A method for combining transaction- and valuation-based data in a property price index

Olof Netzell

Building and Real Estate Economics

Royal Institute of Technology

Stockholm 2010

A method for combining transaction- and valuation-based data in a property price index

Abstract: This paper presents a method for combining transaction- and valuation-based data in a property price index. The methodology is devised for a world where observable transaction prices can be used to construct a price index that constitutes a noisy, unbiased signal of the ”true” price index. It is furthermore assumed that valuations can be used to construct a market value index which does not contain noise but that suffers from so called appraisal ”smoothing”. The valuation-based index is thus assumed to lag the ”true” value index and exhibit lower volatility. The model of the valuation-based index follows Geltner (1993). By regressing the transaction-based index on the valuation-based index (contemporaneous and lagged one period) it is possible to filter out the noise in the observable price index thus estimating the ”true” price index. The method may be seen as a way of ”de-smoothing” a valuation-based index without knowing the smoothing parameter beforehand. The methodology may also be used as a way of estimating the smoothing parameter.

Introduction

Price (or market value) indices for property markets are important for several reasons. Price indices are for example used as benchmarks by property owners and by investors as a means to compare average returns on property and alternative assets such as stocks and bonds. High quality price indices are also important in portfolio allocation decisions (indices can for example be used to calculate correlations between asset classes). Price indices are furthermore important in research on property markets. Research topics where price indices are used include property cycles and the relationship between property markets and other financial markets.

Unfortunately it is not a simple task to construct property price indices of high quality. Two important reasons for this are that properties are heterogeneous - different properties have different characteristics (size, age, technical amenities etc) - and that properties are transacted seldom. This means that there exists relatively few observable property prices during a given time period on a given market and that those prices are not directly comparable.

The difficulty of constructing price indices is less severe for certain types of property. Single-family homes is an example of a property type with relatively many sales where those properties that are sold also are relatively comparable. For this property type it is therefore comparatively easy to design a reliable index. For commercial properties, on the other hand, there may exist only a few transactions in a given year and market. In these conditions it may be impossible to construct a reliable index.

The difficulty of constructing an index is related to the level of aggregation. If the index is intended to capture the price level for properties in Europe we will most likely have enough transaction prices. This is likely also the case if we want to construct an index for Swedish offices. If we however want to construct an index for Stockholm CBD offices or single family homes in a particular parish of Stockholm there may not be enough data to construct a reliable index based on transactions.

One way of circumventing the problem of low liquidity is to make use of valuations instead of transaction prices. This approach depends heavily on the quality of valuations. If valuations are inaccurate this may not be a reliable way of obtaining a price index. As an index is an aggregate of many observations, inaccuracy of individual valuations is not *necessarily* problematic. Errors may cancel out. There is however research that suggests that valuations of properties lag behind and underestimate the volatility of actual value movements

(Geltner et al. 2003). This valuation bias, popularly termed ”appraisal smoothing”, does not cancel out when valuations are aggregated (Geltner et al. 2003).

This paper presents a method for combining transaction- and valuation-based data in a price index. The point of the method is to at least partly provide a remedy for inherent problems in the two types of data: noise in transaction data and smoothing in valuation data. The methodology is devised for a world where there are at least some observable transaction prices that can be used to construct a price index that constitutes a noisy signal of the ”true” price index (an index free of bias and noise). Furthermore, it is assumed that valuations from the population can be used to construct a noiseless but smoothed valuation index. This valuation index is a lagged, smoothed-out version of the ”true” index. By regressing the price index on the valuation index (contemporaneous and lagged one period) it is possible to filter out the noise in the observable price index and hence estimate the ”true” price index.

The nature of price indication data in property markets

An asset price index is an index that measures price movements in a population of assets. For some assets the construction of the index is fairly straightforward. For common stocks for example, we may simply collect price observations of every stock in the population for every time period, add them and divide by the price level in the chosen base period. Price data in property markets is generally not as easily transformed to a reliable index. For some markets there simply are too few transactions for this procedure to be feasible and when transaction data actually is available, heterogeneity of properties often makes it difficult to construct a reliable index.

Unless we control for differences in property characteristics, transaction prices are not comparable. Transaction price A may differ from transaction price B because the two transactions occur at different points in time and prices have changed *or* because property A and B are of different quality (property B may have a nice view for instance). Unless we can control for differing quality, heterogeneity will introduce noise in observed transaction prices. Hence, an index constructed by taking the average of transaction prices will be noisy. Noise will pose less of a problem the more transaction data that is available.

Heterogeneity may also introduce bias in an index. There are two reasons for this. First, the characteristics of properties may change systematically over time. If properties’ technical amenities are improved across a whole market for instance, we should observe price increases due to quality improvement. For a given level of quality however, prices may have been constant.

Second, properties of different characteristics may transact at different points in time. If high quality properties typically transact in certain time periods, failing to control for this may lead us to believe that prices have increased more than they actually have during these periods. Note that if we had continuous price data for every property, this would not be a problem. Heterogeneity and low liquidity thus together make it difficult to create indices.

It should be noted that what we mean by bias may depend on what use the index is intended for. For some applications it may not be necessary or even desirable to control for all types of differences in characteristics. One may for instance want to construct an index for which depreciation and improvements are not controlled for. This is discussed in more detail by Wang and Zorn (1997).

The fact that property markets are search markets is another source of noise in property transaction data. Transaction prices are the outcomes of negotiations between buyers and sellers. For any transaction the outcome of the transaction process is just one realization of many possible outcomes. The actual selling price can be viewed as a random variable distributed around the market value (where I think of the market value as the expectation of the selling price in a normal transaction, i.e. no forced sales for example). To exemplify, the price may end up below market value if the buyer has an exceptionally skilled negotiator at the negotiation.

A substantial literature has addressed index construction methodology and has suggested solutions to the inherent problems. The repeat sales regression (first developed by Bailey et al., 1963) is a method for producing an index that compares prices of houses that have transacted at least twice during the period for which the index is constructed. The regression model is constructed so as to compare the transaction price for the same property at two (or more) transactions. The methodology thus at least in part avoids the problem of heterogeneity.

There are three main problems with this type of index. First, the method requires plenty of transaction data and is therefore not feasible for many property markets. Only properties that have transacted at least twice during the index period can be used. Second, in its simple form, the method does not adjust for the fact that the properties in the index may change over time (depreciation, renovations etc). Later literature has suggested ways of dealing with this problem (Case and Quigley, 1991, is one example). Third, the method necessarily means that we build the index on properties that transact often (properties that have transacted only once during the index period will not enter the regression). These properties may not be representative of the population. One study that investigates this problem is Englund et al. (1999). Their study shows that Swedish single family homes that are transacted often typically are of lower quality (small lots etc).

Another way to design transaction-based indices that controls for differences in characteristics is to use a hedonic regression model. In the hedonic approach, a property is viewed as a composite good: When buying a property one is really buying a *set* of goods. The hedonic approach aims to find the marginal contribution of each of these goods or characteristics on the value of the composite good (in our case a property). This is achieved by regressing the transaction price of a property on a number of its characteristics (location, area, age etc). By introducing time dummies in the regression, it is possible to capture the price level in different time periods while the included property characteristics control for heterogeneity. An alternative approach is to estimate a hedonic regression for each time period and revalue a representative property each time period using each respective period's characteristics prices. Miles et al. (1990) and Webb et al. (1992) are examples of studies where a hedonic methodology is used.

Clapp and Giacotto (1992) suggested an efficient way of controlling for heterogeneity among properties. They argue that valuations of each respective property provide an excellent heterogeneity control. Using valuations as a control for differing characteristics is an attractive idea for two reasons: They are likely to capture very much of the heterogeneity and they are fairly easy to obtain unlike other controls that may require collection of an extensive array of property attributes. Fisher et al. (2007) present a new quarterly index for commercial property that uses this technique. As with repeat-sales methods the hedonic method is only feasible when there is plenty of data. For the hedonic approach not only transaction data is needed but also data on the characteristics of the properties in the index.

A completely different approach to constructing price (or value) indices for property is to use valuations instead of observed transaction prices. A valuation-based index is constructed by revaluing the same sample of properties each time period. Valuation-based indices thus in part avoid the problem of heterogeneity. However, assuming that the properties in the sample change in quality over time, this should be taken into account.

Using valuations as a means of tracking price (or value) movements hinges critically on the nature and quality of valuations. There is a fairly substantial literature that shows that valuations are prone to a certain type of bias. More specifically, a number of articles suggest that valuations tend to lag actual prices and also tend to smooth out actual price movements, so called ”appraisal smoothing” (Geltner et al. 2003, Diaz and Wolverton 1998, Fisher et al. 1999 and Fisher and Geltner 2000). This phenomenon can be shown to be the result of optimal valuer behaviour (Quan and Quigley, 1989 and 1991, Childs et al. 2002) but is not optimal from an index-construction point of view as smoothing in individual valuations is likely to spread to an aggregate index. If smoothing is present in valuations, an index based on valuations will simply not show actual price movements but movements in valuations. The phenomenon will be dealt with in some detail in the following.

A valuer that tries to estimate the market value of a property uses transaction prices from properties that are as comparable as possible to the property that is being valued. Ideally these comparable sales (comps) should (1), come from properties that are identical to the property being valued, (2), the transactions should have occurred very recently (ideally at the same moment that we are making the valuation) and (3), we should have access to plenty of them. This will typically not be the case as properties are heterogeneous and transact seldom. The valuer will have to make do with less perfect data. The data that is available to the valuer will contain noise (due to heterogeneity) and it will not be completely up to date (old transactions). We can think of the value estimate as a simple average of the transaction prices from comparable sales. If we use only very recent comps the value estimate will be up to date but noisy due to the fact that we have very few comps in the average, perhaps only one comparable sale. As we include older and older comps the number of comps in the average will be larger reducing the effect of noise. The value estimate will however be less up to date the farther back in time we go. Thus there will be a trade-off between noise and bias depending on how far back the valuer decides to go. Using only recent comparable sales will give an estimate that contains a lot of noise but very little bias. Using comparable sales from a longer time period will result in less noise but more bias.

How far back it is optimal to go depends on what use the valuation is intended for. If we aim for as small error as possible in the individual value estimate it may be optimal to go quite far back as this will reduce the noise. If we want to have an estimate with as little bias as possible it may be optimal to use only very recent comps. If we want to value an entire portfolio of properties for example it is arguably better to have unbiased but noisy estimates of the individual properties as noise will filter out in the aggregate.

This description of the valuers problem is simplified (valuations are usually not simple averages of comps) and is meant to give an intuitive explanation for appraisal smoothing. The general idea is that valuers use old information and that this behaviour is justified. Quan and Quigley (1991) have studied the valuers problem more formally. They find that, given a number of assumptions, it is optimal[1] for the valuer to behave according to the following model:

$$v_{it} = \alpha_{it} p_{it} + (1 - \alpha_{it}) v_{it-1} \qquad (1)$$

Where v_{it} is the valuation of property i in time t, p_{it} are (noisy) contemporaneous comparable sales and v_{it-1} is the valuation in the previous period. α_{it} is a parameter that tells us how much weight is given to current information relative to how much weight is given to old information. A large α_{it} corresponds to much weight being given to contemporaneous information and vice versa. Note that previous valuations (v_{it-1}, v_{it-2} etc) will follow the same model. It can easily be shown that formula (1) is simply a weighted average of the current and all previous comparable sales (i.e. p_{is} where s=t, t-1, t-2,…..) with lower and lower weights the farther back we go (see formulas (16) and (17) below). α_{it} is usually written without time or individual subscripts but they are included here in order to emphasize that alpha may differ over time as well as for different properties.

An index based on valuations that follow the pattern in formula (1) will be smoothed. For many applications this is problematic and research has therefore been devoted to the question of how to derive unbiased price indices from valuation-based indices. Two groups of solutions are the ”zero-autocorrelation” method and the ”reverse engineering” method. The zero-autocorrelation method builds on the idea that returns in property markets should be unpredictable. Using this assumption it is possible to back out ”true” (non-autocorrelated) returns through a regression where autocorrelated return is filtered out. Once ”true” returns have been calculated these can be used to calculate ”true” price levels. Blundell and Ward (1987) proposed this technique and a number of articles have used and/or developed the method (Fisher et al. 1994, Cho et al. 2003 and Brown and Matysiak, 1998). The main caveat of the method is the problematic assumption of zero-autocorrelation in returns, which may not hold.

The ”reverse engineering” method is related to model (1) and was proposed by Geltner (1993). Geltner (1993) argues that if individual valuations follow the pattern in formula (1), a valuation-based index will be well described by the following model:

[1] It is optimal in the sense that if V_t is chosen in accordance with this formula, V_t will converge to the true market value of the property faster than any other simple linear valuation rule.

$$V_t = \alpha \tilde{P}_t + (1-\alpha) V_{t-1} \qquad (2)$$

where V_t and V_{t-1} are the valuation-based index levels at time t and t-1 respectively and $\tilde{P}_t$ is a price (or market value) index level at time t (the tilde is merely there to distinguish $\tilde{P}_t$ from a different price index P_t below). Note that V_t and V_{t-1} are observable. $\tilde{P}_t$ on the other hand is here regarded as a non-observable component of V_t. If (2) holds and if we know α it is possible to construct a price index by backing out ("reverse engineering") $\tilde{P}_t$ from formula (2):

$$\tilde{P}_t = \frac{V_t}{\alpha} - \frac{(1-\alpha) V_{t-1}}{\alpha} \qquad (3)$$

(3) is just a simple manipulation of formula (2). Geltner furthermore argues that the noise in p_{it} will largely diversify away in their aggregate counterpart $\tilde{P}_t$ so that $\tilde{P}_t$ may be viewed as a "true" (unbiased, noiseless) price index. We may also assume that $\tilde{P}_t$ contains noise and employ some noise-reduction technique.

The i subscripts have been dropped in formula (2) and (3) in order to emphasize that $\tilde{P}_t$, V_t and V_{t-1} are measured at the index level in these formulas and that α when used in this way usually is assumed to be constant (an assumption that may not hold).

One of the main problems with reverse engineering is that we must estimate α, which is inherently difficult as we do not have access to the "true" price index and probably not the valuers' comps (p_{it} in formula (1)) either. One of the few studies on the subject is Clayton et al. (2001). The difficulty of obtaining α is aggravated by the fact that α may vary over time and over different properties (empirical support for this can be found in Brown and Matysiak (1998)) and that α on the individual property level not is necessarily immediately transferable to the aggregate (index) level (Bond and Hwang, 2007).

So far we have discussed how both valuations and transaction prices are imperfect measures of price movements in property markets. There are however, other more indirect indicators of property prices. One prominent example is prices on stocks of listed property companies (or REITS). These prices refer to indirectly owned property which means that they cannot be used as price indicators for the direct property market without adjustment (or at least not without caution). Property stocks are for example usually leveraged assets. This has to be taken into account as we usually create property indices for properties as such, not *leveraged* property holdings (which does not preclude that the properties in indices are owned by leveraged owners). Empirical research has also found that property stock prices move partly independently from the directly owned property market (Chau et al. 2001).

Ling et al. (2000) and Fu (2003) are two examples of articles that present methods of using indirect indicators for computing price indices. Both articles make use of latent variable models. With this type of model it is possible to calculate an unobservable ”latent” variable with the help of a number of observable ”indicator” variables. Applied to property price indices, the latent variable is the ”true” value index while valuations and property stock prices may be used as indicator variables.

Proposed index construction method

In short, the setting is as follows. It is assumed that indications of current market value can be obtained from two sources; transaction prices and valuation data. The transaction prices are assumed to be unbiased estimates of market value but contain a lot of noise. The valuation data on the other hand is assumed to suffer from the effects of appraisal smoothing (lag, lower volatility).

Assume that there are three indices in the market, two of them observable and one unobservable. First we have the unobservable ”true” price index, I_t, that we want to estimate. There is also a transaction-based index, P_t, which is built on noisy transaction price data. It is assumed that the price index P_t is dispersed around the ”true” market value index:

$$P_t = I_t + u_t \qquad (4)$$

Where u_t is a random error distributed around the market value index and $E(u_t) = 0$. u_t is assumed to be uncorrelated with I_s where s =…, t+2, t+1, t, t-1, t-2….. The variance of u_t may differ in different time periods. In words, P_t is a noisy measure of I_t.

Assume furthermore that we have a valuation-based index, V_t. This index is built on individual appraisals. The individual appraisals are assumed to follow the pattern discussed above (formula (1)). It is furthermore assumed that this pattern carries through to the index so that we have

$$V_t = \alpha I_t + (1-\alpha) V_{t-1}, \qquad (5)$$

where α is the smoothing parameter. In words, the valuation series V_t provides a ”smoothed” but noiseless signal of I_t. Regarding the behaviour of I_t and V_t, the presented setting is the same as Geltner (1993). One could use ”reverse engineering” on the valuation based index V_t presuming that we have an idea of the value of α.

After considering the set-up, the following question may arise: Why does the price index P_t contain noise while the signal of ”true” value in the valuation-based index does not? In the presented set-up, the individual valuation is built on noisy price information and the previous valuation, but when we combine valuations in an index, the noise in the price information filters out. Why can we not simply collect the price information that valuers use and create a transaction-based index free of noise? The noise filters out in the valuation-based index – why not in the transaction-based index?

The set-up implicitly assumes that the price information that valuers have access to is richer than the price information available to the person constructing the index. This requires some motivation. First of all, the information available to valuers may be costly or impractical for the index-constructor to acquire. It may for example be the case that the data are not collected in one place or that the raw data needs extensive processing before use. Secondly, valuers may have access to information that simply is not available to the index-constructor. Some transaction prices may not be disclosed publicly but leak to valuers. Some transactions are part of a larger deal that includes other assets as well. In this type of deal the implicit transaction price of the property may not be known to the public but to valuers. Furthermore, the noisy price information that valuers use may not be actual transaction prices. Knowledge of deals that did not happen, rumours etc may be seen as part of the noisy price information used by valuers. Despite this argument one may argue that the ”true” price index component in (5) should include an error term. The effects of allowing for this are discussed in a subsequent section (equations (15) and (16)).

Simulation (A) in Figure 1 shows visually how I_t, P_t and V_t relate to each other. In this simulation I_t is assumed to follow a random walk:

$$I_t = I_{t-1} + v_t \qquad (6)$$

$$v_t \sim N(0,1) \qquad (7)$$

I_t was constructed by generating 25 random numbers (v_t) and then using formula (6). P_t was generated using formula (4) where $u_t \sim N(0,4)$. V_t was constructed using formula (5). α was set equal to 0.4. The figure illustrates that P_t is a noisy (more volatile) version of I_t and that V_t is a smoothed (less volatile, lagging) version of I_t.

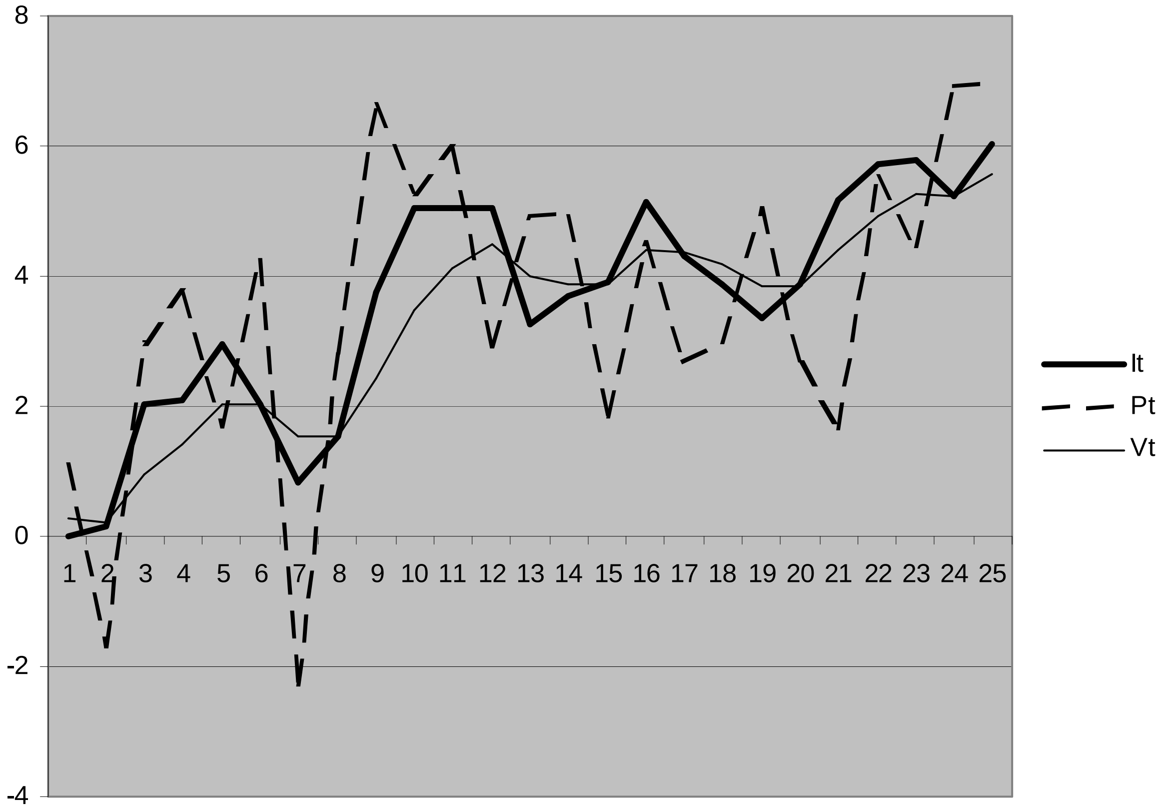


Figure 1. Simulation (A) of a "true" value index (I_t), an index constructed with observable transactions (P_t) and an index constructed with valuations (V_t).

Equation (5) is equivalent to:

$$I_t = \frac{V_t}{\alpha} - \frac{(1-\alpha)V_{t-1}}{\alpha} \qquad (8)$$

Equation (8) is a description of how I_t is related to V_t and V_{t-1} where I_t is expressed as a linear function of V_t and V_{t-1}. Of course, I_t is not literally *driven* by V_t and V_{t-1}. (8) merely shows how variation in I_t can be captured with V_t and V_{t-1} if we assume that equation (5) holds. Assuming that we can observe the three variables we could estimate (8) by OLS. If we were to regress I_t on V_t and V_{t-1} we would be able to capture all variation in I_t since I_t only

”depends” on V_t and V_{t-1}. The coefficient for V_t would equal $1/\alpha$ and the coefficient for V_{t-1} would equal $-(1-\alpha)/\alpha$. If we included an intercept in the regression it would equal zero. I use the word depend here in the sense that the variation in I_t can be captured by V_t and V_{t-1}.

Now, we can observe V_t and V_{t-1} but not I_t. We can however observe P_t which is just a noisy measure of I_t:

$$P_t = \frac{V_t}{\alpha} - \frac{(1-\alpha)V_{t-1}}{\alpha} + u_t \tag{9}$$

I have simply inserted the right-hand side of equation (8) instead of I_t in equation (4) in order to arrive at (9). Model (9) is possible to estimate since we have assumed that P_t and V_t are observable. We would then run the following regression model:

$$P_t = \beta_0 + \beta_1 V_t + \beta_2 V_{t-1} + e_t \tag{10}$$

where we know from (9) that the true parameters are $\beta_0 = 0$, $\beta_1 = 1/\alpha$, $\beta_2 = -(1-\alpha)/\alpha$ and that $e_t = u_t$. Assuming that u_t is uncorrelated with V_t and V_{t-1} the coefficients for the explanatory variables will be unbiased. In other words, their expected values are their respective true population counterparts:

$$E\left(\hat{\beta}_0\right) = 0, \tag{11}$$

$$E\left(\hat{\beta}_1\right) = 1/\alpha, \tag{12}$$

$$E\left(\hat{\beta}_2\right) = -(1-\alpha)/\alpha \tag{13}$$

We can obtain predicted P_t:

$$\hat{P}_t = \hat{\beta}_0 + \hat{\beta}_1 V_t + \hat{\beta}_2 V_{t-1} \tag{14}$$

The expected value of $\hat{P}_t$ given V_t and V_{t-1} is:

$$E\left(\hat{P}_t \middle| V_t, V_{t-1}\right) = E\left(\hat{\beta}_0 + \hat{\beta}_1 V_t + \hat{\beta}_2 V_{t-1} \middle| V_t, V_{t-1}\right) \tag{15}$$

$$= \beta_0 + \beta_1 V_t + \beta_2 V_{t-1}$$

$$= \frac{V_t}{\alpha} - \frac{(1-\alpha)V_{t-1}}{\alpha}$$

$$= I_t$$

In words, predicted P_t is an unbiased estimate of I_t. As the number of observations increases, the coefficients are better and better estimated and the predicted P_t will come closer and closer to I_t.

Figure 2 shows simulation (B) which is similar to simulation (A) in Figure 1 but in which I have also included $\hat{P}_t$ which is predicted P_t from a regression where P_t is regressed on V_t and V_{t-1} (regression model (10)). As is evident from the figure, the predicted P_t comes close to I_t.

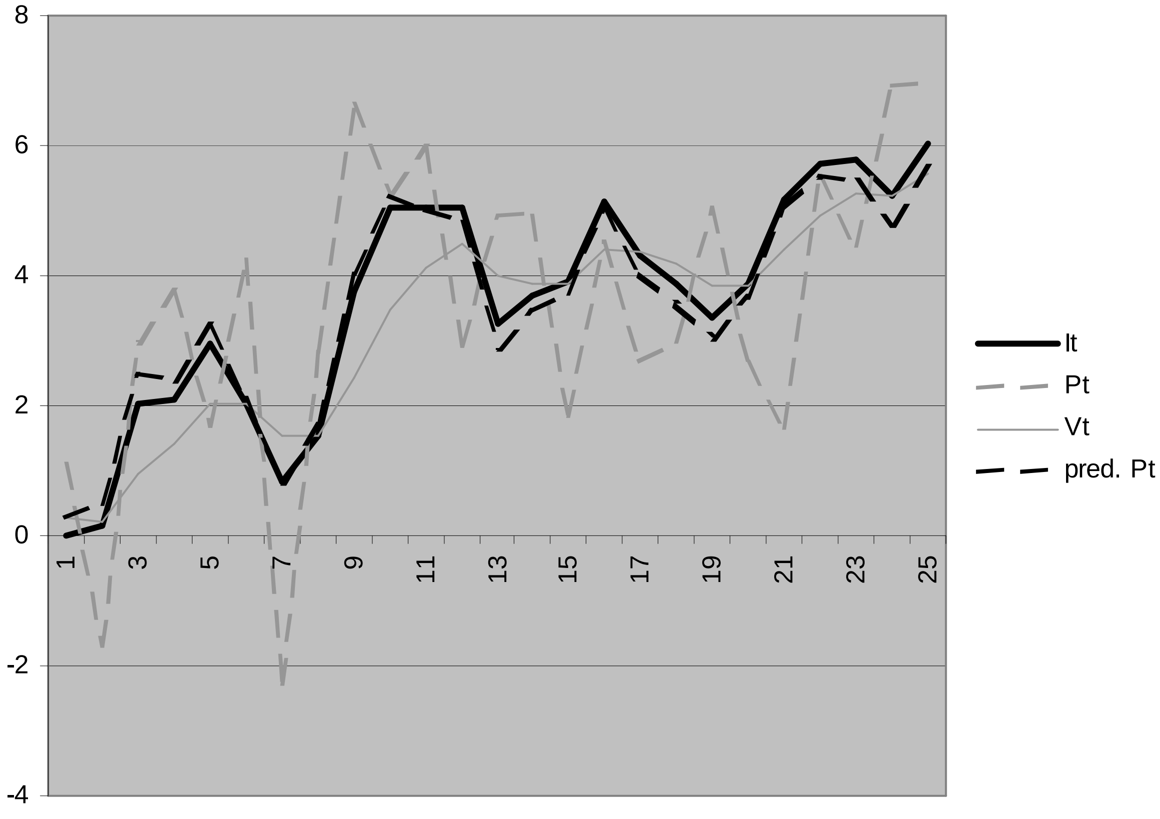


Figure 2. Simulation (B) of a ”true” value index (I_t), and an estimation of I_t (predicted P_t) using transactions-based and valuation-based indices.

We do not actually have to assume that u_t is uncorrelated with V_t and $V_{t\text{-}1}$. It follows from previously made assumptions: (i) the assumption that u_t is uncorrelated with I_t in all time periods and (ii) the assumed model of the appraisal-based index, equation (5). To see this, note that equation (5) implies that V_t can be expressed as a function of the current and lagged values of I_t. We have (equation (5) restated):

$$V_t = \alpha I_t + (1-\alpha)V_{t-1} \tag{16}$$

Insertion of $\alpha I_{t-1} + (1-\alpha)V_{t-2}$ instead of V_{t-1}, $\alpha I_{t-2} + (1-\alpha)V_{t-3}$ instead of V_{t-2} and so on yields:

$$V_t = \alpha I_t + (1-\alpha)\alpha I_{t-1} + (1-\alpha)^2 \alpha I_{t-2} + (1-\alpha)^3 \alpha I_{t-3} + ... \tag{17}$$

Equation (17) shows that V_t is a function of I_s where $s = t$, t-1, t-2…. which are all uncorrelated with u_t by assumption. Hence, u_t is uncorrelated with V_t. The same argument holds for $V_{t\text{-}1}$.

The reader may object that estimating P_t on V_t and $V_{t\text{-}1}$ results in biased coefficient estimates due to simultaneity (the argument might be that prices drive valuations, not the other way round). Then we have to remember what we are trying to achieve with regression equation (10). The point of the regression is not to test a causal relationship. The point is instead to reduce the noise in the P_t observations (or to get rid of the lagging/smoothing behaviour in V_t if you will). β_1 and β_2 should not be thought of as measuring causal effects but rather the linear relationship between P_t, V_t and $V_{t\text{-}1}$. We know from the assumptions that we have made that this relationship follows formula (9).

How can valuations completely capture ”true” price movements in this setting? In order to give an intuitive explanation why this may be the case let us start with the basic model of how valuations relate to the ”true” price:

$$V_t = \alpha I_t + (1-\alpha)V_{t-1} \tag{18}$$

The formula shows that V_t contains both the ”true” price I_t scaled down by a factor α and the previous valuation $V_{t\text{-}1}$. Thus, by scaling up the ”true” price component and getting rid of the V_{t-1} component we have the ”true” price. This is exactly what happens when we regress P_t on V_t and V_{t-1}. From above (equation (15)) we have that:

$$E\left(\hat{P}_t \middle| V_t, V_{t-1}\right) = \frac{V_t}{\alpha} - \frac{(1-\alpha)V_{t-1}}{\alpha} \tag{19}$$

The first term in this expression may be thought of as the term that scales up the I_t component of V_t. To see this note the following:

$$\begin{aligned} \frac{V_t}{\alpha} &= \frac{1}{\alpha}\left(\alpha I_t + (1-\alpha)V_{t-1}\right) = \\ &= I_t + \frac{(1-\alpha)V_{t-1}}{\alpha} \end{aligned} \tag{20}$$

Subtracting the ”previous-valuation-component”, $\frac{(1-\alpha)V_{t-1}}{\alpha}$, from $\frac{V_t}{\alpha}$ we get:

$$\frac{V_t}{\alpha} - \frac{(1-\alpha)V_{t-1}}{\alpha} = I_t + \frac{(1-\alpha)V_{t-1}}{\alpha} - \frac{(1-\alpha)V_{t-1}}{\alpha} = I_t \tag{21}$$

Relaxing assumptions

The proposed method relies on a number of assumptions. If these assumptions are fulfilled, the index construction method works well in the sense that it produces an unbiased estimate of the ”true” index series. Of course, the assumptions may not be fulfilled or at least may not be completely fulfilled. The rest of the paper discusses how the results are affected if the assumptions are not fulfilled.

Price process

In the presentation of the methodology, the process of the ”true” price index was not discussed and no assumptions were made about what it looks like. In other words, the index construction method is not dependent on a particular process of the ”true” price. Simulation (C) was made to illustrate this. I_t is assumed to follow an ARMA(1,1) process:

$$I_t = 0.5 I_{t-1} + v_t + v_{t-1} \tag{22}$$

$$v_t \sim N(0,1) \tag{23}$$

V_t and P_t are constructed in the same way as in simulation (A) and (B) but α is assumed to be 0.3 in this simulation.

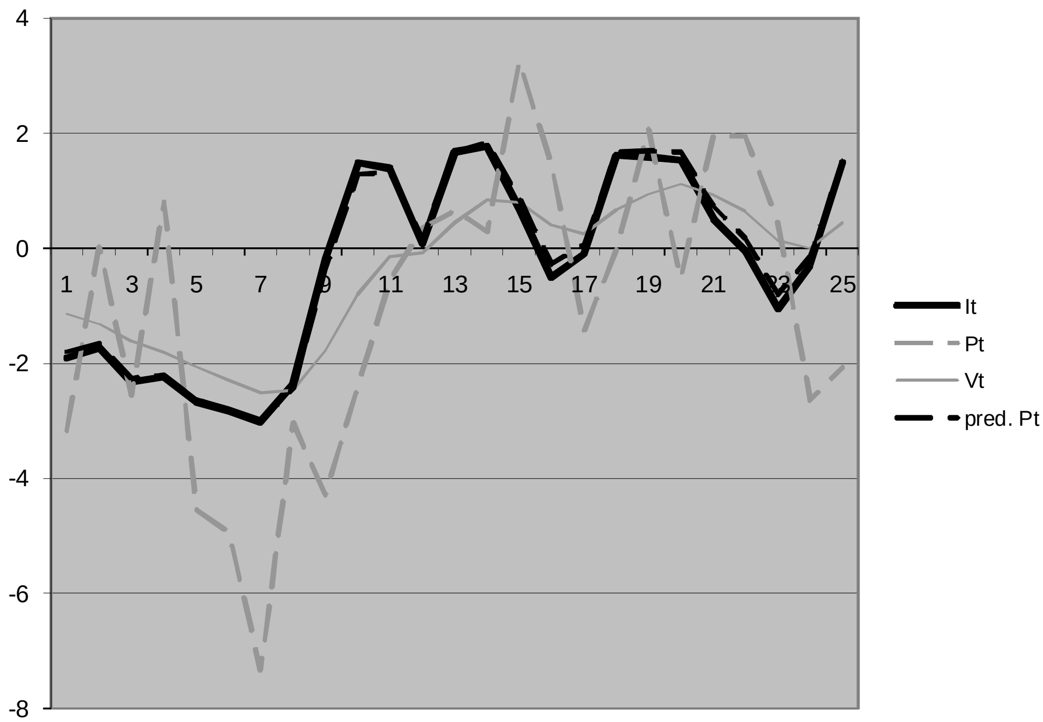


Figure 3. Simulation (C) of a "true" value index (I_t), and an estimation of I_t (predicted P_t) using transactions-based and valuation-based indices. I_t is assumed to follow an ARMA(1,1) process.

As in simulation (B), predicted P_t follows I_t closely: the methodology is not sensitive to the process of the "true" price index. The simulation serves a second purpose. In this simulation, 200 observations were generated instead of 25 observations as in simulation (B). This means that when regressing P_t on V_t and V_{t-1} in this simulation, coefficients are estimated with more accuracy. Consequently, predicted P_t follows I_t more closely than in simulation (B) illustrating the fact that the more observations, the better the proposed methodology works.

Valuer model

The assumption of how the valuation index behaves, equation (5), is explicitly used in the derivation of the index construction method. In general, therefore, the method does not work unless this assumption holds. The method may however still work as an approximation even if equation (5) does not hold in a strict sense. Whether the approximation is reasonable or not depends on exactly how reality deviates from equation (5). As the true behaviour of V_t may deviate from equation (5) in countless ways it is impossible to give an exhaustive discussion of what happens when model (5) is invalid. This section will discuss some possible deviations.

First, one may think of several models that share important traits with model (5) but deviate in some sense. Model (24) is one such example:

$$V_t = \alpha_1 I_t + \alpha_2 I_{t-1} + (1 - \alpha_1 - \alpha_2) I_{t-2} \quad (24)$$

This model will lag the ”true” index and will smooth out its movements just like model (5). The difference between the models is the weights and the fact that model (5) goes further back in time. Model (24) is motivated for example if we think that valuers do not go as far back in time as suggested by model (5).

A simulation was run where the ”true” price index is assumed to follow a random walk as in simulation (B), P_t is generated as in simulation (B) and V_t is now assumed to follow model (24) with weights chosen to be $\alpha_1 = \alpha_2 = 1 - \alpha_1 - \alpha_2 = 1/3$. The results of simulation (D) are shown in figure 4. As expected, the results are not as good as in the previous simulations. The methodology does however not collapse completely. There is little lagging and much of the noise is eliminated. If we have more observations the results are even better. Simulation (D) was made with 25 observations. Appendix A shows the results when the simulation is made with 1000 observations. While the results for model (24) are encouraging, they cannot be generalized. Simulation (D) does however show that the methodology does not *necessarily* collapse if model (5) is not true.

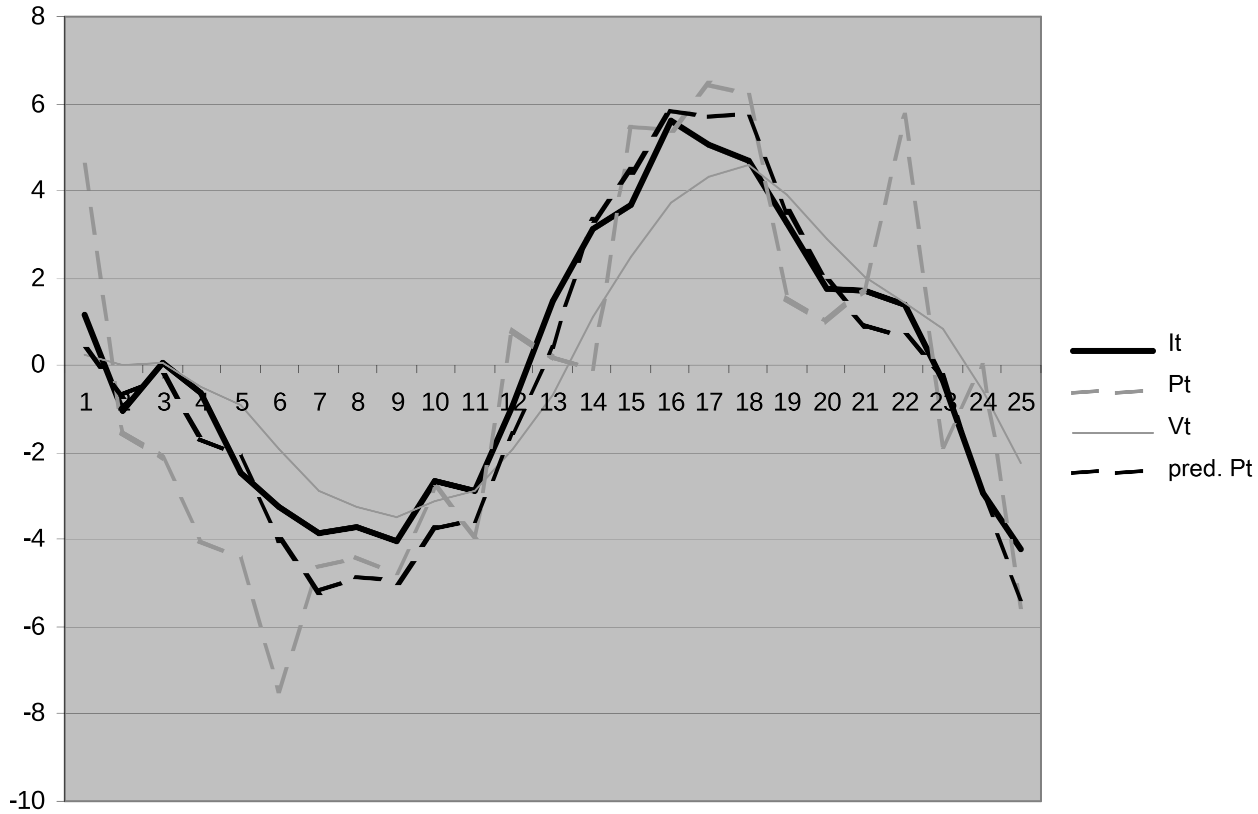


Figure 4. Simulation (D) of a "true" value index (I_t), and an estimation of I_t (predicted P_t) using transactions-based and valuation-based indices. V_t is assumed to follow model (24).

An alteration to model (5) that makes sense intuitively is to assume that instead of I_t in model (5) we have I^*_t which is I_t plus random noise n_t:

$$V_t = \alpha I^*_t + (1-\alpha)V_{t-1} \tag{25}$$

$$I^*_t = I_t + n_t \tag{26}$$

The rationale for this model is that maybe not all of the noise from the individual valuations is filtered out when valuations are aggregated into an index. If (25) holds the true population model of P_t is:

$$P_t = \frac{V_t}{\alpha} - \frac{(1-\alpha)V_{t-1}}{\alpha} - n_t + u_t \tag{27}$$

If we regress P_t on V_t and V_{t-1} when the true population model is equation (27) the coefficient estimates will be biased as V_t is correlated with the error term in equation (27). This can be seen from equation (25) and (26): V_t is a function of n_t. In general therefore, this type of deviation from the assumptions is problematic. Three simulations were made in order to see *how* problematic. The simulations are all similar to simulation (B) except that V_t is constructed using formula (25) and (26). They differ between each other in how large the variance of n_t is. Simulation (E) has the lowest variance of n_t, 0.0625, which can be compared with each time periods innovation in I_t which has a variance of 1. When the variance of n_t is this low the problem associated with this type of deviation is relatively small (see figure 5).

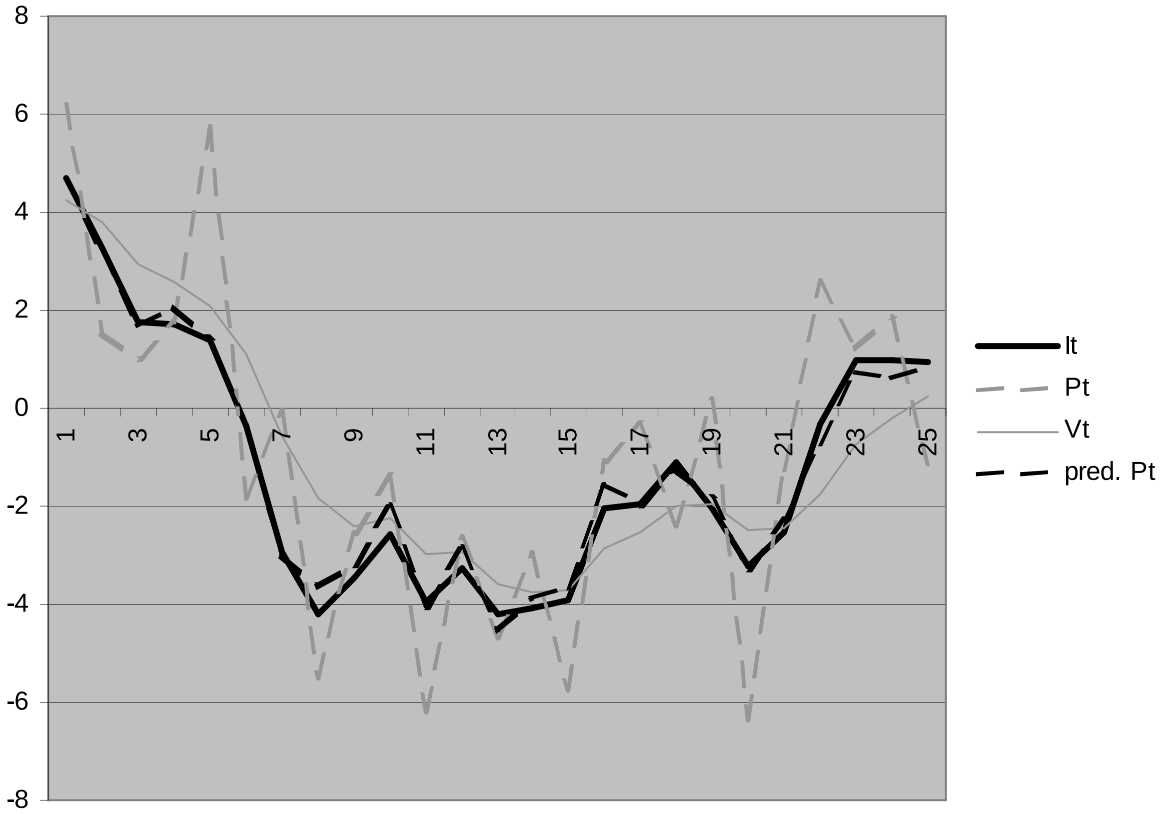


Figure 5. Simulation (E) of a ”true” value index (I_t), and an estimation of I_t (predicted P_t) using transactions-based and valuation-based indices. V_t is assumed to follow model (25) and the variance of n_t is 0.0625.

If the variance of n_t is 0.5625 as in simulation (F) there are bigger problems as can be seen from figure 6. Appendix B shows the results when the variance of n_t is 6.25. When the variance is this high, the predicted P_t follows V_t rather than I_t. This simulation is however not included as a practical example but rather to show that the estimate of P_t is biased towards V_t.

The results show that the effect of this type of noise depends critically on the variance of the noise.

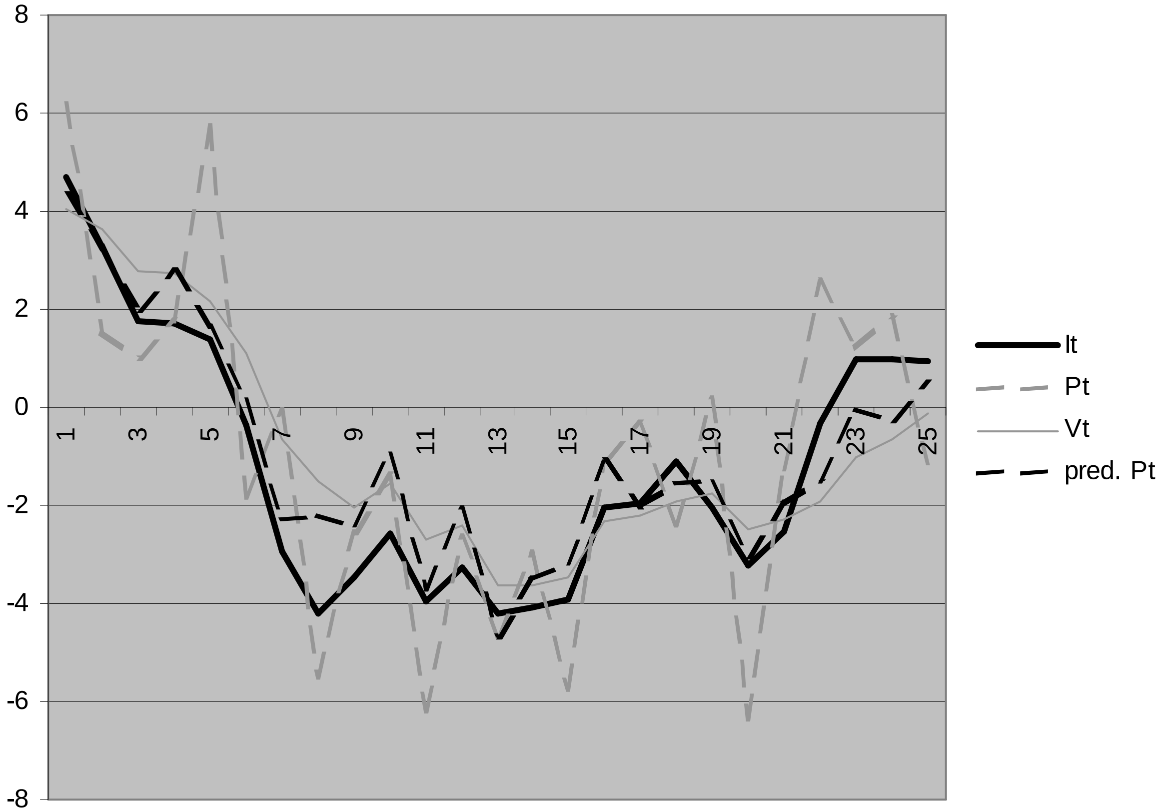


Figure 6. Simulation (F) of a ”true” value index (I_t), and an estimation of I_t (predicted P_t) using transactions-based and valuation-based indices. V_t is assumed to follow model (25) and the variance of n_t is 0.5625.

Constant alpha

The proposed model implicitly assumes that the smoothing parameter α does not change over time. Quan and Quigley (1991) showed in a theoretical model that α can be expected to be different in different market conditions. This is intuitively appealing since different periods exhibit differences in transaction volume and hence the number of comps that valuers can use. Brown and Matysiak (1998) show empirical evidence that α differs over time and circumstances. A simulation (G) was made in order to see what happens when α changes over time. In the simulation, α follows a simple process: for the first 13 time periods,

α is 0.4, for the latter 12 time periods α is 0.2. Except for the changing α the simulation is similar to simulation (B).

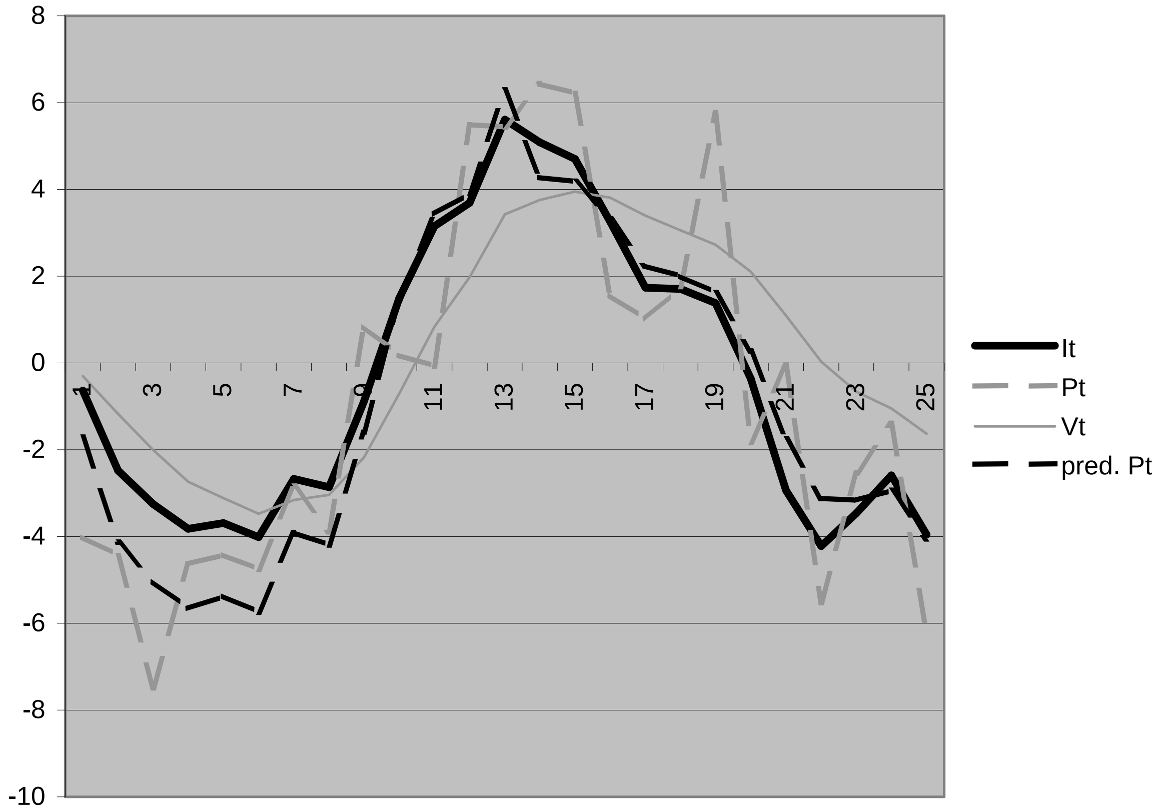


Figure 7. Simulation (G) of a ”true” value index (I_t), and an estimation of I_t (predicted P_t) using transactions-based and valuation-based indices. The smoothing parameter α shifts over time in this simulation.

The simulation shows that the method is sensitive to changing α: For the first part of the index, true market movements are exaggerated while the opposite is true for the latter part. This stems from the fact that α is estimated at 0.3 or the average α over the time period. Consequently α is underestimated for the first half of the period and overestimated for the second half. This in turn has the effect that movements in I_t is exaggerated in the first half and the other way round in the second half. Simulation (G) has shown but one way in which α may change but has demonstrated that the method is sensitive to this assumption. A feasible remedy to this problem is to use a rolling regression technique.

Conclusion

This paper presents a method for combining transaction- and valuation-based data in a price index. The point of the method is to at least partly provide a remedy for inherent problems in the two types of data: noise in transaction data and smoothing in valuation data. The methodology is devised for a world where the observable transaction prices can be used to construct a price index that constitutes a noisy signal of the ”true” price index. Furthermore, it is assumed that valuations can be used to construct a market value index which is a noiseless but smoothed version of the ”true” index.

By regressing the observable price index on the valuation index (contemporaneous and lagged one period) it is possible to filter out the noise in the observable price index. If there are many observations, the predicted observable price index comes very close to the ”true” price index. The method may be seen as a way of ”de-smoothing” a valuation-based index. The advantage that this method gives compared to earlier de-smoothing techniques is that it does not require us to know the smoothing parameter beforehand. On the contrary, the methodology may be seen as a way of estimating the smoothing parameter.

The paper discusses some of the assumptions made. It is shown that the method is insensitive to the ”true” price process. The model of the valuation index is a more crucial assumption but it is demonstrated that deviation from the model assumed is not necessarily critical. It is furthermore pointed out that over time varying smoothing of the valuation index is problematic. This may however be remedied by a rolling regression technique.

References

Bailey, M. J., Muth, R. F. & Nourse, H. O. (1963). "A Regression Method for Real Estate price index construction", *Journal of American Statistical Association, 58*, 933-942.

Blundell, G. & Ward, C. (1987). "Property portfolio allocation: A multifactor model", *Land Development Studies, 4*(2), 145-146.

Bond, S., & Hwang, S. (2007). "Smoothing, nonsynchronous appraisal and cross-sectional aggregation in real estate price indices", *Real Estate Economics, 35*(3), 349-382.

Brown, G. R. & Matysiak, G. A. (1998). "Valuation smoothing without temporal aggregation", *Journal of Property Research, 15*(2), 89-105.

Case, B. & Quigley, J.M. (1991). "The dynamics of real estate prices", *The Review of Economics and Statistics, 73*(1), 50-58.

Chau, K.W., Macgregor, B. D. & Schwann, G. M. (2001). "Price discovery in the Hong Kong real estate market", *Journal of Property Research, 18*(3), 187-216.

Childs, P., Ott, S. & Riddiough, T. (2002). "Optimal valuation of noisy real assets", *Real Estate Economics, 30*(3), 385-414.

Cho, H., Kawaguchi, Y. & Shilling, J.D. (2003). "Unsmoothing commercial property returns: A revision to Fisher-Geltner-Webb's unsmoothing methodology", *Journal of Real Estate Finance and Economics, 27*(3), 393-405.

Clapp, J. & Giacotto, C. (1992). "Estimating price indices for residential property: A comparison of repeat sales and assessed value methods", *Journal of the American Statistical Association, 87*, 300-306.

Clayton, J., Geltner, D. & Hamilton, S. (2001). "Smoothing in commercial property valuations: Evidence from individual appraisals", *Real Estate Economics, 29*(3), 337-360.

Diaz, J. & Wolverton, M. (1998). "A longitudinal examination of the appraisal smoothing hypothesis", *Real Estate Economics, 26*(2), 349-358.

Englund, P., J. M. Quigley & C. L. Redfearn (1999). ”The choice of methodology for computing housing price indexes”, *Journal of Real Estate Finance and Economics, 19*(2), 91-112.

Fisher, J. & Geltner, D. (2000). ”De-lagging the NCREIF index: transaction prices and reverse-engineering”, *Real Estate Finance, 17*(1), 7-22.

Fisher, J., Geltner, D. & Webb, B. (1994). “Value indices of commercial real estate: A comparison of index construction methods”, *Journal of Real Estate Finance and Economics, 9*(2), 137-164.

Fisher, J., Geltner, D. & Webb, B. (2007). “A quarterly transactions-based index (TBI) of institutional real estate investment performance and movements in supply and demand”, *Journal of Real Estate Finance and Economics, 34*(1), 5-33.

Fisher, J., Miles, M. & Webb, R. (1999). “How reliable are commercial property appraisals? Another look”, *Real Estate Finance, 16*(3), 9-15.

Fu, Y. (2003). “Estimating the lagging error in real estate price indices”, *Real Estate Economics, 31*(1), 75-98.

Geltner, D. (1993). “Estimating market values from appraised values without assuming an efficient market”, *Journal of Real Estate Research, 8*(3), 325-345.

Geltner, D., Macgregor, B. D. & Schwann, G. M. (2003). “Appraisal smoothing and price discovery in real estate markets”, *Urban Studies, 40*(5-6), 1047-1064.

Ling, D. C., Naranjo, A. & Nimalendran, M. (2000). ”Estimating returns on commercial real estate: A new methodology using latent variable models”, *Real Estate Economics, 28*(2), 205-231.

Miles, M., Cole, R. & Guilkey, D. (1990). “A different look at commercial real estate returns”, *Journal of the American Real Estate and Urban Economics Association, 18*(4), 403–430.

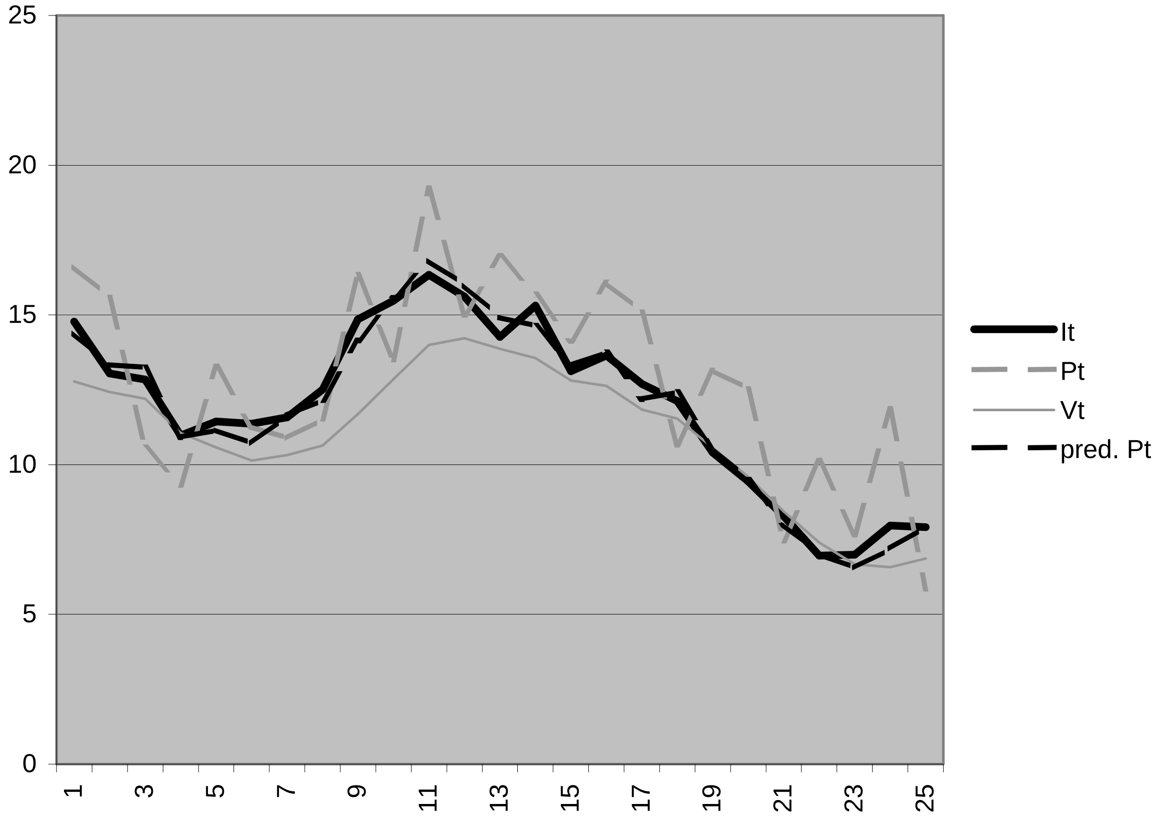
Quan, D. C. & Quigley, J. M. (1989). “Inferring an investment return series for real estate from observations on sales”, *Journal of the American Real Estate and Urban Economics Association, 17*(2), 218-230.

Quan, D. C. & Quigley, J. M. (1991). “Price formation and the appraisal function in real estate markets”, *Journal of Real Estate Finance and Economics, 4*(2), 127-146.

Wang, F.T., & Zorn, P.M. (1997). “Estimating House Price Growth with Repeat Sales Data: Whats the Aim of the Game?”, *Journal of Housing Economics, 6*(2), 93-118.

Webb, R. B., Miles, M. & Guilkey, D. (1992). “Transactions-Driven Commercial Real Estate Returns: The Panacea to Asset Allocation Models?”, *Journal of the American Real Estate and Urban Economics Association, 20*(2), 325–357.

Appendix A



Simulation of ”true” value index (I_t), and estimation of I_t (predicted P_t) using transactions-based and valuation-based indices. V_t is assumed to follow model (24). The simulation is based on 1000 observations.

Appendix B

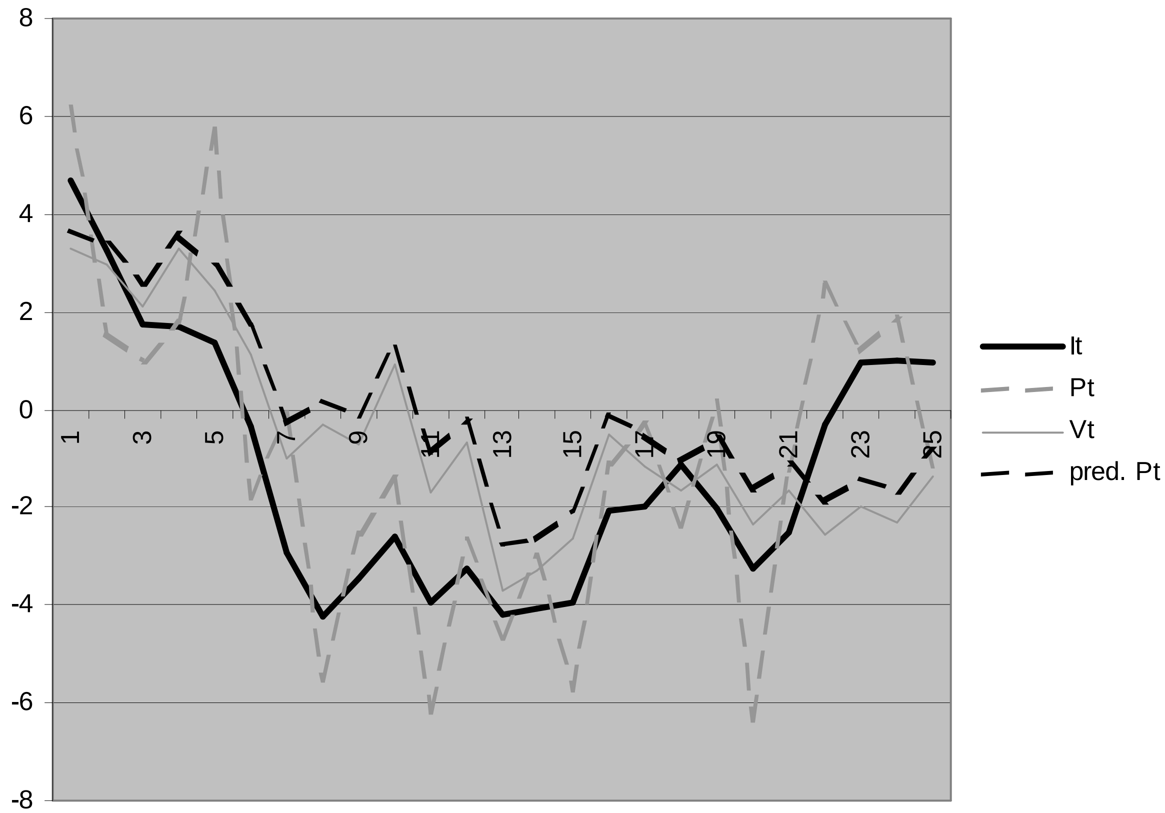


Figure 6. Simulation of a ”true” value index (I_t), and an estimation of I_t (predicted P_t) using transactions-based and valuation-based indices. V_t is assumed to follow model (25) and the variance in n_t is 6.25.