Differential measurement error in house price indices

Citation for published version:
Hayunga, D, Pace, RK, Zhu, S & Calabrese, R 2024, 'Differential measurement error in house price indices', *Journal of Real Estate Finance and Economics*, pp. 1-44. https://doi.org/10.1007/s11146-024-09994-z

Digital Object Identifier (DOI):
10.1007/s11146-024-09994-z

Published In:
Journal of Real Estate Finance and Economics

Differential Measurement Error in House Price Indices

Darren Hayunga[1]
Department of Insurance, Legal Studies, and Real Estate
University of Georgia
Athens, Georgia 30602
hayunga@uga.edu

R. Kelley Pace
LREC Endowed Chair of Real Estate
Department of Finance
Louisiana State University
Baton Rouge, LA 70803-6308
kelley@spatial.us

Shuang Zhu
Department of Finance
Kansas State University
Manhattan, KS 66506
shuangzhu@ksu.edu

Raffaella Calabrese
Business School
University of Edinburgh
raffaella.calabrese@ed.ac.uk

[1]Corresponding author. Hayunga gratefully acknowledges financial support from the Terry-Sanford Research Award.

Abstract

This article investigates measurement errors when using indices to model house prices over time. Our analysis, comparing index prices to actual transaction values, reveals that in many cases, widely-used indices display measurement errors correlated with the index values. Measurement error correlated with predictors constitutes "differential measurement error" at the level of the data generating process (DGP). We further explore the presence of differential measurement error within the context of mortgage lending. Our findings uncover substantial measurement errors in mortgage data, which not only diminish the predictive accuracy of models but also introduce notable biases in the coefficient estimates of variables.

1 Introduction

Numerous academic studies in real estate and economics rely on empirical specifications that incorporate house price indices as explanatory variables. For instance, a substantial body of literature on mortgage lending focuses on modeling decisions related to loan default and prepayment, contingent upon the house price relative to the debt's value.[1] Beyond mortgages outcomes, homeowners use indices to monitor general price movements, and to inform decisions related to consumption, maintenance, and retirement planning.[2]

Because the true price for each property across time is unobservable, researchers and homeowners typically use repeat sales indices at the ZIP code or MSA levels as a proxy.[3] The use of an index as a proxy raises the concern of measurement error. Econometrically, when a proxy equals the true value plus a measurement error that is unrelated to the other variables in the model, this error is termed non-differential measurement error or classical measurement error (Carroll et al., 2006). In a housing context, if the estimated price $\hat{P}$ equals the true price P plus a random measurement disturbance ϵ where none of the other model variables have a relation with ϵ, then this is an instance of non-differential measurement error. However, differential measurement error is when

[1] Various empirical mortgage studies track the price of underlying houses over time. Relevant references include Kau et al. (1992), Deng, Quigley, and Van Order (2000), Ambrose, Capone, and Deng (2001), Ambrose, LaCour-Little, and Sanders (2004), and Balla et al. (2024).

[2] See Case, Quigley, and Shiller (2005) and Carroll, Otsuka, and Slacalek (2011), along with their respective citations, regarding the wealth effects of housing. Goda, Shoven, and Slavov (2011) examine changes in retirement aspirations due to house prices and the Great Recession, while Hayunga, Pace, and Zhu (2019) delve into home investment.

[3] For example, Rosenthal (2014) examines the filtering of the housing stock to low-income households. Mian, Suri, and Trebbi (2015) and Harding, Rosenblatt, and Yao (2009) investigate foreclosures. Davis and Heathcote (2007) and Liu, Wang, and Zha (2013) explore land-price dynamics. Guerrieri, Hartley, and Hurst (2013) consider gentrification. Davidoff (2013) studies house supply elasticity. In addition, a number of studies such as Inklaar and Wang (2013) and McMillen and Singh (2020) use house price indices to deflate other economic variables.

the measurement error exhibits a relation with one or more model variables. A differential measurement error thus consists of a systematic component that varies with model variable(s) and a random component. This systematic component results in the coefficient of the proxy variable in the data generating process (DGP) underlying the estimation regression differing from the DGP using the true variable. Even if the random portion of the differential measurement error has a very low magnitude, the estimated parameters from the regression involving the proxy will be biased. This is true with a proxied dependent variable as well as a proxied independent variable. The bias introduced by the differential measurement error from the DGP can either attenuate or amplify the coefficient estimates in a multivariate model, depending on the correlation(s) between the measurement error and the variable(s). In contrast, non-differential measurement error will lead to no bias for a proxied dependent variable. For a proxied explanatory variable, non-differential measurement error can lead to attenuation bias in estimation. Although, in a multivariate setting with correlations among the variables, non-differential measurement error may also attenuate some estimates and amplify others (Judge et al, 1985, p. 708).[4]

Using almost 600,000 transactions within the Dallas-Fort Worth (DFW) Metroplex, we investigate the measurement errors of the popular indices from S&P Corelogic Case-Shiller (SPCCS) and the Federal Finance Housing Agency (FHFA). We find that these repeat sales indices are not overly accurate. The differences between transaction prices and predicted index values exhibit standard deviations of approximately 30%.

[4] "Thus, if a variable is subject to measurement error, it will not only affects its own parameter estimate, but may also affect the parameter estimates of other variables that are measured without error."

When examining the measurement error from the indices over 10 deciles of predicted prices, we find that repeat sales indices tend to underpredict price appreciation for higher-priced properties in deciles 7–9 while overpredicting for lower-cost properties from Deciles 2–5. There is almost no bias for Decile 6. Conversely, the lowest-cost homes in Decile 1 exhibit a substantial underprediction for Case-Shiller and FHFA at the MSA (but not zip-code) regions while the highest priced houses in Decile 10 vary from underprediction for Case-Shiller to almost no bias for FHFA. Overall, the predicted price deciles explain anywhere from 7.1% to 9.2% of the measurement errors from the repeat sales indices.

Our main contribution to the academic literature and decision-makers is the finding that price indices demonstrate differential measurement errors, meaning that the errors vary systematically with the predicted house prices. In addition, we have shown the impact of these measurement errors on model estimations.

Because there exists a vast mortgage literature that uses repeat sales indices, we investigate for measurement errors in a mortgage setting. The results show that there is a prevalence of positive differential measurement error associated with both predicted house prices and credit scores in mortgage data. A positive differential measurement error indicates that borrowers who own more (less) expensive homes and have higher (lower) credit scores tend to own properties that appreciate faster (slower) than indicated by the price indices. More importantly, the results demonstrate that the overall measurement error not only significantly diminishes the predictive power of mortgage models but also introduces notable biases in the coefficient estimates of those variables correlated with the measurement error.

To our knowledge, this article highlights an estimation bias in house price indices that has not been discussed previously in the residential real property literature. With regards to remediation of the issue, we offer four insights. First, researchers can use an internal validation study similar to our tests below to obtain estimates of the bias and correct for the problem. Second, although they do not have the universal availability of repeat sales indices, we find hedonic equations help mitigate differential measurement error. Third, avoiding strong non-linear transformations of estimated prices reduces the sensitivity of the results to measurement error. Fourth, studies using samples of the same types of housing as the repeat sales indices may help diminish estimation bias. For instance, researchers can use price-tier repeat sales indices wherever possible, such as the ones available from SPCCS.

The remainder of this manuscript is as follows. Section 2 presents illustrative theory concerning the possible biases that can result from using house price indices. Section 3 conducts an empirical validation study to investigate differential measurement error in house price indices. Section 4 conducts a further empirical study to demonstrate the presence of differential measurement error in mortgage studies and illustrate its impact on model performances. Lastly, Section 5 summarizes the key findings from our analysis.

2 Measurement Error and the Estimation Impact

In this section we examine the effects of measurement error in house prices when researchers do not have actual property values but use repeat sales or hedonic-model indices to estimate property values. To begin, let δ_t represent the value of the logged index at period t. In Equation (1), p_{it} is the natural logarithm of the house price and P_{it} denotes the untransformed house price level. The difference $(\delta_t - \delta_0)$

between the logged price index at periods t and 0 represents the log of the cumulative appreciation in the market between origination and time t. In a repeat sales context, updating the house price at origination with the cumulative market-wide estimate of appreciation yields an estimate of the logged current house price $\hat{p}_{it}$ as shown in (2).

$$p_{it} = \ln(P_{it}) \tag{1}$$

$$\hat{p}_{it} = p_{i0} + (\delta_t - \delta_0) \tag{2}$$

Alternatively, researchers may use a hedonic method (3) that predicts prices based on the characteristics of the house X_{it} which could involve size, quantity, age, location, and so forth.

$$\hat{p}_{it} = X_{it} \cdot \hat{\beta} \tag{3}$$

In either case, the methods result in estimated prices $\hat{p}_i$ for all i observations in the sample.

To simplify notation we now examine the associated $(n \times 1)$ vectors p and $\hat{p}$. The measurement error from fitting the true price vector p less the prediction $\hat{p}$ appears in (4). Ideally, any measurement errors will collectively equal a white noise vector ϵ so that $E(\epsilon) = 0$. However, the predicted price $\hat{p}$ could explain some of measurement errors. In other words, the measurement errors decompose into the vector of estimated or proxy prices $\hat{p}$ times the scalar bias parameter α plus a random error vector ϵ. If $\hat{p}$ unbiasedly estimates p, then $E(\alpha) = 0$. Also, we can rewrite the residual equation (4) in terms of the level of p as in (5).

$$p - \hat{p} = \hat{p} \cdot \alpha + \epsilon \tag{4}$$

$$p = \hat{p} \cdot (1 + \alpha) + \epsilon \tag{5}$$

Suppose the researcher wishes to explain some dependent variable y as a function of a true price p as in (6) with the additional set of explanatory variables in the $(n \times k)$ matrix X.

$$y = p \cdot \tau + X \cdot \beta + \varepsilon \tag{6}$$

As in many empirical cases, the researcher only has $\hat{p}$ available and wishes to estimate the equation (6) using the proxy $\hat{p}$. What is the consequence of this? To determine, substitute the decomposition from (5) into (6), which results in (7) and further simplifies into the DGP in terms of the proxied variable $\hat{p}$ in (8).

$$y = ((1 + \alpha)\hat{p} + \epsilon) \cdot \tau + X \cdot \beta + \varepsilon \tag{7}$$

$$y = \hat{p} \cdot (1 + \alpha)\tau + X \cdot \beta + \epsilon \cdot \tau + \varepsilon \tag{8}$$

If the measurement error is independent of the explanatory variables $\hat{p}$ and X then $E(\alpha) = 0$. This is the case of non-differential measurement error. In such case, as can be seen from (6) and (8), the underlying DGPs using the true variable and the proxied variable remain the same (apart from the increased purely random error). So that the non-differential measurement error leads to no bias associated with the true parameter of house price (τ) in the DGP. In actual estimation, the attenuation bias associated with non-differential measurement errors comes from the magnitude of ϵ and in a multivariate setting the correlations among the explanatory variables. If the variance of ϵ is low relative to that of its own variable (e.g., $\hat{p}$), the attenuation bias will be less pronounced.

Conversely, when the error is correlated with $\hat{p}$, it constitutes a form of differential measurement error. The differential measurement error encompasses a systematic component ($\hat{p} \cdot \alpha$) and a random component (ϵ). The systematic component will introduce a bias through the DGP based on the proxied variable. Specifically, instead of estimating τ, even as ϵ goes to 0, the model in (8) will produce an estimate of $\tau + \alpha\tau$, introducing a bias of $\alpha\tau$ (assuming no relation between measurement error and X). In actual estimation, a large magnitude of the random component ϵ could attenuate the estimate of $\tau + \alpha\tau$ (with both the potential for attenuation and amplification for the estimates associated with X emerging with correlations between $\hat{p}$ and X).

Differential measurement error arises not only when the measurement error systematically varies with its own variable, but also when it systematically varies with other variables. For example, if the measurement error in predicted house price varies systematically with X and is not correlated with $\hat{p}$, it is still a form of differential measurement error and will lead to a DGP with different coefficients associated with X than the DGP with the true house price. If the measurement error varies systematically with both $\hat{p}$ and X, biases associated with the true τ and β will exist in the DGP.

Note, beyond attenuation bias, non-differential measurement error still has the potential to create a bias through the DGP that either underestimates or overestimates the true effect in a non-linear context. To see this, suppose the housing price model yields a predicted value

$\hat{p}$ of the log-price p, but the researcher needs to use the untransformed house price $P = \exp(p)$. However, the researcher only has access to $\hat{p}$. What would be the consequence of using $\exp(\hat{p})$? Assuming the measurement error ϵ_i is distributed $N(0, \sigma_\epsilon^2)$,

$$\exp(\hat{p}_i) = \exp(p_i)\exp(\epsilon_i) \tag{9}$$

$$E(P_i) = E[\exp(\hat{p}_i)] = \exp(p_i)\exp(0.5\sigma_\epsilon^2) \geq \exp(p_i). \tag{10}$$

Similarly, many mortgage models use the estimated loan-to-value ratio which equals $L \cdot (1/\hat{P})$ where L is the loan amount

$$E(L_i/\hat{P}_i) = L_i \cdot E[\exp(-\hat{p}_i)] = (L_i/P_i)\exp(-0.5\sigma_\epsilon^2) \leq L_i/P_i \tag{11}$$

Therefore, using a loan-to-value based on predicted prices will result in a loan-to-value ratio biased downwards and an associated regression coefficient that is biased upwards. In our results below, we find a standard deviation of $\sigma = 0.3$ in measurement errors, which will lead to an upward bias in predicted price of 4.6% and a downward bias in the loan-to-value of -4.4%. Note, the use of a bias adjustment for the regression estimates in models with a logged dependent variable has been a longstanding practice in economics.[5]

[5] In the log-model case, it is not correct to obtain predictions of $\ln(y)$ and then compute the exponential of the predicted value without an adjustment. It is well known that given a mean (μ) and variance (σ^2) of the parent normal variate, then the mean (δ) of a log-normal variate is $\delta = \exp(\mu + 1/2\sigma^2)$. See, for instance, Heien (1968) or Crow and Shumizu (1988).

3 House Price Internal Validation Study

We now empirically investigate the measurement errors resulting from using repeat sales indices and hedonic price models. We describe the data in section 3.1 and calculate the measurement errors in section 3.2. Section 3.3 investigates for differential measurement error and determines the estimation biases.

3.1 House Price Data

To examine the possible existence of differential measurement error, we require a sample of homes with sufficient properties that sell at least twice to form the repeat sales index along with many observable structural and transactional characteristics to fit a hedonic model. To meet both requirements as well as the ability to measure against a commonly used price index like SPCCS, we use broker assisted transactions from the DFW multiple listing service (MLS) for the period from 10/2002 to the 4/2013. The DFW Metroplex's population is more than 6 million people during our sample period, ranking it in the top 10 of largest metropolitan areas in the US. During this same time, there were almost one million real property transactions of various types within the DFW Metroplex.

Since we are examining housing indices, our initial sample removes transactions involving the sales of non-residential property types (e.g., farms). To better generalize our results, we apply additional filters. We require that the homes have between one and fourteen bedrooms, have one or more bathrooms, be built post-1800, and sell for more than \$20,000. Consistent with typical housing indices, we do not include new construction sales as builders can have different price-liquidity preferences than other sellers. Because so-called flipped properties generally include major renovations and thus can exhibit considerable levels of appreciation that can create a bias in the index, we also require the holding period between repeat sales to be at least six months.

We use three repeat sales indices to track house prices in the DFW area. The first, the SPCCS index, draws on sales data from ten counties within the DFW Metroplex, providing nearly 600,000 observations.[6] A notable advantage of the SPCCS index is its inception in 2000, aligning closely with the starting period of our sample. This alignment aids in mitigating potential selection bias issues stemming from sample properties with short average holding periods being disproportionately represented as "winners." Short holding periods may indicate winning scenarios for various reasons: on the demand side, sellers might opt for quicker sales when offered bids surpassing market expectations, while from a supply perspective, shorter holding periods could suggest unobservable superior property quality. It could be argued that our repeat sales sample predominantly consists of winners due to its 13-year timeframe compared to an index with a much longer history of data. However, the SPCCS starting at a similar time as our sample period mitigates this concern within our analysis.

The second repeat sales index we employ is the FHFA House Price Index, a quarterly indicator derived from data sourced from seven major cities in the DFW Metroplex.[7] These data are compiled from repeat mortgage transactions on single-family properties whose mortgages have been purchased or securitized by either Fannie Mae or Freddie Mac. The FHFA index encompasses nearly 250,000 observa-

[6] The counties are Collin, Dallas, Delta, Denton, Ellis, Hunt, Johnson, Kaufman, Parker, Rockwall, Tarrant, and Wise.

[7] The cities are Dallas, Irving, Plano, Richardson, Fort Worth, Arlington, and Denton.

tions fitting its geographical definition. Beyond providing an alternative comparison of measurement errors, the FHFA index enables us to use price values specific to the two metropolitan divisions within the DFW Metroplex. We assess measurement errors using one index for the Dallas-Plano-Irving division and another for the Fort Worth-Arlington-Grapevine division.

To examine a more granular spatial dimension, the third repeat sales index we use is the FHFA values at the five-digit ZIP code level. Since the ZIP code data are annual, we improve the accuracy of the index prediction by appending more frequent data. We extrapolate to the quarter of sale in both the first and second transaction using the FHFA House Price Index for the West South Central Census Division, which encompasses the states of Texas, Oklahoma, Louisiana, and Arkansas.

Table 1: Descriptive Statistics of the Dallas-Fort Worth Samples

	SPCCS areas		FHFA areas	
Variables	Mean	Median	Mean	Median
Transaction Price ($k)	188.692	140.901	195.329	140.401
Square Feet (k)	2.163	1.920	2.101	1.870
Age (years)	24.179	19.000	31.128	27.000
Pool (yes=1)	0.163	0.000	0.163	0.000
Garage Spaces	1.808	2.000	1.727	2.000
Owner Occupied (yes=1)	0.626	1.000	0.613	1.000
Tenant Occupied (yes=1)	0.019	0.000	0.021	0.000
Vacant (yes=1)	0.355	0.000	0.365	0.000
Cash Financing (yes=1)	0.150	0.000	0.164	0.000
Conventional Financing (yes=1)	0.599	1.000	0.607	0.000
REO Seller (yes=1)	0.195	0.000	0.185	0.000
Number of Observations	596,778		249,772	

Table 1 provides descriptive statistics detailing primary transaction characteristics within the SPCCS and FHFA geographical areas. At this juncture, the data exhibit comparability and adhere to standard conventions.

As just mentioned, we consider the implications of holding periods. Our analysis reveals no significant disparities, as similar holding periods are observed across all three repeat sales indices. Specifically, for the SPCCS index, closely aligned with the inception of our data, mean and median holding periods are 3.71 and 3.39 years, respectively. Similarly, at the metropolitan division level, the FHFA index demonstrates comparable figures, with mean and median periods of 3.70 and 3.39 years, respectively. With the FHFA ZIP code index operating on an annual basis, a minimum one-year holding period is requisite. Consequently, mean and median holding periods for this index are 3.62 and 3.97 years, respectively. Notably, maximum holding periods across all three indices are consistent, at 10.58, 10.57, and 10.57 years, respectively.

Another data aspect we consider is the Global Financial Crisis (GFC). As depicted in Figure 1 using the SPCCS index, house prices within the Metroplex remained generally stable over our sample period, marked by the vertical lines indicating the beginning and end dates. Notably, the GFC had a subdued effect on the local housing market compared to other regions such as Boston or Atlanta. This resilience allows our findings to offer broader applicability and more generalized insights that can be extrapolated to various markets. Additionally, we confirm that our repeat sales samples are not disproportionately influenced by second sales occurring immediately after the GFC. Such dominance could potentially bias our price returns downward due to lower prices post-crisis. Considering the mean and median holding periods, and the fact that our sample begins in October 2002, we observe an expected increase in repeat sales observations over the initial four years from 2003 to 2006, followed by relatively stable occurrences of second sales in sub-

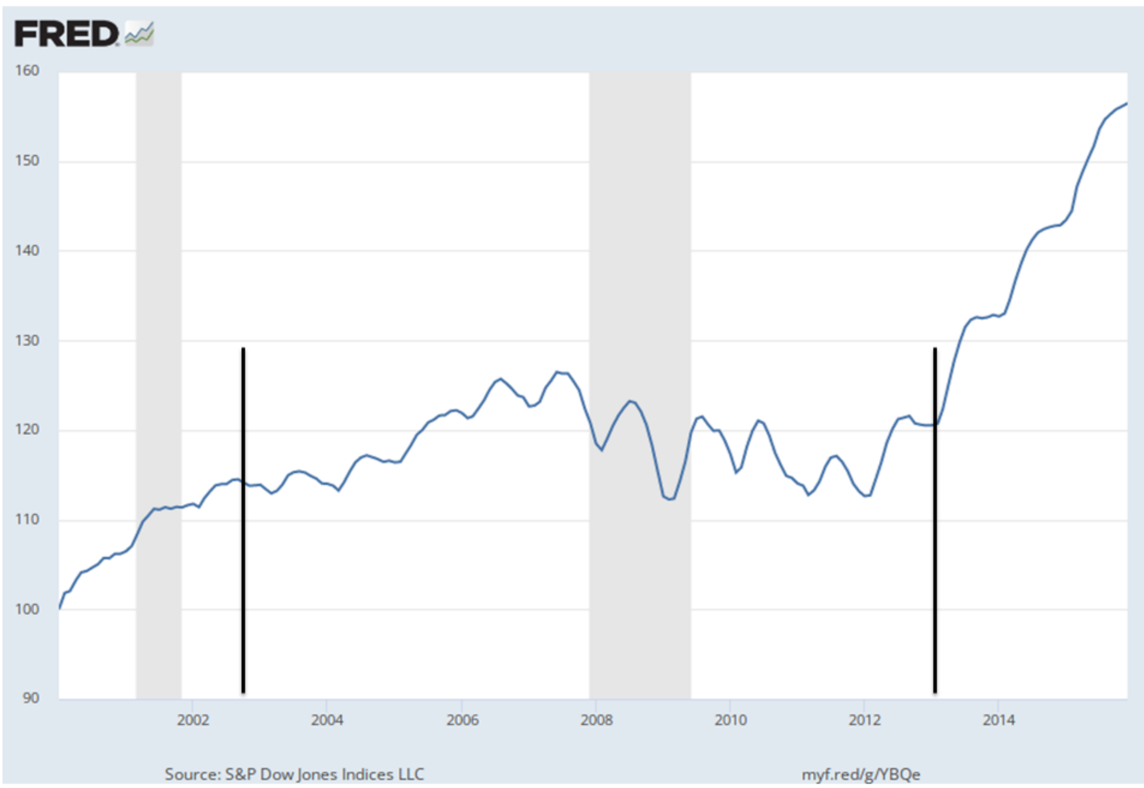

Figure 1: DFW House Prices from the SPCCS Index

sequent years. For instance, of the more than 106,000 repeat sales using the SPCCS geographical area, the percentage of second sales from 2007 to 2012 are 11.7, 13.0, 13.2, 13.3, 13.1, and 17.1.

In addition to the repeat sales indices, we model hedonic specifications, leveraging transactions within corresponding geographical areas for the FHFA and SPCCS comparisons. The detailed hedonic models, outlined in Appendix A, encompass various common and unique independent variables. For instance, the models control for attributes like vacancy and financing. While these features may not be discernible in house price indices, their inclusion in our study offers a means of conducting comparative analysis and provides a technique to mitigate potential differential measurement error, as revealed by our empirical results.

3.2 Differences in Transaction Prices and the Indices

Turning our focus to assessing potential measurement errors between actual transaction prices and predicted values derived from various indices, we present summary statistics of the log measurement errors in Table 2. Both the hedonic specifications and the SPCCS index reveal minimal mean errors. The baseline mean and median values of the hedonic models align with our expectations, as these models are designed to minimize errors by constructing the coefficient vector in the least squares specification (denoted by $\hat{\beta}$ in Equation 3), ensuring that residuals are orthogonal to the regressors. This performance can be attributed not only to the inclusion of more variables but also to the ability to calculate distinct growth rates for each property. However, while mean and median errors remain modest, the standard deviations of hedonic residuals mirror those of the repeat sales indices, hovering around 30% across all five indices. Furthermore, the measurement errors exhibit considerable magnitudes in absolute values, accompanied by significant kurtosis.

Table 2: Descriptive Statistics of Log Measurement Errors and Log Prices

Variables	SPCCS MSA	FHFA MSA	FHFA ZIP Code	SPCCS Hedonic MSA	FHFA Hedonic MSA
Mean e	−0.0001	−0.0239	−0.0480	0.0000	−0.0000
Median e	0.0205	−0.0139	−0.0164	−0.0027	−0.0043
$\sigma(e)$	0.3017	0.3283	0.3044	0.2870	0.2870
Mean $\|e\|$	0.2024	0.2184	0.2005	0.1724	0.2037
Median $\|e\|$	0.1192	0.1223	0.1118	0.1261	0.1491
Kurtosis e	6.2301	6.7570	7.5460	10.9023	10.2008
Mean $\ln(P)$	11.9496	12.0088	12.0293	11.8948	11.8846
Median $\ln(P)$	11.9184	11.9954	12.0015	11.8558	11.8523
$\sigma(\ln(P))$	0.6962	0.7377	0.7947	0.6645	0.7458
n	106,168	44,424	42,527	596,778	249,772

This table details summary statistics of the log measurement errors produced when comparing the actual transaction prices to the predicted prices using repeat sales indices (SPCCS or FHFA) as well as hedonic regressions. The hedonic models use the observations within the geographical area defined by the SPCCS and FHFA repeat sales indices and the independent variables listed in Appendix A. The table includes summary statistics of the log transactions prices.

While the measurement error of SPCCS, on average, hits the center, the FHFA indices are slightly off with the average measurement errors of −0.0239 and −0.0480. One potential reason for this is that the SPCCS sample might be closer to the actual sample of repeat sales, which includes both transactions with mortgages and cash transactions. FHFA sample, however, consists only transactions with mortgages. This makes FHFA sample different than the repeat sales sample. To investigate whether the difference in samples between FHFA and repeat sales contributes to the off center mean of measurement error, we provide the FHFA statistics excluding cash transactions in repeat sales. The results are reported in Table 8 in Appendix B. Without cash sales, the baseline mean and median measurement errors are close to the center in Table 8 and comparable to the SPCCS indices in Table 2.[8] However, the other characteristics remain consistent with those in Table 2.

It should be noted that we focus on measurement error with repeated sales including cash transactions in the main paper. Given the likelihood that researchers and decision-makers lack insight into which trades involve cash, these analyses are more informative to typical users who apply the index to their studies.

To determine if there exists a meaningful pattern to the errors beyond these descriptive statistics, we next bin the sample in deciles based upon the predicted prices. We then regress the log measurement errors against binary variables set to one if the predicted house price fits within that specific bin, or zero otherwise. The intercept is suppressed to avoid rank deficiency. Table 3 details the parameter estimates within each decile.

The results demonstrate that the repeat sales methods can exhibit large differences between the true prices and the index values. This is especially true at the lower price levels in Decile 1–3 for the repeat sales measures. In addition, with the exception of Deciles 1 and 10, the repeat sales methods demonstrate a noticeable monotonic increase in the level of measurement errors across the deciles. Taking the FHFA MSA

[8] The standard deviations in Table 2 are calculated around the means, so an off-center mean should not affect this measure of dispersion. However, an off-center mean could impact the mean squared error (MSE). In calculating the MSE for Table 2, we find that the SPSCC sample has a slightly lower MSE (SPSCC MSA: 0.0910; FHFA MSA: 0.1083; FHFA ZIP Code: 0.0950). After removing the cash sales and recalculating the MSE for Table 8, we observe lower MSEs for the FHFA samples (FHFA MSA: 0.0647; FHFA ZIP Code: 0.0596).

index as an example, the magnitude of the mean measurement errors is −0.1871 for the prices in Decile 2. The slope coefficients increase through Decile 9 to 0.0402. The monotonic relations across deciles exhibited by the repeat sales indices suggest possible differential measurement error, which we will examine below.

Compared to the hedonic models, there is a marked increase in the R^2 for the repeat sales indices, which is the proportion explained by the means in each predicted price range (i.e., subtracting the means of the predicted prices within each decile explains some of the variance in the measurement errors). The means of the predicted prices explain more than 7% of the variance in the measurement errors using the two MSA definitions and more than 9% using the FHFA ZIP codes. In contrast, the means of the predicted prices in the hedonic models explain less than 1% of the variance. Since the desired result is a lack of correlation between the predicted prices and the measurement errors, the lower R^2s are an improvement. The betterment in the hedonic regressions is not necessarily unexpected as they not only contain a number of independent variables that are not controlled for in repeat sales indices but also calculate different growth rates for each property.

3.3 Differential Measurement Errors

The fact that measurement errors exist within each decile in Table 3 is not in and of itself extremely problematic because a proxy can deviate from the true value. However, the measurement errors do not appear to randomly vary with the predicted values, and exhibit a largely monotonic increase across Deciles 2–9 when using the repeat sales indices. To determine if this relation results in differential measurement error, we regress the log errors on the log predicted prices, respectively (i.e.,

Table 3: Log Mean Error Coefficients by Decile of Predicted Price

Decile	SPCCS MSA	FHFA MSA	FHFA ZIP Code	SPCCS Hedonic	FHFA MSA Hedonic
1	0.1506	0.1379	−0.0606	−0.0184	−0.0244
	53.3741	29.1600	−13.4578	−18.6878	−13.5049
2	−0.1266	−0.1871	−0.2196	0.0239	0.0422
	−44.8564	−39.5693	−48.7978	24.2607	23.3092
3	−0.1056	−0.1406	−0.1432	0.0262	0.0320
	−37.4201	−29.7334	−31.8244	26.6264	17.6877
4	−0.0755	−0.1051	−0.1037	0.0132	0.0102
	−26.7517	−22.2332	−23.0533	13.4163	5.6589
5	−0.0302	−0.0459	−0.0411	−0.0017	−0.0035
	−10.7082	−9.7048	−9.1267	−1.7003	−1.9570
6	0.0094	−0.0020	−0.0014	−0.0118	−0.0156
	3.3441	−0.4166	−0.3178	−11.9967	−8.6240
7	0.0329	0.0237	0.0239	−0.0214	−0.0295
	11.6413	5.0100	5.3156	−21.7536	−16.3267
8	0.0509	0.0386	0.0285	−0.0277	−0.0320
	18.0487	8.1742	6.3333	−28.1785	−17.7191
9	0.0502	0.0402	0.0301	−0.0106	−0.0107
	17.8066	8.5064	6.6956	−10.8032	−5.9101
10	0.0426	0.0012	0.0069	0.0283	0.0314
	15.0838	0.2628	1.5333	28.8122	17.3860
R^2	0.0710	0.0788	0.0918	0.0070	0.0082
n	106,168	44,424	42,527	596,778	249,772

This table reports the parameter estimates from a regression modeling the log measurement errors against binary variables set to one if the predicted house price is in that particular decile bin, and zero otherwise. The measurement errors are the difference between the actual transaction price and the estimated house values using the different indices. The numbers below the slope coefficients are t-statistics.

$p - \hat{p} = \text{intercept} + \hat{p} \cdot \alpha + \epsilon$). If the measurement errors $(p - \hat{p})$ exhibit pure randomness, the regression estimate of α will be statistically insignificant in a sufficiently large sample.

Table 4 presents the estimates of α of log errors on log predicted prices across several samples. We focus on the repeat sales data since the hedonic equations exhibit muted levels of measurement errors in Table 3. For the full sample, the parameter estimate on the FHFA ZIP Code index is 0.0621 with a t-statistic of 29.44. However, the FHFA MSA index does not exhibit strong systematic biases, considering the large sample size. The R^2 of both MSA indices is very small. This suggests that differential measurement error might not be a significant concern when dealing with the full sample in those circumstances. The inclusion of Deciles 1 and 10, which do not follow the monotonic relation displayed in the other deciles, could contribute to the overall weak results. However, it is worth noting that at the price level, the full sample exhibits statistically significant systematic biases for all three indices as presented in Table 9 in Appendix C.[9]

Because the measurement errors in Deciles 1 and 10 in Table 3 do not follow the monotonic relation displayed in the other deciles, we next explore possible differences across the house price distribution. We first investigate whether the measurement errors in Decile 10 can be attributed to the fact that these highest cost homes may use more jumbo-loan financing that does not conform to Fannie Mae and Freddie Mac requirements. To investigate this aspect, we assume a 20% down payment and remove homes from the full sample that have purchase

[9]Motivated by the practice observed in influential and recent studies, which employ index levels to gauge changes over time (e.g., Deng, Quigley, and Van Order, 2000; Case, Quigley, and Shiller, 2005; Demyanyk and Van Hemert, 2011; Hayunga, Pace, and Zhu, 2019; and Balla et al., 2024), Table 9 in Appendix C provides a detailed breakdown of regression coefficients and additional statistics for the various samples using price levels. This way we provide both linear and non-linear analyses, thus accommodating the preferences of empirical researchers and decision-makers.

Table 4: Regressions of Log Measurement Errors

	Coefficient	t-stat	R^2	n
SPCCS MSA Full	0.0059	4.00	0.0002	106,168
FHFA MSA Full	0.0043	1.83	0.0001	44,424
FHFA ZIP Code Full	0.0621	29.44	0.0200	42,527
SPCCS MSA Conforming	−0.0047	−2.64	0.0001	101,216
FHFA MSA Conforming	0.0033	1.14	0.0000	42,027
FHFA ZIP Code Conforming	0.0855	31.41	0.0244	39,488
SPCCS MSA Trimmed	0.1758	73.08	0.0592	84,936
FHFA MSA Trimmed	0.1951	52.11	0.0710	35,540
FHFA ZIP Code Trimmed	0.1905	57.04	0.0873	34,023
SPCCS MSA Lowest Cost	−0.5912	−39.35	0.1273	10,616
FHFA MSA Lowest Cost	−0.6816	−29.23	0.1613	4,442
FHFA ZIP Code Lowest Cost	−0.5794	−25.05	0.1287	4,252

The table presents regressions of the log measurement errors as the dependent variable against the independent variable of the natural log of predicted house prices. The coefficients stated in the first numerical column measure the level of *differential* measurement errors. The larger the coefficient in absolute value the greater the potential bias in using the index for that specific sample.

prices which exceed the conforming loan limit for each year, factoring in the down payment (i.e., conforming loan limit $<$ purchase price $\times$ 0.80). The reduction in the number of observations across the subsamples is detailed in the last column of Table 4. Upon removing potential jumbo loans, the SPCCS MSA sample decreases by 4,952 transactions (4.7%), the FHFA MSA sample decreases by 2,397 observations (5.4%), and the FHFA ZIP code sample decreases by 3,039 transactions (7.1%).

The parameter estimates using the remaining conforming loan sample detailed in Table 4 are similar to the full sample. The three price-level indices as well as the FHFA ZIP code index demonstrate significant differential measurement errors.

We next consider the bulk of the repeat sales data that display the monotonic increase across Deciles 2–9 in Table 3. Because empirical researchers often drop data to avoid the possible effect of outliers, we create a trimmed sample by removing all transactions in Deciles 1 and 10. The results in Table 4 demonstrate positive parameter estimates across the three indices. The systematic biases from differential measurement errors increase relative to the other samples detailed thus far, with magnitudes that are approximately 20%. This arises because the differential error in the lowest decile more than offsets the differential error in the 10th decile.

To provide additional perspective of the estimation biases, we plot the measurement errors across the predicted house prices in Figure 2. For the sake of caution, we use the SPCCS as it exhibits the smallest differential measurement error for our sample. This choice ensures a conservative approach, recognizing that empirical findings in other studies may encounter more substantial biases. For easier interpretation, we present the data in terms of price levels. The band around the

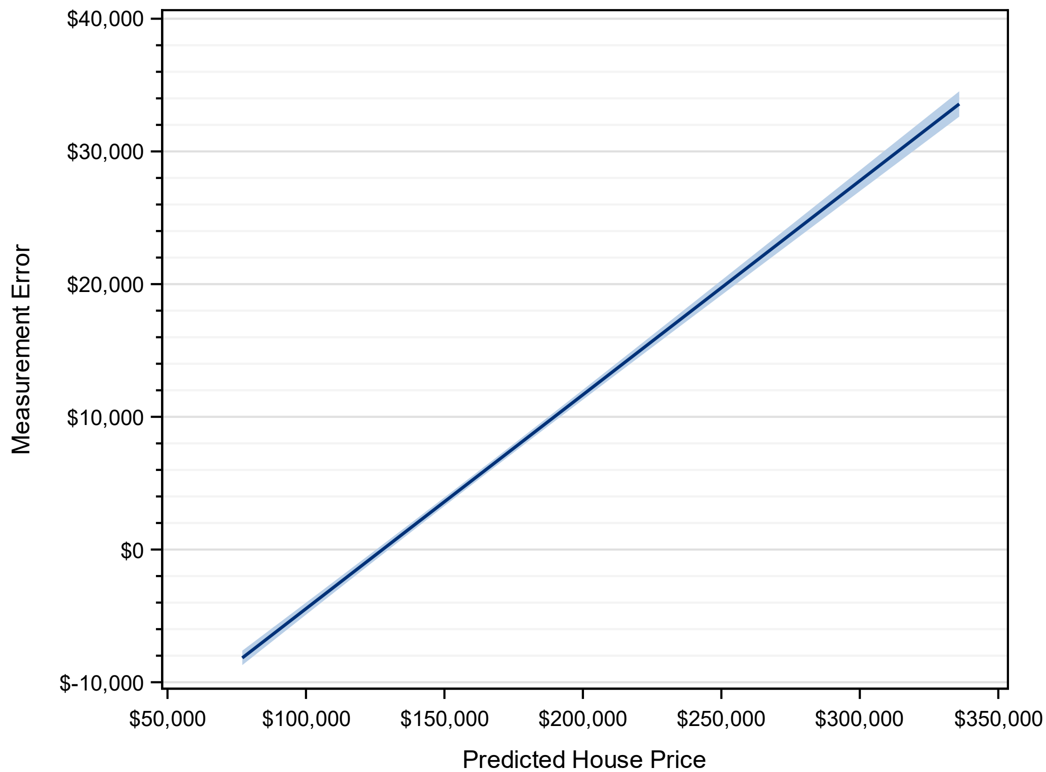

Figure 2: Differential Measurement Errors from SPCCS across Deciles 2–9

regression line in Figure 2 is the 95% confidence limits for the mean predicted measurement errors.

Looking at the data spanning Deciles 2–9, depicted in Figure 2, the mean home value within this sample is \$161,977. The predicted measurement error for this average house amounts to \$5,534, translating to an actual price of \$167,512. Notably, owing to the positive slope between the measurement errors and predicted house prices, lower-priced homes tend to exhibit lower prices than indicated by the index. For example, a home predicted to be valued at \$80,000 using the SPCCS index would have a transaction value of \$72,317.

While the absolute values of the measurement errors are not huge for the lower portion of the data in Figure 2, the systematic bias becomes more problematic at the upper end of the predicted-house-price distribution. For example, a $325,000 home based upon the index will have an actual value of slightly more than $356,000, which is a measurement error of approximately $31,000.[10]

We note in Table 3 that the values in Decile 1 do not follow the general trend line. We consequently focus on these lowest cost homes next. Restricting the sample to only the homes within Decile 1, we again regress the measurement errors against the predicted house prices. The last set of values in Table 4 reports the findings. The estimates of α for this subsample are large and negative. The estimation biases range from -0.5794 to -0.6816. Given the large negative values we examine the frequency distribution across the year of the second sale to ensure that our finding is not driven by the lowest priced properties being sold predominately during or immediately after the GFC. Using the SPCCS index, we find the percentages of second sales within Decile 1 from 2007 to 2012 are 11.6, 13.4, 14.2, 12.6, 12.1, and 17.4. In comparison, the percentages across the same time periods for Decile 5, for example, are 12.3, 13.8, 15.1, 14.7, 12.0, and 11.9.

To again provide further perspective, we graph the measurement error levels against the predicted house prices. Using price levels, Figure 3 presents the values for the SPCCS index. For the higher priced properties within Decile 1, the measurement errors are somewhat less problematic. For instance, a $70,000 properties exhibits an error of $7,723. However, the estimation biases become large as the home prices decrease. The average home price within this subsample is $55,804. Us-

[10]Due to higher priced homes exhibiting greater measurement errors, we confirm in untabled regressions that the measurement errors in Decile 10 does not exhibit much correlation with the predicted price.

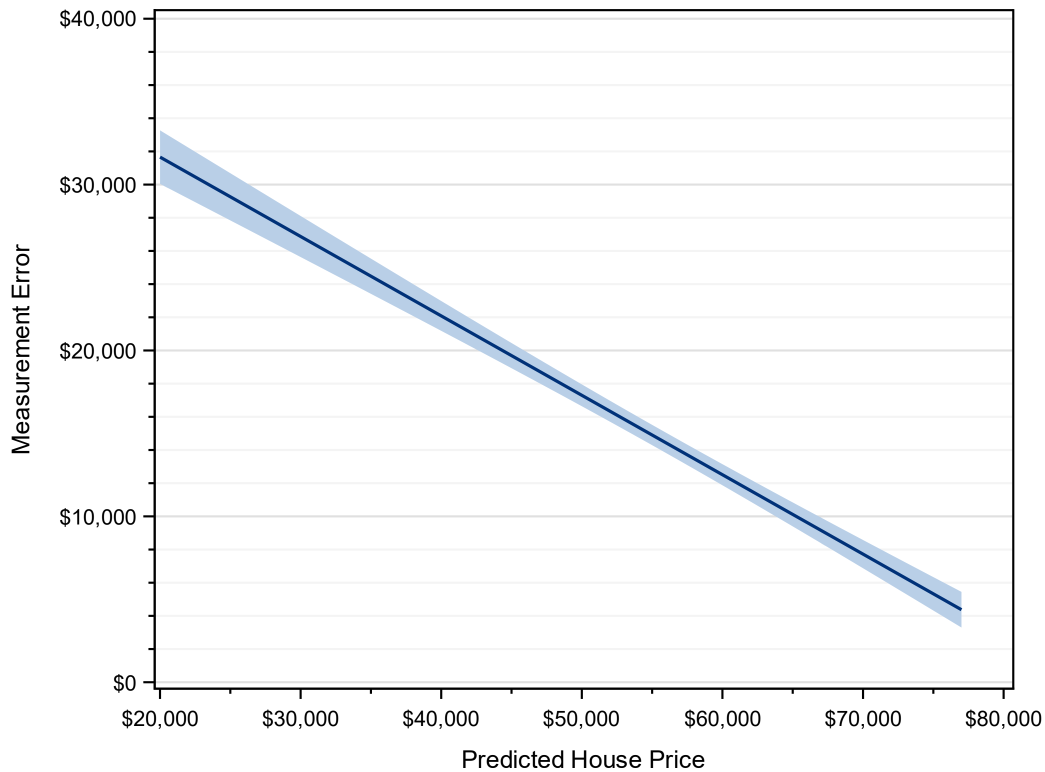

Figure 3: Differential Measurement Errors in the Lowest Cost Homes

ing this magnitude as the index price indicates a measurement error of \$14,518, which equates to true price of \$70,322. For an index price of \$25,000, the measurement error is \$29,264 or a transaction value of \$54,264.

To summarize the findings thus far, the regressions demonstrate significant differential measurement errors which can bias the findings of models used by researchers that include repeat sales indices. The biases are not entirely monotonic. The largest issue is with the lowest cost homes, which exhibit large negative slope coefficients in Table 4. For the bulk of the data in Deciles 2–9, the measurement errors correlate positively and significantly with the predicted house prices.

4 Differential Measurement Errors in a Mortgage Context

Given the widespread use of house price indices in mortgage literature, our next focus is on investigating measurement errors within a mortgage context. To achieve this, we augment loan data onto the MLS transactions, as detailed in Section 4.1. Section 4.2 investigates potential differential measurement errors in mortgage data, and Section 4.3 illustrates their impact on the estimations of the mortgage models.

4.1 Mortgage Data

The mortgage data come from Blackbox Logic BBx, which covers over 95% of US residential privately securitized loans and has detailed origination and monthly performance information. We restrict the sample to first liens with 30-year terms.[11] To merge the MLS and mortgage data, the first sale from MLS and the initial loan origination information from BBx are required to have the same zip codes, same transaction dates, and the ratio between the original appraised house value (BBx) to the original sales price (MLS) is between 0.99 and 1.10. We find these constraints are somewhat restrictive and our sample decreases markedly, nevertheless, we err on the side of accurately appending the correct mortgage and MLS records. To mitigate the influence of flippers, we again require a minimum of one year holding period like in the prior tests of predicted house prices. Next we examine the presence of differential measurement errors in mortgage data.

[11]Since our study focuses on the DFW Metroplex, we further mitigate the issue in mortgage modeling of unobserved second liens as Texas has a unique set of laws which discourage obtaining second mortgages after origination (Kumar, 2015). Individuals can have second mortgages, but the amounts tend to be lower than other jurisdictions.

4.2 Differential Measurement Errors in Mortgage Data

As outlined in Section 2, differential measurement error arises when it exhibits systematic variation with its own and/or some other model variables. Within a mortgage context, other than the predicted house price itself, we consider whether measurement error in house prices systematically varies with credit scores. If the measurement error associated with predicted house price varies systematically with credit scores, regardless of its correlation with the predicted house price, it will introduce bias in the coefficient estimate of credit score through the DGP. Moreover, if measurement error systematically varies with both predicted prices and credit scores, it can result in biased coefficient estimates for both variables through the DGP.

Using the merged sample along with the SPCCS/FHFA repeat sales index, we now estimate Equation (12) which extends (4) with the addition of credit scores. The regression in (12) explains the measurement error in log-prices $(p-\hat{p})$ as a function of the estimated log-price $\hat{p}$, the credit score at origination s_0 (logged), an intercept parameter κ_p, and the disturbances ϵ.

$$p - \hat{p} = \hat{p} \cdot \alpha + s_0 \cdot \pi + \kappa_p + \epsilon. \tag{12}$$

We are motivated to explore the variation in measurement errors concerning not only house prices but also credit scores for several reasons. Firstly, individuals with higher credit scores typically have greater liquidity, higher incomes, and own more valuable homes, indicating greater wealth. With increased liquidity and wealth, borrowers may allocate more funds towards home improvements and maintenance. If the market positively values these enhancements, the unrecorded housing investments by borrowers with higher credit scores are likely to result in higher home prices and appreciation levels than what the house price index indicates. Although the return on investment for maintenance and home improvements may be negative for homeowners, as long as it exceeds -100%, it will still contribute to an increase in market value.

Furthermore, Hayunga, Pace, and Zhu (2019) propose an alternative explanation for how measurement errors could vary based on credit scores and house prices. They argue that a higher likelihood of default leads to decreased investment in the home. Consequently, lower credit scores, associated with a higher probability of default, may lead to reduced investment in the home, resulting in lower future prices. Thus, credit scores and property values are interconnected through the investment channel, with both factors exhibiting measurement bias in the same direction.

The results of estimating (12) appear in Table 5. We report the results for both SPCCS and FHFA measurement errors. As previously mentioned, if the measurement error $(p - \hat{p})$ demonstrates pure randomness, the right-hand side variables of the predicted price $\hat{p}$ or initial credit score s_0 should not exhibit statistical significance. Conversely, significant slope coefficients would indicate the presence of differential measurement errors.

In Table 5, both of the SPCCS and FHFA indices exhibit positive correlations between the measurement errors and initial credit scores as well as predicted prices. Notably, our previous analysis, based on the full sample, did not reveal a strong relationship between measurement errors and log prices using the SPCCS MSA index, as shown in Table 4. However, these updated findings, derived from the merged sample, underscore that homes purchased by borrowers with higher (lower) initial credit scores, investing in more (less) expensive properties, tend to appreciate faster than predicted by the SPCSS index. This observation holds true even when credit scores are not factored in. The R^2 values highlight that predicted prices and credit scores collectively explain 13.25% and 12.63% of the variation in measurement errors in SPCCS and FHFA, respectively. Such substantial coefficients of determination emphasize the potential bias introduced when employing house price indices to model mortgage events like foreclosure sales, prepayments, and defaults.

Table 5: Measurement Error as a Function of Predicted Price and FICO

	SPCCS MSA		FHFA MSA	
Variable	Estimate	t	Estimate	t
Predicted Price	0.1216	8.43	0.1124	7.59
FICO	0.6567	8.76	0.7053	9.16
Intercept	−2.7926	−14.78	−2.8451	−14.55
R^2	0.1325		0.1263	

This table reports the regression results from Equation (12). The dependent variable is the measurement error, which equals the logged difference between the actual transaction price and the estimated house price. The independent variables are the estimated house prices (logged) and initial credit scores (logged). We report the results for the SPCCS and FHFA MSA indices.

4.3 Impacts of Differential Measurement Errors on Mortgage Models

This section examines the influence of differential measurement errors on mortgage studies, with a specific focus on foreclosure/REO sales. By comparing regression outcomes derived from observed transaction prices with those from estimated prices sourced from FHFA or SPCCS indices, we can discern the effects of measurement errors. As discussed in Section 2, differential measurement error has a systematic part and a random component. The systematic part of the differential measurement error leads to a different DGP of the estimation model than those using the true variable. The random component of the differential measurement error can attenuate the "biased" coefficient estimate, with more complex effects in a multivariate setting with correlations among the variables. So the effect we estimated here represents the total effect from those two resources.

Table 6 presents the results of linear probability regression models for foreclosure/REO sales. The dependent variable indicates whether a property sale is a foreclosure, coded as 1 for true and 0 otherwise. Models 1 to 3 employ observed sale prices (P), FHFA updated prices (computed by multiplying the price at origination by the FHFA index at property sale and dividing it by the FHFA index at loan origination), and SPCCS updated prices (computed similarly using the SPCCS index) to calculate the corresponding current loan-to-value ratio (LTV). Conversely, Models 4 to 6 express the current LTV ratios in log form (log(loan amount/price)). The independent variables include initial credit scores (FICO), dummy variables for fixed-rate mortgages (FRM), full documentation (FullDoc), and exotic mortgages (Exotic), which takes the value of 1 if a mortgage features a teaser rate or negative amortization, and a dummy variable indicating owner-occupied properties. All regressions incorporate fixed effects for loan origination year, property sale year, and county. Coefficient estimates and corresponding t-statistics are provided for each model.

Table 6 offers several insights. Firstly, models using the observed price demonstrate significantly stronger predictive power for foreclosure sales, with an R^2 of 38% compared to 18% when using the HPI

Table 6: LPRs of REO Sales Using Observed versus Estimated Prices

Model	(1)	(2)	(3)	(4)	(5)	(6)
Variable	P	$\hat{P}_{\text{FHFA}}$	$\hat{P}_{\text{SPCCS}}$	$\log(P)$	$\log(\hat{P}_{\text{FHFA}})$	$\log(\hat{P}_{\text{SPCCS}})$
FICO	−0.0918	−0.1752	−0.1760	−0.0733	−0.1740	−0.1749
	−4.78	−8.10	−8.14	−3.95	−8.04	−8.08
Current LTV	0.6626	0.5826	0.6193	0.8736	0.5167	0.5112
	20.66	4.24	4.24	23.77	4.37	4.35
FRM	0.0003	−0.0337	−0.0343	0.0172	−0.0333	−0.0338
	0.01	−1.10	−1.12	0.67	−1.08	−1.10
FullDoc	−0.0449	−0.0570	−0.0577	−0.0483	−0.0565	−0.0572
	−1.82	−1.98	−2.00	−2.04	−1.97	−1.99
Exotic	0.0452	0.0178	0.0185	0.0483	0.0172	0.0179
	1.55	0.53	0.55	1.73	0.51	0.53
OwnerOccupied	−0.0167	−0.0505	−0.0498	−0.0214	−0.0490	−0.0486
	−0.50	−1.30	−1.28	−0.66	−1.26	−1.26
R^2	0.3834	0.1808	0.1808	0.4308	0.1816	0.1814

This table reports the outcomes of linear probability regression models focusing on REO sales. The dependent variable indicates whether a property sale corresponds to a foreclosure sale, with a code of 1 denoting true and 0 otherwise. Models 1 to 3 employ different price metrics: observed sale price (P), FHFA updated price (calculated by multiplying the price at origination by the FHFA index at property sale and dividing it by the FHFA index at loan origination), and SPCCS updated price (similarly computed with the SPCCS index) to determine the corresponding current loan-to-value ratio (LTV). Meanwhile, Models 4 to 6 express the current LTV ratios in log form. Additional independent variables include initial credit scores (FICO), a binary variable representing fixed-rate mortgages (FRM), a dummy variable indicating full documentation (FullDoc), and another binary variable for exotic mortgages (Exotic), assigned a value of 1 if a mortgage features a teaser rate or negative amortization. A dummy variable also denotes owner-occupied properties (OwnerOccupied). All regressions incorporate fixed effects for loan origination year, property sale year, and county. Coefficient estimates and corresponding t-statistics are provided for each model.

estimated price. Secondly, the coefficient estimates using the estimated price exhibit significant bias for both credit score and current LTV coefficients. For instance, the magnitude of the coefficient estimate for FICO in Model 2 using the FHFA estimated price is 90% larger than in Model 1 using the observed price. Similarly, the coefficient estimate for current LTV in Model 2 using the FHFA estimated price is 12% lower than in Model 1 using the observed price. Additionally, the t-statistics of the coefficient estimates for current LTV and FICO vary significantly between Model 1 (or 4) versus 2 and 3 (or 5 and 6). Thirdly, the bias becomes even more pronounced when using logged prices. For instance, the coefficient estimates for FICO in Model 5 using logged HPI updated prices are 137% larger than in Model 4 using logged observed prices. Lastly, the estimation bias appears to be similar when using FHFA and SPCCS indices.

In summary, this section demonstrates that (1) there is a prevalence of differential measurement errors in predicted house prices as a function of both predicted house prices and credit score in mortgage data, and (2) these measurement errors (including both the systematic part and the random component together) not only diminish the predictive power of mortgage models but also introduce significant biases in the coefficient estimates of those variables correlated with the measurement errors.

5 Conclusion

A myriad of research topics within the housing literature rely on home prices as an explanatory variable, spanning inquiries into mortgage behavior, migration patterns, housing affordability, and the effects of housing wealth on the economy. Often, these models use estimates of housing value movements through price indices, which can introduce measurement error. The extent of this error can significantly impact coefficient estimates in regression models.

This study explores the effects of measurement errors stemming from the use of house price indices. We first show that the measurement error associated with indexed prices varies systematically with the predicted house prices. This is a form of differential measurement error that will introduce estimation bias through the DGP. Next, in a mortgage setting, our results unveil substantial differential measurement error in indexed house prices that varies systematically with both credit scores and predicted house prices. The differential measurement error, which includes both a systematic part and a random component, leads to significantly lower predictive power of the regression models and large biases in the coefficient estimates.

To mitigate biases from using house price indices, we propose several methods. Firstly, researchers can employ internal validation studies to estimate and correct bias. Secondly, leveraging hedonic models proves effective in alleviating much of the differential measurement error. Thirdly, avoiding highly non-linear transformations of estimated variables can reduce dependency on measurement error. Lastly, avoiding mismatches between overall market appreciation and price changes in specific submarkets can mitigate bias. Some geographies offer low, medium, and high price indices, which may further alleviate estimation bias. By adopting these strategies, researchers can enhance the accuracy and reliability of their analyses, ensuring more robust conclusions and informed policymaking within the realm of housing research.

6 References

Ambrose, Brent W., Charles A. Capone Jr., and Yongheng Deng (2001), "Optimal Put Exercise: An Empirical Examination of Conditions for Mortgage Foreclosure," *Journal of Real Estate Finance and Economics*, Vol. 23, No. 2, pp. 213–234.

Ambrose, Brent W., Michael LaCour-Little, and Anthony B. Sanders (2004), "The Effect of Conforming Loan Status on Mortgage Yield Spreads: A Long Level Analysis," *Real Estate Economics*, Vol. 32, No. 4, pp. 541–569.

Balla, Eliana, Raymond Brastow, Daniel Edgel, and Morgan Rose (2024), "The Effect of Regulatory Oversight on Nonbank Mortgage Subsidiaries," *Journal of Real Estate Finance and Economics*, Vol. 68, pp. 523–575.

Carroll, Christopher D., Misuzu Otsuka, and Jiri Slacalek (2011), "How Large are Housing and Financial Wealth Effects? A New Approach," *Journal of Money, Credit and Banking*, Vol. 43, No. 1, pp. 55–79.

Carroll, Raymond, David Ruppert, Leonold Stefanski, and Ciprian Crainiceanu (2006), *Measurement Error in Nonlinear Models*, Second edition, CRC Press.

Case, Karl. E., John M. Quigley, and Robert J. Shiller (2005), "Comparing Wealth Effects: The Stock Market versus the Housing Market," *Advances in Macroeconomics*, Vol. 5, No. 1.

Crow, Edwin L. and Kunio Shimizu (1988), *Lognormal Distributions*, New York, Marcel Dekker.

Davidoff, Thomas (2013), "Supply Elasticity and the Housing Cycle of the 2000s," *Real Estate Economics*, Vol. 41, No. 4, pp. 793–813.

Davis, Morris A. and Jonathan Heathcote (2007), "The Price and Quantity of Residential Land in the United States," *Journal of Monetary Economics*, Vol. 54, No. 8, pp. 2595–2620.

Demyanyk, Yuliya and Otto Van Hemert (2011), "Understanding the Subprime Mortgage Crisis," *Review of Financial Studies* Vol. 24, No. 6, pp 1848—1880.

Deng, Yongheng, John M. Quigley, and Robert Van Order (2000), "Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options," *Econometrica*, Vol 68, No. 2, pp. 275–307.

Goda, Gopi Shah, John B. Shoven, and Sita Nataraj Slavov (2011), "What Explains Changes in Retirement Plans during the Great Recession?" *American Economic Review: Papers and Proceedings*, Vol. 101, No. 3, pp. 29–34.

Guerrieri, Veronica, Daniel Hartley, and Erik Hurst (2013), "Endogenous Gentrification and Housing Price Dynamics," *Journal of Public Economics*, Vol. 100, pp. 45–60.

Harding, John, John Knight, and C.F. Sirmans (2003), "Estimating Bargaining Effects in Hedonic Models: Evidence From the Housing Market," *Real Estate Economics*, Vol. 31, No. 4, pp. 601–622.

Harding, John P., Eric Rosenblatt, and Vincent W. Yao (2009), "The Contagion Effect of Foreclosed Properties," *Journal of Urban Economics*, Vol. 66, No. 3, pp. 164–178.

Hayunga, Darren K., R. Kelley Pace, and Shuang Zhu (2019), "Borrower Risk and Housing Price Appreciation," *Journal of Real Estate Finance and Economics*, Vol. 58, No. 4, pp. 544–566.

Heien, Dale M. (1968), "A Note on Log-Linear Regression," *Journal of American Statistical Association*, Vol. 63, No. 323, pp. 1034–1038.

Inklaar, Robert and J. Christina Wang (2013), "Real Output of Bank Services: What Counts is What Banks Do, Not What They Own," *Economica*, Vol. 80, No. 317, pp. 96–117.

Kau, James B., Donald C. Keenan, Walter F. Muller, James F. Epperson (1992), "A Generalized Valuation Model for Fixed-rate Residential Mortgages," *Journal of Money, Credit and Banking*, Vol. 24, No. 3, pp. 279–299.

Kumar, Anil (2015), "Do Restrictions on Home Equity Extraction Contribute to Lower Mortgage Defaults? Evidence from a Policy Discontinuity at the Texas Border," *American Economic Journal: Economic Policy*, Vol. 10, No. 1, pp. 268–297.

Judge, Griffiths, Hill, Lütkepohl, and Lee (1985), *The Theory and Practice of Econometrics*, 2nd edition, John Wiley and Sons, New York.

Liu, Zheng, Pengfei Wang, and Tao Zha (2013), "Land-price Dynamics and Macroeconomic Fluctuations," *Econometrica*, Vol. 81, No. 3, pp. 1147–1184.

Mian, Atif, Amir Sufi, and Francesco Trebbi (2015), "Foreclosures, House Prices, and the Real Economy," *The Journal of Finance*, Vol. 70, No. 6, pp. 2587–2634.

McMillen, Daniel and Ruchi Singh (2020), "Assessment Regressivity and Property Taxation," *Journal of Real Estate Finance and Economics*, Vol. 60, No. 1–2. pp. 155–169.

Rosenthal, Stuart S. (2014), "Are Private Markets and Filtering a Viable Source of Low-Income Housing?" *American Economic Review*, Vol. 104, No. 1. pp. 687–706.

7 Appendix A

Since the amount of time a home is on the market is determined simultaneously with the transaction price, the hedonic models use two stage least squares. The first stage instruments time on the market (TOM) and the predicted value is used in the second specification modeling prices.

The specification includes many typical possible determinants as well as some unique independent variables. Some common explanatory variables are the natural log of home size (square feet), lot size, structure age, construction type (e.g., brick), and the presence of a pool. We also include HOA fees, cash or conventional financing, whether the home is occupied, vacant, or rented, and the types of sellers as owners, foreclosures, and relocation firms. To control for property quality, we include myriad amenities such as a security and sprinkler systems, vaulted ceilings, dual vanities, walk-in closets, etc. Since they are informative, we also include the number of photographs included in the listing at sale and how the square-feet measure is determined (e.g., tax records, appraisal, or building plan).

We follow Harding, Knight, and Sirmans (2003) to compute binary variables of atypicality, which represent the extreme one percent of the distribution for various structural features. We define a new home as 1 year old while an old home is equal to or greater than 86 years old. A large home is greater than 5,300 square feet and a small home is less than 800 square feet. A home has many bathrooms if there are 5 or more and many bedrooms if 5 or more. Lastly, the specifications use spatial and temporal fixed effects. The spatial controls account for up to 150 locally-defined submarkets within the DFW Metroplex. The temporal fixed effects are both quarterly for seasonality and annually for macroeconomic movement in the housing market.

Table 7: Hedonic models using 2SLS

Variables	SPCCS Log Prices	FHFA Log prices
Predicted Log TOM	−0.055**	−0.032
	(0.010)	(0.016)
Log square feet	0.840**	0.891**
	(0.007)	(0.008)
Lot lot size	0.010**	0.011**
	(0.000)	(0.001)
Log home age	−0.028**	−0.020**
	(0.001)	(0.001)
Garage spaces	0.103**	0.082**
	(0.001)	(0.001)
Carport spaces	0.021**	0.020**
	(0.001)	(0.001)
Areas count	−0.020**	−0.022**
	(0.001)	(0.001)
Stucco construction	0.175**	0.238**
	(0.004)	(0.007)
Rock construction	0.130**	0.159**
	(0.002)	(0.003)
Siding construction	−0.064**	−0.037**

Continued on next page

Variables	SPCCS Log Prices	FHFA Log prices
	(0.001)	(0.002)
Wood construction	−0.029**	−0.035**
	(0.001)	(0.002)
Vacant	−0.035**	−0.043**
	(0.001)	(0.003)
Tenant occupied	−0.025**	−0.027**
	(0.003)	(0.005)
HOA fees	0.000**	0.000**
	(0.000)	(0.000)
Photocount	0.008**	0.010**
	(0.000)	(0.000)
Cash financing	−0.159**	−0.191**
	(0.001)	(0.003)
Conventional financing	0.011**	0.022**
	(0.001)	(0.002)
Owner seller	0.079**	0.095**
	(0.002)	(0.003)
REO seller	−0.172**	−0.179**
	(0.003)	(0.006)
Large home	0.263**	0.327**
	(0.008)	(0.012)
Small home	0.079**	−0.042**
	(0.007)	(0.011)
New construction	0.028**	0.050**
	(0.004)	(0.008)
Old home	0.100**	0.046**
	(0.008)	(0.013)
Many bedrooms	−0.058**	−0.067**
	(0.007)	(0.012)
Many bathrooms	0.142**	0.144**
	(0.007)	(0.011)
Sqft from tax records	−0.026**	−0.031**
	(0.002)	(0.003)
Sqft from appraisal	0.001	0.037**
	(0.002)	(0.004)
Pool	0.092**	0.072**
	(0.001)	(0.002)
Security system	0.037**	0.058**
	(0.001)	(0.001)

Continued on next page

Variables	SPCCS Log Prices	FHFA Log prices
Deck	0.030**	0.088**
	(0.001)	(0.002)
Vaulted ceilings	0.002**	−0.013**
	(0.001)	(0.001)
Hot tub	0.010**	0.021**
	(0.001)	(0.002)
Dual vanities	0.019**	0.012**
	(0.001)	(0.001)
Workshop	0.011**	−0.014**
	(0.001)	(0.002)
Sprinkler system	0.052**	0.066**
	(0.001)	(0.002)
Wet bar	0.008**	−0.009**
	(0.001)	(0.002)
Window treatments	0.013**	0.022**
	(0.001)	(0.001)
Intercom	0.020**	0.008**
	(0.002)	(0.003)
Thermopane windows	0.015**	0.010**
	(0.001)	(0.001)
Tile floors	0.045**	0.029**
	(0.001)	(0.001)
Constant	5.346**	4.742**
	(0.028)	(0.041)
Observations	596,778	249,772
Adjusted R^2	0.862	0.852

Robust standard errors in parentheses.
** denotes p-value< 0.01 and * is p-value< 0.05

8 Appendix B

To maintain consistency and accommodate the likelihood that users lack insight into the cash trades, Table 2 includes cash sales across all cases. However, since the FHFA index derives its measures from mortgages, cash transactions are not incorporated at either the first or second trades of a repeat sale. This distinction may influence the statistics presented in Table 2. Consequently, we exclude cash sales and recalculate the descriptives and report the results in Table 8. Columns 1 and 3 repeat the magnitudes from Table 2, while columns 2 and 4 present results using the sample excluding cash transactions.

In accordance with a more representative sample, the baseline central tendency measures of mean and median indicate reduced baseline mean and median measurement errors in columns 2 and 4. However, the standard deviation and other statistical measures of measurement errors remain largely consistent with previous findings. Excluding cash sales from the sample results in an increased kurtosis of measurement errors in columns 2 and 4.

Table 8: Remove Cash Sales from FHFA Data

	FHFA MSA	FHFA MSA	FHFA ZIP Code	FHFA ZIP Code
Mean e	-0.0239	-0.0044	-0.0480	-0.0124
Median e	-0.0139	-0.0068	-0.0164	-0.0037
$\sigma(e)$	0.3283	0.2544	0.3044	0.2439
Mean $\|e\|$	0.2184	0.1647	0.2005	0.1552
Median $\|e\|$	0.1223	0.0958	0.1118	0.0907
Kurtosis e	6.7570	6.9138	7.5460	11.3301
Mean $\ln(P)$	12.0088	12.1322	12.0293	12.1414
Median $\ln(P)$	11.9954	12.0951	12.0015	12.0856
$\sigma(\ln(P))$	0.7377	0.6432	0.7947	0.6993
Include Cash Sales	Yes	No	Yes	No
n	44,424	32,197	42,527	32,521

Like in Table 2, the results in this table report summary statistics of the measurement errors incurred using the FHFA index to model repeat sales at the MSA and ZIP Code levels. The first and third numerical columns are the same as in Table 2. The second and fourth numerical columns do not include cash transactions at either the purchase or sale of homes in the sample. The data includes summary statistics of log transactions prices.

9 Appendix C

Housing, mortgage, and broader economic research consistently rely on house price levels, rather than the natural log of prices, to model changes over time. Studies, such as Demyanyk and Van Hemert (2011), Hayunga, Pace, and Zhu (2019), and Balla et al. (2024), are testament to this trend. In line with this practice, we compute measurement error regressions by analyzing the differences in price levels between the indices and the actual transaction prices. Table 9 provides the slope coefficients and additional statistics across the different samples same as in Table 4, shedding light on the magnitude and implications of these measurement errors.

Table 9: Measurement Error Regressions using Price Levels

	Coefficient	t-stat	R^2	n
SPCCS MSA Full	0.0986	63.80	0.0369	106,168
FHFA MSA Full	0.0787	29.93	0.0198	44,424
FHFA ZIP Code Full	0.1164	43.06	0.0418	42,527
SPCCS MSA Conforming	0.0812	40.81	0.0162	101,216
FHFA MSA Conforming	0.1381	37.41	0.0322	42,027
FHFA ZIP Code Conforming	0.1743	43.12	0.0450	39,488
SPCCS MSA Trimmed	0.1612	61.07	0.0421	84,936
FHFA MSA Trimmed	0.2149	44.82	0.0535	35,540
FHFA ZIP Code Trimmed	0.2112	43.31	0.0533	34,023
SPCCS MSA Lowest Cost	−0.4787	−22.50	0.0455	10,616
FHFA MSA Lowest Cost	−0.4680	−14.41	0.0447	4,442
FHFA ZIP Code Lowest Cost	−0.6104	−20.49	0.0900	4,252

This table presents regressions of the measurement errors levels as the dependent variable against the independent variable of predicted house prices. The coefficients stated in the first numerical column measure differential measurement errors. The larger the coefficient in absolute value the greater the potential bias in using the index for that specific sample.