

A Dual Interpretation of the Case-Shiller Index and Its implications to Home Appraisals

Eliezer Z. Prisman
eprisman@yorku.ca
Schulich School of Business
York University
4700 Keele St
Toronto ON M3J 1P3
Canada

ABSTRACT

The S&P/CASE-SHILLER® HOME PRICE index is designed to measure change in the total value of all existing single-family housing stocks. This paper utilizes duality in linear programming to explore a close connection between the index's methodology and the classic no-arbitrage condition (NA) in financial markets. In essence, the interpretation of the NA is the "dual" problem to the primal minimization problem by which a regression is used to estimate the index. The variables of the dual of the NA maximization problem present a term structure of discount factors, the reciprocal of which, are backward looking indices. The insight induces a new methodology for appraising single-family housing and exposes its connection to the commonly used appraisal method. It is well known that the pricing of derivative securities is based on arbitrage arguments. In view of the index being the underlying asset of the recent home price derivatives, the intimate connection between the index's estimation methodology

EMF qualifications: 310, 780

Introduction

The S&P/CASE-SHILLER® HOME PRICE index, (henceforth the index), is designed to measure changes in the total value of all existing single-family housing stocks. This paper utilizes duality in linear programming to explore a close connection between the index's methodology and the classic no-arbitrage condition (NA) in financial markets. In essence, the no-arbitrage interpretation is the "dual" problem to the primal minimization problem by which a regression is used to estimate the index. The variables of the dual of the NA maximization problem present a term structure (TS) of discount factors, the reciprocal of which, are backward looking indices. The insight induces a new methodology for appraising single-family housing. It is well known that the pricing of derivative securities is based on arbitrage arguments. In view of the index being the underlying asset of recent home price derivatives, the intimate connection between the index's estimation methodology and the classical definition of the no-arbitrage condition is a point of interest.

This paper explores the dual motivation of the index and the minimization of squared errors by the regression estimation. It shows that the index is induced by imposing a no-arbitrage condition on realized home price transactions and that the regression error minimization is akin to minimization of deviations from no-arbitrage prices. This interpretation holds even when the method of instrumental variables is utilized for the estimation. The insight induces a new methodology for appraising single-family housing and exposes its connection to the commonly used appraisal method. The new point of view leads to a deeper understanding and intuitive motivation of the estimation. Thereby some new features and methods that were not appreciated prior to the dual interpretation are introduced. This new motivation may help overcome the psychological barrier (discussed in Shiller 2008) to trade in the recent derivatives that are contingent on the index.

The index of home prices has been dealt with in the literature starting with the work of Bailey, Muth, and Nourse (1963) and later modified by Case and Shiller (1987). A survey and a compression of methods can also be found in Case, Pollakowski, and

Wachter (1991) and Case and Szymanoski (1995). The Index's methodology samples repeated sales transactions of relevant homes in different urban geographical areas to generate indices for these counties. From these respective indices, composite indices for cities are established and subsequently a national index is generated. The index has been developed from an econometrical point of view and over the years its statistical properties and qualities have been refined and examined from a statistical point of view. See Abraham and Schauman (1991), Dreiman and Anthony Pennington-Cross (2002), and Gao and Wang (2005).

The index is also the underlying asset of new home price derivatives traded on the Chicago Mercantile. It is well known that the pricing of derivative securities is based on arbitrage arguments. In view of the index being the underlying asset of new home price derivatives, the intimate connection between the index's estimation methodology and the classical definition of the no-arbitrage condition is a point of interest. The estimated index was shown by Shiller (1991) to be the value of a portfolio of existing houses divided by its value at the base period. This conclusion was reached algebraically based on the structure of the matrices in the regression used to estimate the index. However, a direct link between the essence of the index and the no-arbitrage condition, to the best of our knowledge, has not been established.

This direct link between the index and the no-arbitrage condition exposes the index as being based on lagrangian coefficients (shadow prices or discount factors) of an arbitrage maximization problem. Thereby another feature of the index is highlighted. The minimization of the errors in the regression used for its estimation is virtually a minimization of deviations from no-arbitrage prices. Consequently, another facet of the index is being revealed: the justification of the minimization of the deviations is not only to obtain estimation with some statistical properties but has its roots also in classical arbitrage arguments. This property of the index might be useful when comparing different alternatives suggested for designing home price indexes as well as in studies of the efficiency of the real estate market *vis a vis* the magnitude of the errors. Furthermore, the shadow price interpretation induces a new methodology for appraising single-family

housing based on the shadow prices and exposes its connection to the commonly used appraisal method.

In order to present the arbitrage interpretation of the index we will address a single index which measures the changes in the prices of homes in a certain geographical area. Furthermore, we will suppress the features of the index that are introduced from econometrical considerations in order to produce a more accurate, unbiased estimator. Specifically, we will ignore weighting of observations, price tiers, and assumptions regarding the structure of the errors.

The rest of this paper is organized as follows. The next section stipulates the basic data structure from which an index for a county is constructed. It is followed by a section that sets the stage for the arbitrage interpretation imbedded in the index. The next section discusses new appraisal methods and conclusions are offered in the final section.

The index's basic data structure

The index is built by identifying repeated sales of the same property from a certain time, time 0 (or the base time) to time V . For each home's sales transaction in the time interval $[0, V]$, a search is conducted to find information regarding any sales for the same home in the above time period. The set of identified transactions for the same property are then "paired" in the following manner. Starting with the earliest transaction, possibly at time 0, the adjacent, time-wise transaction is paired with the next transaction. If additional transactions of the same property during the time frame are presented they are paired in the same manner. Let Ω be the (ordered) set of all dates on which a sales transaction¹ in a pair of matched transactions is recorded, e.g., $\Omega = \{0, 1, \dots, T\}$. Consequently, a set of pairs of repeated sales of all existing homes in the time interval $[0, V]$, is constructed. The elements of this set are

¹ For reasons explained in the description of the methodology, see S&P 2007, some observations are omitted from the estimation.

$$(P_{i,t_1}, P_{i,t_2}), \text{ for } i = 1, \dots, N \text{ and } t \in \Omega \quad (1)$$

where;

N is the number of identified pairs across all existing homes,

P_{i,t_1} is the price of the sale at time t_1 of pair i

P_{i,t_2} is price of the sale at time t_2 of pair i , $t_2 > t_1$.

Note that P_i and P_j for $i \neq j$ may or may not correspond to the same property.

The data used to estimate the index is therefore summarized by the matrix, A of order N by $T+1$, where $a_{i,j}$ is defined below

$$a_{i,j} = \begin{cases} -P_{i,j} & \text{if } j = t_1 \quad i = 1, \dots, N, \quad j = 0, \dots, T \\ P_{i,j} & \text{if } j = t_2 \quad i = 1, \dots, N, \quad j = 1, \dots, T \\ 0 & \text{Otherwise} \end{cases} \quad (2)$$

Each row of the matrix represents a “pair” as defined above. The matrix A can be interpreted as a payoff matrix of N strategies in a financial market:

- the i^{th} row of the matrix represents the cash flow that results from the strategy of buying and selling the property that corresponds to the i^{th} pair; and
- the first column of A specifies the cash flow at time zero that is needed to execute the i^{th} strategy .

Alternatively, one can think about each row as outlining the cash flow from an asset in this market and about the first column A as the vector of prices of these assets.

Arbitrage and the Home Price index

Consider an artificial market with N assets and a payoff matrix A as defined above. The price of the i^{th} asset is a_{i_0} (which could be zero) and the cash flow from asset i at time j , is a_{ij} for $i = 1, \dots, N$ and $j = 1, \dots, T$. A negative a_{ij} , $j = 1, \dots, T$, is interpreted as cash outflow, buying the property, and vice versa for a positive a_{ij} . An asset with $a_{ij_i} < 0$

represents a property that was purchased at time $j_1 = 0, \dots, T$, which must be followed by a positive a_{ij_2} , for one $j_2, j_2 = (j_1 + 1), \dots, T$. Each asset $i, i = 1, \dots, N$, in this market therefore has a negative cash flow at some time, followed by a positive cash flow at a later time.

Let x be a column vector of order N , defining a strategy of selling and buying properties in this market. A positive x_i denotes purchasing x_i units² of the property that corresponds to row i at time j_b and selling it at time j_s . The indices j_b and j_s correspond to the columns with the first and the second nonzero elements of the matrix A in row i , respectively. By the definition of A , $a_{ij_b} < 0$ and $a_{ij_s} > 0$. A negative x_i simply reverses the order of buying and selling the property in question. That is, it represents selling short x_i units of the property corresponding to row i at time j_b and buying (closing the short position) at time j_s , where j_b and j_s are defined as above.

Keeping in mind the interpretation of the vector x and the definition of the matrix A , $[x'A]_j$ the j^{th} element of the vector $[x'A]$ is the cash (inflow) obtained at time j from following the strategy stipulated by x . Hence, in this imaginary market, if $[x'A]_j \geq 0$ for every j , and $[x'A]_j > 0$ for at least one j , x is an arbitrage strategy. Tracking the buy and sell orders stipulated by x produces cash inflows that are nonnegative at all times, since $[x'A]_j \geq 0$ for every j , and strictly positive at least at one time. Hence, such a strategy is self financing since $[x'A]_0 \geq 0$ does not require the use of out-of-pocket cash and imposes no “future” liability. Furthermore, since $[x'A]_j > 0$ for at least one j , it produces a cash inflow in at least one time.

The no-arbitrage condition in this market rules out the existence of a strategy x that is self-financing, generates positive cash inflow at some future time and imposes no “future” liability. Thus the NA can be defined by requiring that the system of inequalities

² In fact only one such property exists and thus x should have been constrained to satisfy $|x| < 1$. This issue is addressed henceforth.

$x'A \geq 0$ can only be satisfied as $x'A=0$. Of course, this is an imaginary market in which x represents a strategy of purchasing and selling properties at certain times that could not be carried out in reality. This strategy cannot be carried out in real markets as the future prices are not known with certainty. The database on which A is built is of realized transactions, and has hence already occurred.

We will refer to x interchangeably as a portfolio or a strategy, as its role within our interpretation (mathematically) is akin to a portfolio role in the classical definition of a no-arbitrage condition, e.g., in Ross (1976). While x there represents a buy and hold portfolio in a classical one-period model, here it represents a dynamic strategy. A strategy of buying and selling properties, executed during the time interval $[0, T]$ on the discrete times specified in Ω .

It is well known, a variant of Farka's Lemma, Mangasarain (1994), that $x'A \geq 0$ can only be satisfied as $x'A=0$ if and only if there exists a vector $d > 0$ such that

$$Ad = 0. \quad (3)$$

Denote the negative of the first column of A by P , the matrix A without its first column by X , and the vector $\frac{d}{d_0}$, without its first component, by β . Using these notations, the system of equations in (3) can be written as

$$X\beta = P \quad \beta > 0. \quad (4)$$

Hence the NA in this imaginary market is satisfied if and only if the set of equations in (4) is consistent. In the realm of arbitrage arguments for financial markets, β is a vector of discount factors or a valuation operator. The meaning of equation (4) is that in order to avoid arbitrage, the price of each asset at time 0 must be the present value of its future cash flow calculated based on β . This observation yields intuitively the conclusion that

β_i can be written as $\beta_i = \frac{1}{1+r_j}$ $j=1, \dots, T$, where r_j is the rate of return of home prices

from time 0 to time $j, j=1, \dots, T$.

Equation (4) (albeit without the positivity constraint) is the regression equation based on which the index is defined. The index estimation utilizes the least square solution of equation (4) namely,

$$\beta = [X'X]^{-1}X'P. \quad (5)$$

Obviously $X'X$ can be invertible only if $T > N$, and indeed empirically the case is that $N > T$. Hence β is the optimal solution of

$$\begin{aligned} & \text{Min } \varepsilon'\varepsilon \\ & \text{s.t. } X\beta = P + \varepsilon \end{aligned} \quad (6)$$

where ε is a vector of error terms, the deviations of equation (4) from its theoretical satisfactions when the NA holds.

For P to satisfy the NA there must exist a β satisfying (4). If such a β does not exist the index is estimated based on a β that generates a vector of prices that satisfies the NA, $X\beta$, which is as close as possible, in the mean square sense, to observed prices.

It is therefore apparent that an index which is based on β has its root in minimizing the squared deviations from prices that satisfy the NA. The meaning of a β satisfying equation (4) can be more formally exposed. Looking at two pairs of properties one which was purchased at time 0 for a price of P_0 and sold at time t_1 for a price of P_1 , and one which was purchased at time t_2 , at a price of P_{t_2} , and was sold at time t_3 ($t_3 > t_2$) at a price of P_{t_3} . From equation (4) it follows that:

$$P_0 = \beta_{t_1} P_{t_1} \quad (7)$$

and hence

$$1 + r_{t_1} = \frac{1}{\beta_{t_1}} = \frac{P_{t_1}}{P_0} \quad (8)$$

where r_{t_1} is the rate of return on this transaction, which is used to estimate the index from time 0 to time t_1 . Equivalently β_{t_1} can be interpreted as the discount factor by which the present value, P_0 , of P_1 is calculated.

Similarly for the second property equation (4) implies that,

$$0 = -\beta_{t_2} P_{t_2} + \beta_{t_3} P_{t_3}. \quad (9)$$

Keeping in mind the interpretation of β from equation (8), equation (9) expresses the fact that β_{t_2} and β_{t_3} are the discount factors that calculate the present value of P_{t_2} and P_{t_3} at time 0, respectively. It states that the present value of buying the property at time t_2 and selling it at time t_3 is 0. Consequently we have that

$$\frac{\beta_{t_2}}{\beta_{t_3}} = \frac{P_{t_3}}{P_{t_2}} = \frac{1+r_{t_3}}{1+r_{t_2}} = 1+r_{t_2,t_3} \quad (10)$$

where r_{t_2,t_3} , is the rate of return on this transaction which is used to estimate the index from time t_2 to time t_3 .

The connection between the index and the NA is however, even stronger. It will be portrayed utilizing duality theory which makes the meaning of β even more intuitive. A strategy that maximizes the cash inflow at time 0 and presents no future liability can be identified by solving the arbitrage maximization problem below:

$$\begin{aligned} & \text{Max } -x'P \\ & \text{s.t. } x'X \geq 0 \\ & \quad -1 \leq x_i \leq 1 \quad i = 1, \dots, N \end{aligned} \quad (11)$$

Note that the formulation of problem (11) acknowledges the fact that in this market there is only one unit of each property (asset). This is the meaning of the last constraint, $|x| \leq 1$,

were 1 is used loosely to represent also the vector $(1, \dots, 1)'$. If the NA is satisfied, without the $|x| \leq 1$ constraint, the optimal value of (11) must be zero³.

The dual of problem (11) is a minimization, in the spirit of the minimization in (6), which is defined below:

$$\begin{aligned} \text{Min } & \sum_{i=1}^N |\varepsilon_i| \\ \text{s.t. } & X\beta = P + \varepsilon \\ & \beta \geq 0 \end{aligned} \tag{12}$$

The optimal solution of the optimization in (6) generates the estimated β so that the no-arbitrage prices, $X\beta$, are as close as possible, in the mean square sense, to the observed prices. The minimization in problem (12) has the same meaning. Its optimal solution β^* induces NA prices $X\beta^*$ that are as close as possible, in the absolute deviation sense, to the observed prices. Of course if the NA is satisfied then the optimal value of problem (11) is zero, and by duality in linear programming that occurs, if and only if, equation (4) is satisfied. In this case the optimal values of both problems (6) and (12) are zero.

A new interpretation of the estimated β^* , the optimal solution of (12) is thus revealed. By virtue of β^* being the variable in problem (12), which is the dual of problem (11), it is also the vector of shadow prices (lagrangian multipliers) of the optimization problem in (11). As such β_i^* measures the increase in the objective function of (11) due to an infinitesimal relaxation of the i^{th} constraint, or roughly speaking due to the 0 in the right hand side of the i^{th} constraint being replaced by 1. This exposes, again, the meaning of β_i^* as being the present value, at time 0, of a \$1 obtained from selling the i^{th} property. Consequently, the link of the index to the NA by which β^* exists is reinforced.

³ Indeed in the financial literature it is common to define the NA by requiring that the optimal value of (11), without the last constraint, is zero. This definition is referred to as the weak NA relative to the way the NA was defined above.

In many cases, numerically the optimal solution of problem (12), β^* , and of problem (6), β , are not much different. Hence the index that is estimated by the optimal value of (6) is likely not to differ that much from an index that is estimated by the optimal value of (12). We therefore see that the foundation of the estimation⁴ of the index is built upon the assumption that if the properties could be traded at the realized prices, no-arbitrage would be possible.

In practice, the index is estimated using the method of instrumental variables as well as a weighted regression technique. The use of the instrumental variables method in building the index can also be interpreted in terms of arbitrage arguments.

The instrumental variable method is implemented by multiplying, from the left, equation (4) by a matrix Z' of order. The elements of Z are defined as follows: $Z_{ij} = 1$ if X_{ij} is positive, -1 if X_{ij} is negative and zero otherwise. The estimated β , of equation (4), is given by

$$\hat{\beta}_I = [Z'X]^{-1}Z'P. \quad (13)$$

Each column of Z can be interpreted as a portfolio or as a strategy of buying and selling properties in this market at certain times. Consequently, the k^{th} row of the matrix $Z'X$ represents the cash flow obtained from the k^{th} fund or strategy (the k^{th} column of Z). Thus $\hat{\beta}_I$ can be interpreted in the same way as β , but in another imaginary market. In this new market the payoff matrix is $Z'X$ (the counterpart of X) and the price vector is $Z'P$ (the counterpart of P). Hence, the NA in this new market holds if and only if equation (4) is satisfied when $Z'X$ and $Z'P$ are substituted for X and P , respectively. That is, the NA holds if and only if, there exists a vector $\hat{\beta}_I$ satisfying

$$Z'X\hat{\beta}_I = Z'P, \quad \beta_I > 0. \quad (14)$$

⁴ It would have been also possible to analyze the relation of the index utilizing duality in convex programming via the optimization problem in (6) following the guideline in Prisman (1990). However given the closeness of the numerical solutions of problems (6) and (12) and the environment, which is better modeled with the constraint $|x| \leq 1$, we decided to use problem (6)

Consequently, our interpretation and explanation of the rule of β , and its connection to the index in the market defined by X and P , holds for β_I in the market defined by $Z'X$ and $Z'P$.

There is an intimate connection between the market defined by (P, X) and the new market that is defined by $(Z'X, Z'P)$. Since $N > T$ the rank of X cannot exceed T , and must equal it when $Z'X$ is invertible. The rank of $Z'X$ is assumed to be T , as is supported by the empirical data. This means that there must be redundant assets⁵ (strategies) in the original imaginary market defined by (P, X) . That is, the cash flow of some assets (strategies) in this market can be obtained as a linear combination of other assets (strategies). Hence, there exists T assets (portfolios) in the (P, X) market that can be chosen from the N primary assets, so that the cash flows that are spanned by the T assets are the same as those spanned by the primary assets.

It is easy to show that that the NA is satisfied in the (P, X) market, if and only if, it is satisfied in the $(Z'X, Z'P)$ market. In that case the two markets are equivalent in the sense of Ohlson and Garman (1980). That is, $\hat{\beta}_I = \hat{\beta}$ and identical cash flows must have the same price (present values), in both markets. The market $(Z'X, Z'P)$ is complete and there are no redundant assets in it. Thus, given any vector of cash flow (a vector in R^T) there exists a unique strategy or a portfolio (also a vector in R^T) generating it, i.e.,

$$\forall v \in R^T \exists \lambda \in R^T \ni \lambda'(Z'X) = v \quad (15)$$

The results summarized by equations (16) to (19) below, are well known. They are reviewed here since the next section uses them in a non-conventional way, similar to Ioffe (2002), for markets in which the NA is not satisfied. Furthermore, they suggest a new method of home appraisal and also highlight its connection to the present way home appraisal is commonly done.

⁵ The empirical evidence confirm that $Z'X$ is invertible, the positivity constraint is not binding and that $N > T$.

Assuming the NA holds in the $(Z'X, Z'P)$ market, the value of each cash flow $v \in R^T$ in this market can be obtained by applying two dual ways. A primal way that is virtually stipulated by equation (15) but can be better described by an optimization problem in the form of problem(11). Consider the two optimization problems below:

$$\begin{array}{ll} \text{Max } x'Z'P & \text{Min } x'Z'P \\ \text{s.t. } x'Z'X \leq v & \text{s.t. } x'Z'X \geq v \end{array} \quad (16)$$

Since the market is complete and the NA is satisfied, these problems must have the same optimal value, which equals that of each of the two problems below:

$$\begin{array}{ll} \text{Max } x'Z'P & \text{Min } x'Z'P \\ \text{s.t. } x'Z'X=v & \text{s.t. } x'Z'X=v \end{array} \quad (17)$$

It is also clear that the optimal solutions of these four problems are the same as well as their optimal solutions which equal the λ specified in equation (15). Let us denote the optimal value of the problems in (17) by⁶ $F(v)$. Thus

$$F(v) = \text{Max}_x \{x'Z'P \mid x'Z'X=v\} = \text{Min}_x \{x'Z'P \mid x'Z'X=v\}. \quad (18)$$

Applying the duality of linear programming we have another way to calculate the present value of v as

$$F(v) = \text{Max}_x \{v' \beta \mid Z'X\beta = Z'P\} = \text{Min}_x \{v' \beta \mid Z'X\beta = v\} = \beta'_I v. \quad (19)$$

Home Appraisals

The appreciation of the intuition behind the suggested appraisal methods, hinges on the index's practical estimation using arbitrage arguments. If $(Z'X, Z'P)$ would have been a real financial market in which the NA holds, we would have the following: Consider, a property that was last sold at time $t_1 \leq T$ for a price of P_1 , and should be appraised at time

⁶ The function F is termed the perturbation function and it is convex (concave) for a convex (concave) maximization (minimization) problem, see Rockefeller (1970). Generally in linear programming it is piecewise linear convex or concave, but in our case it is linear and equals $\beta'_I v$.

$t_2, T \geq t_2 \geq t_1$, where $t_1, t_2 \in \Omega$. Since the NA holds, $\hat{\beta}_I$ would have satisfied equation (14), and the value of this property at time 0 would be given by

$$P_1 \hat{\beta}_I(t_1). \quad (20)$$

where $\hat{\beta}_I(t_1)$ is the component of β_I that corresponds to time t_1 .

Based on equation (19) this is also the price of a strategy of buying and selling properties, including short selling, that generates the cash flow of P_1 at time t_1 and zero at all other times. The property value at time t_2 can be calculated based on its value at time 0 and multiplying it by $\frac{1}{\beta_I(t_2)}$. Suppressing the sub index of β_I the appraised value of

the property at time t_2 is given by

$$P_1 \frac{\hat{\beta}(t_1)}{\hat{\beta}(t_2)} = P_1 \frac{1+r_{t_2}}{1+r_{t_1}} = P_1(1+r_{n/2}). \quad (21)$$

The appraisal method suggested by equation (21) therefore is equivalent to finding the price of a strategy, including short positions, that replicates the cash flow P_1 at time t_1 . This strategy includes properties that sold in the market over a period of time and takes into account the index appreciations over this period. Empirically, the data base from which the imaginary market is constructed does not satisfy the NA. Thus the optimization problem in (16) is unbounded and the optimization problem in (17) is infeasible as equation (14) is not satisfied. Thus in order to execute this appraisal method $\hat{\beta}_I$ is estimated using a second best solution of equation (14). Both the index and $\hat{\beta}_I$ are built on a second best solution in the sense of the mean square sense, e.g. regression, of equation (14).

The common method of properties appraisals is based on prices of transactions in neighboring properties. The pricing of specific features of the property at hand is then done on an ad hoc basis by adding or subtracting certain (subjective) amounts to account for the presence (or absence) of these features. Furthermore, in most cases neighboring

properties that are used for the estimation in the common method, are not sold or bought at the same time. Thus an additional bias is introduced to the common estimation, since it is implicitly assumed that no appreciation or depreciation occurred during the period that the transactions in the neighborhood properties took place.

The common appraisal is thus in the spirit of trying to replicate the property at hand with similar properties that have been sold during the recent time. However, these replications are built only on long positions. Hence, a property that has been sold with an overvalued price is used for the replications with this price. Furthermore, only properties that are sold recently are used in the valuation and the time dimension is overlooked. Hence, the appreciation or the depreciation of the index is not accounted for. In contrast the appraisal methodology suggested by equation (21) does not suffer from this caveat. Moreover, it is built on a much richer environment as the replication of the property at time t_1 is built on properties that were bought and sold during the time interval $[0, T]$. Furthermore, this method allows short positions and it is not ignoring the index variations.

There is, however, a distinct difference between the imaginary (real estate) market $(Z'X, Z'P)$ and a financial market. In a financial market the assumption is that the NA holds but that observed prices may violate it due to noises or non synchronization of trades. Thus a second best solution is used to solve for β in equation (14). Such is the case when the vector of discount factors (the term structure) is estimated from prices of bonds and their cash flows. The estimated factors are then justifiably used to calculate preset values of future cash flows.

In the imaginary (real estate) $(Z'X, Z'P)$ market there is no theoretical justification for equation (14) or for the NA to hold over realized prices. Different properties, as opposed to cash flow, are not a perfect substitute for one another even when the index depreciation/appreciation (the time value of money) is taken into account. It is conceivable therefore that those properties' prices are different because of the different features of the properties. This is the same phenomenon that occurs in incomplete financial markets. In such markets, pricing by replication creates only bonds on the valued property/assets since not all the attributes of the property/assets can be

replicated. Some attributes cannot be matched but can only be compared to property/asset with attributes that dominates the one at hand. Hence, only upper and lower bounds on the value of the property at hand can be placed. These bounds are not due to noises in the recorded prices but due to the incompetence of the market. Consequently the violation of equation (14) cannot and is not, attributed to noises only.

The estimation of the index based on the CASE-SHILLER methodology, as shown here, amounts to identifying an index (it's reciprocal actually) that satisfies the NA in the second best solution. This is, of course, very reasonable and an elegant idea for an index estimation. However, there is a way to incorporate and reflect the distinctions between the imaginary market and a financial market in appraising properties. As we shall see, incorporating the deviations from NA prices, not regarding them just as noises, produces an interval estimation of the price of a property. Numerically the cost of producing this interval is however, greater than the point estimation.

Since the NA does not hold, the optimization problem in (11) without the constraint $|x| \leq 1$ is unbounded. Consequently, the problems in (17) are not feasible and thus the valuation in (18) and (19) is not applicable. When the realistic constraint $|x| \leq 1$ is incorporated, the optimization problem in (17) is not feasible for every v . There might be cash flow v for which there exist no $|x| \leq 1$ such that $x'Z'X=v$. Such a v could not be replicated in this market and hence cannot be valued uniquely (by a point estimation). However the problems in (16) are feasible for every v and will possess optimal values. Their optimal values, as outlined in (22) below, convey an upper bound, the optimal value of the minimization problem, and a lower bound, the optimal value of the maximization problem, on the case flow v at time 0.

$$\begin{array}{ll}
 \text{Max } x'Z'P & \text{Min } x'Z'P \\
 \text{s.t. } x'Z'X \leq v & \text{s.t. } x'Z'X \geq v \\
 |x| \leq 1 & |x| \leq 1
 \end{array} \tag{22}$$

A dual way of specifying the bounds is possible. Calling again on the duality in linear programming the dual of the maximization and minimization problem are:

$$\begin{array}{ll}
 \text{Min } v'\beta + \sum_{i=1}^N |\varepsilon_i| & \text{Max } v'\beta - \sum_{i=1}^N |\varepsilon_i| \\
 \text{s.t. } Z'X\beta + \varepsilon = P & \text{s.t. } Z'X\beta + \varepsilon = P \\
 \beta \geq 0 & \beta \geq 0
 \end{array} \quad (23)$$

The optimal value of the Max in (22) equals that of the Min in (23) and the optimal value of the Min in (22) equals that of the Max in (23). Thus, the lower bound is given by the optimal value of the minimization problem in (23) and the upper bound by that of the maximization problem in (23). When the only element of v that is different from zero is P_1 at time t_1 , an upper and lower bound on the value of P_1 at time zero will be obtained by solving (23). Namely the lower bound is $P_1\beta^{Min}(t_1) + \sum_{i=1}^N |\varepsilon_i^{Min}|$ and the upper bound is $P_1\beta^{Max}(t_1) - \sum_{i=1}^N |\varepsilon_i^{Max}|$, where, $\beta^{Min}, \varepsilon^{Min}$ and $\beta^{Max}, \varepsilon^{Max}$ are the optimal solutions of the minimization and maximization problems in (23), respectively. The bounds on the value of the property at time t_2 , P_2 can be thus estimated by

$$\frac{P_0\beta^{Min}(t_1) + \sum_{i=1}^N |\varepsilon_i^{min}|}{\beta^{Min}(t_2)} \leq P_2 \leq \frac{P_0\beta^{Max}(t_1) - \sum_{i=1}^N |\varepsilon_i^{max}|}{\beta^{Max}(t_2)}. \quad (24)$$

When β is estimated by β^* , the optimal solutions of problem (23) are β^*, ε^* , and the present value of P_0 is estimated as $P_0\beta^*$. It is possible to show (see the proof and the discussion in the Appendix) that

$$P_1\beta^{Min}(t_1) \leq P_1\beta^*(t_1) \leq P_1\beta^{Max}(t_1) \quad (25)$$

and if, the NA is assumed to hold in both markets, (X, P) $(Z'X, Z'P)$, that

$$P_1\beta^{Min}(t_1) + \sum_{i=1}^N |\epsilon^{\min}_i| \leq P_1\beta^*(t_1) \leq P_1\beta^{Max}(t_1) - \sum_{i=1}^N |\epsilon^{\max}_i| \quad (26)$$

As pointed out above, producing the upper and lower bounds in (24) is costlier than the appraisal advocated by (21). The latter requires solving once an optimization problem which produces the estimated β that can be used to value any property. In contrast the bounds displayed in (24) requires solving two optimization problems for each property to be appraised.

In fact if the cost of solving two optimization problems for each appraisal is accepted, the bounds in (24) should be modified slightly. Rather than solving for the value of the property at time t_0 and using $\beta^{Min}(t_1)$ and $\beta^{Max}(t_1)$ to generate the appraisal value at t_1 , the value at t_1 can produced directly and more accurately. If in the optimization problems in (22) and (23) :

- a) P is replaced with $-X_{.t_2}$ where $X_{.t_2}$ is the column of X that corresponds to time t_2 , and
- b) $X_{.t_2}$ is omitted from the matrix X and -P is add to it as a first column to generate a new matrix denoted by \bar{X} ,

then optimal values of in (22) and (23) will produce the requested bounds.

Of course the modified optimization in (23), induces also a modified interpretation for its optimal solution β : $\beta(t)$ is the value, as of time t_2 , of a dollar received from a property at time t_0 . That is $\beta(t)$ can be a future ($t > t_2$) or preset ($t < t_2$) value coefficient.

The assumption of the discrete times, i.e., $t_1, t_2 \in \Omega$ can be relaxed. A continuous estimation of $\beta(t)$ can be derived given the value of $\beta(t)$ for $t \in \Omega$ using a polynomial approximation. Therefore a continuous function, $d(t)$ which specifies the discount factor from time 0 to time t is generated. Hence, the index from time t_1 to t_2 ($t_1 < t_2$) is given

by $\frac{d(t_1)}{d(t_2)}$ and the continuous approximation of the index thus facilitates the generation,

at time T , of the index I_{t_1, t_2} for every t_1 and t_2 such that $t_1 < t_2$.

While being suppressed here, the index I_{t_1,t_2} is in fact dependent on T also, and without abuse of notation, should have been noted as ${}_T I_{t_1,t_2}$. The set of observations (pair-wise sales) at time T_1 may not necessarily be a subset of the set of observations at time T_2 ($T_1 < T_2$). This can happen if a certain property was sold once at time $t_1 < T_1$ and the second time at time t_2 such that $T_1 < t_2 < T_2$. Hence the matrix X associated with time T_1 is not necessarily a sub matrix of the matrix X associated with time T_2 . Consequently, the vector β associated with time T_1 and T_2 may have different values even for a time (and a pair) that appear in both data bases. Of course this could be the case even if the matrix X associated with time T_1 is a sub matrix of the matrix X associated with time T_2 . Consequently, as time progresses the number of observations increases and we may update our estimate of the index I_{t_1,t_2} as of time T_1 , when it is being estimated at time T_2 ($T_1 < T_2$).

Being presented at time T with a property that is needed to be appraised, and was sold last at time $t < T$ for a price P_t , the appraisal value is calculated as ${}_T I_{t,T} P_t$. There may be cases that an estimate of a property's value at some past time is done due to some legal dispute, at two distinct times, say at time T_1 and T_2 where ($T_1 < T_2$). The property's value at time t , ($t < T_1 < T_2$), might be different when it is estimated at time T_1 compared to at time T_2 .

Conclusions

The Case-Shiller index is constructed by running a regression identifying discount factors, which are only time dependant. These discount factors equate, at the base period, the present value of a buying price to the present value of an adjacent selling price of the same property. Since the discount factors are only functions of time, the same factor is applied to different transactions of different properties that occurred at the same time.

Consequently, it is not likely that the present value of the selling and buying prices of all the properties will be equated.

A regression is thus used to estimate the discount factors that make the present value of adjacent transactions as close as possible to zero. This paper explored the dual motivation of the index and the minimization of squared errors by the regression. It showed that the index is induced by imposing a no-arbitrage condition on the prices of realized transactions and the error minimization is akin to minimization of the deviations from no-arbitrage prices. Using this interpretation, two methods of property appraisals are suggested: A method which requires resolving an estimation-like optimization problem for each property and better accounts for the different features of the property at hand and a method which is based on the calculated index and easier to implement.

It is shown that the interpretations provided here are invariant even with the use of the instrumental variables technique. This interpretation is then used to motivate appraisal methods for real estate properties. The method reflects the actual features of the appraised property since it is built on the change in the index, and an historical price of the same property rather than on prices of neighboring properties in which transactions took place in recent periods. The new interpretations of the index may shed a new light on the index and make investors more comfortable with its use. Consequently, it may induce investors to be more active in the new derivative market for which the index is the underlying asset and thus provide, perhaps, a partial answer to some of the questions raised in Shiller (2008).

Note that during each period the index estimates reports only the marginal change in the index over the last period. It does not re-estimate the value of the index for each period since the base period. With each additional period an additional variable, the last element of the new β vector, and some equalities are added to the system of equations in (13). When the change of the index is estimated over the last period, all the values of the components in β but the last one are left as they were and only the last component is solved. The value of the vector β , so obtained, is of course not necessarily the value that would be obtained by resolving the system of equations in (13). However re-solving the system of equations in (13) is appropriate when the index is used for property

evaluations as it takes into account the new information and not only the trend in the index. A similar problem with the index that is produced by OFHEO in the USA is addressed in the paper by Deng, Yongheng and Quigley (2008).

Finally, having a term structure of the index can serve as the base for a fix for float swap agreement in which the floating rate is based on the index. Note that there is a difference between a standard fix for float interest rate swap agreement and a swap in which the floating rate is based on the index. In a standard interest rate swap agreement the realization of the interest rate spanning a certain period is known at the beginning of each period while with this new swap the realized floating rate is known at the end of each period.

References

- Abraham, J.M. and Schauman, W.S., "New Evidence on Home Prices from Freddie Mac Repeat Sales," *AREUEA Journal*. 19(3), 1991, 333-352.
- Case, Karl E., and Robert J. Shiller, "The Efficiency of the Market for Single Family Homes," *American Economic Review* 79, 1989, 125- 137.
- Case, B., and Szymanoski, E., 1995, "Precision in House Price Indices: Findings of a Comparative Study of House Price Index Methods", *Journal of Housing Research*, 6(3), 1995, 483-496.
- Case, B., Pollakowski, H. O., and Wachter, S. M., "On Choosing Among House Price Index Methodologies", *AREUEA Journal*, 19(3), 1991, 286-307.
- Case, K., Shiller, R., 1987, "Prices of Single-Family Homes Since: New Indexes for Four Cities", *New England Economics Review*, September-October, 1987, 45-56.
- Case, B., and Szymanoski, E., "Precision in House Price Indices: Findings of a Comparative Study of House Price Index Methods", *Journal of Housing Research*, 6(3), 1995, 483-496.
- Dreiman, Michelle H and Anthony Pennington-Cross, "Alternative Methods of Increasing the Precision of Weighted Repeat Sales House Prices Indices" OFHEO WORKING PAPERS Working Paper 2002.
- Deng, Yongheng and Quigley, John M., "Index Revision, House Price Risk, and the Market for House Price Derivatives" *Journal of Real Estate Finance and Economics*, Vol. 37, No. 3, 2008.
- Gao, Andre H and Wang George H. K ., "Multi Transactions Model for Constructing Housing Market Index", Andre H. Gao ,George H. K. Wang, paper presented at the American Real Estate & Urban Economics Association Mid-Year Meeting 2005, Washington, DC.
- Ioffe, Ioulia, "Arbitrage Bounds (in Markets with Noisy Prices) and the Puzzle of Negative Option Prices Implicit in Bonds", *Journal of Banking & Finance*, 26 (2002) 1199-1228.
- Ohlson, James A. and Garman Mark B., " A Dynamic Equilibrium for the Ross Arbitrage Model" *The Journal of Finance*, Vol. 35, No. 3 (Jun., 1980), pp. 675-684
- Goetzmann, William, "The Accuracy of Real Estate Indexes: Repeat Sales Estimators," *Journal of Real Estate Finance and Economics*, 1992, 5, 5-53.

Mangasarian, Olvi L. "Nonlinear Programming" Academic Press, 1971

Rockafellar, R.T. "Convex analysis", Princeton University Press , 1970

Ross, S., "Return, Risk, and Arbitrage," in *Risk and Return in Finance*. Friend and Bicksler (eds.), Cambridge, Ballinger, 1976.

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

Shiller, Robert J., "Derivatives Markets for Home Prices" (March 2008). Cowles Foundation Discussion Paper No. 1648 Available at SSRN:
<http://ssrn.com/abstract=1114102>

S&P/CASE-SHILLER® HOME PRICE INDICES: INDEX METHODOLOGY, S&P
November 2007

Appendix

Claim: For a time $t_0 > 0$ in Ω , we have that $P_0\beta^{Min}(t_0) \leq P_0\beta^*(t_0) \leq P_0\beta^{Max}(t_0)$

It is easy to verify that the following relations hold $v'\beta^{Min} + \sum_{i=1}^N |\varepsilon_i^{Min}| < v'\beta^* + \sum_{i=1}^N |\varepsilon_i^*|$ and $v'\beta^* - \sum_{i=1}^N |\varepsilon_i^*| < v'\beta^{Max} - \sum_{i=1}^N |\varepsilon_i^{Max}|$.

But also by the minimization in (23), $\sum_{i=1}^N |\varepsilon_i^*| \leq \sum_{i=1}^N |\varepsilon_i^{Min}|$ and $\sum_{i=1}^N |\varepsilon_i^*| < \sum_{i=1}^N |\varepsilon_i^{Max}|$.

Hence, $v'\beta^{Min} + \sum_{i=1}^N |\varepsilon_i^{Min}| - \sum_{i=1}^N |\varepsilon_i^*| \leq v'\beta^* \rightarrow v'\beta^{Min} \leq v'\beta^*$,

and

$$v'\beta^* - \sum_{i=1}^N |\varepsilon_i^*| + \sum_{i=1}^N |\varepsilon_i^{Max}| < v'\beta^{Max} \rightarrow v'\beta^* < v'\beta^{Max}$$

thus the claim follows.

Claim: if the NA (without the $|x| \leq 1$) holds in the $(Z'X, Z'P)$ market the following inequalities hold

$$P_0\beta^{Min}(t_0) + \sum_{i=1}^N |\varepsilon_i^{\min}| \leq P_0\beta^*(t_0) \leq P_0\beta^{Max}(t_0) - \sum_{i=1}^N |\varepsilon_i^{\max}|.$$

It is obvious that if the NA holds

$$P_0\beta^{Min}(t_0) + \sum_{i=1}^N |\varepsilon_i^{\min}| = \text{Max } x'Z'P \quad \leq \quad \text{Max } x'Z'P$$

$$\text{s.t. } x'Z'X \leq v \quad \text{s.t. } x'Z'X \geq v$$

$$|x| \leq 1$$

and that

$$P_0 \beta^{Max}(t_0) - \sum_{i=1}^N |\varepsilon_i^{max}| = \begin{array}{l} \text{Min } x'Z'P \\ \text{s.t. } x'Z'X \geq v \\ |x| \leq 1 \end{array} \geq \begin{array}{l} \text{Min } x'Z'P \\ \text{s.t. } x'Z'X \geq v \end{array}$$

but since

$$\begin{array}{l} \text{Max } x'Z'P \\ \text{s.t. } x'Z'X \leq v \end{array} = \begin{array}{l} \text{Min } x'Z'P \\ \text{s.t. } x'Z'X \geq v \end{array} = v' \beta^*$$

where $\beta^* = \beta^*$, and both satisfy $Z'X\beta = Z'P$ the result follow.

Note that the minimization (maximization) problem in (22) may not be feasible. In this case the upper bound (lower bound) will be ∞ ($-\infty$). This could be the case since under the constraint $|x| \leq 1$ only cash flows such that $|[Z'X]^{-1}v| \leq 1$ can be produced.