YALE
INTERNATIONAL
CENTER
FOR FINANCE

Yale ICF Working Paper No. 03-05
April 2003

Estimating Indices in the Presence of Seller Reservation Prices

William Goetzmann
Yale School of Management

Liang Peng
University of Cincinnati

This paper can be downloaded without charge from the
Social Science Research Network Electronic Paper Collection:
http://ssrn.com/abstract_id=387901

Estimating Indices in the Presence of Seller Reservation Prices*

William Goetzmann
Yale School of Management

Liang Peng
College of Business, University of Cincinnati
Box 210195
Cincinnati, OH 45221-0195
E-mail: liang.peng@uc.edu

Current Draft: April 2003

Abstract

This paper documents the potential bias induced in an index of asset prices when sellers use reservation rules that may include some component of private value. We develop a model in which the seller's asking price is determined by private valuation while the buyer's bid price is determined by the market valuation, and a transaction takes place only if the bid is higher than the ask. Therefore, the trading volume and the observed transaction prices are both affected by the ratio of seller's private valuation to the market valuation, which is called the seller reserve ratio. The higher the seller reserve ratio, the lower is the trading volume and the larger is the difference between the market valuations estimated using observed prices and the actual market valuations.

To address the estimation problem posed by the bias, we propose a three-step econometric procedure. We first estimate the index using observed prices. We next use residuals from the first step and observed trading volume to estimate the latent series of seller reserve ratio and the unconditional population variance of pricing errors. We then use these estimates in step two to correct for the bias in the index based on observed prices in either a hedonic regression or a repeat sale regression. We call the unbiased index a reserve-conditional index.

Simulations show that this remedy effectively mitigates the bias, and the reserve-conditional indices are more accurate than traditional hedonic and repeat sale indices. We apply this technique to Los Angeles housing market, and show that the reserve-conditional index substantially differs from a traditional repeat sale index. In our application, the reserve-conditional index is more volatile, has a much smaller autocorrelation, and appears to capture market downturns in a more timely fashion than the conventional repeat sale index.

JEL classification: C51, G10

---

* We thank John Clapp, Norman Miller, and Matthew Spiegel for helpful suggestions on this paper. All errors are ours alone. We thank MRAC for the use of the Los Angeles data for research. Part of this paper circulated previously as "The Effect of Seller Reserves on Market Index Estimation."

Introduction

Virtually all financial indices are constructed from transactions prices. For very liquid markets such as the New York Stock Exchange, in which assets trade every day, the weighted average price adequately captures daily market dynamics. But for less liquid markets, such as the bond market, the housing market and the art market, the average observed transaction price might not be as informative. In fact, even the construction of an intra-day index of the NYSE poses a challenge due to illiquidity. Not all securities transact in each interval of time, and the structure of the market is such that limit orders as well as market orders affect estimates of the time-series of index returns. In short, price indices in virtually all asset markets are conditional upon liquidity – it is simply the definition of the interval of observation that determines the magnitude of the estimation problem.

Sellers of all sorts – rational or irrational, informed or uninformed – set reservation prices. For purely common value goods, the reserve indicates the seller's assessment of the economic value of the property. For instance, for a share of stock, the reserve might represent an assessment of the net present value of the future sales price, plus the discounted dividend stream over the time until sale. As such, the bid and ask prices for stocks can reflect differing individual opinions about economic value. For private value goods, the seller reserve also reflects the personal satisfaction the agent gets from ownership and use. For instance, certain things such as a lock of hair have sentimental value to the owner, but virtually no value to a potential buyer. Other goods, like paintings, are a combination of private and public valuation. While many artworks are purchased for investment, certain owners derive extraordinary pleasure from owning particular pieces, and might set a reserve higher than the going "price" for the painting. The same

can be said for single-family houses. Much home improvement actually detracts from the resale value of the house, while enhancing the private value to the owner.[1]

Seller reserves are integral to the auction process. For instance, virtually all paintings offered at major auctions have a secret "reserve" price known only to the consignor and the auction house.[2] Since houses, art, antiques, stamps, coins, and other collectibles have a private value component that is difficult to quantify, there are no clear economic rules about how to set such reserves. Reservation prices might be chosen based upon the seller's deep-felt conviction about how painful it would be to part with the piece, or they might be based upon what the latest example of such work happened to go for. It is important to point out that a seller's choice to set a reserve based upon his or her private value for the good is not necessarily irrational. Indeed, the seller may be setting the reserve in such a way as to maximize utility based upon the future enjoyment of the artwork and upon revenues derived from sale.

Although the real estate markets and the art markets are natural settings to study the effects of private valuation and seller reserves on index construction, much of the recent behavioral finance literature suggests that private valuations may play a significant role in investor choice. For example, the disposition effect is well documented in empirical studies of investor behavior.[3] Disposition-prone investors are adverse to selling shares at a loss, where loss is defined as the original purchase price less the current sales price. While this may not seem rational behavior, it is consistent with a mixed private value/common value market, and recent empirical evidence suggests that it may have an effect on realized returns of individual securities (e.g. Grinblatt and Han, 2002) and the index as a whole (Goetzmann and Massa, 2003).

---

[1] Homes are rarely auctioned in the typical sense; however, sellers often set reserves and bids not meeting the asking price may be refused. Other types of real estate may be auctioned with reserve. See for instance, the Ashenfelter and Genovese (1992) study of condominium auctions.

[2] Some private value goods are auctioned without reserve. Stamp auctions, studied by Taylor (1983), for instance, were without reserve.

[3] For example, see: Shefrin and Statman (1985), DeBondt and Thaler (1985), Barber and Odean (2000, 2001, 2002), Odean (1998, 1999), Weber and Camerer (1998), Weber and Zuchel (2001), Oehler et al.

In this paper, we demonstrate that if reserves are set by investors with a significant component of private valuation, then the index of returns estimated from transaction prices is typically biased away from the actual market value index. The sample setting for our analysis is the comparatively illiquid housing market; however, we believe that our results and methods are potentially applicable to all markets in which private-values-based seller reserves may be salient in the price formation process. The intuition for our results is straightforward. Regardless of whether seller behavior is rational or irrational, seller reservation rules "censor" observed market transactions. For some simple rules such as a known price threshold, it may be possible to estimate the unconditional distribution. However, in most settings it is not. Thus we need a flexible empirical methodology to estimate the actual market value index with seller reserve controlled, which we call a reserve-conditional index.

We document the index estimation problem through simulation. Next, we develop a model that incorporates seller reserves. We then estimate this model on housing transactions data for the Los Angeles housing market. We find that our index estimation model generates significantly different results from standard procedures.

Simulations

The results of our simulations are particularly striking in light of the increasing reliance by the mortgage industry and housing markets on on-line, transactions-based local housing indices. We find that, when an index is based upon the average price of works *that sell*, the fluctuation in that index might be statistically unrelated to the index based upon an average across all properties in the market. In fact, under certain conditions, we find no correlation between the true, uncensored indices and the estimated indices. The results of these simulations have a direct bearing on performance measures used in a number of asset markets. When trend estimates are based upon average transaction prices, they may be incorrect.

(2002), Dhar and Zhu (2002), Ranguelova (2001), Grinblatt and Keloharju (2001) and Genesove and Mayer (2001).

Model

To address the censoring problem, we develop a simple trading model of illiquid assets. The model makes useful predictions about the relationship between seller reserve levels and the volume of transactions. In our model, the ask price is determined by the seller's private valuation of the underlying asset and the bid is determined by the market valuation of the asset. A transaction takes place if the bid is higher than the ask, and the transaction price equals the bid. The seller's private valuation may differ from the market valuation. The difference is measured by the ratio of seller's private valuation to the market valuation, which is called seller reserve ratio.

The seller reserve ratio in the model essentially serves as a threshold. It affects not only the trading volume but also the bias in the observed prices. Holding constant all other economic variables, the higher is the seller reserve ratio, the less likely a transaction takes place. Consequently, there is a negative relation between the seller reserve ratio and the trading volume in the market. Furthermore, conditional upon a transaction being observed, the higher the seller reserve ratio, the larger the positive deviation of the observed transaction price from the market valuation. Therefore, we predict a positive relationship between the seller reserve ratio and the upward bias in observed prices.

Since the seller reserve ratio may change over time, both the trading volume and the magnitude of the bias of the observed prices may change over time. For instance, if the seller reserve ratio increases from one period to the next, the trading volume would decrease, and the magnitude of the bias of the observed prices would increase. Since the bias of the observed prices is larger in the second period than in the first period, indices constructed using observed prices will estimate the return from the first period to the second period with an upward bias.

Thus, our model makes unequivocal predictions regarding the direction and magnitude of the bias of market indices constructed using observed transactions. When trading volume increases (decreases), returns are estimated with downward (upward) bias. Moreover, the bias is

more significant if the change in the trading volume is more dramatic. The predictions have significant economic implications. For example, high returns may accompany increasing trading volume when markets go up, and low returns may accompany decreasing volume when markets go down. Without controlling for seller reserves, returns are underestimated in booming markets and overestimated when markets go down. Hence, the variance and volatility of returns may also be underestimated.

Econometric solution

Using our model, we propose a remedy for the bias caused by the time-varying seller reserve ratio. The remedy has the following merits. First, it is compatible with both the hedonic regression and the repeat sale regression, which are commonly used to estimate indices in housing markets, art markets and to some limited extent, bond markets. Second, the remedy makes no assumption regarding the generation of the seller reserves, and does not require information about the variables that may help determine seller reserves. Thus, it does not implicitly assume that investors are prone towards the disposition effect, however it allows for it. The only required information includes the transaction time and prices, and the time series of trading volume measured by the ratio of traded assets to the number of all assets on the market. Simulation verifies that our remedy effectively mitigates the bias of market indices and dramatically improves the accuracy of index estimation.

Application to the housing market

We apply our reserve-conditional estimation procedure to a repeat sale index of Los Angeles housing from 1976 to 2001, and contrast it with an index estimated with the conventional repeat sale regression using the same data. The reserve-conditional index demonstrates very different risk and return characteristics compared to the traditional repeat sale index. First, the reserve-conditional index has lower returns than the traditional index when trading volume decreases. Thus, in periods when liquidity dries up, returns estimated using transactions are likely

to be positively biased: investors are not able to realize the returns. Second, the first order autoregressive coefficient is not significant different from 0 for the reserve-conditional index, while it is 0.27 for the traditional repeat sale index. Therefore, the serial correlation of illiquid asset index returns may be an artifact caused by censored transactions, and the latent actual market value index is unpredictable. Finally, the reserve-conditional index returns are more volatile than the conventional index returns – this is particularly noticeable in periods of market declines. The raw index smoothes out drops, while the reserve-conditional index reflects drops more immediately. The standard deviation of semiannual returns is 4.50% for the reserve-conditional index, and 3.82% for the traditional index, which is a substantial difference from a risk control perspective.

Related literature

This paper is related to the literature of sample selectivity bias, which is pioneered by Heckman (1974, 1976, and 1979), Hausman and Wise (1977), and Lee (1982), among others. The sample selectivity bias has long been recognized a serious problem in many areas of economics such as labor economics. The bias and its corrections are also important for the estimation of illiquid asset indices. For example, Case, Pollakowski and Wachter (1991), Gatzlaff and Haurin (1997, 1998), and Meese and Wallace (1997), among others, show that house indices constructed based on a sample of sold houses (or repeated sold houses) may be subject to substantial sample selectivity bias. Various methods, such as the Heckman two-step approach, have been applied to mitigate the bias in index estimation.

This paper contributes to this literature by proposing a novel econometric remedy to the sample selectivity bias in index estimation. The remedy has some important merits. First, it is conceptually intuitive and computationally convenient. In addition, it is fully compatible with the hedonic and repeat sale regressions, which are classic methods of index estimation for illiquid assets. Moreover, in contrast to the Heckman two-step approach, our method does not depend on a specific generating process of seller reserves; therefore, misspecification and latent variables are

less likely a problem. More importantly, our method is less data-demanding than the Heckman approach: our method only needs trading time, prices, and volume to estimate unbiased indices.

This paper is also related to the literature of constant liquidity indices. Fisher et al. (2003) propose the economic concept and estimation methodologies of indices with constant liquidity thus constant ease of sales. Constant liquidity indices facilitate more economically sensible comparison of asset values across time. The reserve-conditional index proposed in this paper is essentially a constant liquidity index because we control for the seller reserves and thus the probability of transaction. More over, the reserve-conditional index represents market valuation of assets instead of seller's private valuation.

Structure

This paper is structured as follows. Section one provides some simulation evidence of biases in estimated indices due to some plausible reservation models. Section two develops the trading model in which trading volume and transaction prices are jointly determined by the interaction between the market and private valuation of underlying assets. It also investigates the biases of the traditional hedonic and the repeat sale indices. Section three proposes a remedy for the biases of indices estimated with hedonic and repeat sale regressions. Section four conducts simulations to verify the bias pattern predicted by our model, and investigate the performance of the proposed remedy. Section five applies the remedy to construct a house index of Los Angles from 1976 to 2001, and contrast it with a traditional repeat sale index. Section six concludes.

I. Simulations of Seller Reservation Rules

To examine the effects of different reservation rules upon the average price index, we repeatedly simulate an asset market over a period of time in the following manner. For each simulation, prices for 100 assets in the initial period are distributed uniformly over the interval [0,1]. Returns in the subsequent 40 periods are generated by a multiplicative market model, in which all assets have the same sensitivity to the market. For simplicity, returns are lognormal,

and simulations are performed in logs. Log market returns are distributed normally with positive drift of 0.05 and standard deviation of 0.1. Residuals are distributed normally with zero drift and standard deviation of 0.4.

Reservation Rules

Seller reserves are expressed as a conditioning rule at time t. $P_{i,t}$ is observed [denoted as $p_{i,t}$] conditional upon it exceeding a seller threshold. Rule one called "Start" in the table, conditions upon price at time 0. That is $P_{i,t} = p_{i,t} \mid P_{i,t} > P_{i,0}$. This assumes that all prices are observed in the initial period, and that a sale is only made if the price increases. The rationale for rule one is that a sale occurs only when the price increases beyond its beginning value. Rule two, called "Max", conditions upon the maximum past observed value: $P_{i,t} = p_{i,t} \mid P_{i,t} > \max\left\{P_{i,0}...P_{i,t-1}\right\}$. The rationale for rule two is that the seller will wait until the asset price exceeds its historical high. Rule three, called "Random" in the table, conditions upon a previous purchase price, chosen randomly from earlier observed transactions for the asset, assuming that all assets trade in the first period: $P_{i,t} = p_{i,t} \mid P_{i,t} > random\left\{P_{i,0}...P_{i,t-1}\right\}$. The rationale for rule three is that a sale occurs only when the price exceeds the buyer's purchase price, where the buyer may have acquired the asset at a transaction chosen at random from the asset's transaction history. This rule simulates the disposition effect. Rule four, called "Quality," conditions upon the N transaction prices for assets observed last period, and scales them according to the known quality differences observed in the first period: $P_{i,t} = p_{i,t} \mid P_{i,t} > \left(P_{i,0}\right)\max\left\{P_{i,t-1}...P_{N,t-1}\right\}$. The rationale for rule four is that the seller observes the best asset sold for the highest price last period, and then scales the asking price for quality variation.[4] When the condition cannot be satisfied, the price is not observed. We investigate the

[4] Notice that rule four implicitly assumes that the seller takes the maximum price observed last period as the price for the highest quality good. This price is then scaled down by the quality differential observed in

effects of seller reserve rules by calculating the simple average price indices using conditional transactions, and compare them with the actual price indices. We also estimate two additional indices, based upon rule three transactions, via a maximum likelihood procedure called the *repeat-sales regression*. This procedure is discussed in the following section.

Repeat-Sales Regression

The repeat-sales regression is used in both real estate and art market research to address the problem of infrequent transactions in asset markets, and to control for quality variation in transacting assets from auction to auction (see Bailey, Muth and Nourse (1963), Case and Shiller (1987) for real estate market examples, and Anderson (1974) and Goetzmann (1993) for art market examples). The regression uses only matched buy prices and sale prices to infer changes in returns from period to period. When prices for all assets are observed each period the regression is equivalent to a simple average across log returns each period, however in relatively illiquid markets such as the art market or the real estate market, asset prices are not observed each period.[5] For such markets, when the time-series of returns of a given asset are log-normally distributed with i.i.d. errors around a market index, the repeat-sales regression represents the maximum-likelihood estimate of the equal-weighted index of all assets in the sample, calculated each period, regardless of whether the prices are observed or not. The model has also been adapted to conditions where returns contain a non-temporal component. Goetzmann and Spiegel (1995), for instance, include an intercept term in the regression to control for the components of asset returns unrelated to the stochastic errors or the trend in the index. There are reasons to expect that the selection bias due to seller reserves is non-temporal in nature. For instance, the magnitude of the "jump" from one transaction to the next caused by the selection bias is unrelated to the interval between sales. Thus, it is possible that the Goetzmann and Spiegel (1995)

---

the first period between the good held by the buyer and the highest quality good. This assumes that the buyer knows exactly the relative quality of his or her good. Taylor (1983) shows how uncertainty about quality can lead to the actual choice to bring a good to market.

methodology may mitigate the selection bias. We report the results based upon the repeat-sales regression with and without the intercept term.

For purposes of study, we apply the repeat-sales regression to the conditional transactions generated under the third rule – the disposition effect rule, i.e. owners only sell if the bid exceeds their own purchase price. The regression is performed in the following manner. Note in equation one that the log return over any period, say, from time b to time s, for $s > b$, is specified as the a log price difference $P_{i,s} - P_{i,b}$ (which we denote as $R_{i,s,b}$). The log market returns for each time period are estimated via a regression of the form $R = X\mu + \varepsilon$, where R is a vector of all available returns formed from observed repeated asset sales, X is a dummy matrix with the number of columns equal to the number of time periods over which market data is available. Indicator variables in columns of X identify the time periods from the purchase date to the sale date. μ is a vector of market log returns to be estimated by the regression. ε is the regression error, which is proportional to the number of time periods between the purchase and sale dates. The advantage of the repeat-sales regression is that it does not limit the estimation of the mean to only those assets transacting in a given period t. Sales in periods later than t may contain information about the period t return, and the regression makes use of this information.

On additional consideration is what to do about first period prices. The simulation implicitly assumes that prices in the first period are unconditional. The return from an unconditional price to a price conditional upon exceeding a lower bound is certainly upwardly biased. Because of the effect that first period prices will have on the repeat-sales regression, we omit all observations for which the purchase price occurs in the first period.

Simulation Results

Deviations of the conditional series from the unconditional series are calculated by subtracting the unconditional return from the conditional return each period, for each simulation.

[5] Repeat sale regressions that estimate average of actual returns instead of log returns are also developed by

Table I reports the summary statistics of these deviations for each reservation rule. Mean bias represents the annual average deviation from the unconditional index. Note that this number is positive for each rule, including the repeat-sales regression. The Goetzmann and Spiegel (1995) specification of the repeat-sales regression reduces the bias only slightly. This positive deviation results in an upward bias in price indices that is of the same order of magnitude as the mean annual return of the index itself, about 5%. In general, the magnitude of the bias will depend upon the cross-sectional variation in the asset market. If all assets moved closely with the market with little residual variation, then the conditioning rules would have little effect. Thus, the bias in Table I is a function of the residual variation of 40%. The bias in returns, as well as the standard deviation of the series also appears to be related to the average percentage of the market that transacts each period. Certain rules appear to censor the market more than others. For instance, the "Random" and "Quality" rules reduce the percentage of asset prices observed each period to around 30%. Table I also reports the first order autoregressive coefficient for the conditional estimates. Note that they are negative and significant for the "Random" and "Quality" reserve rules. In other words, average prices will appear to be mean-reverting from period to period. This is not the case, however, for the repeat-sale regression estimate. After adjusting for quality variation, the reversion disappears.

How well do the conditional indices capture the dynamics of the unconditional index? To answer this question, we regress the unconditional return series on the conditional return series. Table II reports the results of one hundred such regressions. The slope coefficients are typically below one, and the t-statistics indicate that most coefficients are significantly different from zero, indicating that the estimates provide some information about the actual market. The exception is the "Max" rule which appears to result in a completely uninformative index. The R-squares reported in the table indicate that less than half of the variation in the unconditional index can be explained by variation in the conditional indices. This suggests that information about the

---

Shiller (1991) and Peng (2002).

actual index behavior is lost as a result of the censoring process. Changes in the conditional indices due to quality variation mask fluctuations in the unconditional indices due to the stochastic process of returns. Note that the repeat-sales regression improves dramatically upon all of the conditional series. Not only is the slope coefficient closer to one, but the R-square is about three times greater than that of the "Random" rule. Clearly, the repeat-sales regression filters out a considerable amount of the noise due to quality variation. Given that the repeat-sales regression controls for quality variation by exactly matching purchase and sales prices of the assets, why doesn't it do better? Part of the answer is that the procedure only uses prices of 30% of the assets in the market each period. Thus, the remaining error in the index is due to small sample variation and truncation.

Little can be done to address the small sample variation since this is dependent upon the reserve rule, however it may be possible to address the truncation. Heckman (1979), for instance, suggests an estimation procedure when the probability of observing the truncated dependent variable can be estimated. Gatzlaff and Haurin (1997) apply the Heckman procedure to estimating a repeat-sales index of housing prices. In the current framework, the results in table I suggest that market volume or percent of properties transacting in a given period might contain some information about the probability of observing an increase in the index. This is the motivation for the model we develop in the next section.

II. Seller Reserves, Market Liquidity, and Transaction Prices

A trading model of illiquid assets

Let us define the market value of the underlying asset to be the expected price that the winning bidder would be willing to pay for the asset. Without loss of generality, we assume the market value is a function of asset attributes, and both the function and the attributes may vary

across time. Denote by $V_{i,t}$ and $X_{i,t}$ the market value and attributes of asset i in time period t, respectively, the assumption is as follows.

$$\log(V_{i,t}) = f_t(X_{i,t}) \quad (1)$$

Denote by $Bid_{i,t}$ the price the buyer is willing to pay for asset i in period t. Denote by $Ask_{i,t}$ the seller's ask price. We assume that $Bid_{i,t}$ equals the market value multiplied by an error.

$$Bid_{i,t} = V_{i,t}u_{i,t} \text{ or } \log(Bid_{i,t}) = f_t(X_{i,t}) + \log(u_{i,t}) \quad (2)$$

In addition, we assume that $Ask_{i,t}$ systematically differs from the market value.

$$Ask_{i,t} = V_{i,t}S_t v_{i,t} \text{ or } \log(Ask_{i,t}) = f_t(X_{i,t}) + \log(S_t) + \log(v_{i,t}). \quad (3)$$

S_t in equation (3) is the ratio of seller's private valuation, $V_{i,t}S_t$, to the market value, $V_{i,t}$. For example, $S_t = 1.1$ implies that the seller's private valuation is 10% higher than the market value of asset i. We call S_t the seller reserve ratio in time period t. Certainly, the seller reserve ratio is influenced by the seller's opportunity cost of selling the underlying asset. For rational sellers, S_t should be greater than 1 because there are hardly any reasons for sellers to charge less than the market value. We assume that S_t is determined by an unknown exogenous data generating process and is uncorrelated with $X_{i,t}$. To simplify the model, we further assume that $u_{i,t}$ and $v_{i,t}$ are log normal with mean 0 and variance δ^2, and independent across both time and assets, from each other, and from $X_{i,t}$. Although we adopt a first-price auction structure, rather than second price it makes no difference to our formulation because we are only modeling the seller reserve, not buyer reserve. Instead, we place an error term on the bid, however the model maps easily into a dual reserve structure, since we work with the spread between the two.

We assume that a transaction takes place if $Bid_{i,t} \geq Ask_{i,t}$, and the transaction price $P_{i,t}$ equals $Bid_{i,t}$. Therefore, we can describe the data generating process of observed transactions using a bivariate process that consists of $Bid_{i,t}$ and a latent variable $Z^*_{i,t}$ defined as follows.

$$\log\left(Bid_{i,t}\right) = f_t\left(X_{i,t}\right) + \log\left(u_{i,t}\right) \tag{4}$$

$$Z^*_{i,t} = \log\left(u_{i,t}\right) - \log\left(S_t\right) - \log\left(v_{i,t}\right) \tag{5}$$

The actually observed variables are $P_{i,t}$ and $Z_{i,t}$. They are related to $Bid_{i,t}$ and $Z^*_{i,t}$ as follows:

$$P_{i,t} = Bid_{i,t} \text{ if } Z^*_{i,t} > 0;\ P_{i,t} = 0 \text{ otherwise;} \tag{6}$$

$$Z_{i,t} = 1 \text{ if } Z^*_{i,t} > 0;\ Z_{i,t} = 0 \text{ otherwise.} \tag{7}$$

Apparently, in this process, $Z_{i,t}$ is a dummy variable that equals 1 if asset i is traded in period t, and 0 otherwise.

Three points are worth noting. First, S_t in equation (3) works as an unobserved threshold that determines the probability of transaction as well as the liquidity in the market, which can be measured with either the time on the market of assets before being sold or the turnover ratio – the ratio of assets sold to all assets on the market – during a fixed period of time. When the threshold is high, the probability of $Z^*_{i,t} > 0$ is low. Hence, it would take more time for an asset to sell, and there are a smaller portion of assets sold in a fixed period. On the other hand, when the threshold is low, the probability of $Z^*_{i,t} > 0$ is high, so assets would be sold quickly and a larger portion of assets may be sold in a fixed-length period. This property of our model is consistent with empirical evidence, such as that asking a higher (lower) price relative to value may result in a slower (faster) sale shown by Miller and Sklarz (1987). Second, the model assumes that S_t is not correlated with $X_{i,t}$. The implication is that, during a given period, all

assets have the same probability of trade regardless their attributes. Therefore, in this model, market liquidity is not correlated with asset attributes. Third, a more general assumption regarding the transaction price is that the transaction price lies between the bid and the ask. For example, we could assume that $P_{i,t} = \left(1-\lambda_{i,t}\right) Ask_{i,t} + \lambda Bid_{i,t}$ with $\lambda \in [0,1]$. The more general assumption would not alter the predictions made by our model regarding the bias of market indices constructed using observed transaction prices, which is discussed in detail later.

The bias in the hedonic regression

A well-known method to estimate indices of illiquid and/or differentiated goods is the hedonic regression. The regression is justified by the theory of equilibrium in markets for differentiated products by Tinbergen (1959) and Rosen (1974) among others. The regression was first applied by Grilliches (1961), and later becomes a widely used method to estimate indices for illiquid or differentiated assets [e.g., Witte, Sumka, and Erikson (1979) among many others]. A general form of hedonic regression is as follows.

$$\log\left(P_{i,t}\right) = f_t\left(X_{i,t}\right) + \varepsilon_{i,t} \tag{8}$$

Assume the functional form in equation (8) is correctly specified, if we run the hedonic regression in equation (8) using transaction data generated by the process described in equation (1) to (7), it is clear that $\varepsilon_{i,t} = \log\left(u_{i,t}\right)$. The expectation of $\varepsilon_{i,t}$ conditional upon observed transactions is

$$E\left(\varepsilon_{i,t} \mid Z_{i,t} = 1\right) = E\left(\log\left(u_{i,t}\right) \mid \log\left(u_{i,t}\right) - \log\left(S_t\right) - \log\left(v_{i,t}\right) > 0\right). \tag{9}$$

In Figure 1, the value of the conditional expectation equals the integrals over the area above the line $\log\left(u_{i,t}\right) = \log\left(v_{i,t}\right) + \log\left(S_t\right)$. If the conditional expectation were 0, the hedonic regression would be the same as the population regression, and least squares estimators would estimate $f_t(\ )$ without bias. However, the private evaluation may be greater than or equal to the

market evaluation, so $\log(S_t)$ is positive and thus the conditional expectation $E(\varepsilon_{i,t} \mid Z_{i,t} = 1)$ is positive, which is easy to verify numerically or using simulations. Consequently, the hedonic regression estimates $f_t(\ )$ with upward bias. Denote by $\hat{f}_t(\ )$ the estimated hedonic pricing function,

$$bias_t(x) = E\left(\hat{f}_t(x)\right) - f_t(x) > 0. \tag{10}$$

S_t in equation (9) affects the conditional expectation and thus the magnitude of the bias. It is easy to verify that both $E(\varepsilon_{i,t} \mid Z_{i,t} = 1)$ and the magnitude of the upward bias increase with S_t.

$$\frac{d\left(bias_t(x)\right)}{dS_t} > 0 \tag{11}$$

If S_t is constant over time, though $\hat{f}_t(\ )$ is still biased, the magnitude of the bias would be constant over time.

$$bias_t(x) = bias_{t-1}(x) \tag{12}$$

Hence, the hedonic regression estimates the expected *change* in the log price, i.e. the expected *return* per period, without bias.

$$E\left[\hat{f}_t(x) - \hat{f}_{t-1}(x)\right] = f_t(x) + bias_t(x) - f_{t-1}(x) - bias_{t-1}(x) = f_t(x) - f_{t-1}(x) \tag{13}$$

However, if S_t increases (decreases) over time, $bias_t$ increases (decreases) over time, and the hedonic regression estimates the expected appreciation rates with an upward (downward) bias. In short, the *change* of S_t determines the bias of the estimated *returns*.

Another implication of our data generating process is that the hedonic regression in equation (8) underestimates the volatility of the actual returns when the market valuation leads the private valuation. For example, if the market valuation goes down before private valuation

does, S_t increases, which has two effects. First, the market liquidity decreases. Second, the market return is estimated with an upward bias using observed transactions. Therefore, instead of observing a negative market return with usual liquidity, we may observe a market with a non-negative return and decreasing liquidity. On the other hand, if the market valuation goes up before private valuation does, S_t decreases, which also has two effects. The market liquidity increases, and the market return is estimated with a downward bias using observed transactions. As a result, we may observe a market with increasing liquidity and a moderately increasing return, instead of a market with usual liquidity and a dramatically increasing return. In short, the estimated return is lower (higher) than the actual one when the market goes up (down). Consequently, estimated indices are less volatile than actual market value indices.

The date generating process described by equations (1) to (7) also suggests that the conditional variance, $\text{var}\left(\varepsilon_{i,t} \mid Z_{i,t} = 1\right)$, changes over time. It is easy to verify that the conditional variance is a function of the seller reserve ratio S_t and the unconditional variance δ^2. Since S_t changes over time, the conditional variance changes over time. The relation between the conditional variance and the values of S_t and δ^2 is important because it facilitates the estimation of S_t and δ^2 as well as the correction of the sample selectivity bias of hedonic and repeat sale regressions.

Note that a more general assumption regarding the transaction price would be

$$P_{i,t} = \left(1 - \lambda_{i,t}\right) Ask_{i,t} + \lambda Bid_{i,t} \text{ with } \lambda \in [0,1]. \quad (14)$$

This assumption does not alter the positive relationship between the change in seller reserve ratio and the upward bias of return estimates, which is based on the fact that the conditional expectation of the error in equation (8) increases with the seller reserve ratio. To illustrate this, note that two extreme cases of equation (14) are $P_{i,t} = Ask_{i,t}$ and $P_{i,t} = Bid_{i,t}$. Since we have

investigated the characteristics of the model under the assumption $P_{i,t} = Bid_{i,t}$, now we study the assumption $P_{i,t} = Ask_{i,t}$. Under the assumption, the conditional expectation of the regression error in equation (8) is

$$E\left(\varepsilon_{i,t} \mid Z_{i,t} = 1\right) = E\left(\log\left(v_{i,t}\right) + \log\left(S_t\right) \mid \log\left(u_{i,t}\right) - \log\left(S_t\right) - \log\left(v_{i,t}\right) > 0\right). \tag{15}$$

It is easy to show numerically or using simulations that the conditional expectation still increases with S_t. Since for any $\lambda \in [0,1]$, the price is a weighted average of the prices in the two extreme cases, as long as the transaction price is between the bid and the ask price, the positive relation between the change of seller reserve ratio and the bias of return estimates is not altered.

The bias in the repeat sales regression

The repeat sale regression described above can be reconciled with a general form of the hedonic regression. To illustrate this, since hedonic variables are typically unavailable in a repeat sale data set, without loss of generality, we assume the unobserved asset attributes change over time as follows,

$$X_{i,t} = X_{i,t-1} + C_{i,t}, \tag{16}$$

where $C_{i,t}$ measures the change of asset attributes. Apply equation (16) and assume $f_t(\ )$ satisfies $f_t(x+y) = f_t(x) + f_t(y)$, the hedonic regression in equation (8) changes to

$$\begin{aligned}
&\log\left(P_{i,t}\right) = f_t\left(X_{i,t}\right) + \varepsilon_{i,t} \\
&= f_t\left(X_{i,0} + \sum_{s=1}^{t} C_{i,s}\right) + \varepsilon_{i,t} \\
&= f_t\left(X_{i,0}\right) + \sum_{s=1}^{t} f_t\left(C_{i,s}\right) + \varepsilon_{i,t} \\
&= E\left(f_t\left(X_{i,0}\right)\right) + f_t\left(X_{i,0}\right) - E\left(f_t\left(X_{i,0}\right)\right) + \sum_{s=1}^{t} f_t\left(C_{i,s}\right) + \varepsilon_{i,t}
\end{aligned} \tag{17}$$

Define $V_t = E\left(f_t\left(X_{i,0}\right)\right)$, which measures the expected value of a representative asset in time period t with constant asset attributes. Define $d_{i,t} = f_t\left(X_{i,0}\right) - E\left(f_t\left(X_{i,0}\right)\right)$, which is the difference between the value of asset i and the expected asset value. Equation (17) changes to

$$\log\left(P_{i,t}\right) = V_t + d_{i,t} + \sum_{s=1}^{t} f_t\left(C_{i,s}\right) + \varepsilon_{i,t}\,. \tag{18}$$

For a pair of transactions, equation (18) leads to

$$\log\left(P_{i,sell}\right) - \log\left(P_{i,buy}\right) = V_{sell} - V_{buy} + d_{i,sell} - d_{i,buy} + \sum_{s=buy+1}^{sell} f_t\left(C_{i,s}\right) + \varepsilon_{i,sell} - \varepsilon_{i,buy}\,. \tag{19}$$

Equation (19) is essentially a very general form of repeat sales regression.

Further simplification leads to a familiar version of the repeat sale regression. Assume $d_{i,t}$ is constant over time, so the time subscript disappears and $d_{i,t} = d_i$. Assume $E\left(f_t\left(C_{i,t}\right)\right) = 0$, equation (19) changes to

$$\log\left(P_{i,sell}\right) - \log\left(P_{i,buy}\right) = V_{sell} - V_{buy} + \sum_{s=buy+1}^{sell} f_t\left(C_{i,s}\right) + \varepsilon_{i,sell} - \varepsilon_{i,buy}\,. \tag{20}$$

The price index of assets with constant attributes, $\{V_t\}_{t=1}^{T}$, can be estimated using equation (20) without knowing the attributes and their changes.[6]

Note that the error in equation (20) includes three components: $\sum_{s=buy+1}^{sell} f_t\left(C_{i,s}\right)$, $\varepsilon_{i,sell}$, and $\varepsilon_{i,buy}$. Assume the variance of $f_t\left(C_{i,t}\right)$ is constant across time and denoted by σ^2, the variance of $\sum_{s=buy+1}^{sell} f_t\left(C_{i,s}\right)$ is $\left(sell - buy\right)\sigma^2$, which increases linearly with the length of time between two transactions. On the other hand, the variance of $\varepsilon_{i,sell} - \varepsilon_{i,buy}$ is constant across time if the regression uses randomly observed asset values so the sample regression would be the same with

[6] The level of the log index at the base level is not identified, and is usually set to be 0.

the population regression. The structure of the error terms in equation (20) provides a new interpretation of the well known fact that, in the case of real estate assets, one component of the error variance grows with the length of the holding period [see e.g. Case and Shiller (1987)]. Equation (20) suggests that the increasing component corresponds to changing attributes and the constant component corresponds to price noise.

Once again, if transaction data are generated from the process described in equation (1) to (7), estimating equation (19) or (20) using observed transactions is problematic. For example, the conditional expectation of the error terms in equation (20) is as follows.

$$\begin{aligned} & E\left(\sum_{s=buy+1}^{sell} f_t\left(C_{i,s}\right) + \varepsilon_{i,sell} - \varepsilon_{i,buy} \mid Z_{i,sell}=1, Z_{i,buy}=1 \right) \\ & = E\left(\varepsilon_{i,sell} \mid Z_{i,sell}=1\right) - E\left(\varepsilon_{i,buy} \mid Z_{i,buy}=1\right) \\ & = E\left(\log\left(u_{i,sell}\right) \mid Z_{i,sell}=1\right) - E\left(\log\left(u_{i,buy}\right) \mid Z_{i,buy}=1\right) \end{aligned} \tag{21}$$

The conditional expectation is generally not zero, so the repeat sale regression is biased. Since the repeat sale regression estimates the appreciation rates instead of price levels, the magnitude of the bias of estimated appreciation rates is determined by the difference between $E\left(\log\left(u_{i,sell}=1\right) \mid Z_{i,sell}=1\right)$ and $E\left(\log\left(u_{i,buy}\right) \mid Z_{i,buy}=1\right)$. Specifically, if S_t of the selling period is higher than that of the buying period,

$$E\left(\log\left(u_{i,sell}\right) \mid Z_{i,sell}=1\right) - E\left(\log\left(u_{i,buy}\right) \mid Z_{i,buy}=1\right) > 0 . \tag{22}$$

Consequently, the repeat sale regression estimates the return with an upward bias. In addition, similar to the hedonic regression, the repeat sale regression underestimates volatility of actual market indices if market valuation leads private valuation in a dynamic setting.

Since the repeat sale regression uses observed transactions, the variance of errors in equation (20) is conditional variance. The conditional variance is

$$\begin{aligned}&\operatorname{var}\left(\sum_{s=buy+1}^{sell} f_t\left(C_{i,s}\right)+\varepsilon_{i,sell}-\varepsilon_{i,buy} \mid Z_{i,sell}=1, Z_{i,buy}=1\right)\\&=\operatorname{var}\left(\sum_{s=buy+1}^{sell} f_t\left(C_{i,s}\right)\right)+\operatorname{var}\left(\varepsilon_{i,sell} \mid Z_{i,sell}=1\right)+\operatorname{var}\left(\varepsilon_{i,buy} \mid Z_{i,buy}=1\right)\end{aligned} . \tag{23}$$

The first component of the conditional variance, $\operatorname{var}\left(\sum_{s=buy+1}^{sell} f_t\left(C_{i,s}\right)\right)$, increases linearly with the number of periods between buying period and selling period. The second and the third components are conditional variance that are determined by the seller reserve ratio S_t and the unconditional variance δ^2.

III. A Remedy

If S_t is a function of exogenous variables and both the function and the exogenous variables are known, the Heckman two-step method estimates the model defined by equation (4) to equation (7) without bias. However, the functional form of S_t is generally unknown and the exogenous variables may be unavailable. On the other hand, maximum likelihood may estimate the model. The log likelihood function of this model is

$$\sum_{t=0}^{T}\left(\sum_{Z_{i,t}=0} \log\left(\Pr\left(Z_{i,t}=0\right)\right)+\sum_{Z_{i,t}=1} \log\left(\Pr\left(Z_{i,t}=1, P_{i,t}=p_{i,t}\right)\right)\right), \tag{24}$$

where $p_{i,t}$ is the sample value of $P_{i,t}$. However, the consistency of the ML estimators depends critically on assumptions regarding error distributions, and maximization of equation (24) is burdensome due to the calculation of conditional probabilities and a large number of parameters including time series of market valuations and seller reserve ratios.

We propose a simple three-step approach to eliminate the bias, which works for both the hedonic regression and the repeat sale regression. The remedy is based on the fact that the seller reserve ratio $\{S_t\}_{t=0}^{T}$ and the unconditional variance of the pricing error δ^2 jointly determine the

market liquidity, the conditional variance, and the conditional mean of error terms in each period. Therefore, we are able to use the time series of market liquidity and residuals of the hedonic or repeat sale regression to estimate $\{S_t\}_{t=0}^{T}$ and δ^2. Then, we could use $\{S_t\}_{t=0}^{T}$ and δ^2 to calculate the conditional expectation of pricing errors in each period and correct for it in the hedonic or repeat sale regression, which generates unbiased estimation of indices. Note that the three-step approach is much easier to implement than the estimation of the likelihood function in equation (24). The likelihood function in equation (24) is a function of not only δ^2 and $\{S_t\}_{t=0}^{T}$ but also other parameters, so it is much more difficult to estimate.

Specifically, the bias of the hedonic regression and the repeat sale regression can be eliminated if the conditional expectation of the error terms, $E\left(\varepsilon_{i,t} \mid Z_{i,t}=1\right)$, is known for each time period t. Once the conditional expectation is known, the hedonic regression can be adjusted as follows to estimate the price index without bias.

$$E\left(\log\left(P_{i,t}\right)\right)-E\left(\varepsilon_{i,t} \mid Z_{i,t}=1\right)=V_t\left(X_{i,t}\right) \tag{25}$$

The repeat sale regression can be adjusted as follows.

$$E\left(\log\left(P_{i,sell}\right)\right)-E\left(\varepsilon_{i,sell} \mid Z_{i,sell}=1\right)-E\left(\log\left(P_{i,buy}\right)\right)+E\left(\varepsilon_{i,buy} \mid Z_{i,buy}=1\right)=V_{sell}-V_{buy} \tag{26}$$

Both (25) and (26) can be easily estimated with OLS or GLS.

We need $\{S_t\}_{t=0}^{T}$ and δ^2 to calculate the conditional expectation $E\left(\varepsilon_{i,t} \mid Z_{i,t}=1\right)$ for any period t. Note that $\{S_t\}_{t=0}^{T}$ and δ^2 determine not only the market liquidity but also the conditional variance of errors, $\mathrm{var}\left(\varepsilon_{i,t} \mid Z_{i,t}=1\right)$, in each period; therefore, the MLEs of $\{S_t\}_{t=0}^{T}$ and δ^2 should maximize two types of likelihood functions simultaneously: the market liquidity as a function of $\{S_t\}_{t=0}^{T}$ and δ^2 as well as the variance of conditional errors as a function of $\{S_t\}_{t=0}^{T}$ and δ^2.

The market liquidity can be measured with the turnover ratio - the ratio of sold assets to all assets on the market. A likelihood function of the turn over ratio in period t is as follows.

$$\prod_{Z_{i,t}=0} \Pr\left(Z_{i,t}=0\right) \prod_{Z_{i,t}=1} \Pr\left(Z_{i,t}=1\right). \tag{27}$$

Denote by O_t the number of sold assets in period t, denote by N_t the total number of assets on the market, equation (27) changes to

$$F(0)^{N_t-O_t}\left(1-F(0)\right)^{O_t}, \tag{28}$$

where F is the cumulative distribution function for $Z_{i,t}^*$, which has a mean of $-\log\left(S_t\right)$ and a variance of $2\delta^2$. We can assume $\log\left(u_{i,t}\right)$ and $\log\left(v_{i,t}\right)$ are normally distributed, then F is the cumulative distribution function of normal distribution. Maximizing equation (28) generates

$$\hat{F}(0)=\left(N_t-O_t\right)/N_t\,. \tag{29}$$

The MLEs of S_t and δ^2 should satisfy equation (29).

To establish the relations between $\{S_t\}_{t=0}^{T}$, δ^2, and residuals of the hedonic or the repeat sale regression, note that the error terms in the hedonic regression for time t satisfy the following equation,

$$E\left(\varepsilon_{i,t}^2\right)=\operatorname{var}\left(\log\left(u_{i,t}\right)\mid Z_{i,t}=1\right) \tag{30}$$

and the error terms in the repeat sale regression for observation i satisfy the following equation.

$$E\left(\varepsilon_{i,t}^2\right)=\operatorname{var}\left(\log\left(u_{i,buy}\right)\mid Z_{i,buy}=1\right)+\operatorname{var}\left(\log\left(u_{i,sell}\right)\mid Z_{i,sell}=1\right)+\left(Sell-Buy\right)\sigma^2 \tag{31}$$

The economic implication of σ^2 in equation (31) is the variance of the value of the change of asset attributes: $\sigma^2=\operatorname{var}\left(f\left(C_{i,t}\right)\right)$. Using regression residuals for actual error terms, equation (30) and (31) are likelihood functions of S_t and δ^2 because the conditional variance is a function of S_t and δ^2.

The maximum likelihood estimators of S_t and δ^2 can be obtained as follows. First, for a given value of the unconditional variance δ^2, we grid search S_t that satisfies equation (29) for each period. Searching for S_t is straightforward because, first, given δ^2, there is only one S_t that satisfies equation (29) for time t; second, F is a monotonically increasing function. After obtaining $\{S_t\}_{t=0}^{T}$, we use δ^2 and $\{S_t\}_{t=0}^{T}$ to calculate the population conditional variance $\text{var}\left(\log\left(u_{i,t}\right) | Z_{i,t}=1\right)$ for each period. After that, we calculate the distance between squared regression residuals and the population conditional variances. Denote by $e_{i,t}$ the hedonic regression residual for asset i at time t. For the hedonic regression, the distance is simply

$$\sum_{\{i,t|Z_{i,t}=1\}} \left(e_{i,t}^2 - \text{var}\left(\log\left(u_{i,t}\right) | Z_{i,t}=1\right)\right)^2. \tag{32}$$

For the repeat sale regression, denote by e_i the repeat sale regression residual for repeat sale observation i. Since $\sigma^2 = \text{var}\left(f\left(C_{i,t}\right)\right)$ is unknown, we treat it as a parameter, and run the following regression and calculate the sum of squared regression residuals, which measures the distance between squared repeat sale regression residuals and population conditional variances.

$$E\left(e_i^2 - \text{var}\left(\log\left(u_{i,buy}\right) | Z_{i,buy}=1\right) - \text{var}\left(\log\left(u_{i,sell}\right) | Z_{i,sell}=1\right)\right) = \left(sell - buy\right)\sigma^2 \tag{33}$$

The MLEs of δ^2 and $\{S_t\}_{t=0}^{T}$ should minimize the distance defined by equation (32) for the hedonic regression and equation (33) for the repeat sale regression.

IV. Simulations of the Model and the Remedy

The simulations of the model have two goals. First, we use the simulations to investigate the bias in the hedonic regression and the repeat sale regression due to the time-varying seller reserve ratio, which now conform to the model developed above. Second, we examine whether the remedy we proposed in the last section effectively mitigates the bias.

As before, we conduct 100 runs of each simulation. Each round consists of the following steps. First, we generate history of a market with 500 properties and 51 periods (from 0 to 50). Each property has an attribute, denoted by X_i, which is randomly generated from a uniform distribution [1,4]. We keep the attribute constant over time so both hedonic and repeat sale regressions are correctly specified and the estimation is straightforward. The attribute can be interpreted as the squared foot if the property is considered a house. The market value, bid price, and ask price of property i in time period t are generated as follows.

$$\begin{aligned} V_{i,t} &= X_i \beta_t \\ Bid_{i,t} &= V_{i,t} u_{i,t} \\ Ask_{i,t} &= V_{i,t} S_t v_{i,t} \end{aligned} \tag{34}$$

In equation (34), β_t is the price index of the market, and can be interpreted as the price per square foot.

We generate β_t, $u_{i,t}$, $v_{i,t}$, and S_t as follows. First, since many illiquid assets demonstrate positive serial correlation of returns and positive long-term average return, β_t is generated as follows.

$$\beta_t = \beta_{t-1} + 0.2(\beta_{t-1} - \beta_{t-2}) + 0.03 + \xi_t \tag{35}$$

ξ_t is a i.i.d. error with 0 mean and standard deviation of 0.05. β_{-2} is set to be 1. β_{-1} is set to equal $\beta_{-2} + 0.03 + \xi_{-1}$. We generate i.i.d. $\log(u_{i,t})$ and $\log(v_{i,t})$ from normal distribution with 0 mean and 0.1 standard deviation. S_t is randomly generated from a uniform distribution [1,1.15]. We assume a transaction takes place if $Bid_{i,t} > Ask_{i,t}$, and the observed transaction price equal $Bid_{i,t}$. When $Bid_{i,t} \le Ask_{i,t}$, no transaction takes place, and both the bid and the ask are not observed.

Second, using the observed transactions, we estimate the price index, $\{\beta_t\}_{t=1}^{50}$, with the hedonic regression and the repeat sales regression respectively. The hedonic regression is as follows.

$$E\left(\log\left(P_{i,t}\right)-\log\left(X_i\right)\right)=\log\left(\beta_t\right) \tag{36}$$

The repeat sale regression is as follows.

$$E\left(\log\left(P_{i,sell}\right)-\log\left(P_{i,buy}\right)\right)=\log\left(\beta_{sell}\right)-\log\left(\beta_{buy}\right) \tag{37}$$

Third, we use the transaction history and residuals of regression (36) and (37), respectively, estimate the MLE of S_t and the unconditional variance of $\log\left(u_{i,t}\right)$. Then, we calculate the conditional expectation $E\left(\log\left(u_{i,t}\right)\mid\log\left(u_{i,t}\right)-\log\left(S_t\right)-\log\left(v_{i,t}\right)>0\right)$ using the estimated S_t and the estimated unconditional variance of $\log\left(u_{i,t}\right)$. After that, we estimate the reserve-conditional hedonic return estimators,

$$E\left(\log\left(P_{i,t}\right)-\log\left(X_i\right)-E\left(\log\left(u_{i,t}\right)\mid Z_{i,t}=1\right)\right)=\log\left(\beta_t\right), \tag{38}$$

and the reserve-conditional repeat sale return estimators.

$$\begin{aligned}&E\left(\log\left(P_{i,sell}\right)-\log\left(P_{i,buy}\right)\right)-E\left(\log\left(u_{i,sell}\right)\mid Z_{i,sell}=1\right)\\&+E\left(\log\left(u_{i,buy}\right)\mid Z_{i,buy}=1\right)=\log\left(\beta_{sell}\right)-\log\left(\beta_{buy}\right)\end{aligned} \tag{39}$$

Table 3 reports our simulation results. Some important relations are also plotted using data from the first run of the simulations to graphically provide intuition. The simulations first verify the negative relation between the level of seller reserve and trading volume, which is shown in Panel A of Table 3 and Figure 2. This is consistent with the simulations in Table 1. An OLS regression of trading volume upon seller reserve generates an average slope coefficient of -2.27 with *t*-statistic of -32.2. The adjusted R^2 of the regression is 0.96 on average.

Second, the simulation verifies the negative relation between the change in the trading volume and the bias of return estimators. We regress the estimation error of the market return,

which is defined as $\hat{\beta}_t / \hat{\beta}_{t-1} - \beta_t / \beta_{t-1}$, upon the change of trading volume, $O_t / N_t - O_{t-1} / N_{t-1}$, for the traditional hedonic and repeat sale return estimates respectively. For the hedonic return estimates, the average slope coefficient of the change of trading volume is -0.25, with average *t*-statistic of -20.64. The average adjusted R^2 is 0.89. For the repeat sale estimates, the average slope coefficient of the change of seller reserve is -0.25, with *t*-statistic of -17.98. The average adjusted R^2 is 0.86. Figure 3 suggests that when trading volume increases, returns are estimated with a downward bias; when trading volume decreases, returns are estimated with an upward bias.

Third, the simulations demonstrate that the remedy we propose effectively mitigates the bias of the hedonic and repeat sale regressions. Regressions of estimation errors of reserve-conditional hedonic and repeat sale regressions upon the change of trading volume do not generate any statistically significant slope coefficients. The average t-statistics are 0.33 and 0.34 respectively. The average adjusted R^2 is 0.03 for both regressions. Figure 4 plots the regression corresponding to the first run of simulations. Regressions of estimation errors upon the change of seller reserve do not generate any significant slope coefficients either, though we do not report the results. Apparently, the remedy proposed in the last section effectively eliminates the sample selection bias due to time-varying seller reserves.

Finally, the simulations show that the reserve-conditional hedonic and repeat sale regressions are much more accurate in terms of having much smaller average Mean Squared Error (MSE). The average MSE of the estimation errors decreases from 0.0014 to 0.0001 for the hedonic regression, from 0.0014 to 0.0002 for the repeat sale regression. The accuracy of estimation is achieved by eliminating the bias due to the time-varying seller reserve.

It is worth noting that the relation between trading volume and seller reserves, the relation between the change of seller reserve and the bias of return estimators, and the relation between the change of trading volume and the bias of return estimators are ultimately governed

by equation (20) and are generally not linear. The relations seem linear in the simulations because we draw S_t from a small interval around 1, and the relation determined by equation (20) is approximately linear when S_t is close to 1.

V. An Application: The Los Angeles Housing Market 1976 to 2001

In this section, we estimate a reserve-conditional index and a conventional repeat sale index for Los Angeles housing market from 1976 to 2001, and illustrate the distinctions between the two indices. This is an important market to perform the analysis, because it is one of the largest housing markets in the United States and represents a non-trivial proportion of mortgages guaranteed by agency bonds. The office of Federal Housing Enterprise Oversight (OFHEO) is the agency charged with overseeing the risks of Fannie Mae and Freddie Mac. OFHEO explicitly uses the repeat sales regression to calculate indices used to evaluate loan to value ratios and expected agency default risks. Given the governmental reliance on standard repeat-sales measures for risk analysis of major agencies that are often referred to as "too big to fail," it is worth examining a major housing market to determine whether it is possible to significantly improve the index measure.

The Los Angeles data include 379,296 repeat sale observations. Each repeat sale observation contains the dates and prices of the two consecutive trades, as well as zip code, address, physical attributes, and financing information of the property. However, data are often missing for variables other than the trading date and price, so hedonic regression is not feasible.

We prepare the variables for the estimation of the reserve-conditional index as follows. First, we estimate the series of the number of transactions, O_t, and the number of all houses on the market, N_t, because the dataset does not provide the information directly. To estimate O_t, we first estimate *unique sales.* Since a transaction may appear

as the first transaction in a repeat sale observation or the second transaction in another repeat sale observation, the unique sales include all first transactions of the repeat sales, and the second transactions of the repeat sales that do not have the same date, zip code, and price with any first transactions. After having the unique sales, we calculate the number of unique sales in each time period, and use them as proxies of O_t. There seems to be a structural break in period 25 (1988), when the number of transactions increases by about 6 times. The break almost certainly reflects a change in the data collection method, a change in the scope of data, or a change in the housing market itself. We do not have enough information to identify the source of this break. Therefore, we divide the whole sample period into pre and post 1989 when we estimate the series of trading volume. We measure the trading volume with O_t/N_t where N_t is the number of houses on the market at t. To estimate N_t, we first calculate the number of repeat sale observations with $buy \le t$ and $sell \ge t$ in each time period t, which is denoted by n_t. Then we assume $N_t = \max\left(n_t\right)_{t=-7,..,8}$. Therefore, the trading volume in a period is estimated with O_t/N_t, which is plotted in Figure 5.

After preparing the variables, we use OLS to estimate the following equation.

$$\log\left(P_{i,sell}\right) - \log\left(P_{i,buy}\right) = V_{sell} - V_{buy} + \psi_i \tag{40}$$

We normalize the market index by letting $V_1 = 0$. Then, we use the regression residuals and the estimated trading volume measured by O_t/N_t to estimate the MLEs of $\{S_t\}_{t=0}^{T}$, the series of seller reserve ratio, and δ^2, the unconditional variance of $\log\left(u_{i,t}\right)$, as described by equations (29) to (33). We then use the estimated δ^2 and $\{S_t\}_{t=0}^{T}$ to calculate the expectation of the error term conditional upon transactions taking place. This calculation is easily done with Monte Carlo

experiments. Using the conditional expectation, we adjust the repeat sale regression as follows and estimate the reserve-conditional repeat sale index.

$$E\left(\log\left(P_{i,sell}\right)\right) - E\left(\varepsilon_{i,sell} \mid Z_{i,sell} = 1\right) - E\left(\log\left(P_{i,buy}\right)\right) + E\left(\varepsilon_{i,buy} \mid Z_{i,buy} = 1\right) = V_{sell} - V_{buy} \quad (41)$$

We also estimate a traditional repeat sale index for comparison.

$$E\left(\log\left(P_{i,sell}\right) - \log\left(P_{i,buy}\right)\right) = V_{sell} - V_{buy}\,. \quad (42)$$

When estimating (41) and (42), we down-weight each observation using the conditional variance corresponding to that observation to improve efficiency. Table 4 reports time series of the estimated trading volume, the estimated seller reserve, the conventional repeat sale index, and the reserve-conditional repeat sale index. Figure 6 plots the indices estimated by (41) and (42) respectively.

Table 5 reports the geometric averages, the standard deviations, and the first order autoregressive coefficients of semiannual returns of the traditional and the reserve-conditional indices, respectively. The reserve-conditional index has a higher average return. The geometric average return is 2.36% for the reserve-conditional index, and 2.28% for the conventional repeat sale index. At the same time, the arithmetic average is 2.45% for the reserve-conditional index, and 2.35% for the traditional index. The reserve-conditional index is also more volatile. The standard deviation of semiannual returns is 4.50% for the reserve-conditional index, while it is 3.82% for the conventional one. Perhaps most relevant to the issue of risk measurement, the reserve-conditional index estimated a steeper drop in housing values in the early 1980's and around 1993 than did the standard repeat-sale measure. Furthermore, the first order autoregressive coefficient is significantly different for the traditional and the reserve-conditional indices. The coefficient is 0.27 for the traditional index, while it is 0.08 for the reserve-conditional index. This is very interesting because it poses the question that to what extent the predictability of real estate index returns is due to the predictability of the latent actual market value index returns. If our reserve-conditional index estimates the actual market value index

without bias, the significant autoregressive coefficient of the traditional index is merely a symptom of the presence of seller reserves in the generation of transactions. In sum, the reserve-conditional index has a higher average return and is more volatility than the traditional repeat sale index. In addition, the reserve-conditional index has a virtually zero autoregressive coefficient, in contrast to the traditional index which has a positive serial correlation.

Our model predicts that the traditional return estimates are upward biased when trading volume decreases, and downward biased when trading volume increases. Therefore, a reserve-conditional index should have higher (lower) returns than a traditional index when trading volume increases (decreases). To verify this, we regress the differences between the reserve-conditional return estimators and the traditional repeat sale return estimators upon the changes of trading volume as follows.

$$E\left(\frac{\hat{V}_t^{RSR}}{\hat{V}_{t-1}^{RSR}}-\frac{\hat{V}_t^{Unbiased}}{\hat{V}_{t-1}^{Unbiased}}\right)=\pi\left(\frac{O_t}{N_t}-\frac{O_{t-1}}{N_{t-1}}\right) \tag{43}$$

An OLS regression generates $\hat{\pi}$ that equals -1.20 with t-statistic -5.39. The adjusted R^2 is 0.37. The regression apparently confirms the systematic difference between the traditional repeat sale index and the reserve-conditional (or actual) price index. However this is not entirely surprising since returns in our model are partial functions of volume.

Readers should take our results with caution for several reasons. First, we lack the true trading volume data, and our estimation of trading volume may be inaccurate; therefore, the seller reserve estimation may be inaccurate. Second, we estimate a semiannual index, while in our model the length of a time period should be the average time the market needs to observe a bid. Hence, the estimated seller reserves may be proportional to instead of precisely equal the actual seller reserve. Despite these problems, the construction of the reserve-conditional index reveals the significant distinctions between the traditional repeat sale indices and the reserve-conditional indices proposed in this paper.

VI. Conclusions

This paper documents the potential bias induced in an index of asset prices when sellers use reservation rules that may include some component of private value. We develop a model in which the seller's asking price is determined by private valuation while the buyer's bid price is determined by the market valuation, and a transaction takes place only if the bid is higher than the ask. In this model, the trading volume and the observed transaction prices are both affected by the ratio of seller's private valuation to the market valuation, which is called the seller reserve ratio. The higher the seller reserve ratio, the lower is the trading volume and the larger is the difference between the market valuation estimated using observed prices and the actual market valuation. We assume the seller reserve ratio is a time-varying exogenous variable. When the ratio increases (decreases), trading volume decreases (increases), and the change of the market valuation estimated using observed prices is upward (downward) biased.

To correct for the bias and estimate the index of actual market valuation, we propose a remedy that first estimates the seller reserve ratio and the unconditional population variance of pricing errors, and then uses them to correct for the bias in the index estimated using observed prices in either a hedonic regression or a repeat sale regression. Simulations show that this remedy effectively mitigates the bias. Moreover, the reserve-conditional indices appear to be more accurate than traditional hedonic and repeat sale indices. We apply this technique to Los Angeles housing market, and the reserve-conditional index indeed substantially differs from a traditional repeat sale index. In our application, the reserve-conditional index has a higher average return, is more volatile, and is essentially not serially correlated.

This paper reveals a simple relation between trading volume and the bias in the traditional hedonic and repeat sale indices estimated using observed transactions. Both indices underestimate returns when trading volume increases, and overestimate returns when trading volume decreases. The estimated indices are less volatile than the actual ones when market

valuations lead private valuations; therefore, the estimated indices underestimate the volatility of the assets. These relations are based on very general assumptions thus are generally applied. Another major contribution is the technique we propose to mitigate the bias. The technique is conceptually simple, computational feasible, and requires much fewer data than alternatives such as the Heckman approach.

Challenges remain about how to generalize our model by allowing more complicated data generating process, as well as how to eliminate the bias under weaker assumptions. For example, our model assumes homoskedastic errors, while heteroskedasticity may be more appropriate for certain data sets. Also, we assume that seller reserve is not correlated with asset attributes, which may not always be a good assumption. We hope to address these challenges in future research.

References:

Anderson, R. C. (1974): "Paintings as Investment," *Economic Inquiry*, 12, 13-26.

Ashenfelter, O. and D. Genovese (1992): "Testing for Price Anomalies in Real-Estate Auctions," *American Economic Review*, 82 (2), 501-505.

Bailey, M.J., R.F. Muth and H.O. Nourse (1963): "A Regression Method for Real Estate Price Index Construction," *Journal of American Statistical Association*, 58, 933-942.

Barber, B. M. and T. Odean (2000): "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors," *Journal of Finance*, 55, 773-806.

Barber, B. M. and T. Odean (2001): "Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment," *Quarterly Journal of Economics*, 116, 261-292.

Barber, B. M. and T. Odean (2002): "Online Investors: Do the Slow Die First?" *Review of Financial Studies*, 15, 455-487.

Case, B., H. Pollakowski, and S. Wachter: (1991): "On Choosing Among House Price Index Methodologies," *AREUEA Journal*, 19 (3), 286-307.

Case, K. and R. Shiller (1987): "Prices of Single Family Homes since 1970: New Indexes for Four Cities," *New England Economic Review*, 46-56.

DeBondt, W. and R. Thaler (1985): "Does the Stock Market Overreact?" *Journal of Finance*, 40, 793-805.

Dhar, R. and N. Zhu (2002): "Up Close and Personal: An Individual Level Analysis of the Disposition Effect," Working paper, Yale International Center for Finance.

Fisher, J., D. Gatzlaff, D. Geltner, and R. Haurin (2003): "Controlling for the Impact of Variable Liquidity in Commercial Real Estate Price Indices," *Real Estate Economics*, 31 (2), forthcoming.

Gatzlaff, D. and R. Haurin (1997): "Sample Selection Bias and Repeat-Sales Index Estimates," *Journal of Real Estate Finance and Economics*, 14, 33-50.

Gatzlaff, D. and R. Haurin (1998): "Sample Selection and Biases in Local House Value Indices," *Journal of Urban Economics*, 43, 199-222.

Geltner, D. and W. Goetzmann (2000): "Two Decades of Commercial Property Returns: A Repeat-Measures Regression-Based Version of the NCREIF Index," *Journal of Real Estate Finance and Economics*, 21 (1), 5-21.

Genesove, D. and C. Mayer (2001): "Loss Aversion and Seller Behavior: Evidence from the Housing Market," *Quarterly Journal of Economics*, 116 (4), 1233-1260.

Goetzmann, W. (1993): "Accounting for Taste: Art and the Financial Markets Over Three Centuries", *American Economic Review*, 83 (5), 1370-1376.

Goetzmann, W. and M. Massa (2003): "Disposition Matters: Volume, Volatility and Price Impact of a Behavioral Bias," Yale International Center for Finance working paper.

Goetzmann, W. and M. Spiegel (1995): "Non-Temporal Components of Residential Real Estate Appreciation," *Review of Economics and Statistics*, 77 (1), 199-206.

Griliches, Z. (1961): "Hedonic Price Indexes for Automobiles: An econometric analysis of Quality Change," U.S. Congress, Joint Economic Committee, *Government Price Statistics*, 173-196.

Grinblatt, M. and B. Han (2002): "The Disposition Effect and Momentum," NBER Working Paper, 8734.

Grinblatt, M. and M. Keloharju (2001): "What Makes Investors Trade?" Working paper, Anderson School, UCLA.

Hausman, J. A. and D. A. Wise (1977): "Social Experimentation, Truncated Distributions, and Efficient Estimation," *Econometrica*, 45, 919-938.

Heckman, J. (1974): "Shadow Wages, Market Prices, and Labor Supply," *Econometrica*, 42, 679-684.

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

Heckman, J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47 (1), 153-161.

Lee, L. F. (1978): "Some Approaches to the Correction of Selectivity Bias," *Review of Economic Studies*, 49, 355-372.

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

Miller, N. and M. Sklarz (1987): "Pricing Strategies and Residential Property Selling Prices," *Journal of Real Estate Research*, 2 (1), 31-40.

Odean, T. (1998) "Are Investors Reluctant to Realize Their Losses?" *Journal of Finance*, 53, 1775-1798.

Odean, T. (1999): "Do Investors Trade Too Much?" *American Economic Review*, 89, 1279-1298

Oehler, A., K. R. Heilmann, L. Volker, and M. Oberlande (2002): "Dying Out Or Dying Hard? Disposition Investors in Stock Markets," Working paper, University of Bamberger.

Peng, L. (2002): "GMM Repeat Sales Price Indices," *Real Estate Economics*, 30 (2), 239-261.

Ranguelova, E (2001): "Disposition Effect and Firm Size: New Evidence on Individual Investor Trading Activity," Working paper, Harvard University.

Rosen, S. (1974): "Hedonic Price and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82, 34-55.

Shefrin, H. and M. Statman (1985): "The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence," *Journal of Finance*, 40 (3), 777-791.

Shiller, R. (1991): "Arithmetic Repeat Sales Price Estimators," *Journal of Housing Economics*, 1 (1), 110-126.

Taylor, W. (1983): "The Estimation of Quality-Adjusted Rates of Returns in Stamp Auctions," *Journal of Finance*, 38 (4), 1095-1110.

Tinbergen, J. (1959): "On the Theory of Income Distribution," *Selected Papers of Jan Tinbergen*, edited by L. H. Klaassen, L. M. Koyck, and H. J. Witeveen. Amsterdam: North-Holland.

Weber, M. and C. Camerer (1998): "The Disposition Effect in Securities Trading: An Experimental Analysis," *Journal of Economic Behavior & Organization*, 33, 167-184.

Weber, M. and H. Zuchel (2001): "How Do Prior Outcomes Affect Risky Choice? Further Evidence on the House-Money Effect and Escalation of Commitment," Working paper, Universitat Mannheim.

Witte, A. D., H. Sumka, and H. Erekson (1979): "An Estimate of a Structural Hedonic Price Model of the Housing Market: An Application of Rosen's Theory of Implicit Markets," *Econometrica*, 47, 1151-1173.

Table 1
Summary Statistics of Deviations from Unconditional Average Series

Summary statistics in this table are calculated over 100 simulations of market histories of forty years, in a market comprised of 100 assets of quality varying continuously from zero to one. Log price relatives for each year are generated according to the model

$$r_{i,t} = r_{m,t} + error_{i,t},$$

where $r_{m,t}$ is distributed normally with mean of 0.05 and standard deviation of 0.1, and $error_{i,t}$ is distributed normally with zero mean and standard deviation of 0.4. The unconditional series of mean for each simulation is the equal-weighted index of returns for all 100 assets. The repeat-sales regression is applied to the transactions observed conditional upon selling only when the price exceeds the owner's purchase price. The conditioning rules are: rule "Start" is $P_{i,t} = p_{i,t} \mid P_{i,t} > P_{i,0}$; rule "Max" is $P_{i,t} = p_{i,t} \mid P_{i,t} > \max\left\{P_{i,0}...P_{i,t-1}\right\}$; rule "Random" is $P_{i,t} = p_{i,t} \mid P_{i,t} > random\left\{P_{i,0}...P_{i,t-1}\right\}$; rule "Quality" is $P_{i,t} = p_{i,t} \mid P_{i,t} > \left(P_{i,0}\right)\max\left\{P_{i,t-1}...P_{N,t-1}\right\}$. The "Mean Bias" is the annual average of the return differences between the indices estimated using observed transactions and the actual indices. The "Std. Deviation" is the average standard deviation of the estimated indices. The "Auto. Coefficient" is the averages of autoregressive coefficients are calculated across statistics from 100 regressions of the form.

$$\bar{r}_{j,t} = \alpha_j + \beta_j \bar{r}_{j,t-1} + \varepsilon_{j,t}$$

	Start	Max	Random	Quality	RSR	GS-RSR
Mean Bias	0.024	0.012	0.048	0.050	0.047	0.042
Std. Deviation	0.106	0.158	0.168	0.346	0.177	0.177
Auto. Coefficient	-0.079	-0.192	-0.342	-0.353	0.042	0.044
(t-statistic)	(-0.505)	(-1.22)	(-2.26)	(-2.27)	(0.268)	(0.279)
#Trades / #Assets	0.71	0.75	0.32	0.27	0.32	0.32

Table 2
Regressions of Unconditional Return Series on Conditional Return Series

Averages are calculated across statistics from 100 regressions of the form:

$$r_{m,t} = \alpha_j + \beta_j \bar{r}_{j,t} + \varepsilon_{j,t}$$

j indexes the 100 simulations. "Mean" represents the average beta coefficient, "t-stat" represents the average t-statistic, "Std" represents the standard deviation of the regression coefficient distribution and "Med R^2" represents the median R^2.

	Start	Max	Random	Quality	RSR	GS-RSR
Mean	0.524	0.302	0.250	0.019	0.577	0.577
t-stat	4.05	3.50	2.59	0.36	5.93	5.93
Std	0.178	0.093	0.121	0.059	0.115	0.114
Median R-squared	0.281	0.235	0.162	0.015	0.483	0.484

Table 3
Model Simulations

This table summarizes and reports the simulation results. The data generation process of the simulation is in text. Regression 1 is a regression of trading volume (measured by the ratio of traded assets to all assets on the market) upon the level of seller reserve ratios. Regression 2 is a regression of deviations of the hedonic return estimators from actual returns upon the changes of trading volume. Regression 3 is a regression of deviations of the repeal sale return estimators from actual returns upon the changes of trading volume. Regression 4 is a regression of deviations of the reserve-conditional hedonic return estimators from actual returns upon the changes of trading volume. Regression 5 is a regression of deviations of the reserve-conditional repeat sale return estimators from actual returns upon the changes of trading volume. Panel A reports the averages of the slope coefficients, t-statistics, and R squared over 100 simulations. Panel B reports the average mean squared error of the return estimators by the hedonic, reserve-conditional hedonic, repeat sale, and reserve-conditional repeat sale regressions over 100 simulations.

Panel A: Regression Results

	Regression 1	Regression 2	Regression 3	Regression 4	Regression 5
Slope	-2.27	-0.25	-0.25	0.00	0.00
t-statistics	-32.20	-20.64	-17.98	0.33	0.34
R squared	0.96	0.89	0.86	0.03	0.03

Panel B: Mean Squared Errors

	Traditional Hedonic	Reserve-conditional Hedonic	Traditional RSR	Reserve-conditional RSR
MSE	0.0014	0.0001	0.0014	0.0002

Table 4

Trading Volume, Reserve Ratio Estimates, and Indices of Los Angeles 1976 – 2001

This table reports semiannual time series of the estimated trading volume measured by the ratio of traded houses to all houses on the market, the estimated seller reserve ratio, the traditional repeat sale index, and the reserve-conditional repeat sale index.

	Volume	Seller Reserve	Repeat Sale Index	Reserve-Conditional Index
1976:1	0.09	1.86	1.00	1.00
1976:2	0.09	1.84	1.02	1.02
1977:1	0.10	1.83	1.16	1.16
1977:2	0.09	1.87	1.25	1.24
1978:1	0.09	1.86	1.37	1.36
1978:2	0.09	1.86	1.42	1.43
1979:1	0.08	1.93	1.49	1.47
1979:2	0.08	1.93	1.57	1.54
1980:1	0.05	2.20	1.51	1.43
1980:2	0.05	2.10	1.66	1.58
1981:1	0.03	2.34	1.54	1.43
1981:2	0.03	2.39	1.54	1.41
1982:1	0.03	2.44	1.61	1.45
1982:2	0.03	2.44	1.80	1.65
1983:1	0.03	2.34	1.88	1.75
1983:2	0.04	2.30	1.92	1.78
1984:1	0.04	2.25	2.01	1.90
1984:2	0.05	2.20	1.99	1.90
1985:1	0.05	2.15	1.98	1.90
1985:2	0.06	2.05	2.08	2.03
1986:1	0.06	2.10	2.17	2.13
1986:2	0.10	1.78	2.26	2.30
1987:1	0.09	1.88	2.29	2.29
1987:2	0.11	1.76	2.42	2.49
1988:1	0.11	1.78	2.55	2.59
1988:2	0.13	1.70	2.68	2.77
1989:1	0.10	1.80	2.78	2.83
1989:2	0.10	1.81	2.86	2.91
1990:1	0.08	1.95	2.89	2.88
1990:2	0.07	1.98	2.87	2.85
1991:1	0.09	1.88	2.83	2.86
1991:2	0.10	1.82	2.84	2.90
1992:1	0.09	1.86	2.79	2.83
1992:2	0.09	1.86	2.76	2.79
1993:1	0.05	2.15	2.71	2.64
1993:2	0.08	1.93	2.67	2.68
1994:1	0.10	1.80	2.65	2.72
1994:2	0.10	1.81	2.64	2.70
1995:1	0.09	1.85	2.62	2.67
1995:2	0.10	1.83	2.61	2.66
1996:1	0.10	1.81	2.60	2.66
1996:2	0.11	1.77	2.61	2.69
1997:1	0.11	1.76	2.63	2.72
1997:2	0.12	1.71	2.67	2.79
1998:1	0.12	1.71	2.72	2.82
1998:2	0.13	1.70	2.78	2.90
1999:1	0.14	1.66	2.81	2.94
1999:2	0.17	1.56	2.85	3.02
2000:1	0.15	1.61	2.92	3.08
2000:2	0.16	1.61	3.00	3.16
2001:1	0.15	1.64	3.08	3.24
2001:2	0.12	1.71	3.15	3.28

Table 5
Summary Statistics of LA Home Value Index

This table reports the geometric average, the standard deviation, and the first order autoregressive coefficient of semiannual returns of the traditional and the reserve-conditional repeat sale indices for Los Angles from 1976 to 2001.

Indices	Geometric Mean	Arithmetic Mean	Std. Deviation	Autoregressive (Stand. Dev.)
Traditional	2.28%	2.35%	3.82%	0.27 (0.14)
Reserve-Conditional	2.36%	2.45%	4.50%	0.08 (0.14)

Figure 1
This figure shows the relation between seller reserve ratio and the error terms in the bid and the ask.

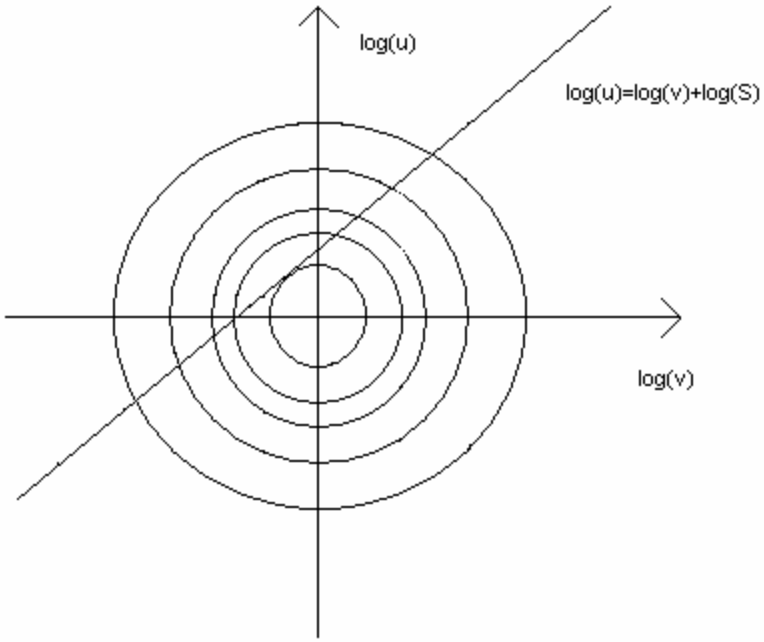

Figure 2
This figure shows the negative relation between the seller reserve ratio and trading volume.

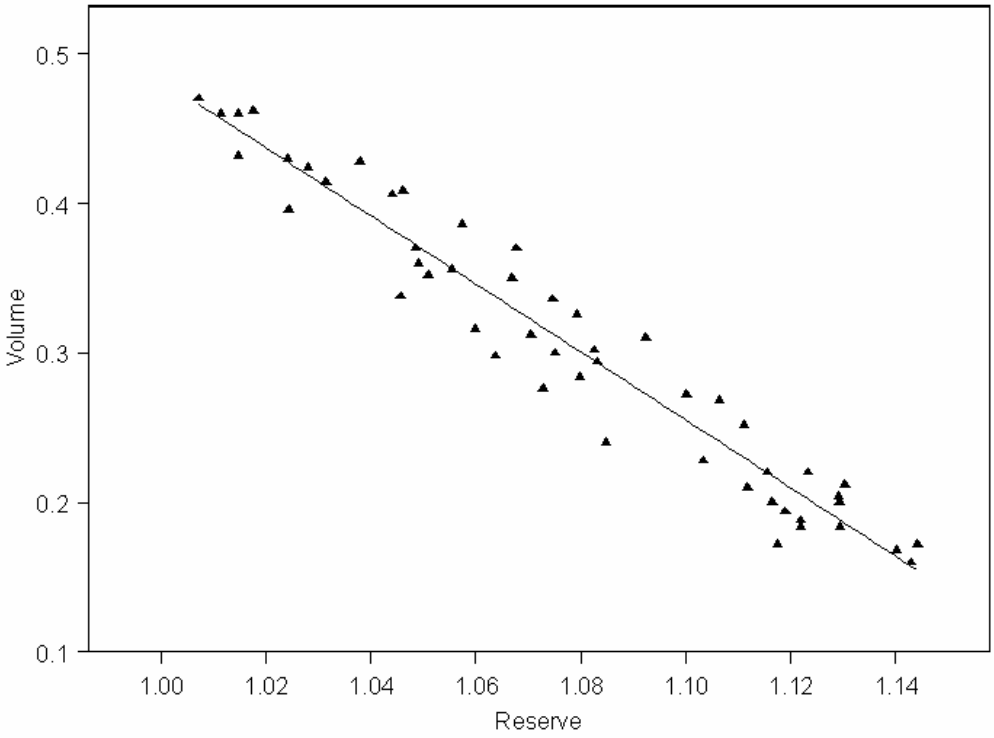

Figure 3
This figure shows that index returns estimated by the hedonic regression and repeat sale regression using observed transactions are downward biased if the trading volume increases, and upward biased if trading volume decreases.

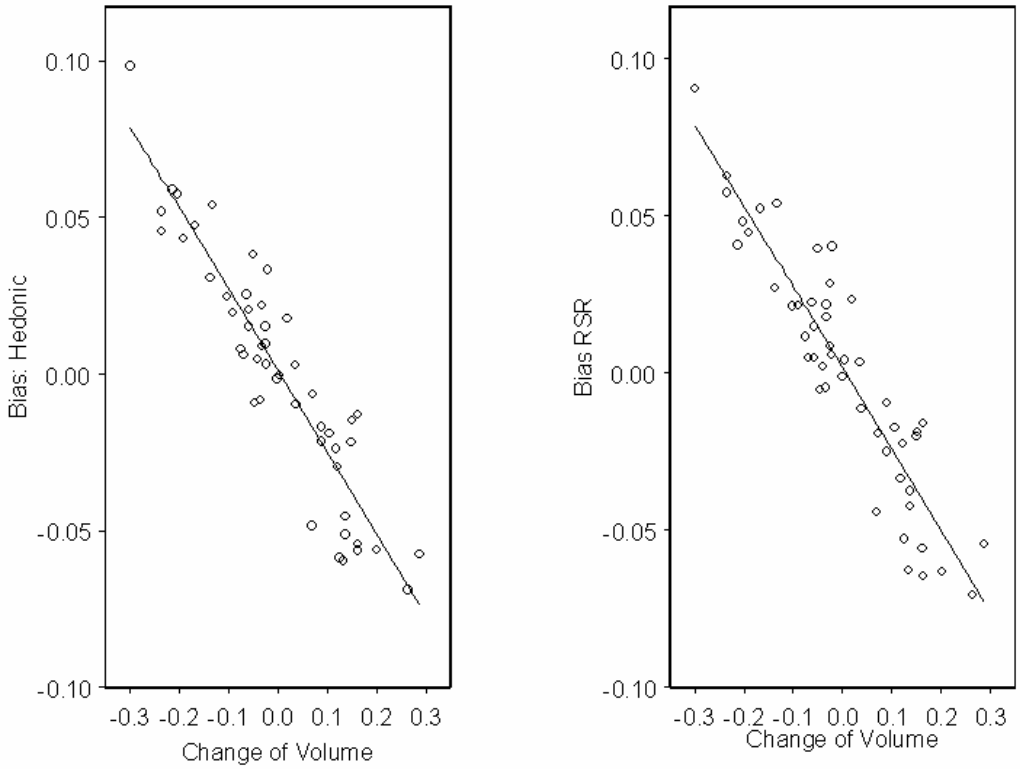

Figure 4
This figure shows that, after controlling for the time-varying seller reserve ratio, there is no statistically significant relation between estimation errors and the changes of the trading volume.

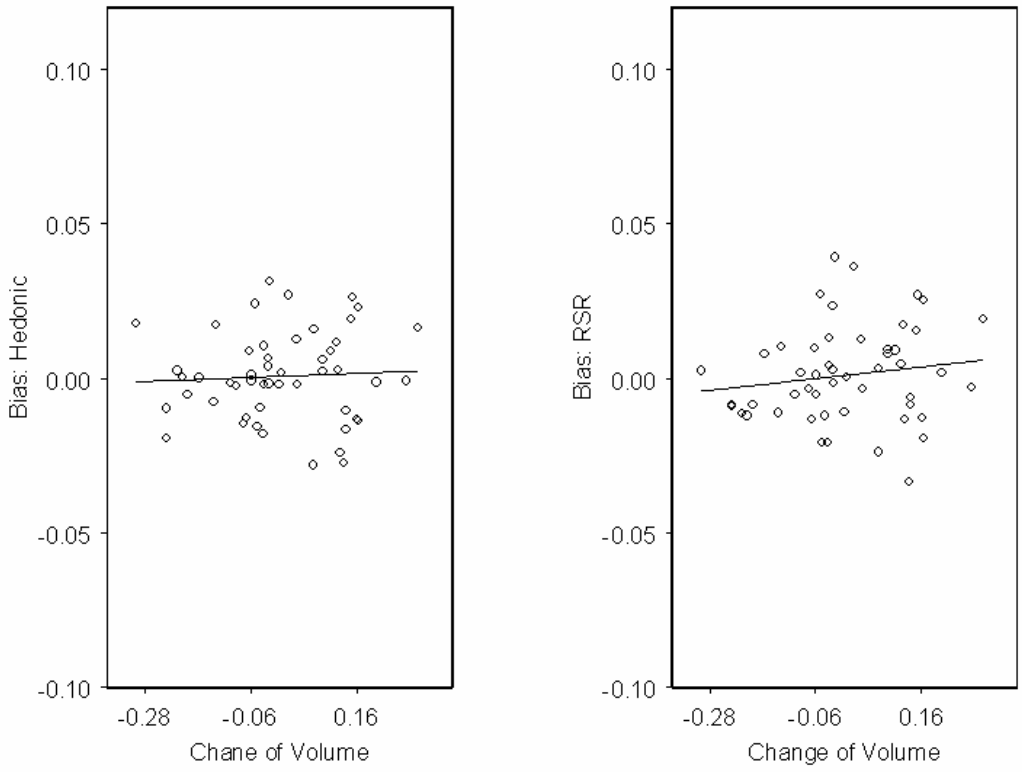

Figure 5
This figure plots the estimated trading volume measured by the turnover rate in Los Angles housing market from 1976:1 to 2001:2.
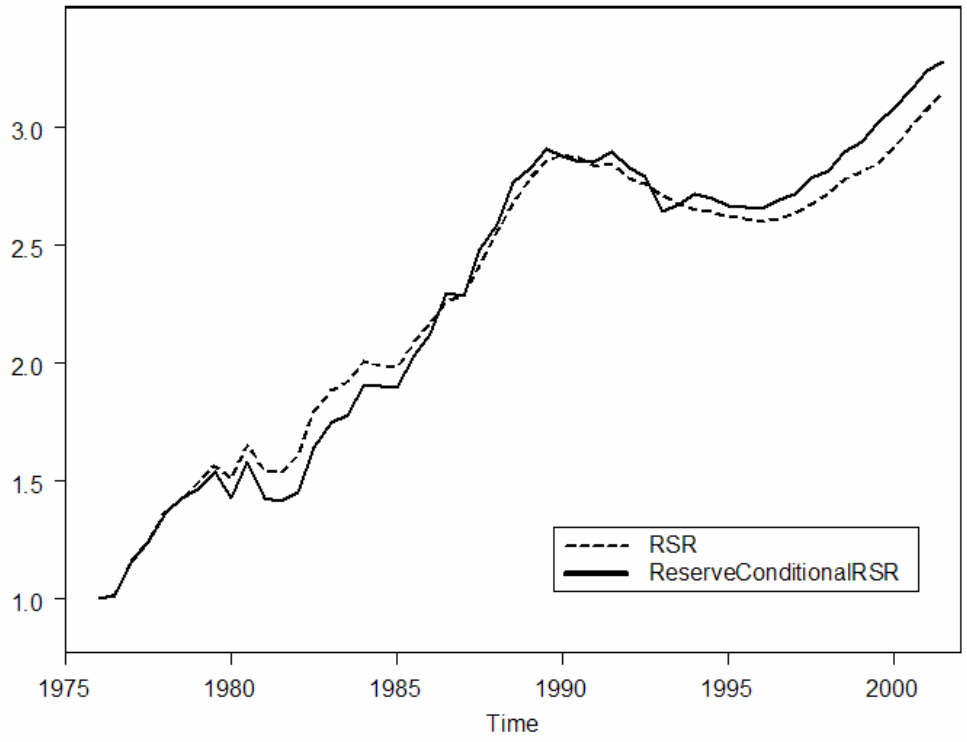
Figure 6
This figure plots the reserve-conditional (unbiased) housing price index and the traditional repeat sale housing index of Los Angles from 1976:1 to 2001:2.