Online Appendix to “A More Timely House Price Index” by Elliot Anenberg & Steven Laufer

A Online Appendix

A.1 Data Cleaning

This section provides an overview of how the listings and sales data are cleaned, and connected together. The codes available on the journal website show the precise steps taken.

First, the address variables in the listings data are cleaned for consistency. E.g., all suffixes for the word “street” are made to be consistent. Then we assign each property a unique id. Altos Research provides a propertyid, but we find that it does not always identify all listings of the same property. For example, sometimes a home that is delisted and then relisted a few weeks later with a slight variation on the address will result in two different Altos Research propertyid. We correct these cases to the extent possible using the geocoding information provided by Altos Research. For example, two homes with the same latitude and longitude, the same house and unit number are given the same houseid. Duplicate listings are also removed.¹

The sales data are in two parts: an assessor dataset and a deeds dataset. The assessor data provide house characteristics such as square feet, property address, as well as other information, for each property. The deeds data contain information about the sale transaction such as the sale price and date of sale.

The address variables in the assessor data are very neat and standardized so we do not need to do any preliminary cleaning. We drop all sales that are not arms length and other

¹A small number of weeks have a large number of duplicate listings.
problematic sales (e.g. prices below 1000 dollars). We also restrict the sales and assessor data to single family homes only.

We first combine each delisting in the listings data with the assessor data using the house number, the first four letters of the street name, the first four numbers of the zip code, and the unit number.\textsuperscript{2} In cases where a delisting matches to multiple properties in the assessor data, we give priority to stronger address matches. Now we have the Dataquick propertyid, which identifies properties in both the assessor and deeds data, merged onto each listing. To get the previous sale associated with each delisting, we merge the Dataquick propertyid to all potential sales in the deeds data and keep the sale for which the original listing date minus the closing date is the smallest, excluding sales for which this difference is less than ninety days. Delistings that cannot be linked to a previous sale are excluded from the analysis. To see if the current delisting results in a sale, we again merge on all potential sales from the deeds data and keep merges for which the closing date minus the delisting date is less than 365 days, excluding sales for which this difference is less than -30 days. In the case where a sale matches to more than one delisting, we give priority to delistings that are not relisted within six months, delistings with closing date minus the delisting date that is greater than zero, and finally, smaller values of closing date minus the delisting date. Multiple delistings of the same house can link to the same previous sale, however, we do not allow two delistings to link to the same current sale.

Overall, 82 percent of the sales in our data can be linked to a delisting within a twelve month window prior to the sale. Because of differences in coverage across MSAs and the share of foreclosures over time, there is variation in the share of sales that can be linked to delistings across both time and MSAs. We show this variation in Figure A1.

A.2 The problem of time aggregation considered in Working [1960]

To see how time aggregation mechanically introduces persistence, consider the following example related to Working [1960]. Let the price of housing follow a random walk: $p_t = p_{t-1} + \epsilon_t$ where $\epsilon$ is iid according to some distribution $F$.

\textsuperscript{2}In a small number of cases, we permit a match on house number, square feet, year built, and zip code to account for cases where the street name in the listings data is missing.
Consider an AR(1) regression of first differences
\[ p_t - p_{t-1} = \alpha_0 + \alpha_1(p_{t-1} - p_{t-2}) + \nu_t. \] (1)

Note that
\[ \hat{\alpha}_1 = \frac{\text{cov}(p_{t-1} - p_{t-2}, p_t - p_{t-1})}{\text{var}(p_{t-1} - p_{t-2})} \] (2)

and that \( \text{cov}(p_{t-1} - p_{t-2}, p_t - p_{t-1}) = \text{cov}(\epsilon_{t-1}, \epsilon_t) = 0 \implies \hat{\alpha}_1 = 0. \)

Next, consider a price index that is an average of prices this period and the period before.
The AR(1) regression of first differences is now
\[ \frac{p_t + p_{t-1}}{2} - \frac{p_{t-2} + p_{t-3}}{2} = \beta_0 + \beta_1(\frac{p_{t-2} + p_{t-3}}{2} - \frac{p_{t-4} + p_{t-5}}{2}) + \eta_t. \] (3)

Note \( \text{cov}(\frac{p_{t-2} + p_{t-3}}{2} - \frac{p_{t-4} + p_{t-5}}{2}, \frac{p_{t-2} + p_{t-3}}{2}) = \text{cov}(\epsilon_{t-2} + 2\epsilon_{t-3} + \epsilon_{t-4}, \epsilon_{t-2} + 2\epsilon_{t-3} + \epsilon_{t-2}) > 0 \)
due to the presence of \( \epsilon_{t-2} \) in both the dependent and independent variables of the regression
and implies that \( \hat{\beta}_1 > 0. \)

### A.3 Additional Forecasting Results

#### A.3.1 Motivation for Forecasting Exercise

We first present 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, as we discuss in more detail in Section A.3.3. Figure A2 shows the results of an event study relating surprises (relative to market expectations) in the 10-city Case-Shiller index to changes in the stock price of six different home building companies.\(^3\) For a sample of 25 Case-Shiller index release days for

---

\(^3\)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 included.
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.
For this result and in Figure A2, 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. In each
estimation, we cluster standard errors by release date.

A.3.2 Simulated Repeat-Sales Index Construction

This section describes the construction of our simulated repeat-sales index that we use to
generate forecasts based on the list-price index. The construction of the 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. The forecast is based on the list-price index, where the estimated transaction price
is simply the final listing price $p_{L}^{T}$ plus an average sale-to-list price ratio $\mu$. We construct
this forecast as follows:

1. For each observed delisting $i$ at time $t_{d}^{i}$, draw $R$ random realizations of the time to
closing: $l_{itr}$ for $r = 1...R$, which gives a simulated transaction that closes at date
$t_{ir} = t_{d}^{i} + l_{itr}$ with price $p_{itr} = p_{L}^{T} + \mu$.
2. To mimic the smoothing approach of Case-Shiller, generate three copies of each sim-
ulated 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 $\delta_{t}$ from

$$p_{itr} - p_{it'} + \delta_{t}^{CS} = \delta_{t_{ir}} + \eta_{itr}$$ (4)

using weighted least squares. Again following the Case-Shiller methodology, the weight-
ing depends on the elapsed time between the two transactions and also the initial sales
price, as described in Section 3.
A.3.3 Relative Forecasting Performance

Next, we compare performance of the list-price index to the performance of a standard forecasting equation that does not use listings data. To this end, we estimate the following AR(3) specification:

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

(5)

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. We estimate equation (5) separately for each city, and so the parameters in (5) depend on $j$. The estimation sample is the full sample of index values available for each city prior to the start of our sample period (typically 1988-2008). Like the list price index, the alternative forecast uses only information that would be available in real time and is not subject to any look-ahead bias.

Panel A of Table A3 shows that the gains in performance by forecasting with the list-price index are meaningful and statistically significant.\(^4\) Compared to the AR(3) model, the list-price index delivers 29 percent and 23 percent improved performance in terms of RMSE and MAE, respectively, for an estimate of the Case-Shiller index five months in advance. While the list-price index does not compare as favorably with the AR(3) model for 2 and 3 month horizons, the adjusted list price index that we describe below does significantly better than the same AR(3) model at these shorter forecasting horizons.

Finally, we would like to test whether the informational content of the listings data that we exploit in our index is already known to market participants. To this end, 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.

\(^4\)Our test statistic is a panel version of the Diebold-Mariano test statistic with a bartlett kernel (see Diebold and Mariano [2002]).
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 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 A2, which shows that surprises in the index level measured relative to these futures prices shift around stock prices in the expected way. Furthermore, Table A3 shows that the RMSE associated with 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. However, one caveat is that during our sample period, these future contracts were thinly traded and so the futures prices may not always adjust to reflect changing expectations, especially at high frequencies.

Panel B of Table A3 summarizes the performance of the list-price index compared to the performance of the futures market over our sample period. At a forecasting horizon of five weeks, the RMSE from our list-price index represents a 20 percent improvement over the forecast implied by the CME futures. For all of the forecast horizons considered in Table A3, we can reject the null hypothesis of no improvement in favor of the alternative hypothesis that the performance of the list-price index is superior. 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.

---

5Seven 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.

6In 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.
A.4 Adjusted List-Price Index

A.4.1 Discussion of the List-Price Index

The list-price index is attractive because it exploits the timely nature of listings data to construct a forecast. The connection between this version of the list-price index and the repeat-sales index, however, relies on several additional assumptions. In this section, we first identify those assumptions. In the following sections, we will develop an alternative “adjusted” 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 $\kappa_{it}$ 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 price level $\delta_t$ from equation (3).

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

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

where recall that $\varepsilon_{it}$ is the error term form the standard Case-Shiller equation and $\mu_{it}$ is the sale-to-list price ratio. We can break up this expression into several terms:

$$E(\delta_t^L \cdot (\varepsilon_{it} - \varepsilon_{i0} - \tilde{\mu}_{it})) = 0 = E(\delta_t^L \cdot (\varepsilon_{it} - \varepsilon_{i0})|\kappa_{it} > 0) \cdot Pr(\kappa_{it} > 0)$$

$$+ E(\delta_t^L \cdot (\varepsilon_{it} - \varepsilon_{i0})|\kappa_{it} < 0) \cdot Pr(\kappa_{it} < 0) - E(\delta_t^L \cdot \tilde{\mu}_{it}).$$

Equation (7) 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,

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

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,

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

...
requires that the error terms of delistings that do not result in transactions, satisfy the same exogeneity restrictions as the error terms for the observations of houses that do sell (equation (8)). 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 (7) is

\[ E(\delta_t L \cdot \tilde{\mu}_{it}) = 0, \]  

which says that the sale-to-list price ratio cannot co-vary with the time effects. In order for equation (10) to hold for each value of \( t \), the average sale-to-list price ratio must be time invariant (i.e. \( E_t(\tilde{\mu}_{it}) = 0 \)). Intuitively, movements in the sale-to-list price ratio would lead to variation in sale prices that we would not be able to identify by looking only at list prices.\(^7\)

If these three conditions discussed above are satisfied, then \( \delta_t L \) can be consistently estimated from equation (3). 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. However, the evidence presented in Section 2 suggests that selection bias arising from the types of homes that are listed on the MLS should not have a large effect on the performance of the list-price index.

In the following section, we examine the empirical relevance of each potential issue with the list-price index. First, we show that the sale-to-list price ratio varies somewhat with the housing cycle so that the final list price is a good, but not unbiased, predictor of the final sales price. Second, we show that using all delistings rather than only the ones that result in closed transactions likely creates selection bias.

### A.4.2 Empirical Evidence on the List-Price Index Assumptions

We first examine the empirical relevance of each potential issue with the list-price index outlined in Section A.4.1.

We first examine trends in the sale-to-list price ratio. Figure A3 shows the 25th, 50th, 7

---

\(^7\)This discussion assumes that that sale-to-list price ratio \( \mu_{it} \) is defined for all delistings, including those that do not lead to sales. For delistings that are not associated with sales, we think of the sale price \( p_{it} \) as an unobserved market price at which the seller could have sold the property.
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 list-price index is the inclusion of all delistings 
rather than just those that lead to sales. Figure A4 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.\footnote{In comparing list prices across properties, we normalize each list price by \( p_{it'} - \delta_{it'} \) to control for differences in house quality.} Figure A5 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 A4 
and A5 suggest that including all delistings, rather than only the ones that result in sales, 
will bias the index due to selection. 

To summarize, the empirical evidence suggests two problems with the 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. In the following sections, we explore whether we can use other information 
available at the time of delisting – such as time on market and the list price history – to 
improve upon these issues. 

A.4.3 Model 

We next 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.\(^9\) We keep the 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 \(\beta\) 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.

\(^9\)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 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 following section. Instead, we move on to a set of empirical results, all of which are consistent with the model’s predictions.

Column 1 of Table A4 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. 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 A4. 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

---

10See 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.
to sell, again consistent with the idea that these sellers have lower reservation values.\footnote{The regressions in Table A4 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].}

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 A4 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 forecasting performance of the simple list-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).

**Model Predictions: Additional Details**  This section 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. Formal proofs of these statements are available from the authors upon request.

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 $w > (1 - \beta)v$, 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.

Formal proofs of these results are available from the authors upon request.

**A.4.4 Adjusted List-Price Index: Methodology**

In this section, we outline the methodology for our adjusted list-price index, which takes advantage of the additional information in the listings data in a way that is consistent with
the evidence presented in Section A.4.3.

**Step 1: Estimate Probability of Sale and Expected Sale Price**  From the set of previous observations that are available at time $\tau$, 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 $\tau$.

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

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

   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 $\varepsilon_{it}^s \sim N(0, 1)$. The expected probability of sale conditional on $X_{it}^s$ is then $\Phi(\beta_{\tau}^s X_{it}^s)$ where $\Phi$ is the standard normal c.d.f.

2. For the sample of delistings that did sell, estimate the equation for the expected sale-to-list price ratio

   $$ p_{it} - p_{it}^L = \alpha_t^P + \beta_{\tau}^P X_{it}^P + \gamma_{\tau}^P \lambda(\hat{\beta}_{\tau}^s X_{it}^s) + \varepsilon_{it}^p $$

   using OLS, where $X_{it}^P$ is the vector of observables that explain variation in the ratio and $\alpha_t^P$ is a time fixed effect that captures additional time-series variation in the sale-to-list price ratio. Because we are looking at the sale-to-list price ratio conditional on a sale and we expect the error terms in equations (11) and (12) to be correlated, we follow Heckman [1979] and control for selection by including the term $\lambda(\hat{\beta}_{\tau}^s X_{it}^s)$, where $\lambda(\cdot)$ is the inverse Mills ratio and $\hat{\beta}_{\tau}^s$ is the probit estimate from equation (11).

**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 $\beta^p X_{it}$, we also find evidence that there is predictable time-series variation in the time effects $\alpha^p_t$, 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 $\alpha^p_t$ 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 $\hat{\alpha}^p_t$ from the equation:

$$\hat{\alpha}^p_t = \rho_0 + \rho_1 \hat{\alpha}^p_{t-1} + e_t$$

(13)

using OLS where $t$ denotes the month and $\rho_1$ 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 $\hat{\alpha}^p_t$ for $t \leq \tau - L$. Then we can estimate

$$\hat{\alpha}^p_\tau = \hat{\rho}_0 (1 + \hat{\rho}_1 + \hat{\rho}_1^2 + \ldots + \hat{\rho}_1^{L-1}) + \hat{\rho}_1^L \hat{\alpha}^p_{\tau-L}.$$  

(14)

This expression results from iteratively substituting into the right hand side of equation (13) until we get back to the observable (as of time $\tau$) estimate $\hat{\alpha}^p_{\tau-L}$. In this equation, $\hat{\rho}_0$ and $\hat{\rho}_1$ denote the OLS estimates from (13).

**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 $\tau$. We then use these estimates to calculate our price index.

1. From Step 1, the final estimate of the probability of sale is

$$\hat{\pi}^p_{it} = \Phi(\hat{\beta}^s_{it} X^s_{it}).$$

(15)

and the final estimate of the sale price for each delisting is given by

$$\hat{p}^p_{it} = p^L_{it} + \hat{\alpha}^p_{it} + \hat{\beta}^p_{it} X^p_{it} + \hat{\gamma}^p_{it} \lambda(\hat{\beta}^s_{it} X^s_{it}).$$

(16)
2. From these estimated transaction prices and probabilities, estimate price levels $\delta^{L,A}_\tau$ using the same estimating equation as we used for the simple list-price index. In this case, the equation takes the form

$$\hat{p}_{i\tau} - p_{it'} + \delta_{v'} = \delta^{L,A}_\tau + \eta_{i\tau}. \quad (17)$$

We estimate this equation using weighted least squares with weights that depend on the elapsed time between the two transactions, as described in Section ???. We further weight each observation by the estimated probability of sale, $\hat{\pi}_{i\tau}$, so that our index places greater weight on delistings that are more likely to result in sales.

The full set of regressors $X^p$ and $X^s$ used to estimate equations (12) and (11) are those shown in Table A4.\textsuperscript{12} 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 (12) 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 (12) and adjusting for serial correlation, we include instead the two-month lag of the average sale-to-list price ratio. Just to clarify, neither the lagged average sale-to-list price ratio nor the inverse mills ratio are included in the specification shown in Table A4, but they are included in the adjusted list-price index that we now construct.

Constructing a simulated repeat sales index based on the adjusted list-price index follows the same steps as for the simple list-price index presented above, but with two adjustments:

1. Rather than using the marked-up final listing price $p_{it\tau} = p^L_{it} + \mu$, use the expected sale price $p_{it\tau} = p^L_{it} + \hat{\alpha}^p_{i\tau} + \hat{\beta}^p_{i\tau} X^p_{it}$ that we estimate in the construction of the adjusted list-price index.

\textsuperscript{12}The one addition to the regressors in Table A4 is the inverse mills ratio, which has an associated coefficient estimate of -.04 and is statistically significant.
2. Multiply the weight on each simulated transaction by the probability of sale: $\hat{\pi}_{it} = \Phi(\hat{\beta}_s X_{it}^s)$

A.4.5 Performance of the Adjusted List-Price Index

Table A5 compares the performance of the adjusted list-price index to the list-price index presented in the main text. 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 A6-A7 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 significantly in early 2012 in Chicago, the list-price index did not predict that prices would fall nearly as much.

Table A6 shows that the gains in performance relative to the AR(3) model described in the main text. Compared to the AR(3) model, the adjusted list-price index delivers 50 percent and 48 percent improved performance in terms of RMSE and MAE, respectively, for an estimate of Case-Shiller five months in advance.

Table A7 summarizes the performance of the adjusted list-price index compared to the performance of the futures market over our sample period. The detail for a few select forecasting horizons is presented in Figures A8 and A9. At a forecasting horizon of five weeks, the RMSE from our adjusted list-price index represents a 47 percent improvement over the forecast implied by the CME futures.

References


Table A1: Summary Statistics for Delistings

This table shows summary statistics for the 1.9 million houses in our sample from the MLS databases that are delisted between 2008-2012. The initial and final list prices refer to the first and final list prices observed, while the “list price” is the final list price in nominal dollars. Houses are counted as “relisted” if they subsequently appear in the MLS databases within 6 months of having being delisted. $I[\cdot]$ denotes the indicator function.

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Final List/Initial List Price</th>
<th>Number of List Price Changes</th>
<th>Days on Market</th>
<th>$I[\text{House Relisted Within 1 Month}]$</th>
<th>$I[\text{House Relisted Within 2 to 6 Months}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.86</td>
<td>0</td>
<td>110000</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>0.94</td>
<td>0</td>
<td>170000</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>1.00</td>
<td>0</td>
<td>282500</td>
<td>56</td>
<td>0</td>
</tr>
<tr>
<td>75</td>
<td>1.00</td>
<td>1</td>
<td>469000</td>
<td>119</td>
<td>0</td>
</tr>
<tr>
<td>90</td>
<td>1.00</td>
<td>3</td>
<td>775000</td>
<td>196</td>
<td>1</td>
</tr>
</tbody>
</table>
Table A2: Summary Statistics for Time Series Variables

This table shows the distribution of the time series variables used in the analysis in Section 4. “Stock prices” refers to the S&P 500 stock price index. News surprises refers to Citigroup U.S. Economic Surprise Index. To construct their surprise index, Citi measures the surprises in units of standard deviation, weights them according to their importance in terms of their previous impact on market prices, and then constructs a moving average of the surprises over the past 90 days using a (roughly) exponential weighting scheme. The level of each variable is computed at a weekly frequency with no smoothing across weeks. For stock prices, news surprises and treasury rates, we compute the weekly value as a simple average of the daily values. With the exception of the surprise index (where the table shows the level), the table reports the weekly change in the log of each variable.

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Change in Contract Dated Index</th>
<th>Change in Closing Dated Index</th>
<th>Change in Stock Prices</th>
<th>News Surprises</th>
<th>Change in 10yr Treasury Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>-0.033</td>
<td>-0.027</td>
<td>-0.041</td>
<td>-0.083</td>
<td>-0.074</td>
</tr>
<tr>
<td>25</td>
<td>-0.020</td>
<td>-0.017</td>
<td>-0.015</td>
<td>-0.039</td>
<td>-0.046</td>
</tr>
<tr>
<td>50</td>
<td>-0.002</td>
<td>-0.003</td>
<td>0.008</td>
<td>0.009</td>
<td>-0.004</td>
</tr>
<tr>
<td>75</td>
<td>0.013</td>
<td>0.012</td>
<td>0.023</td>
<td>0.036</td>
<td>0.030</td>
</tr>
<tr>
<td>90</td>
<td>0.029</td>
<td>0.024</td>
<td>0.036</td>
<td>0.058</td>
<td>0.069</td>
</tr>
</tbody>
</table>
Table A3: Relative Forecasting Performance of List-Price Index

This table compares the ability of the simulated repeat-sales index (based on the list-price index) to forecast the Case-Shiller house price index compared to (i) a forecast regression using an AR(3) in house price changes and (ii) a forecast inferred from Chicago Mercantile Exchange (CME) futures prices. The simulated repeat sales index simulates closing dates for delistings and assigns them to calendar months. 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. In Panel A, we compare this forecast to an alternative AR(3) forecast, which includes seasonal dummies, national mortgage rates, and state level unemployment rates, and is estimated separately for each MSA. In Panel B, we compare our list-price forecast to the forecast inferred from CME futures prices 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 list-price index. In both panels, *, **, *** denote 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 respectively according to the Diebold-Mariano test.

<table>
<thead>
<tr>
<th># Months in advance of Case-Shiller</th>
<th>Root Mean Square Error</th>
<th>Mean Absolute Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Alternative Forecast Model</td>
<td>List-Price Index</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel A: Alternative Forecast model is AR(3)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.022</td>
<td>0.023</td>
</tr>
<tr>
<td>3</td>
<td>0.038</td>
<td>0.031</td>
</tr>
<tr>
<td>4</td>
<td>0.048 **</td>
<td>0.038</td>
</tr>
<tr>
<td>5</td>
<td>0.064 **</td>
<td>0.045</td>
</tr>
<tr>
<td>6</td>
<td>0.079 **</td>
<td>0.053</td>
</tr>
<tr>
<td>7</td>
<td>0.096 **</td>
<td>0.059</td>
</tr>
<tr>
<td><strong>Panel B: Alternative Forecast Model is CME Futures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.027 **</td>
<td>0.021</td>
</tr>
<tr>
<td>3</td>
<td>0.041 **</td>
<td>0.032</td>
</tr>
<tr>
<td>4</td>
<td>0.051 **</td>
<td>0.040</td>
</tr>
<tr>
<td>5</td>
<td>0.063 **</td>
<td>0.048</td>
</tr>
<tr>
<td>6</td>
<td>0.072 **</td>
<td>0.057</td>
</tr>
<tr>
<td>7</td>
<td>0.075 **</td>
<td>0.063</td>
</tr>
</tbody>
</table>


Table A4: Variation in Sale-to-List Price Ratio and Probability of Sale

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, \( I[\text{Sell}] \) 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. Final List Price/Initial List Price is the list price at week of delisting divided by the list price in the week the home was first listed. Previous Sales Price denotes the sale price associated with the home’s most recent sale prior to the current listing episode. “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, “\( I[\text{Sell}] \)”, 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. All specifications include MSA and seasonal dummies.

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Log (Sale Price) - Log (List Price)</td>
<td>( I[\text{Sell}] )</td>
<td>Log (Sale Price) - Log (List Price)</td>
<td>( I[\text{Sell}] )</td>
</tr>
<tr>
<td>Months on Market</td>
<td>-0.0143***</td>
<td>-0.0191***</td>
<td>-0.0021</td>
<td>0.0930*</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.0009)</td>
<td>(0.0059)</td>
<td>(0.0526)</td>
</tr>
<tr>
<td>Months on Market Squared</td>
<td>0.0014***</td>
<td>-0.0027***</td>
<td>0.0012</td>
<td>-0.0286*</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0002)</td>
<td>(0.0015)</td>
<td>(0.0153)</td>
</tr>
<tr>
<td>( I[\text{Months on Market}&gt;6]) * Months on Market</td>
<td>0.0089***</td>
<td>-0.0056***</td>
<td>0.0219***</td>
<td>-0.0686</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.0008)</td>
<td>(0.0069)</td>
<td>(0.0639)</td>
</tr>
<tr>
<td>( I[\text{Months on Market}&gt;6]) * Months on Market Squared</td>
<td>-0.0013***</td>
<td>0.0032***</td>
<td>-0.0023</td>
<td>0.0285*</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0002)</td>
<td>(0.0015)</td>
<td>(0.0157)</td>
</tr>
<tr>
<td>( \text{Final List Price}/\text{Initial List Price} ) * ( \text{Change List Price}=1 )</td>
<td>-0.0227***</td>
<td>-0.2064***</td>
<td>0.4089***</td>
<td>-1.6316*</td>
</tr>
<tr>
<td></td>
<td>(0.0026)</td>
<td>(0.0080)</td>
<td>(0.1369)</td>
<td>(0.9260)</td>
</tr>
<tr>
<td>( \text{Final List Price} &gt; \text{Initial List Price} ) * ( \text{Change List Price}=1 )</td>
<td>0.0117***</td>
<td>0.0417***</td>
<td>0.5486***</td>
<td>4.3955***</td>
</tr>
<tr>
<td></td>
<td>(0.0008)</td>
<td>(0.0032)</td>
<td>(0.1001)</td>
<td>(0.6948)</td>
</tr>
<tr>
<td>( \text{Change List Price}=1 )</td>
<td>0.0121***</td>
<td>0.2228***</td>
<td>-0.4605***</td>
<td>0.7166</td>
</tr>
<tr>
<td></td>
<td>(0.0025)</td>
<td>(0.0074)</td>
<td>(0.1235)</td>
<td>(0.9033)</td>
</tr>
<tr>
<td>( \text{Final List Price} &lt; \text{Previous Sales Price} )</td>
<td>0.0334***</td>
<td>0.0393***</td>
<td>-0.0070</td>
<td>-0.1259*</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.0008)</td>
<td>(0.0085)</td>
<td>(0.0662)</td>
</tr>
<tr>
<td>Foreclosure Dummy</td>
<td>0.0012***</td>
<td>0.2131***</td>
<td>0.0445***</td>
<td>0.0641</td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.0012)</td>
<td>(0.0168)</td>
<td>(0.1474)</td>
</tr>
<tr>
<td>( \text{Property Delisted} &lt; 1 \text{ month ago} )</td>
<td>-0.0203***</td>
<td>-0.0434***</td>
<td>-0.0498***</td>
<td>-0.1573***</td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.0010)</td>
<td>(0.0100)</td>
<td>(0.0596)</td>
</tr>
<tr>
<td>( 2 \text{ months ago} &lt; \text{Property Delisted} &lt; 6 \text{ months ago} )</td>
<td>-0.0106***</td>
<td>-0.0118***</td>
<td>0.0008</td>
<td>0.2537***</td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.0011)</td>
<td>(0.0116)</td>
<td>(0.0687)</td>
</tr>
<tr>
<td>( \text{List Price Level} = \text{&quot;Common&quot;} )</td>
<td>-0.0079***</td>
<td>-0.0220***</td>
<td>0.0709**</td>
<td>-1.1189***</td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.0010)</td>
<td>(0.0329)</td>
<td>(0.2473)</td>
</tr>
<tr>
<td>( \text{Hundredths Digit of List Price Level} &gt; 0 )</td>
<td>0.0053***</td>
<td>0.0522***</td>
<td>-0.0872***</td>
<td>0.3314**</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.0009)</td>
<td>(0.0195)</td>
<td>(0.1471)</td>
</tr>
<tr>
<td>( \text{Ten Thousandths Digit of List Price Level} &gt; 0 )</td>
<td>-0.0009**</td>
<td>0.0065***</td>
<td>-0.1925***</td>
<td>1.0707**</td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
<td>(0.0015)</td>
<td>(0.0496)</td>
<td>(0.4272)</td>
</tr>
<tr>
<td>( \text{Thousandths Digit of List Price Level} = \text{&quot;Precise&quot;} )</td>
<td>0.0051***</td>
<td>0.0343***</td>
<td>0.3344***</td>
<td>1.1817***</td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.0013)</td>
<td>(0.0304)</td>
<td>(0.2184)</td>
</tr>
<tr>
<td>( \text{Thousandths Digit of List Price Level} = 4 \text{ or } 9 )</td>
<td>-0.0031***</td>
<td>0.0269***</td>
<td>0.1250***</td>
<td>0.4974***</td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.0010)</td>
<td>(0.0243)</td>
<td>(0.1709)</td>
</tr>
<tr>
<td>( 3 &lt; \text{Months on Market} ) * ( \text{Change List Price} )</td>
<td>-0.0442***</td>
<td>0.0442***</td>
<td>0.5416**</td>
<td>(0.2296)</td>
</tr>
<tr>
<td></td>
<td>(0.0016)</td>
<td>(0.0016)</td>
<td>(0.0243)</td>
<td>(0.1659)</td>
</tr>
<tr>
<td>( \text{Days since last price change} &lt; 30 )</td>
<td>-0.0447***</td>
<td>0.0368**</td>
<td>0.3568**</td>
<td>(0.1659)</td>
</tr>
<tr>
<td></td>
<td>(0.0013)</td>
<td>(0.0032)</td>
<td>(0.0304)</td>
<td>(0.2184)</td>
</tr>
<tr>
<td>( \text{Days on Market} = 180 )</td>
<td>-0.0814***</td>
<td>0.3387</td>
<td>(0.0023)</td>
<td>(0.3627)</td>
</tr>
</tbody>
</table>

Observations: 933627
R-squared: 0.063

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

23
Table A5: Forecasting Performance of Adjusted List-Price

This table shows the forecasting performance of the simulated repeat-sales index (based on the adjusted list-price index) at different forecast horizons. The simulated repeat sales index simulates closing dates for delistings and assigns them to calendar months. 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.

<table>
<thead>
<tr>
<th>Forecast Horizon (Weeks)</th>
<th># Months Ahead of Case Shiller</th>
<th>List-Price Index</th>
<th>Adjusted List-Price Index</th>
<th>Adjusted Index/List-Price Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RMSE  MAE  R-squared</td>
<td>RMSE  MAE  R-squared</td>
<td>RMSE    MAE</td>
</tr>
<tr>
<td>-3</td>
<td>2</td>
<td>0.023  0.018  0.146</td>
<td>0.015  0.012  0.616</td>
<td>0.671  0.646</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>0.031  0.026  0.233</td>
<td>0.020  0.016  0.668</td>
<td>0.657  0.629</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>0.038  0.032  0.288</td>
<td>0.025  0.020  0.688</td>
<td>0.662  0.640</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>0.045  0.038  0.289</td>
<td>0.032  0.025  0.652</td>
<td>0.700  0.673</td>
</tr>
<tr>
<td>13</td>
<td>6</td>
<td>0.053  0.043  0.261</td>
<td>0.041  0.032  0.562</td>
<td>0.770  0.743</td>
</tr>
<tr>
<td>17</td>
<td>7</td>
<td>0.059  0.047  0.248</td>
<td>0.050  0.039  0.468</td>
<td>0.841  0.825</td>
</tr>
<tr>
<td>21</td>
<td>8</td>
<td>0.065  0.050  0.230</td>
<td>0.058  0.044  0.387</td>
<td>0.892  0.877</td>
</tr>
<tr>
<td>25</td>
<td>9</td>
<td>0.070  0.052  0.128</td>
<td>0.065  0.048  0.252</td>
<td>0.926  0.920</td>
</tr>
<tr>
<td>29</td>
<td>10</td>
<td>0.073  0.053  0.040</td>
<td>0.070  0.050  0.130</td>
<td>0.952  0.949</td>
</tr>
<tr>
<td>33</td>
<td>11</td>
<td>0.076  0.054  -0.038</td>
<td>0.074  0.052  0.022</td>
<td>0.971  0.960</td>
</tr>
<tr>
<td>37</td>
<td>12</td>
<td>0.080  0.056  -0.137</td>
<td>0.078  0.054  -0.086</td>
<td>0.978  0.965</td>
</tr>
<tr>
<td>41</td>
<td>13</td>
<td>0.085  0.059  -0.286</td>
<td>0.082  0.057  -0.208</td>
<td>0.969  0.968</td>
</tr>
</tbody>
</table>
Table A6: Forecasting Performance of Adjusted List-Price Index Relative to Alternative Forecasts

This table compares the forecasting performance of the simulated repeat-sales index (based on the adjusted list-price index) to a forecast regression using an AR(3) in house price changes. The simulated repeat sales index simulates closing dates for delistings and assigns them to calendar months. The alternative AR(3) forecast includes seasonal dummies, national mortgage rates, and state level unemployment rates, and is estimated separately for each MSA. 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.

<table>
<thead>
<tr>
<th># Months in advance of Case-Shiller</th>
<th>Root Mean Square Error</th>
<th>Mean Absolute Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Forecasting Regression</td>
<td>Adjusted List-Price Index</td>
</tr>
<tr>
<td>2</td>
<td>0.022 **</td>
<td>0.015</td>
</tr>
<tr>
<td>3</td>
<td>0.036 **</td>
<td>0.020</td>
</tr>
<tr>
<td>4</td>
<td>0.048 **</td>
<td>0.025</td>
</tr>
<tr>
<td>5</td>
<td>0.064 **</td>
<td>0.032</td>
</tr>
<tr>
<td>6</td>
<td>0.079 **</td>
<td>0.041</td>
</tr>
<tr>
<td>7</td>
<td>0.096 **</td>
<td>0.050</td>
</tr>
</tbody>
</table>

*,, **, *** 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 A7: Forecasting Performance of Adjusted List-Price Index Relative to CME Futures

This table compares the forecasting performance of the simulated repeat-sales index (based on the adjusted list-price index) to the forecast inferred from Chicago Mercantile Exchange (CME) futures prices. The simulated repeat sales index simulates closing dates for delistings and assigns them to calendar months. 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.

<table>
<thead>
<tr>
<th>Forecast Horizon (Weeks)</th>
<th># Months Ahead of Case Shiller</th>
<th>Root Mean Square Error</th>
<th>Mean Absolute Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Adjusted List-Price</td>
<td>CME Futures</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Horizon Ahead</td>
<td></td>
</tr>
<tr>
<td>-3</td>
<td>2</td>
<td>0.014</td>
<td>0.027</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>0.020</td>
<td>0.041</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>0.026</td>
<td>0.051</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>0.034</td>
<td>0.063</td>
</tr>
<tr>
<td>13</td>
<td>6</td>
<td>0.044</td>
<td>0.072</td>
</tr>
<tr>
<td>17</td>
<td>7</td>
<td>0.053</td>
<td>0.075</td>
</tr>
</tbody>
</table>

*, **, *** 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.
Figure A1: Share of Sales Transactions Appearing in the Listings Data

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

The MSAs are Chicago, Washington DC, Denver, Los Angeles, Phoenix, San Diego, San Francisco, Seattle and Las Vegas.
Figure A2: 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 A3: 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 A4: 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 A5: 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 A6: Minus Three-weeks-ahead Forecast of Adjusted 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 A7: Five-weeks-ahead Forecast of Adjusted List-Price Index

The darker lines in the figure show 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 A8: Forecast Errors of Adjusted 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 A9: Forecast Errors of Adjusted 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).