A MORE TIMELY HOUSE PRICE INDEX

Elliot Anenberg and Steven Laufer*

Abstract—Using listings data, we construct a new repeat-sales house price index that describes house values at the contract date when the price is determined rather than the closing date when the property is transferred. We show that this difference in timing helps explain several puzzles about house prices, including their strong short-term serial correlation and their weak correlation with stock prices and macroeconomic news shocks. In addition, we show that a variant of our index that relies exclusively on listings data for recent transactions accurately reveals trends in house prices several months before existing price indexes like Case-Shiller become available.

1. Introduction

Changes in house prices have important consequences for the real economy because they affect both households’ wealth and their ability to borrow. However, the measurement of house prices differs from those of other important assets, such as stocks and bonds, in ways that could distort our understanding of how and why house prices change over time. For financial assets, prices describe the value of the assets at the trade 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. The difference between the two dates is significant, as the delay between contract and closing dates is typically one to two months and exhibits substantial variation.

The timing of home sales to the closing date rather than the contract date is a consequence of the available data. Analysts collect information about home sales from local government offices, where, in most states, housing transactions must be recorded at the time of closing. While local jurisdictions require that these documents contain the closing date, the contract date is not recorded. As a result, analysts building house price measures based solely on these documents are forced to associate the sale with the closing date.

In this paper, we develop a new house price index that exploits the informational content of listings data to overcome this difficulty in the measurement of house prices. Unlike the sales records, listings data contain information about the date a property is no longer listed for sale. If we observe that the house is sold soon after this delisting date, we assume that the property was delisted because the seller found a buyer and the two parties reached agreement on a sale price. In this case, we associate the contract date with this delisting date and identify the sale price as reflecting the market value of the house on that date. Aside from this difference in timing, our index is constructed using the same methodology as the Case-Shiller index and other conventional repeat-sales indexes. This allows us to isolate the role of this timing in generating some of the puzzling time series properties of house prices.

We demonstrate the advantages of our contract-dated house price index relative to a conventional closing-dated index using a sample of 2 million listings drawn from nine major U.S. metropolitan areas from 2008 to 2012. First, we take advantage of the fact that our contract-dated index for any particular time period describes a set of transactions negotiated in that time period. In contrast, a standard closing-dated index describes 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 contract-dated index is positively correlated with contemporaneous movements in equity prices, whereas the conventional closing-dated index is not. The latter index is instead correlated with lagged equity prices and the strength of this correlation is attenuated, consistent with the expected effect of measurement error in the timing of when house prices are established. We find similar results in studying the relationship of house prices to macroeconomic news shocks.

We also provide evidence that the timing convention for conventional closing-dated measures explains much of the positive short-term serial correlation in house prices at high frequencies. Intuitively, the timing of the closing-dated 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 among the growth rates over this time period. For example, while the two-week growth rates in a standard repeat-sales index are positively autocorrelated, our contract-dated index exhibits 0 autocorrelation at this frequency. Nonetheless, at longer frequencies, consistent with the fact that the lag between agreement and closing is rarely longer than several months, we reproduce the puzzling and well-documented sizable positive autocorrelation in house prices even using our new index.

In addition to containing information about the contract date, listings data have a second advantage that they are available almost immediately, while sales data are often not...
available for several additional months. This delay in sales data reporting arises 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, typically a couple of months following the sale agreement, there is another delay of a couple of additional months before the public record becomes available.\(^1\) 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 (see Chen & Rosenthal, 1996). Thus, information on listing prices is disseminated on Internet platforms such as Multiple Listing Service (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.

To take advantage of the timely availability of listings data, we construct an alternative list-price index that relies exclusively on listings data for information on recent transactions instead of waiting for the associated sales records to become available. A key aspect of our methodology is that while we look at delistings for information about recent sales, we use sales data as a source of information about the previous sale of that property. Linking a recent delisting to a prior sale 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 indexes. This is important because it allows us to provide a more timely index of house price trends without 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 (a) 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 (b) as in our main contract-dated index, the price of that second sale is associated with the delisting date rather than the eventual closing date.\(^2\)

Because listings data reveal information about house prices sooner than sales records, we are able to use them to forecast more standard house price measures, which depend on the availability of the sales records. We show that our list-price index (a) accurately forecasts the Case-Shiller index several months in advance, (b) outperforms forecasting models that do not use listings data, and (c) outperforms the market’s expectation as inferred from prices on Case-Shiller futures contracts. We find that adjusting for variation in sale-to-list price ratios and propensity to sell using other observable information on seller behavior in the listing data, such as the time on market (TOM) and the history of list price changes, reduces our forecasting errors by approximately 30% relative to the simpler model in which we neither adjust list prices nor weight delistings differently.\(^3\)

Our paper contributes to a large literature that has studied the time series properties of house prices. The positive autocorrelation in house prices was famously introduced by Case and Shiller (1989). 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. 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, Ludvigson, and van Nieuwerburgh (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 span 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 standard house price indexes leads to a downward bias in its measurement. Perhaps most similar to our paper in terms of its goal is Wallace (2011), who discusses measurement problems associated with several of the most popular U.S. house price indexes 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 indexes and studies its effects on the measurement of house price dynamics.

We also contribute to the literature on house price forecasting (Gallin, 2008; Malpezzi, 1999; Rapach & Strauss, 2009; Carrillo et al., 2015, among others), which is motivated in part by the importance of future house price movements to the broader economic outlook for real economic activity. The existing literature mostly focuses on the explanatory power of variables that measure macroeconomic conditions like rents, income, unemployment rates, and mortgage rates. Our paper is unique in that we exploit the timeliness of listings data relative to transaction data and use the microdata on listings, rather than aggregates, to tie each individual list price to a previous sale price.\(^4\)

\(^1\) For example, the Case-Shiller house price index summarizing sales prices that close in month \(t\) is not released until the end of month \(t + 2\).

\(^2\) 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 III.

\(^3\) In exploiting data on seller behavior to forecast the final sales price, we are related to the literature that studies the connections between seller behavior and selling outcomes. See Han and Strange (2015) for a review of this literature.

\(^4\) Recently, CoreLogic started using listings data to forecast house prices. Based on our reading of their published materials, their measure does not fully exploit the informational content of the listings data as we do in this study.
II. Data

Our methodology uses two types of data. The first data requirement is the type of microdata on housing transactions used to produce standard house price indexes such as the Case-Shiller index. We obtained these data from Dataquick, which collects records from local governments throughout the United States on home transactions (which in most cases are required to be publicly disclosed by law). For each home sale, the data include the sale price; the closing date; the precise address of the home; home characteristics; the loan amounts originated; 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 to be recorded in the public record.

Our second data requirement is microdata on home listings, which we purchased from Altos Research. For the universe of homes listed for sale on the MLS, the dominant platform through which homes for sale are advertised in the United States, 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; 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 sale 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 property address and some house characteristics. Importantly, and in contrast to sales data, listings data can be obtained in real time.

In order to construct our new indexes, we obtain sales and listings data for the following MSAs: Chicago, Denver, Las Vegas, Los Angeles, Phoenix, San Diego, San Francisco, Seattle, and Washington, DC.⁵ Our Dataquick sample generally runs from 1988 to 2012 for each metro area; the Altos Research sample spans 2008 to 2013.⁶ The construction of our contract-dated index requires linking a sample of the delistings to the associated transaction record, which we do using the address, which is common to both data sets. 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.⁷ As we describe below, our index further requires linking each home in the listing data directly to its previous sales record in the transaction data, which we again do using the address. From our merged sample, we drop sales that do not merge to a previous transaction, sales where the length of time since the last transaction is less than six months, and properties that are not single-family as identified using the flag in the sales data.

In the end, our sample consists of 1.9 million single-family home listings that we can merge to a previous transaction record and are delisted during our sample period. The median property spends about eight weeks on the market before being delisted. Of those properties that are delisted, approximately half of them result in sales.⁸ Many of the delistings that do not result in sales are relisted soon after delisting, which may be due to sales agreements that fall through because the mortgage contingency or home inspection fails. However, our listings data do not provide the specific reason. Figure 1 shows the distribution of lag times between the delisting and closing dates for sales during our sample. The average closing time is about six weeks. While approximately one-third of properties close within two weeks, 16% of properties take more than ten weeks to close. We find that transactions closing within a week are disproportionately associated with no accompanying loans at the date of purchase, suggesting that investors or other buyers who pay in cash can close at or quickly after the contract date.⁹ Overall, 82% of the sales in our data can be linked to a delisting within a twelve-month window prior to the sale. This is consistent with reports from the National Association of Realtors showing that a vast majority of home sales are broker assisted and would thus involve marketing the home for sale on the MLS.¹⁰ Sales may not merge to a delisting

---

⁶ 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.
⁷ We also allow for a small number of sales where the delisting date is up to thirty days after the closing date, which can occur due to recording error in the closing date or the delisting date.
⁸ Carrillo and Williams (2015) find a similar share in their data, which come from a different data provider.
⁹ A full set of summary statistics for our sample is shown in table A1 in the appendix.
¹⁰ For example, the 2012 NAR Profile of Home Buyers and Sellers, which reports that 88% of home sales are broker assisted.
background for the construction of these indexes, we begin with a stylized presentation of the Case-Shiller repeat sales methodology. Our contract-dated and list-price indexes build off of the equations and notation introduced in this section.

The Case-Shiller regression equation is

\[ p_{it} = v_i + \delta_t + \epsilon_{it} \]  

(1)

where \( p_{it} \) is the log sales price of house \( i \) sold in month \( t \), \( v_i \) is a house fixed effect, \( \delta_t \) is a month effect that captures the citywide level of house prices at month \( t \), and \( \epsilon_{it} \) is the unexplained portion of the house price. Case and Shiller (1989) interpret \( \epsilon_{it} \) 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 \( \delta_t \), which we denote \( \delta^F_{CS} \), are the basis for the Case-Shiller index. For example, \( \delta^F_{CS} - \delta^L_{CS} \) 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

\[ p_{it} - p_{i't} = \delta_t - \delta_{t'} + \epsilon_{it} - \epsilon_{i't}, \]

(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 \( \epsilon_{it} - \epsilon_{i't} \) (“interval weighting”). Case and Shiller drop homes that cannot be matched to previous sales (e.g., new construction), home pairs where the interval between sales is less than six months, and all non-single-family homes. In practice, Case and Shiller also effectively weight each sale pair by the level of the first sale price, \( p_{0t} \), 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 \( \epsilon_{it} - \epsilon_{i't} \). 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 on by a month or two on average. In contrast, our contract-dated index, which we discuss next, reflects the value of housing at the time the sale price is negotiated. 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. Our list-price index, which we present in section V, also describes house prices at the time the sale price is determined and has

III. Case-Shiller Sales Price Index

The remaining sections of the paper describe the construction and properties of our contract-dated and list-price indexes, which can be constructed by supplementing sales data with listings data for information about recent sales. As

---

11 Because of differences in coverage across MSAs and changes in share of foreclosures over time, there is some variation across time and MSAs in the share of sales that can be linked to delistings. We show this variation in the appendix.

12 This is similar to the methodology used in the national CoreLogic repeat sales index and differs somewhat from the construction of the composite Case-Shiller indexes in which the index for each MSA is weighted by the aggregate value of the housing stock in that metro area.

the further advantage that it can be computed essentially in real time.

IV. Comparing the Contract-Dated and Closing-Dated Indexes

This section explores how our contract-dated measure of house prices compares to a more standard closing-dated repeat-sales price index. The contract-dated index is computed using the same repeat sales methodology just presented. The only difference is that for the second sale of each transaction pair, we include each house in the period it is delisted rather than the period in which it closes. To perform the comparison between the two indexes, 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, just as we did in figure 2. Figure 3 plots the contract-dated index and the closing-dated index over our sample period. The contract-dated index appears to lead the closing-dated index because the delisting date tends to precede the closing date. This is confirmed more formally in table 1, which reports the results of a Granger causality test in which we regress the closing-dated index on both its own lags and lagged values of the contract-dated index. Lagged values of the contract-dated index are statistically significant, causing us to reject the null hypothesis of no causality between the contract-dated index and the closing-dated index. The second column shows the results when the contract-dated index is included as the dependent variable instead of the closing-dated index. As expected, the closing-dated index does not appear to lead to the contract-dated index. Figure 3 also shows that the contract-dated index displays greater volatility than the closing-dated index. This effect arises because the closing-dated index is effectively a moving average of the contract-dated index due to variation in the lag between delisting and closing.

In the following sections, we investigate the implications of using the contract-dated index instead of the more standard closing-dated index for our understanding of the relationship between house prices, stock prices, news shocks, and interest rates; the predictability of house price changes; and the seasonality of house prices.

14 With a longer time series of listings data than the sample we have available, one could also use the delisting date for the first sale of each transaction pair instead of the closing date.
15 We include two lags because we found that additional lags of the sale price index are not statistically significant.
16 Some of the outlier price changes in fall 2011 shown in figure 3 appear to be associated with weeks in which our listings data provider experienced bugs in their data collection process for certain metro areas. Our results that follow are robust to excluding these outlier weeks.
17 Table A2 in the appendix provides some summary statistics for the variables we use in these exercises.

![Figure 3. Comparison of Contract-Dated and Closing-Dated Indexes](image-url)

**Table 1. Relationship between the Closing-Dated Index and Contract-Dated Index**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Closing-Dated Index</th>
<th>Contract-Dated Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closing-Dated Index&lt;sub&gt;_t&lt;/sub&gt;</td>
<td>0.4371*** (0.0545)</td>
<td>0.0146 (0.0554)</td>
</tr>
<tr>
<td>Closing-Dated Index&lt;sub&gt;_t-1&lt;/sub&gt;</td>
<td>0.3743*** (0.0466)</td>
<td></td>
</tr>
<tr>
<td>Contract-Dated Index&lt;sub&gt;_t&lt;/sub&gt;</td>
<td>0.0986*** (0.0342)</td>
<td>0.5357*** (0.1279)</td>
</tr>
<tr>
<td>Contract-Dated Index&lt;sub&gt;_t-1&lt;/sub&gt;</td>
<td>0.0900*** (0.0336)</td>
<td>0.4042*** (0.1226)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.0021 (0.0588)</td>
<td>0.0340*** (0.0115)</td>
</tr>
<tr>
<td>Observations</td>
<td>244</td>
<td>244</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.985</td>
<td>0.937</td>
</tr>
</tbody>
</table>

This table shows estimates from regressions of the closing-dated and contract-dated indexes on lags of these two series. Each house price index is computed at a weekly frequency with no smoothing across weeks. Standard errors, in parentheses, are adjusted for heteroskedasticity and autocorrelation using the Newey-West method. All indexes are in logs. \( **p < 0.01, *p < 0.05, *p < 0.1 \).

A. Correlation with Stock Prices

First, we consider the co-movement between the two-week growth in house prices and in the S&P 500 stock market index. Such co-movements may arise for a number of reasons such as common shocks (e.g., monetary policy news) or wealth effects. To measure the connection between house prices and stock prices, we compute the impulse response function of both our contract-dated index and the closing-dated index to a 1% rise in stock prices. In estimating the responses, we follow the approach of Jordà (2005) and control for seasonal effects and lagged house price growth. As shown in the left-hand panels of figure 4, the contract-index responds quickly, climbing by...
0.19% in the week following the shock to stock prices and remains slightly higher for approximately ten weeks. In contrast, the response of the closing-dated index is slower, reaching a maximum response after five weeks, and is also smaller in magnitude. We show two versions of the closing-dated index—one computed using the same sample period used to compute the contract-dated index and the other using the entire longer sample of available sales data (“Full Sample”). Both versions of the closing-dated index show similar patterns, suggesting that our main results may hold beyond the relatively short time period in which our listings data are available.

Additional details of the comparison are shown in table 2. In column 1, we report the results for the contract-dated index at a lag of one week, controlling for seasonal effects and lagged house price growth. A 1% increase in stock prices leads to a 0.19% rise in the contract-dated index in the following week. Columns 2 and 3 report results of the same exercise using the closing-dated index for the entire time period available and the same time period used to generate the results in column 1, respectively. The closing-dated index, which reflects prices determined several weeks earlier, on average, is only very weakly correlated with (roughly) contemporaneous changes in stock prices. The slightly stronger correlation in column 3 relative to column 2 likely reflects that during the sample period in column 3, a larger share of transactions were all cash purchases, which tend to have closing dates very close to contract dates, as discussed in section II.

Columns 4 and 5 report additional results using the closing-dated index, this time with lagged changes in stock prices. For both samples, we find the strongest correlation using a five-week lag, roughly consistent with the average time between the contract and closing dates. In addition to the closing-dated 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 contract-dated index than when they are measured with a standard closing–dated repeat-sales index. The differences between these estimates are statistically significant.

The oscillations observed at longer horizons are a common feature of this method for computing impulse response functions (see Ramey & Zubairy, 2014).

Recall that our sales data set begins in 1988, whereas our listings data set begins only in 2008 and the closing-dated index does not make use of the listings data.

22 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.
To understand the intuition for these results, it is useful to imagine an idealized one-time positive shock to the economy that causes an instantaneous rise in both stock prices and 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 caused the measured size of the response in house prices to be biased downward.

The economic implications of these results are important. Naively looking at the contemporaneous correlation between stock and house prices as measured by a standard repeat-sales index may generate the false impression that the marginal home buyer and the marginal stock investor display very different sensitivities to news about the economic outlook or risks, which might influence our conclusions about price formation in the housing market. Looking at properly lagged changes in the stock market would deliver results that are closer to the truth, but measurement error that is introduced by using closing dates might still cause the researcher to draw the wrong conclusions. In addition, while a five-week lag returns results that are closest to the truth in our sample, it is certainly possible that for particular subsamples of transactions or in other countries or cities, the appropriate lag to use might be different, and the researcher may have no way of easily determining this.

### B. Correlation with Surprises in Economic News

Next, we consider how house prices respond to economic news shocks. The right-hand panels of figure 4 show impulse response functions analogous to those shown in the upper panels of the figure except that the explanatory variable of interest is the Citigroup U.S. Economic Surprise Index. This index is available daily beginning in 2003 and captures the extent to which economic data released on a given day are above or below the Bloomberg consensus forecast. To construct the 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 ninety days using a (roughly) exponential weighting scheme. During our sample period, the mean of the index is −0.004, indicating that on average, economic news was weaker than expected, and the standard deviation is 0.05.

As with stock prices, we compute a weekly average surprise index so that the frequency of this variable aligns with that of our house price index. The first column of table 3 shows that a 1 standard deviation weekly average news surprise is associated with an increase in the two-week change in stock prices of 0.7%.

As with stock prices, the response of the contract-data index is larger than for the standard closing-dated repeat sales index. In addition, whereas the contract-dated index peaks one to two weeks after the news shock and subsequently declines, the closing-dated index exhibits a flatter response, again consistent with the measurement error described. The positive effect of news shocks on house prices at positive lags is likely a consequence of the smoothing used to compute the news shock index.

Table 3 reports regressions analogous to those reported in table 2 for stock prices. In column 2 of table 3 and in figure 4, a unit increase in the index is associated with a 0.13% change in the contract-dated index, controlling for seasonal effects and lagged house price growth. Columns 3 to 6 report results using the closing-dated index instead of the contract-dated index. This index is only weakly correlated with contemporaneous surprises in economic news, and the measured correlation between house prices and surprises in economic news is weaker than when house prices are measured using...
our contract-dated index. The differences between the estimate using the closing-dated index and contract-dated index are again statistically significant. Finally, column 7 reports the response of the contract-dated index to both stock prices and economic news. Both variables appear to have separate explanatory power for house prices.

C. Correlation with Interest Rates

Next, we consider how house prices respond to changes in long-term interest rates. All else equal, we would expect an increase in rates to depress house prices, though the literature tends to find only modest effects of interest rates on house prices (Glaeser, Gottlieb, & Gyourko, 2010; Adelino, Schoar, & Severino, 2012).

Table 4 reports a set of regressions analogous to those reported in table 3, except that the explanatory variable of interest is the yield on the ten-year Treasury. As seen across the six columns, we never find much of a correlation between changes in interest rates and changes in house prices, even after properly measuring house prices at the contract date.

Evidently the lack of correlation between house prices and interest rates at high frequencies does not appear to be explained by the measurement issues that are the focus of this paper. Perhaps the lack of correlation reflects omitted variables bias. For example, increases in interest rates may be associated with positive changes in the economic outlook, which works to offset the direct, negative effect of interest rates on house prices. Our contract-dated index does not solve this identification challenge. Another possible, not mutually exclusive explanation for the lack of correlation is that the contract date may not reflect the relevant interest rate environment, as borrowers can lock in rates at various periods prior (or subsequent) to the contract date.

D. Serial Correlation

Comparing the time-series properties of house prices as measured by the different house price indexes also provides some insight into one of the most striking features of house prices: the strong amount of short-run positive autocorrelation. Figure 5 shows that the contract-dated index displays significantly less positive serial correlation than the standard closing-dates index, especially at higher frequencies. The explanation is similar to the reasoning already presented for why the closing-dated 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.

As shown in figure 5, at very short frequencies, the weekly volatility in house prices offsets any underlying persistence in house price changes, and on net, the contract-dated index displays zero persistence. Nonetheless, as we

23 We obtain similar results if we use the thirty-year mortgage rate instead of the Treasury rate.

24 Consistent with the pattern that we find for the closing-dated index shown in figure 5, Bollerslev, Patton, and Wang (2015) also find that house price changes are less serially correlated at very high frequencies.

25 The intuition for this result is related to the problem of time aggregation first considered in Working (1960) because one can think of using the closing date instead of the contract date as a type of time aggregation. We illustrate a particular example of the Working (1960) problem in the appendix.

26 The volatility in house prices should not be due to estimation error as described in Case and Shiller (1989), as we adjust the time subscripts in our regression equation so that the same individual house sales do not enter both the dependent and independent variables of the regression. When we drop observations where the absolute value of the change in the house price index exceeds 10%, we continue to find that the contract-dated index displays significantly less persistence than the closing-dated index.
approach longer frequencies, the difference in serial correlation between the closing-dated index and the contract-dated index is not as large. This is as expected because the lag between agreement and closing is rarely longer than several months. Thus, while measurement issues can explain much of the very short-term persistence in house price growth, house prices, even when properly timed, still appear highly persistent at longer frequencies, suggesting that other mechanisms also contribute to the persistence in house prices.

Because we find that the timing of conventional closing-dated measures explains essentially all of the very short-term momentum in house prices, our findings are consistent with other explanations for house price momentum that would predict additional positive serial correlation only at slightly longer frequencies. For example, we argue elsewhere in this paper that information about recent house price movements is revealed to the market with lag. Therefore, explanations for house price momentum that rely on buyers using information about recent house prices might imply that correlations in house prices should emerge at a slightly longer frequency. (Examples of such explanations include Glaeser, Gyourk, & Saiz, 2008; Guren, 2013; Glaeser & Nathanson, 2015; and Anenberg, 2016.) In contrast, mechanisms that generate momentum as a gradual response of prices to a one-time shock to fundamentals (e.g., Head, Hwu, & Sun, 2014; Anenberg, 2016) might produce additional serial correlation at very high frequencies. Because we do not find any evidence of additional serial correlation at high frequencies, our findings offer less support for explanations that rely exclusively on such mechanisms, though we cannot rule them out.

E. Seasonality

Finally, we consider the implications of measuring house prices at the contract date for the seasonality of house prices. Ngai and Tenreyro (2014) document substantial seasonality in house prices and note that the predictability and size of seasonal fluctuations in house prices pose a challenge to existing models of the housing market. The stylized facts about seasonality that they and others cite are derived using a standard closing-dated index.

Figure 6 shows that the contract-dated index displays even more seasonality than the closing-dated index. For example, house prices in June relative to January are about 25% higher when prices are measured using the contract-dated

---

27 See section A.3.1 in the appendix.
index relative to the closing-dated index. As with our earlier results, this effect arises because the closing-dated index is effectively a moving average of the contract-dated index due to variation in the lag between delisting and closing, which smooths out seasonal fluctuations. Thus, our results suggest that true seasonality in house prices is an even bigger puzzle than the literature has recognized to date.

V. List-Price Index

We next introduce an alternative house price index, the list-price index. The list-price index uses the same timing convention as the contract-dated index, but rather than waiting for the sales data to become available to obtain the sale price, the list-price index uses the final listing price as a proxy for the sale price. Because listings data are available essentially in real time, the list price reveals trends in house prices earlier than conventional repeat-sales indexes that use sales data. As an application, we use the list-price index to generate a forecast of the Case-Shiller index prior to its release. We tailor our list-price index methodology to track that index specifically because that 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 based on transactions data.

In the appendix, we present a brief event study exercise showing that the stock prices of home-building companies respond to releases of the Case-Shiller index. This helps to motivate our forecasting exercise because 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.

A. Methodology

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

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.\(^{28}\) For convenience, we further define \( \overline{\mu} = E(\mu_{it}) \), the expected sale-to-list price ratio, and \( \delta_{it} = \mu_{it} - \overline{\mu} \) so that \( E(\delta_{it}) = 0 \). Then to obtain the month-\( t \) list-price index value \( \delta_{iT}^L \), we substitute into equation \( (2) \) as follows,

\[
p_{it}^L = \delta_{it}^L + \delta_{it'} + \mu_{it} - \mu_{it'} = \delta_{it}^L + \delta_{it'} + \mu_{it} - \overline{\mu} + \nu_{it}
\]

where \( \nu_{it} = \mu_{it} - \mu_{it'} - \overline{\mu}_{it} \).

Rather than jointly estimating the previous house price level \( \delta_{it'} \) along with \( \delta_{it}^L \), we use an estimate of \( \delta_{it}^L \) calculated from the transaction data alone using the Case-Shiller methodology. This means that when we estimate \( \delta_{it}^L \), we take \( \delta_{it'} \) as given and move it to the left-hand side of the equation.\(^{29}\) Finally, because \( \delta_{it}^L \) 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 the standard repeat-sales index. This gives our estimating equation,

\[
p_{it}^L - p_{it'} + \delta_{it'} = \delta_{it}^L + \nu_{it}.
\]

Our estimate of \( \delta_{it}^L \), which we denote \( \delta_{it}^L \), is the 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 indexes, as described above.

B. Forecasting House Prices with Listings Data

Our list-price index is tightly connected to standard repeat-sales indexes such as Case-Shiller by the simple fact that the delistings that underlie our index will ultimately become the transactions on which these standard indexes 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 be used to generate a forecast of the Case-Shiller index several months ahead of its release.

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 the Case-Shiller repeat-sales index, we must associate a closing date with each transaction. We assume that the

\(^{28}\) 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 forecast the repeat-sales index in section VB.

\(^{29}\) Alternatively, one could estimate \( \delta_{it} \) along with \( \delta_{it'} \).
lag between delisting and closing dates is drawn randomly from a discrete distribution, which we estimate nonparametrically using the empirical distribution of lag times shown in figure 1.\textsuperscript{30} 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.\textsuperscript{31} We use the simulated repeat-sales index as a forecast for what the Case-Shiller index will look like when it is finally released.

In examining the forecast, it is important to keep in mind that 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 exploiting the long lag between when seller behavior is observed and when the corresponding sales price index is released. A direction for future research is to combine our list-price index with prediction methods and extrapolation, which should improve forecasting performance further.

\subsection*{C. Forecasting Performance}

We examine our ability to forecast the Case-Shiller index, as computed from our sales data following the Case-Shiller methodology discussed in section III, 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.\textsuperscript{32} 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 index 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 sales is 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. Further, 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 index 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 relative to the latest available index value. Thus, a forecasting error of $x$ means that our forecasts underestimate or overestimate the percent change in sales prices by $100 \times x$ percentage points.

\begin{table}
\centering
\caption{Forecasting Performance of List Price Index}
\begin{tabular}{|c|c|c|c|}
\hline
\textbf{Forecast Horizon (Weeks)} & \textbf{Number of Months Ahead of Case-Shiller} & \textbf{List-Price Index} \\
\hline
& & \textbf{RMSE} & \textbf{MAE} & \textbf{$R^2$} \\
\hline
$-3$ & 2 & 0.023 & 0.018 & 0.146 \\
1 & 3 & 0.031 & 0.026 & 0.233 \\
5 & 4 & 0.038 & 0.032 & 0.288 \\
9 & 5 & 0.045 & 0.038 & 0.289 \\
13 & 6 & 0.053 & 0.043 & 0.261 \\
17 & 7 & 0.059 & 0.047 & 0.248 \\
21 & 8 & 0.065 & 0.050 & 0.230 \\
25 & 9 & 0.070 & 0.052 & 0.128 \\
29 & 10 & 0.073 & 0.053 & 0.040 \\
33 & 11 & 0.076 & 0.054 & -0.038 \\
37 & 12 & 0.080 & 0.056 & -0.137 \\
41 & 13 & 0.085 & 0.059 & -0.286 \\
\hline
\end{tabular}
\end{table}

This table shows the ability of the simulated repeat-sales index (based on the list-price index) to forecast the monthly Case-Shiller index at different forecast horizons. The simulated repeat sales index simulates closing dates for delistings and assigns them to calendar months. We forecast changes in the Case-Shiller index (i.e., changes in the log of the price level). 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 Case-Shiller price index for the month we are forecasting, which is released with a two-month delay. RMSE abbreviates root mean square error; MAE abbreviates mean absolute error. Each observation is an MSA-month.

Table 5 summarizes the performance of the list-price index at various horizons. The forecast performs well, even at forecasting horizons of up to thirteen weeks, which is six months in advance of the Case-Shiller release. The root mean square error (RMSE) associated with a forecasting horizon of thirteen weeks is $0.53$, the mean absolute error (MAE) is $0.43$, and the list-price index explains $26\%$ 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.

In the appendix, we assess the relative performance of the list price index. In particular, we show that the list-price index outperforms a standard AR forecasting model that does not use listings data. In addition, we show that our index outperforms the market’s expectation as implied by the prices of futures contracts for the Case-Shiller index over our sample period. This suggests that the information we exploit in our index is novel and not already known to the market.

\subsection*{D. Discussion and the Adjusted 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 two additional assumptions. In this section, we first identify those assumptions and then summarize some work we present in the appendix where
we construct an alternative “adjusted” list-price index where these assumptions are relaxed.

First, we note that at the time of delisting, the researcher cannot observe which transactions will close and which will not. As a result, we include all delisted prices in our index, effectively assuming that all these delistings will ultimately result in sales, which we know is not the case. If the list prices of houses that do not sell imply higher or lower values for the level of house prices, then including these observations will bias our estimates. Second, by using list prices rather than sale prices, we are implicitly assuming away any variation in the sale-to-list price ratio across time. 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.33

In the appendix, we examine the empirical relevance of each implicit assumption underlying our 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. Building on these results, we explore whether we can relax these assumptions and use additional information available at the time of delisting, such as time on market and the list-price history, to improve the forecasting performance of the simple list-price index. In particular, we are interested in whether these variables can help us better predict the final sale price and the probability that each delisting corresponds to the property being sold.

We find that using this additional information lets us explain about 80% of the variation in both the ratio of sale to list-price and the sales rate over time. Based on these findings, we construct an adjusted list-price index in which we weight each delisting by an estimate of the probability that it will lead to a sale, and we adjust the final list price toward a prediction of the final sale price. Importantly, all of the variables used to construct the weights and adjustment are available in real time. The improvement in the forecast relative to our baseline list-price index (the simple list-price index) is substantial. Using the adjusted index delivers improved performance by about 20% to 30% for shorter forecasting horizons. Details about the construction and forecasting performance of the adjusted list-price index are contained in the appendix.

VI. Conclusion

In this paper, we have presented new contract-dated and list-price house price indexes that attempt to more 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. Using this alternative timing convention reveals that house prices actually display less short-term correlation, stronger relationships to stock market fluctuations and economic news, and greater seasonality than previously estimated. Second, the listings data are available several months before the records of the actual transactions, allowing us to construct our list-price index as a measure of house prices that is available with almost no delay. In working toward these goals, our methodology attempts to construct new house price indexes with minimal differences from standard repeat-sales indexes. The contract-dated index differs only in the date assigned to the transactions yet reveals significant new insights into the behavior of house prices at higher frequencies. The list-price index takes the additional step of using the more easily accessible list price as a proxy for the final sale price and, as a result, yields improvements in our ability to measure house prices in a more timely manner. In both cases, 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.

REFERENCES


33 Our list-price model makes an additional assumption that all housing transactions first appear as delistings in the MLS. However, the evidence presented in section II 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.


