A COMPARISON OF THE CASE – SHILLER HOUSE PRICE INDEX METHODOLOGY WITH THE ACADATA HOUSE PRICE INDEX METHODOLOGY

This report compares the Case-Shiller Index Methodology (CSI) which employs Repeat Sales Regression (RSR) techniques, as widely used and understood in the USA, with the Acadata House Price Index methodology. The latter employs mix adjustment techniques which, together with hedonic analyses, are widely used and understood in the construction of UK house price indices.

An evaluation of the strengths and weaknesses of each methodology concludes with an explanation of why RSR, appropriate in the USA where the lack of a national property tax mitigates against the collection of consistent national and regional data, is not yet a preferred technique in the UK.

Originally published on 23rd May 2007, this paper was written by Dr Chris Meissner, then Fellow of Kings College and Lecturer in the Faculty of Economics at the University of Cambridge, and Dr Stephen Satchell, then The Reader in Financial Econometrics at the University of Cambridge, and now Economics Fellow Trinity College Cambridge.

28th October 2013

www.acadata.co.uk
Acadata Limited
226 Sheen Lane
London SW14 8LD
United Kingdom
telephone +44 20 8392 9082
fax +44 20 8876 1694
email info@acadata.co.uk

DR CHRIS MEISSNER, FELLOW OF KING'S COLLEGE CAMBRIDGE AND LECTURER IN THE FACULTY OF ECONOMICS AT THE UNIVERSITY OF CAMBRIDGE.

DR STEPHEN SATCHELL, FELLOW OF TRINITY COLLEGE AND THE READER IN FINANCIAL ECONOMETRICS AT THE UNIVERSITY OF CAMBRIDGE.

A COMPARISON OF THE CASE-SHILLER HOUSE PRICE INDEX METHODOLOGY WITH THE ACADATA HOUSE PRICE INDEX METHODOLOGY.

23RD MAY 2007

1. INTRODUCTION

1.1. This paper compares the Case-Shiller Index (CSI) methodology, as used by Standard & Poor's[1], with the mix adjustment methodology used for the Acadata House Price Index (AcadHPI). The Repeat Sales Regression (RSR) techniques upon which the CSI methodology is based are widely understood and used in the USA. Mix adjustment, as in the AcadHPI, and hedonic analysis are widely used and understood in the UK. This note evaluates, from an academic standpoint, the strengths and weaknesses of the AcadHPI mix adjustment methodology and of the RSR methodology, as used in a Repeat Sales Index (RSI) and explains why RSR is not yet a preferred methodology in the UK.

1.2. AcadHPI uses data based upon every transaction to construct a mix adjusted index of housing market prices in England and Wales. The CSI restricts attention to properties sold at least twice within a given period to measure changes in housing market values in 20 metropolitan areas in the USA, as well as a national index. The two procedures use different information and one method is unlikely to be dominant in all markets and situations. What we seek in this analysis is an understanding of the circumstances in which one methodology can be deemed to be better. As we will see, the concept of "better" is not clear-cut.

1.3. In section 2, we discuss the intuition for the different methodological approaches, and include a note on the concept of property characteristics. In section 3, we explore the construction and methodology of the CSI and, in section 4, the construction and methodology of the AcadHPI. In section 5, we consider the question of accuracy in an index. Section 6 concludes and notes the demerits of each methodology.

[1] *S&P Alternative Indices S&P/Case-Shiller® Home Price Indices– Standard and Poor's, December 2006*

2. METHODOLOGICAL APPROACHES

2.1. DATA AND CULTURE

Whilst indices using RSR methodologies are generally regarded as being more accurate than those that use other available methodologies, there is concern that they are not infallible. Both AcadHPI mix adjusted and CSI RSR type indices have their advantages and disadvantages. RSR is particularly viable where data volumes are large and where homes have homogeneous characteristics and are bought primarily for occupation, rather than for speculative improvement, so that price increases occur primarily as a by-product of occupation and resale prices are not biased by any significant "profit" element resulting from a change in the characteristics of a property. These conditions suitable for an RSR index are found to a much greater extent in the USA than in the UK where the property owning culture in much of the country leads to specific data problems in developing a house price index.

2.2. ESSENTIAL DIFFERENCES

The essential difference between the AcadHPI and the CSI is that the former directly measures monthly house prices and thence change; the latter measures change and thence monthly house prices. AcadHPI relies upon using every transaction to minimise the effects of bias; the CSI relies upon using data concerning single properties sold twice to minimise bias. The bias in the first case would arise from a sample selection whilst the bias which the CSI tries to avoid stems from comparing heterogeneous properties over time. But bias implies that there is a true answer and this is not a concept without difficulties. This point is discussed further in section 5.

2.3. PROPERTY CHARACTERISTICS

Bias in an index may result from a number of causes. Take, as an example, change in the characteristics of the properties sold. Suppose that a set of two bedroom houses were all converted to three bedroom houses by their current owners for speculative purposes and that these were the only properties sold in a given period. The higher prices fetched by such houses would, if no allowance was made, cause the index price to rise despite the fact that the true price of two bedroom properties may not have changed. An RSI, based upon price increases for single properties, avoids this problem provided that, to use the same example, a two bedroom property within the dataset had not had a bedroom added to it prior to resale. The CSI is a technique for attempting to eliminate such properties from the data sample used. However, the issue is fully addressed only by hedonic analysis.

In exploring this point in more detail, it will be useful to consider the particular characteristics of a property in abstract. Conceptually, it is helpful to think of a house as a bundle of characteristics. These characteristics would comprise the particular features of a property such as the number of bedrooms, bathrooms and garages, the square footage, the location and the carbon footprint. Many of these features would change over time. As a result, the property quality characteristics within the index would change through time and price increases, resulting from an improvement in quality, will be confounded with price increases resulting from pure house price inflation. Furthermore, any list of characteristics, as well as their implicit values, should change over time. For example, the carbon footprint is a characteristic that is unlikely to be on any current list, but is something that may well appear on a future UK list, despite the current setback to government plans requiring all vendors to provide a Home Information Pack, including a report on energy use, for the use of potential purchasers.

Methods that address the above problems are based on hedonic models that attempt to price the dimension of any of several characteristics of a housing unit. Hedonic modelling thereby allows the effect of changes in the implicit values of qualities, reflecting changes in demand and supply, to be estimated. When considering property characteristics, changes in quantity and price should be considered. The CSI method abstracts from these changes by comparing like for like. The AcadHPI attempts to control for changes in the mix of types of houses sold (detached houses, terraced houses, apartments etc.). Hedonic models attempt to control for the changing characteristics and rely on detailed information on each housing unit – data that are not readily available. Finally, hybrid hedonic RSR methods exist but have not yet gained wide popularity.

3. THE CSI AND RSR TYPE INDICES

3.1. The basic argument in favour of using the CSI RSR type methodology is that housing quality and characteristics vary so substantially at the unit level that any index based on two consecutive samples of data is bound to produce biases in the estimated index. RSR attempts to separate price movements and quality changes; it is *not* a hedonic method. Rather, it simply restricts attention to houses that have sold at least twice within a given period. By looking only at houses that sell more than once, the index eliminates any price fluctuations due to changes in underlying housing characteristics, under an implicit assumption of characteristic constancy for the sold properties.

3.2. The CSI makes a number of adaptations to the original RSR model, as introduced by Bailey, Muth and Nourse (1963). These basically include purging outliers and data errors from the data sets. The CSI version of the RSR methodology also attempts to decrease standard errors; that is, to tighten the confidence intervals around any given point forecasts. We discuss such issues in section 5. Outliers include properties showing large price changes that may be associated with significant quality changes in the houses concerned. Such quality changes might include property improvement or deterioration or even a decline in neighborhood environment, all of which might affect a particular house price.

3.2.1. The RSR and CSI methodology in more detail.

3.2.1.1. Bailey, Muth and Nourse (1963) posited that the ratio of the sale price of a house in period *t* to that in period *t*-1 was equal to the true but unobservable indexes plus an idiosyncratic error term. Thus they wrote

$$R_{it,s} = \frac{B_{it}}{B_{is}} U_{it,s} \qquad (1)$$

where $R_{it,s}$ is the ratio of house price *i* in period *t* to period *s*, B_{it} and B_{is} are the unobservable indexes in period *t* and *s* and $U_{it,s}$ is an idiosyncratic error term. Taking the logarithm of both sides, we have the rate of return as a function of the price index of interest and an error term so that

$$r_{it,s} = b_{it} - b_{is} + u_{it,s} \qquad (2)$$

where the error term $u_{it,t-1}$ is assumed to be zero mean, have a constant variance across observations and over time and to be serially uncorrelated and uncorrelated across observations.

3.2.1.2. in order to recover the unobservable price indices for each year, an OLS regression is run where the dependent variable is the logarithm of the ratio of observed sale prices in period *t* and period *s* and the independent variables are time dummies for each period which the sample covers. For each observation, the time dummy for any particular period can take on one of three values. It takes on 1 if the house was sold in period *s*; -1 if the house was sold in period *t* (*t*< *s*); zero otherwise. From this model, the time dummies are interpretable as estimators of the logarithm of the sample price index and the difference between any two estimated period dummies is the approximate percentage change in the price index between these two periods.

3.2.1.3. the CSI modifies the RSR methodology in a number of ways. In the following, we refer to page references of the Standard & Poor's model description document. The CSI model is in price levels, not in rates of return, as in the method described above; this has the potential disadvantage that the estimated index could become negative, although perhaps this is not a concern in practice.

3.2.1.4. let $P_{i,t}$ be the price of house I at time *t*; let $B_{s,0}$ be the index at time *s*, evaluated relative to base year 0: it follows that

$$B_{t+k,t} = \frac{B_{t+k,0}}{B_{t,0}}$$

3.2.1.5. in order to understand the workings of CSI, we need to understand what the true model is; this is necessary because the CSI concept of truth is based on their notion of the true model. Our provisional true model will be given next, but this will be adapted as the story is told. Thus,

$$P_{i,t+k} = B_{t+k,t} P_{i,t} + \varepsilon_{i,t+k,t} \qquad (1)$$

3.2.1.6. this model says that every property price grows at the rate of the index plus noise and is not controversial. The CSI methodology rightly notes that the noise is heteroscedastic and that the variance is made up of two components, a constant variance plus k times another variance reflecting the period between sales - a sensible assumption.

However, what is actually fitted is the following - see page 21 of the S&P methodology paper,

$$0 = -\frac{P_{i,t}}{B_{t,0}} + \frac{P_{i,t+k}}{B_{t+k,0}} + \varepsilon_{i,t+k,t} \qquad (2)$$

For equation (2) to be valid, we need to divide equation (1) by $B_{t+k,0}$.

But this means that the true model should be the following;

$$\frac{P_{i,t+k}}{B_{t+k,0}} = \frac{P_{i,t}}{B_{t,0}} + \frac{\varepsilon_{i,t+k,t}}{B_{t+k,0}} \qquad (3)$$

3.2.1.7. this will have the following implication; it will change the variance of the noise term. Effectively it will divide variance by the sale period time index value squared. To avoid this unpalatable recognition, we need to adapt the true model in the following way;

$$P_{i,t+k} = B_{t+k,t} P_{i,t} + B_{t+k,0}\, \varepsilon_{i,t+k,t} \qquad (4)$$

3.2.1.8. equation (4) is now compatible with equation (2), so that we have a new true model. This true model now says that the variance of the noise increases with the square of the index timed at the point of sale. If we take it as given that current index values are higher, weighted least squares will underweight recent information. It is not clear that the weighted methods followed address this point even implicitly.

3.2.1.9. the methodology followed up to year 2000 consisted of simultaneous index estimation via instrumental variables (IV); this is all carefully explained in the S&P methodology paper. After 2000, a sequential IV procedure was used. This basically meant that current sales would not change past index estimates (see page 25); this is a sound practical step but it does introduce what is an inefficiency into the statistical calculations and means that the methods used in CSI are not an RSR as it is usually understood.

4. THE ACADATA HOUSE PRICE INDEX (AcadHPI)

4.1. The AcadHPI uses all of the monthly data from the Land Registry (LR) concerning the final transacted purchase prices of residential property sales in England and Wales, as recorded in the Land Register. Effectively, these are a census of property transactions. Neither the monthly data from LR, nor in consequence the AcadHPI, include any data concerning transactions in Scotland or in Northern Ireland.

The AcadHPI index construction involves four phases: mix adjustment to eliminate several biases from the raw data by re-weighting the raw price data to control for housing type and locality and seasonal adjustment; forecasting to address the problem posed by the timing of the release of the data; smoothing to minimise volatility in the monthly index values; updating to include the impact of the emerging LR data.

The AcadHPI index is not a hedonic index. Only limited characteristic information concerning properties sold is available so that the mix adjustment methodology is necessarily limited to those house types for which data are available. The AcadHPI relies upon using the entire dataset to minimise bias. Whilst a number of commentators have noted the effect of using only limited characteristics data, Mr Robert Wood, in his paper[2] for the Bank of International Settlements concerning UK house price indices concluded, from the results of an experimental index using the LR data developed by the Bank of England, that “even simple quality adjustment can have a large impact on measured rates of house price inflation”.

4.2. Mix adjustment: the AcadHPI mix adjustment method is to use all available data at a given point in time to yield a price index based on average house prices as follows:

[2] *“A comparison of UK residential house price indices”, BIS Papers No 21, Robert Wood, Bank of England*

4.2.1. let p_{ct}^{n} be the average price of house variety n in county[3] c (c = 1,…C) for month t (t is an element of the set of months 1,…,T where t =1 is December 1999. Let the latest month for which AcadHPI will produce an estimate be called *T.* For month *T* and all previous months, AcadHPI updates with the latest values released by LR.

4.2.1.1. we inject, here, a note on the timeliness of the LR data: an electronic system for reporting property sales to LR is still under development and the current system is subject to delay caused by long established manual procedures. Consequently, the availability of data from LR is subject to equivalent delay. In practical terms, this means that the LR data for the AcadHPI for a given month (say *T*), provided at the beginning of the following month (T + 1) include only a small proportion of the previous month's transactions. As an example, LR released data for January at the beginning of February and these would include few data on January transactions. We term these initial data (in our example, the data for January) the "immature" data – see 4.2.2. Most of the transactions which took place will not be reported until a further month has elapsed (in our example, a significant number of January transactions will not be reported until the beginning of March), T + 2). The reporting of additional transactions means that the average prices provided by LR for any given month are subject to revisions which can be significant until some three months have elapsed. The AcadHPI results are not termed "final" until a significant volume of LR data is available which is normally after three months have passed (note that AcadHPI uses the full dataset, as opposed to a sample). Rather than employ what are extremely low initial data volumes for *T*, AcadHPI initial results are a forecast which is explained below.

4.3. The AcadHPI provides average house prices for each and every month, being a weighted average of the average house prices within each housing variety:

4.3.1. property type: the housing stock data collected by LR fall into four categories. These comprise detached houses, semi-detached houses, terraced houses and flats (apartments). As stated in 2.3, using raw price information would lead to an inaccurate guide to general price trends since, in particular months, more houses of one variety may be sold than in another. Valid cross-period comparisons will depend upon the weight given to each type of housing being consistent over time. This AcadHPI mix adjustment methodology ensures a constant weighting.

Average house prices AC_{ct} for county c for all months prior to T are calculated as:

$$AC_{ct} = \sum_{n} p_{ct}^{n} \cdot \omega_{c}^{n}$$

[3] *our explanation refers to counties which are sub-divisions of regions; London is sub-divided into boroughs and London borough data are also used in AcadHPI mix adjustment*

where ω_c^n are county specific weights for each type of house *n* (types are: detached, semi-detached, terraced, and flats). The weights reflect historical market volumes in each type of house.

Since the current month's data are relatively sparse a special procedure is used to form an estimate of the current month average price.

Acadata calculates a 'revised average value' PR_{cT}^n for the present month for each county and house variety as follows**:**

$$PR_{cT}^n = \left\{ p_{cT}^n \cdot s_{cT}^n + \left[\sum_{k=1}^{3} \frac{p_{cT-k}^n \cdot s_{cT-k}^{nt}}{3} \right] \right\} \left(\frac{\sum_{k=1}^{3} s_{cT-k}^n}{3} + s_{cT}^n \right)^{-1}$$

where s is the number of sales of house variety *n.* This value highlights the use of both the current immature numbers for the current month released by LR and the use of previous months' information (as revised by LR in any latest release) to estimate the current average month price levels.

Average house prices AC_{cT} for county *c* in the current month *T* are then calculated as:

$$AC_{cT} = \sum_n PR_{cT}^n \cdot \omega_c^n$$

There is now a complete monthly time series of average house prices in each county of England and Wales back to December 1999.

4.3.2. location: in any given month, sales values in each county or London Borough may be abnormally low or high because of variations in the numbers of properties sold. The AcadHPI mix adjustment methodology controls for these fluctuations by ensuring that the total value of home sales in each county or London Borough does not deviate from its average weight amongst the overall sales, despite the inevitable changes in sales volumes on a month to month basis. This is made explicit below but involves using historical weights to form weighted average values for each region (based on county weights where a region comprises several counties) and for the nation as a whole based on regional weights.

4.3.3. Seasonal adjustment: in the UK, transaction volumes and prices regularly rise in the spring and decline in the winter. Seasonal adjustment is designed to eliminate this component from the data so that trends are not confounded with normal seasonal highs and lows.

Seasonally adjusted average house prices $SAAC_{ct}$ for county c for all months are then calculated as

$$SAAC_{ct} = AC_{ct} w_c^m$$

where w_c^m are month and county specific seasonal adjustment factors reflecting historical seasonal volumes for the UK.

4.4. Forecasting: at the national level, some 100,000 actual transactions occur monthly. Of these, only some 15% are promptly reported to LR for the current month – the "immature" data - see 4.2.1.1. Rather than rely upon a small sample of circa 15,000 transactions – one which was found to be highly unrepresentative of the final data – the AcadHPI provides a forecast.

4.4.1. Rough Forecasts: the first step in attaining a measure of the current month house price index and a forecast for the coming months is to generate three rough forecast numbers: two for the current month t and one for the following month t + 1. These are called the revised average price ($RSAAC_T$), the rough forecast ($RFSAAC_T$) and the forecast for the next month ($RFSAAC_{T+1}$). These are used in the construction of the actual index; they are not the final index numbers.

The revised average price for month T is defined as

$$RSAAC_{cT} = \frac{SAAC_{cT-1} + SAAC_{cT}}{2}$$

This is simply the arithmetic average of the previous month's average house price and the current month's average house price.

The rough forecast is defined as

$$RFSAAC_{cT} = \frac{SAAC_{cT-1} \cdot \overline{\Phi}_{cT-1} + RSAAC_{cT}}{2}$$

where $\overline{\Phi}_{cT-1} = \sum_{i=1}^{5} \frac{SAAC_{cT-i}}{SAAC_{cT-i-1}} \cdot \frac{1}{5}$ is the arithmetic average of the price change factors over the preceding five months. In essence *RFSAAC* is an average of two numbers. The first number is a prediction of the average price this month based on a 5-month moving average of price increases multiplied by the previous month's price. The second number is the average price based on the (seasonally adjusted) revised average value data for this month and last month's average price.

Finally, the rough forecast for the next month is calculated in a similar way as

$$RFSAAC_{cT+1} = \left(\overline{\Phi}_{cT}\right) RFSAAC_{cT}$$

where $$\overline{\Phi}_{cT} = \frac{1}{5}\left[\frac{RFSAAC_{cT}}{SAAC_{cT-1}} + \sum_{i=1}^{4}\frac{SAAC_{cT-i}}{SAAC_{cTt-i-1}}\right]$$

4.4.2. Index of Indices Model: the AcadHPI next employs an academic "index of indices" forecasting model, developed at the University of Cambridge and the Sir John Cass Business School. This is used in conjunction with the transformed price data (as per above) to achieve the reported AcadHPI index number.

4.4.2.1. Acadata "forecast" index results: the AcadHPI academic forecasting model relies upon three indices. Two of these are produced by leading UK mortgage lenders. These are the hedonically adjusted Halifax House Price Index and the Nationwide House Price Index – the former is part of the HBOS bank, and the latter is the largest UK Building Society (Savings and Loan). These indices are sample based and are individually unrepresentative; however, they provide agreed prices at loan offer stage for each month at the month end (Nationwide) and a few days later (Halifax). They have been published regularly on a monthly basis since 1952 and 1984 respectively and, as such, have gained a reputation as the UK's leading indicators. The third constituent is the mix adjusted CLG House Price Index, published by Communities and Local Government (ex-the Office of the Deputy Prime Minister) since September 2003 as an experimental government statistic. The CLG index provides prices agreed at loan completion stage and is published in the second month after the month end. The academic model uses all three indices to forecast the LR result for the month concerned. The methodology used is based on optimal forecast combination theory as follows:

4.4.2.1.1. the model takes, as the "truth", the LR average house price and computes forecast errors with respect to the three indices. A history of forecast errors is assembled and these errors are used to estimate a forecast error covariance matrix; denoted by Ω. It then considers a portfolios of weights, denoted by *w* and chooses *w* to have elements that add to 1; the variance of the portfolio forecast error is minimised, given by $w'\Omega w$.This constitutes the optimal portfolio; the weights can be negative and Bayesian priors have been used in development. The weights give some sense of the relative efficacy of the pertinent index.

4.4.2.1.2. the academic model continuously re-weights the lender and CLG indices monthly, according to how well their results accorded with those published, much later, by LR. The resulting forecasts were proved, over a three year development period, to foreshadow the LR results extremely well and to resolve the wide variations in the results provided by other UK indices.

4.4.3. These weights and indexes are then used to produce two versions of forecasts for house price inflation both for the current month and for the next month. One is based on a five-period moving average forecast of CLG and the other is based on an AR process forecast of CLG (note this is required because of delay in producing CLG figures). The simple average of these two forecasts for price changes is used with the national average house price forecasts to generate 'forecast' average house prices. The average forecast of house price inflation is denoted Θ. It is expressed as percentage growth x 100 so that one percent inflation would give rise to Θ = 1.

Formally we have the an intermediary forecast for average national house prices in month *T* and *T* + 1

$$FAC_T = \frac{1}{2}\left[\sum_{r=1}^{R} v_r \sum_{c=1}^{C} \left(RFSAAC_{cT}\right)\alpha_c\right] + \frac{1}{2}\left\{\frac{\left[\sum_{r=1}^{R} v_r \sum_{c=1}^{C} \left(SAAC_{cT-1}\right)\alpha_c\right]\left[100+\Theta_T\right]}{100}\right\}$$

$$FAC_{T+1} = \frac{1}{2}\left\{FAC_T\left[\frac{\sum_{r=1}^{R} v_r \sum_{c=1}^{C} \left(RFSAAC_{cT+1}\right)\alpha_c}{\sum_{r=1}^{R} v_r \sum_{c=1}^{C} \left(RFSAAC_{cT}\right)\alpha_c}\right]\right\} + \frac{1}{2}\left\{\frac{FAC_T\left[100+\Theta_{T+1}\right]}{100}\right\}$$

where *r* identifies the regions or collections of counties and ($r = 1,\ldots,R$), α_c. is a county weight and υ_r is a regional weight so that

$$\sum_{r=1}^{R} v_r \sum_{c=1}^{C} \left(RFSAAC_{cT}\right)\alpha_c$$

is the national (weighted) average of house prices based on the rough forecasts.

FAC_T yields a forecast price at the national level. It is simply the average of two numbers. The first is the average house price at the national level predicted by 'rough' forecast methods which in turn came from trends in the recent data. The second part of the average is the previous month's data times the growth factor calculated from the academic forecasting model. The goal now is to say something about house price inflation.

To this end, we now obtain a price inflation factor, Π, for month T *and* $T + 1$ given by

$$\Pi_T = \frac{FAC_T}{\sum_{r=1}^{R} \nu_r \sum_{c=1}^{C} (RFSAAC_{cT}) \alpha_c}$$

$$\Pi_{T+1} = \frac{FAC_{T+1}}{\sum_{r=1}^{R} \nu_r \sum_{c=1}^{C} (RFSAAC_{cT+1}) \alpha_c}$$

These factors take the ratio of forecast average house prices and the rough forecast of house price inflation based on previous periods' actual house price inflation. Note that, in generating the factor for the current month, an average of recent past house price inflation and predicted house price inflation is also used. These factors are also multiplied by the seasonally adjusted rough forecasts for month t and $t+1$ and a weighted average over all counties within each region is taken to obtain forecasts of average house prices for each county.

4.4.4. In arriving at a final national forecast these inflation factors are used on the seasonally adjusted data at the county and then the regional level. For county c we obtain average forecast house prices for T and $T + 1$ as

$$FTHPI\_F_{cT} = \Pi_T \cdot RFSAAC_{cT}$$

$$FTHPI\_F_{cT+1} = \Pi_{T+1} \cdot RFSAAC_{cT+1}$$

For region r we obtain average forecast house prices for T and $T + 1$ as

$$FTHPI\_F_{rT} = \Pi_T \sum_{c=1}^{C} (RFSAAC_{cT}) \alpha_c$$

$$FTHPI\_F_{rT+1} = \Pi_{T+1} \sum_{c=1}^{C} (RFSAAC_{cT+1}) \alpha_c$$

4.4.5. The national average house prices are defined in a similar way and turn out to be equal to *FAC*, *the* intermediary forecast for average national house prices

$$FTHPI\_F_T = \Pi_T \sum_{r=1}^{R} \nu_r \sum_{c=1}^{C} \left(RFSAAC_{cT}\right)\alpha_c = FAC_T$$

$$FTHPI\_F_{T+1} = \Pi_{T+1} \sum_{r=1}^{R} \nu_r \sum_{c=1}^{C} \left(RFSAAC_{cT+1}\right)\alpha_c = FAC_{T+1}$$

4.5. Smoothing: AcadHPI house prices are smoothed, in order to minimise volatility. Each average house price is calculated as the average of the adjusted average house price for each of three months, ascribed to the centre month.

For any month prior to *T*-2 a simple average along the following lines is used to calculate smoothed average house prices

$$SMAC_{ct} = \sum_{i=t-1}^{t+1} SAAC_{ci}$$

These smoothed numbers are averaged at the county level and then at the regional level as follows

$$SMAC_{rt} = \sum_{c=1}^{C} \alpha_c SMAC_{ct}$$

$$SMAC_t = \sum_{r=1}^{R} \nu_r SMAC_{ct}$$

$SMAC_t$ defines the smoothed average house price at the national level for any month *T* – 2 and beforehand.

4.5.1. the Acadata "forecast" index house price for month *T* is an average of three prices, two of which are forecast. The three average house prices comprise those for:

- the past month, based upon circa 70,000 (see 4.6) transactions
- the current month, as forecast
- the next month, as forecast

For the current month T the smoothed forecast average house price at the national level is obtained as:

$$SMFAC_T = \frac{1}{3}\left[\sum_{r=1}^{R} v_r \sum_{c=1}^{C}\left(SAAC_{cT-1}\right)\alpha_c + \Pi_t\left(\sum_{r=1}^{R} v_r \sum_{c=1}^{C}\left(RFSAAC_{cT}\right)\alpha_c\right) + \Pi_{T+1}\left(\sum_{r=1}^{R} v_r \sum_{c=1}^{C}\left(RFSAAC_{cT+1}\right)\alpha_c\right)\right]$$

We also note that the county level and regional level smoothed forecasts are available based on a similar procedure. For example, the smoothed forecast for period T in county c is given by

$$SMFAC_{cT} = \frac{1}{3}\left[SAAC_{cT-1} + \Pi_T RFSAAC_{cT} + \Pi_{T+1} RFSAAC_{cT+1}\right]$$

and the regional smoothed forecast is given by

$$SMFAC_{rT} = \frac{1}{3}\left[\sum_{c=1}^{C}\left(SAAC_{cT-1}\right)\alpha_c + \Pi_t \sum_{c=1}^{C}\left(RFSAAC_{cT}\right)\alpha_c + \Pi_{T+1}\sum_{c=1}^{C}\left(RFSAAC_{cT+1}\right)\alpha_c\right]$$

4.5.2. similarly, the forecast for month T-1 is an average of

- the month preceding the past month, based on circa 90,000 (see 4.6) transactions
- the past month based on circa 70,000 (see 4.6) transactions
- the current month as a forecast

$$SMFAC_{T-1} = \frac{1}{3}\left[\sum_{r=1}^{R} v_r \sum_{c=1}^{C}\left(SAAC_{cT-2}\right)\alpha_c + \sum_{r=1}^{R} v_r \sum_{c=1}^{C}\left(SAAC_{cT-1}\right)\alpha_c + \Pi_T\left(\sum_{r=1}^{R} v_r \sum_{c=1}^{C}\left(RFSAAC_{cT}\right)\alpha_c\right)\right]$$

Again, county and regional updated indexes are calculated in a similar fashion to those defined above.

4.6. Updating: one month after any given month, LR provides average house prices based upon circa 70% of the eventual total transactions, which are used to replace the Acadata “forecast” result with an Acadata "updated" result. A further month later, LR provides prices based upon circa 90% of the eventual total transactions which are used to replace the first with a second Acadata "updated" result. Three months after any given month, LR provides prices based upon circa 95% of the total transactions for the month. Thus, assuming an average 100,000 monthly transactions in England and Wales, the number available for any given month for the AcadHPI rises to some 70,000, 90,000 and 95,000 in the months following that in which the transactions occurred. Taking the current month as month T, in month T + 4 the AcadHPI results are regarded as sufficiently updated to be described as the Acadata "final" index.

4.6.1. The updates are sequential. The Acadata "forecast" index result is entirely replaced by the first Acadata "updated" index result which, in turn, is entirely replaced by the second Acadata "updated" index result, which is entirely replaced by the Acadata "final" index result, which is entirely replaced every month, thereafter, until any further transaction data received by the Land Registry make no difference to the average house prices which they provide. It may take up to fifteen months before the Land Registry data finally settle, enabling the Land Registry and Acadata index results to reach "ultimate". However, as noted above, Acadata "final" index results are normally only subject to marginal change thereafter.

5. MEASURING ACCURACY IN AN INDEX

In this section we consider issues concerning index accuracy. There are at least two separate notions of what this could mean.

5.1. The first approach is to use some observable measure of the true house price index; this is the approach taken by Acadata where we take as the true value the average house price for all sales in the period. Then we can measure accuracy by deviations from that quantity. Potentially, many different loss functions could be used to assess accuracy. In our work, we have used the variance of the forecast error or the mean square error. Under such a framework, we can assess accuracy by computing this value. This is really an ex-post analysis, as it compares the index under analysis with what it purports to be measuring at points in time after the event. By inspection of such calculations through time we can get a sense of what a 95% confidence interval might be. This procedure is non-parametric and requires no formal model of the index process.

5.2. The second approach is model based, where the unknown parameters of the model are the index, as in the CSI RSR approach, or constituents of the index as in a hedonic model. Because we have statistical estimates of these unknown parameters, we are able to compute, exactly or approximately, the distribution of the forecast error and this can used to compute confidence intervals. Mathematically, we can write,

$$y_t = x_t' \beta + e_t$$

where β is the true index value, but we know only the estimate, in the case of RSR. However, the theory of regression tells us its distribution about the true unknown value β. This will allow us to construct a confidence interval. Other versions of this approach, such as hedonic models, take as the index the estimated value of the left-hand side of the regression,

$$\hat{y}_t = x_t'\hat{\beta},$$

where $\hat{\beta}$ are the estimated regression coefficients. Again, using their distribution, one can get a confidence interval for the index. Yet another procedure is applied when the index is a non-linear function of $\hat{y}_t = x_t'\hat{\beta}$. Here, it is necessary to approximate the index distribution by the delta method. This last situation arises in log-linear hedonic models.

5.3. The CSI and other RSR methods use regression to estimate their indices and standard techniques to improve the reliability of the accuracy (i.e. to reduce any biases in the standard errors of their estimators arising due to heteroscedasticity). The standard error of a particular index estimate is reduced through these corrections; the index will be more accurate and its reliability may be improved such that the confidence that a user may have in any particular figure generated from the methodology would be enhanced. Nevertheless, the particular corrections (such as feasible generalised least squares of weighted least squares) used to ameliorate the biases in the standard errors depend upon parametric assumptions. If these assumptions are invalid, the bias in the standard errors may not be eliminated and hence the reliability of the index may remain lower than its potential. Again, most of the above adaptions are *NOT* about changing the inferred estimated changes but are concerned only with the standard errors or confidence bands around these 'point estimates'.

5.4. In conclusion, RSR/CSI is better equipped to deal with notions of accuracy than is AcadHPI as its notion of the truth is the model itself. However, this point is weakened by ambiguity about what the true model might be: is it equation (1) see 3.2.1.5 or equation (4) see 3.2.1.7? The AcadHPI notion of truth is the LR average sales appropriate to the period. Neither necessarily corresponds to what a user may deem to be true but the former is more amenable to analytic analysis.

6. DEMERITS OF RSR AND MIX ADJUSTMENT METHODOLOGIES

6.1. In general, use of more data is to be preferred to the use of less data**:** RSR eliminates many data points. Methods that combine RSR and hedonic or mix-adjustment methods have been developed by Case and Quigley (1991).

6.2. If RSR is employed to develop indices at local level, it may result in highly volatile results due to the fact that a relatively small number of observations will be available in certain markets, such as the UK. AcadHPI is reliable, in that it has more data, at local levels as well as at the national level, precisely because it does not eliminate data. However, those devising the methodologies used by CSI are well aware of this point so that the calculations are carried out at a relatively high level.

6.3. Controlling for quality. Neither mix adjustment as used in AcadHPI, nor RSR, fully control for quality. AcadHPI relies upon use of all of the data and contains no quality-specific information, as to e.g. the value of additional bedrooms, bathrooms or garages. Neither does RSR completely account for quality. It is claimed for RSR that, by using data for the same properties, over time, price movements are separated from those resulting from quality changes. However, the micro data necessary are generally not available to substantiate this claim. It is likely that houses that undergo renovation will sell more quickly and that RSR indices will be too heavy in these types of properties. Other properties that take a long time to sell may have poor quality. The CSI attempts to control for these issues by down weighting observations that take longer to sell and by discarding observations that show abnormally large or small changes in value. Other exclusions include sales that occur within six months, non-arms length sales (i.e. between relatives), new properties, condominiums and sales by developers. These exclusions should improve matters but some may be adding noise to the index. In the UK, which is likely to experience huge increases in the sale of new properties in the next 10 years, exclusion of new property will be very distorting.

6.4. Accuracy. Neither method gives a perfect notion of accuracy; both are dependent upon some notion of truth. The greater analytic structure of RSR makes it more amenable to analysis but, as is well-known, if that structure is mis-specified then concepts of accuracy break down. The CSI methodology needs a clearer discussion of what their true model is. The AcadHPI has the benefit of being non-parametric and, although less amenable to the quantification of accuracy, it is likely to be the more robust.

6.5. Revisions. AcadHPI is subject to revision: however, revisions are normally small after three months have elapsed. Any RSR based index is subject to constant revision because most houses will eventually sell twice and, hence, become available for inclusion in the data set. RSR estimates all its time coefficients jointly and it can happen that a spate of sales of properties, bought 20 years ago, can influence values of the index long-since passed. For example, there can be no final values of the index for the year 1986 until every property transacted in 1986 has been re-transacted. This could be up to 50 years later. Clearly, this is a possible defect and CSI have dealt with this by their post 2000 estimation method, see discussion in section 3.2.1.9. This resolution is at the cost of statistical efficiency, that is to say, we discard the new information.

Moreover, if the repeat sales rate changes over time, this could lead to periods of substantial bias in the RSR index. This could happen if, during housing booms, houses are more likely to be renovated for market placement. Of course, the same criticism can be levelled at the AcadHPI; thus, in relative terms, it is not clear whether repeat sales are more variable than total sales. However it is likely that the former is the more variable and this could lead to an increase in the fraction of properties within the underlying data that have actually undergone physical changes during such periods and would lead to *overstatement* of a constant quality index.

FOOTNOTE: Dr Meissner was Fellow, Kings College, Cambridge and is now Professor, UC Davis, California. Dr Satchell is Economics Fellow, Trinity College, Cambridge.

ABOUT ACADATA

Acadata is the new name for Acadametrics, an analytics and research consultancy focusing on house prices and property portfolio risk, and with a 23 year co-operation with Dr Stephen Satchell, Economics Fellow at Trinity College, University of Cambridge. We are expert in the measurement and analysis of house prices. Our FTHPI, launched in 2003 by the Financial Times, pioneered the use of Land Registry data in a mainstream house price index. Following a 2010 sponsorship agreement with LSL Property Services PLC, FTHPI was published as LSL Acad E&W HPI, retaining full independence and with a monthly commentary by Dr Peter Williams. Our LSL Acad Scotland HPI was launched in 2011. As FTHPI, the index was chosen by the Chicago Mercantile Exchange for a possible future residential house price derivative, put on hold as a result of the financial crisis.

In addition to our valued work for LSL, we provide data to other significant parties in the housing sector. For example, Hearthstone PLC uses the LSL Acad E&W HPI and LSL Acad Scotland HPI as a benchmark against which to monitor the price performance of their residential property investments.

In 2009, Acadametrics and New York based MIAC Analytics joined forces to work on risk solutions, forming the top-flight consultancy MIAC Acadametrics Ltd (M|A). In October 2013, Acadametrics accepted an offer by MIAC Analytics to acquire the whole of M|A for which Dr Satchell will continue as consultant. Acadata will focus on house price indices and data, with Dr Satchell advising as necessary. For all risk-related work, including stress and scenario testing, collateral valuation and forecasting, please see **MIAC | Acadametrics** (M|A).

In addition to house price indices, Acadata provides the Acadata Prices and Transactions (APAT LGA) data showing property type prices for Local Government Areas from 1995, using Land Registry data* for England & Wales counties, unitary authorities and London boroughs, as well as data from 2003 for local authorities in Scotland*. APAT LGA includes an interactive chart facility.

Acadata also provides APAT POSTCODE data. These comprise average prices plus transactions for postcode districts and optional data for postcode sectors, towns, streets or defined areas of interest to a client. These are used for example by:

- developers considering a residential investment in a particular post code sector
- house builders needing to understand the price and transaction trends in a postcode district or sector
- estate agents considering opening a further branch and wanting to know how much residential property business has been done in the district over a long term period
- branch comparisons against local trends

We prepare indices for third parties. We also forecast house prices, in conjunction with M|A.

Our work has a strong academic foundation and our solutions are developed using our own resources under our "research first" policy. Further detail is provided on our website www.acadata.co.uk.