative to true index value (which is the log of the price level).

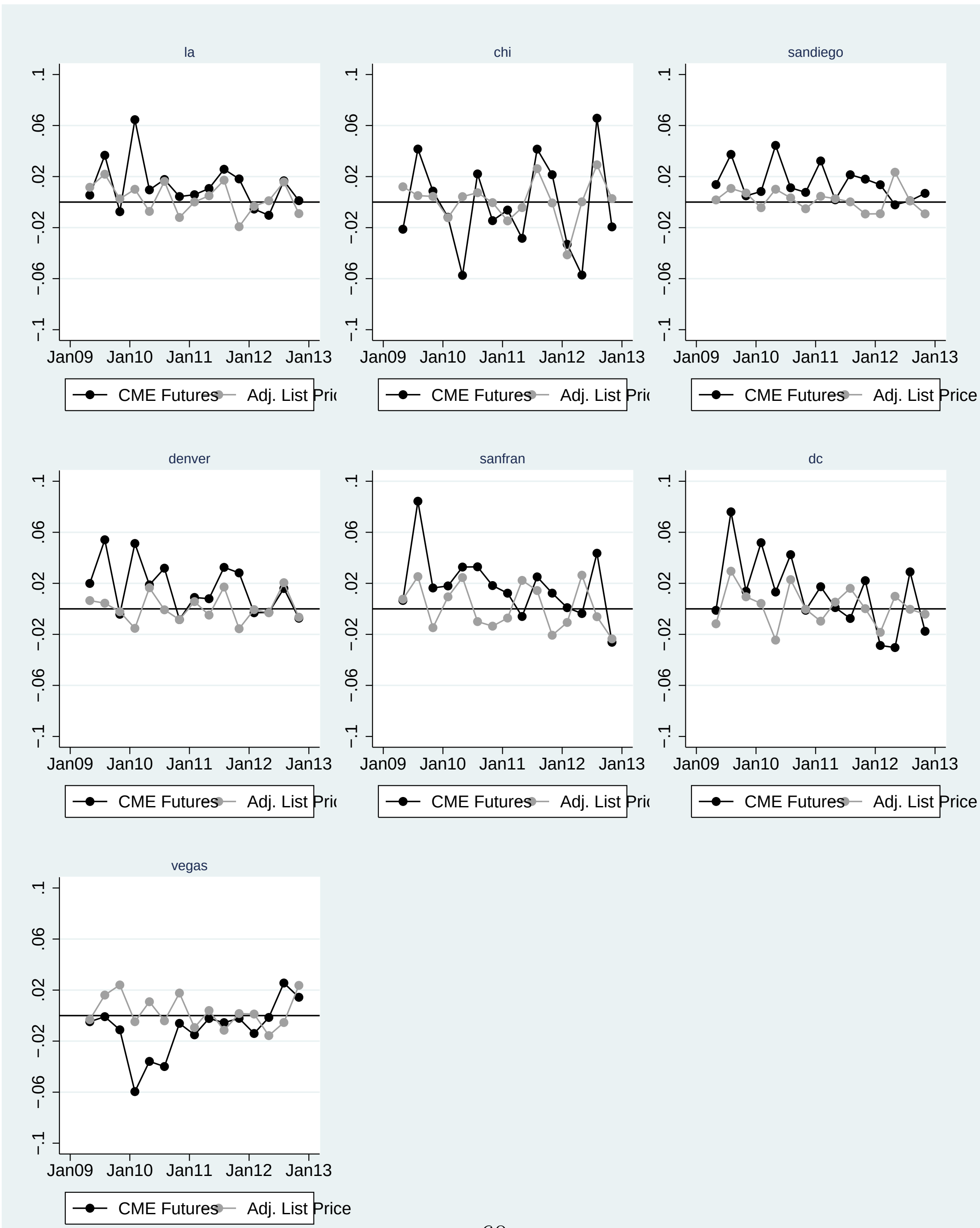

Figure 12: Forecast Errors of List-Price Index and CME Futures (10 Weeks Ahead of Case-Shiller Release)

The darker lines in the figure show the forecast error associated with futures prices on the Chicago Mercantile Exchange (CME) ten weeks ahead of the Case-Shiller release. The lighter lines in the figure show the forecast error associated with the adjusted list-price index ten weeks ahead of the Case-Shiller release. Forecast errors are calculated as the predicted index value (which is the predicted log of the price level) relative to true index value (which is the log of the price level).