# Controlling for Transactions Bias in Regional House Price Indices

Gwilym Pryce and Philip Mason

Department of Urban Studies
University of Glasgow
Glasgow G12 8RS
Email: g@gpryce.com

Working Paper as at 23 June 2006

**Presented at the Conference in Honour of Patric Hendershott**
*(Columbus Ohio, July 2006)*

**Abstract:**

Transactions bias arises because properties that trade are not a random sample of the total housing stock. Price indices are potentially susceptible to this bias because they are typically based on transactions data. Existing approaches to this problem have relied on Heckman-type correction methods where a probit regression is used to capture the differences between properties that sell and those that do not sell in a given period. However, this approach can only be applied where there is reliable data on the whole housing stock. In many countries – such as the UK – no such data exists and so there is little prospect of correcting for transactions bias in any of the regularly updated mainstream house price indices. This paper offers an alternative approach based on information at postcode sector level. The probability of a property transacting is modeled by applying fractional logit regression (FLR) to the proportion of properties that sell in each postcode sector. Transactions data on 1.4 million house sales distributed across 1,200 post code sectors in the South East of England over the period 1996 to 2003 are used to create a correction term for in a simple monthly hedonic house price regression. Corrected and uncorrected price indices are compared.

**Acknowledgements:**

This paper is based on research funded by the UK Department of Communities and Local Government (formerly the Office of the Deputy Prime Minister). We are grateful also to Experian for the generous provision of postcode level data. Neither DCLG nor Experian necessarily agree with our findings.

# Controlling for Transactions Bias in Regional House Price Indices

## Introduction

The existence of sample selection bias in the computation of house price indices is widely acknowledged. Papers that have attempted to correct for this bias, either in repeat sales indices (Gatzlaff and Haurin (1997)) or in hedonic indices (Gatzlaff and Haurin (1998); Hwang and Quigley (2002)) have been published in leading real estate journals are cited frequently[1]. Yet the techniques used in these papers have not been adopted in the mainstream hedonic literature, nor have they changed the way that institutions calculate house price indices. In fact, to our knowledge, there no published attempts at correcting for selection bias in house price indices other than in these three papers, and certainly none that use data from the UK or mainland Europe (with the exception of Sweden). Given that there are vast volumes of hedonic indices published each year and that GHHQ (Gatzlaff, Haurin, Hwang and Quigley) have established the need for selection bias correction, why is there an apparent discrepancy between the demand for and supply of corrected indices?

Unfortunately, the methods used by GHHQ require data are not readily available in most countries. In particular, the selection equation used in these studies to estimate the probability that a property will enter the market requires detailed information not only on the properties that sell but also on those that don't. Our goal in this paper is to develop a method that: (i) can be applied to data that is readily available for all regions in the UK; (ii) can be easily updated; and (iii) is sufficiently straightforward for publishers of official statistics to feasibly adopt as an element of their regular house price index updates.

[1] A search on the Social Science Citation index in June 2006 revealed that these three papers had been cited a total of 47 times.

## UK Price Indices

There are currently eight main providers of house price indices in the UK: Land Registry, Department of Communities and Local Government (DCLG), Nationwide, HBOS/Halifax, Financial Times, Royal Institution of Chartered Surveyors (RICS), Hometrack, and Rightmove.

The Land Registry/Registers of Scotland survey comes out once every three months and is based on the records of all property transactions registered. Since it is a legal requirement for property transactions to be logged with Land Registry, the data should be comprehensive (though certain transactions, such as repossessions and property transfers following a divorce, are usually omitted to avoid misleading results). As a measure of the value of *traded* properties, there is unlikely to be any major sampling bias associated with this index. Only very basic details on properties are recorded (none in Scotland and only house type in England and Wales) and so there is limited scope for mix adjustment.

The DCLG index is based on mortgage origination data from around fifty lenders, collected through the Survey of Mortgage Lenders. Until two years ago, this survey was only a 5% sample of the transactions of those lenders, but this has recently been increased to include nearly all mortgage transactions. Unlike the Land Registry data, this index does not contain information on cash purchases, which account for about a quarter of the market, and so there is potentially a source of sampling bias even as a measure of traded properties. Mortgage origination data typically provides information on the type of dwelling, number of rooms, whether there is a garage etc. and this means that a mix adjusted version of the index is now provided.

The Nationwide and HBOS/Halifax are both major mortgage lenders in the UK and their indices are based on mortgage origination data from their own loan book records. Unlike the Land Registry data, these indices do not contain information on cash purchases, or on mortgage transactions through other lenders. The samples used are therefore potentially biased by variations in the market share of the two lenders across different areas and over time. Both indices use a form of hedonic adjustment to correct for variations in the type of properties traded over time (see Meen and Andrew, 1998, p. 10) but there is no correction for sample selection bias.

The RICS draws on the responses of three hundred surveyors and estate agents in England & Wales who are asked whether they feel prices are falling or rising, along with a number of other questions including whether the number of buyers and sellers rising or falling. The information collected by RICS therefore reflects confidence in the housing market of key market agents, rather than an analysis of actual changes to recorded prices. The results are potentially biased by possible discrepancies between perceptions and reality, and by the possible incentives of respondents (to "talk-up" the market, or play down overheating for fear of interest rate rises). There is no formal adjustment for attribute variation or sample selection bias.

Similar to the RICS, Hometrack base their results on a survey of market agents, but employs a much larger sample. Around 3,500 estate agent offices from all 2,200 postcode districts in England and Wales report whether prices are rising or falling. Again, the results are potentially biased by possible discrepancies between perceptions and reality, and by the possible incentives of respondents (to "talk-up" the market, or play down overheating for fear of interest rate rises). Again, no formal correction is made for variations in the mix of properties that sell or sample selection bias.

Rightmove use information asking prices reported on the Rightmove website over the previous month which they claim represents around 35% of all homes for sale. However, only asking prices are reported, and it is possible that bias could emerge due to variations in the difference between asking and selling prices across areas or over time (see Pryce 2004). There is no correction for mix adjustment or sample selection bias.

The Financial Times house price index is a composite index computed by Acadametrics. It is based on an attempt to combine the Nationwide, HBOS/Halifax and ODPM house price indices to Land Registry records and creates a composite index that attempts to correct for the bias in three component indices. The FT approach is founded on the assumption that LR data is unbiased. The mix adjustment calculation is complex, given that this is an amalgam of indices that have already been mix adjusted. The results are based on “a statistical analysis of the performance of the Nationwide, HBOS/Halifax and ODPM house price indices in respect to any bias (e.g. systematic over or under measurement) or inaccuracy (variation) in measurement of actual house price growth rates as published by HM Land Registry.” FT say that they, “performed recursive analysis of data samples to calculate error, in a number of different ways. [They] next examined the extent to which each individual index contributed to a combined index superior to the individual indices. [They] formed a portfolio of measurement error growth rates and estimated weights to uncover the relative contribution of each index to the construction of a combined index. [They] formed optimal portfolios which are either unbiased or show minimum variance and verified the results"[2]. There is no correction for sample selection bias, however.

These indices, and the methods used to derive them, are important because they are used in a wide spectrum of economic and policy decisions. They are used in planning decisions and in the current debate over demand and supply imbalances at the intra and inter regional level (see for example, the Barker Review of Housing Supply commissioned by Her Majesties Treasury in 2004). They are used in the measurement of affordability and wealth inequalities, the assessment of the impact of new supply, and macro modelling of the relationship between house prices, interest rates and consumer spending. The meaning and reliability of the indices used in each of these respective fields is therefore potentially crucial to the functioning of the market and to efficient policy responses. Distortions in published indices, or confusion over their meaning, could significantly affect personal financial decisions, investment choices, planning and policy

All of these indices are based only on properties that trade (or at least are offered for sale); none of them attempt to correct for biases that result from the fact that traded properties may not be a random selection of all properties. And no caveats are given that these indices may not give an accurate picture of the price appreciation of the housing stock. Of course, for some decisions, this does not matter – estate agents and

[2] www.acadametrics.co.uk

lenders, for example, may only be interested in the price trends of properties that actually sell. In other contexts, particularly the measurement of housing wealth, the potential for equity withdrawal, the impact of intergenerational bequests, it is the value of the entire stock of private housing that is of interest and so there is a prerogative to find ways of measuring and correcting for transactions bias.

## Existence of Transactions Bias in the UK: The Evidence So Far

There are currently no published UK house price studies that correct for transactions bias. There has been some work, however, on the intention to move and on the frequency of sale and the pattern of house price and dwelling characteristics. Pryce (2004) considered the number of properties in each West of Scotland local authority that sold once, twice, three times, four times or five or more times in the 1991 to 2000 period and found evidence of variation in repeat sales even within the West of Scotland. In the City of Glasgow, for example, nearly 30% of properties transacted sold twice, and 10% sold three times. This contrasts with Argyll and Bute where less than 18% sold twice and only 3.6% sold three times. Overall, 63.3% of properties that sold came on the market only once, 25.9% sold twice, 8.4% sold three times, 1.9% sold four times and 0.5% sold five or more times. The second question considered by Pryce (2004) was whether there were different patterns of house price values for different rates of turnover. He found that mean house price tended to be lower for properties that sold frequently.

Kim et al (2005) investigate the impacts of current dwelling, household characteristics and alternative properties on the probability of moving in Oxfordshire, UK. They model to the intention to move by applying a nested logit estimation to stated preference data and find that dwelling characteristics and location factors (such as school performance, density and transport) have a significant role. For example, 'Residents in the most dense neighbourhoods have a 24.1 per cent higher propensity to move than residents in neighbourhoods with the lowest density' and 'the probability of moving decreases by 12.1 per cent with a one stage improvement of school quality'.

Of course, the desire to move is only half the story. In a private housing market, *actual* moves only tend to occur once the owner has found a buyer. So there is a second set of choices, not modelled by Kim et al, with regard to the desire to move *into* particular dwelling types and locations. Some of the variables that affect the desire to sell may have the opposite impact on the desire to buy. For example, school performance reduces the desire to sell in a given neighbourhood, but is likely to increase the desire to buy properties in that area.

## Methods of Correcting Bias in House Price Indices

Gatzlaff and Haurin (1998) argue that 'house value indices derived from the conventional hedonic method are subject to bias if the sample of houses is not a random sample of the stock'. They conclude that, 'Correction requires joint estimation of the probability that a house will sell and the sale price' (Gatzlaff and Haurin, 1998, p.199; see also Quan 1993 and Hwang and Quigley 2004). This form of joint estimation follows in the time honoured tradition of viewing the sample selection problem as one of omitted variable bias (Heckman, 1979) where the omitted variable in the house price equation is the probability of the property coming onto the market. Gatzlaff and Haurin (1998) take this view and use probit regression to estimate the probability of a property coming onto the market. This estimated probability[3] can then be entered into the sale price equation to correct for sample selection bias.

Hwang and Quigley's (2004) study computed the average values of each dwelling characteristic in their sample of traded dwellings according to whether the property had sold once, twice, etc. over the period under consideration. Although the Hwang and Quigley approach is useful, it does not consider the attributes of properties that did not trade at all, neither does it consider the possibility that dwelling types might be clustered across space (their analysis is of individual property characteristics rather than the typical attributes of the neighbourhood). Also, it is not evident from any of

[3] More precisely, the inverse Mills ratio is calculated.

the GHHQ studies how the relationship between probability of sale and dwelling attributes varies over time.

There is a further drawback of the GHHQ research. As noted in the introduction, estimation of the selection equation using probit analysis of whether each dwelling in the housing stock has sold in a given period, will be problematic in most countries. This is because attribute information is not usually available for the whole housing stock at individual dwelling level. Information is, however, often available at neighbourhood or postcode sector level. UK Census information, for example, can provide useful information for relatively small spatial units but not for individual addresses, for confidentiality reasons. Sales data are typically released without full address details for each transaction, but usually the include postcode sector. Since we usually know the total number of dwellings in each postcode sector (from the UK Postal Address File, for example) it is possible to work out the proportion of the housing stock that trades in a given period. By combining this information it is feasible, in principle, to estimate the probability of a property in a given postcode sector selling in a particular time period.

If we use the proportion of sales in each postcode sector as our dependent variable cannot be modelled using probit or logit, however, because it is not a dichotomous variable. OLS is not appropriate either because proportions are bounded at zero and one and so OLS could predict outside of the feasible range. Whilst it is unlikely that there will be much difference between OLS and FLR prediction in the dataset used in our study (because the distribution of proportions lies well within the upper and lower bounds – in all the postcode sectors considered, only between 1% and 3% of properties trade in any one year) in attempting to establish a general procedure that could be applied to a wide variety of datasets, a method needs to be found that would not be vulnerable to this problem.

One solution to the problem of modelling variables bounded between zero and one is to apply the log-odds transformation to the dependent variable ($\log[y/(1-y)]$) which allows OLS to be applied to the estimation of $\mathbf{x}\boldsymbol{\beta}$. According to Wooldridge (2002) this approach has two major drawbacks, however:

> "First, it cannot be used directly if $y$ takes on the boundary values, zero and one. While we can always use adjustments for the boundary values, such adjustments are necessarily arbitrary. Second, even if $y$ is strictly inside the unit interval, $\beta$ is difficult to interpret: without further assumptions, it is not possible to recover an estimate of $E(y|\mathbf{x})$, and with further assumptions, it is still nontrivial to estimate $E(y|\mathbf{x})$." (Wooldridge, 2002, p.662).

Papke and Wooldridge (1996) and Wooldridge (2002) suggest modelling $E(y|\mathbf{x})$ as a logistic function:

$$E(y \mid \mathbf{x}) = \exp(\mathbf{x}\boldsymbol{\beta}) / [1 + \exp(\mathbf{x}\boldsymbol{\beta})]$$

which ensures that "predicted values for $y$ are in (0,1) and that the effect of any $x_i$ on $E(y|\mathbf{x})$ diminishes as $\mathbf{x} \to \infty$." (Wooldridge, 2002, p.662). The technique, labelled Fractional Logit Regression (FLR), has recently been used in the real estate literature by Hendershott and Pryce (2006) to model loan-to-value ratios which are usually bounded between zero and one. The approach also offers an apt solution to modelling the proportion of properties that sell and is the method used here to estimate the probability of non-selection (in the FLR regressions we actually use the probability of selection as the dependent variable in order to simplify interpretation of coefficients; the predicted probability is then subtracted from one to derive the probability of non-selection in each postcode sector for each year).

Note that both the Postal Address File and the Land Registry records of transactions include properties that are owned by social landlords. Tenants of municipal housing in the UK have the 'Right to Buy' which means that such dwellings can potentially enter the set of dwellings that transact. Public ownership of a property is likely to reduce the probability of sale, partly because of the beurocracy associated with privatisation of a public asset, and partly because of the limited demand for housing that is often aesthetically unappealing and often situated in deprived areas. Whether one wants to screen out such properties from the calculation of house price indices depends on whether one wants to value the entire housing stock (public and private), or just private housing. In this paper we assume the latter, so we use information on the proportion of social renting in each area to predict sale probabilities as though the stock comprised only of private housing.

## Temporal Variability in the Non-Selection Probability Coefficient

Low turnover properties may well be more (or less) expensive than high turnover properties due to household, attribute and location factors (GHHQ; Pryce 2004; Kim et al 2005). Ostensibly this means that the probability of non-selection[4] will have a positive (negative) sign in the hedonic price equation. The literature has tended to assume that the effect will remain constant over time. So an important question is whether a good economic rationale exists for the coefficient on this probability to vary? For example, if the prices of low-turnover dwellings rise relative to high-turnover properties, then one would expect the coefficient on this probability to increase.

This is not an implausible scenario. Kim et al (2005) find that the intention to move is much more prevalent in high-density neighbourhoods. Although it is only one side of the story, it does indicate that low-density neighbourhoods may well tend to have a lower turnover of stock. At the same time, there are good reasons to believe (in the UK at least) that the value of low-density housing is likely to rise in value at a faster rate, due, for example, to the combination of rising incomes, and low-density housing having a greater income elasticity of demand than high-density housing. Or because of an ageing population and older households seeking lower-density locations. Or because the majority of new construction is high-density due to planning policy, hence increasing the supply of high-density housing relative to the supply of low-density housing. Thus prices of low-density, low-turnover stock would rise in value at a faster rate than high-density, high turnover dwellings.

[4] This will be hazard of non-selection if a Heckman approach is used, but the implications are the same.

### *Submarkets and Sign Switching*

The existence of submarkets leads to a further possibility: that of the coefficient on the probability of non-selection switching sign. Jones et al (2003) argue that for localities to be considered as separate submarkets, not only must their attribute prices be different at a particular point in time, but also the dynamics of house prices must be independent. They consider 'whether price differences between submarkets have been eroded by a process of arbitrage operating through supply-side responses and/or migration flows' (p.1315) and verify that differences in price dynamics can persist over time between areas in close proximity.

This finding is relevant here because differences in the rate of price appreciation across neighbourhoods will affect the probability of dwellings coming onto the market due to the impact on the absolute difference in the value of housing equity and the transactions costs (see Stein 1995; Genesove and Mayer 1997, 2001). The corollary is that a subregion could temporarily switch from being a low-turnover area to being a high-turnover area simply because it has increased in value at a faster rate than other subregions. The adjustment process could be less than smooth due to tipping points that arise in the volume of subregional transactions due the existence of housing chains (Rosenthal 1997).

Tipping points could also be caused by information imperfections arising from the publication of uncorrected house price indices. For example, suppose low-density housing increases in value over a prolonged period at a rate that exceeds that of other property types. That difference in appreciation rates may not be widely known because house price information may only be presented in the form of averages for all property types (as in the UK). When owners do eventually become aware of the accelerated appreciation of their houses, there may be a rush of low-density dwellings being traded by households keen to access their accumulated equity, purchased by investors newly aware of the favourable long term prospects of this asset class. The damn-burst effect catapults areas of low-density housing from being classified as low-turnover to being high-turnover areas, at least temporarily. This could have the perverse effect of causing the coefficient on the probability of non-selection in the hedonic price equation to change sign: the set of properties with high-probabilities of non-selection temporarily no-longer includes the expensive low-density properties which are experiencing a transactions boom – the set of properties with high probability of non-selection is dominated for a time by those that infrequently trade because they are of particularly low quality (occupants are eager to sell, but no-one wants to buy).

### *A Complex System*

Taken together, these arguments highlight the complex nature of housing transactions. The quality and type of construction of a dwelling along with other factors will determine the desirability of a neighbourhood and the history of planning decisions and economic development will determine the spatial clustering of property types across neighbourhood desirability. As such the relationship between type and duration of stay is likely to be complex, and made all the more so market cycles, trends, and information imperfections. Variation in the types of property that trade may be reinforced by processes of gentrification: clusters of detached dwellings may initially be concentrated in less affluent areas, but the structural attributes of these properties eventually attract more affluent purchasers, leading to bouts of intensive trade in otherwise low turnover stock.

Note also that low demand areas often have low levels of trade, not because people do not want to leave those areas, but because few want to enter. Refusal to sell will, ultimately dominate – even the highest levels of demand can only be realised if the occupants are willing to sell (a buyer cannot *make* someone sell). So one would expect the lowest rates of turnover either in areas where there is both low demand and low supply, or in areas where demand may be high, but where supply is very low because no one wants to move out.

If demand is low, prices will adjust downwards to compensate. However, aversion to making a loss on sale means that most residents will only trade when they have accumulated sufficient equity to cover their mortgage debt. This only occurs when the wider market is at its peak, so trading levels are highly contingent upon current prices (the Stein/Genesove & Mayer effect). Such areas may be characterised by intermittent bursts of transactions – trade is generally low except during the final phase of sustained price appreciation. A stylised set of scenarios is summarised in Table 1, though the permutations considered are far from exhaustive.

**Table 1 Turnover Rate Market Scenarios**

| | **Low Supply*** (high long term satisfaction with dwelling/location and/or high expected capital gain ⇒ few want to sell) | **High Supply** (low long term satisfaction with dwelling/location and/or low expected capital gain ⇒ many want to sell) |
|---|---|---|
| **Low Demand** (low expected satisfaction with dwelling/location and/or low expected long term capital gain ⇒ few want to buy) | Low Turnover | Medium/Intermittent** Turnover |
| **High Demand** (high expected satisfaction with dwelling/location and/or high expected capital gain ⇒ high potential demand) | Medium/Intermittent** Turnover | High Turnover |

* Refusal to sell will always dominate – even the highest levels of demand only become realised if the current occupants are willing to sell. So lowest rates of turnover will either be in areas where there is both low demand and low supply, or in areas where demand may be high, but where supply is very low because no one wants to move out.

** If demand is low, prices will adjust accordingly. However, aversion to making a loss on sale means that most residents will only trade when they have accumulated sufficient equity to cover their mortgage debt. This only occurs when the wider market is at its peak, so trading levels are highly contingent upon current prices (the Stein/Genesove & Mayer effect). Similarly, high demand/low supply areas may experience intermittent clusters of sales as reluctant sellers eventually succumb to the temptation to sell and access their equity gains.

# Econometric Model

We attempt to estimate the following two equation model of house prices and the probability of sale:

$$p = a_0 + a_1\, detached + a_2 semi + a_3 terraced + a_4\, pnonselect \qquad [1]$$

$$pnonselect = f(p, \mathbf{B}, \mathbf{A}, \mathbf{N}, \mathbf{E}, \mathbf{D}) \qquad [2]$$

where:

$p$ = ln(price),

$pnonselect$ = probability of non-selection (i.e. not trading),

$\mathbf{B}$ = barriers to sale, particularly public ownership,

| | | |
|---|---|---|
| **A** | = | attributes of dwellings, |
| **N** | = | neighbourhood quality (e.g. school performance, density, and crime), |
| **E** | = | employment factors, |
| **D** | = | life-cycle factors, such as age of household, and population change. |

The direction of the effect on the probability of sale of variables included in vectors **B**, **A**, **N**, **E**, **D**, will be ambiguous because they affect not only the decision to sell but also the decision by potential purchasers to buy a given property. Given that the demand and supply effects are likely to run in opposite directions, it will be the net effect that will determine the sign of each coefficient in a given period.

While the system as a whole is identified (Rank Condition for identification), equation [1] will clearly be over-identified if the number of variables in [2] that are not in [1] exceeds 1 (Order Condition for identification). Over-identification is something of a non-isssue, however, given that our primary aim is prediction (rather than identifying the values of individual parameters). Second, an exactly identified specification of [1] could be obtained by either (a) reducing the number of exogenous variables in [2] and/or (b) introducing additional endogenous variables into the system. The problem with option (a) is that there is a trade-off with omitted variable bias and explanatory power in the estimation of the reduced form of equation [2]. Option (b) is perfectly feasible – we could, for example, introduce population migration flows as a right-hand-side endogenous variable in [1] but this would move us away from estimating a hedonic price equation towards a deterministic model. A further option would be to use factor analysis or principle component estimation to reduce the number of right hand side exogenous variables in [2] but again this seems somewhat specious given that the predicted values from [2] with principle components on the right-hand-side would presumably be very similar to those obtained from including the original list of exogenous variables.

In summary, we argue that because our overall goal is to derive predicted values for *p* in each time period, the over-identification issue does not preclude estimation, particularly since the remedies for over-identification are either likely to make only superficial changes in the actual estimates or because they introduce additional complications. Our plan, therefore, is to use FLR to estimate the reduced form equation for the probability of non-selection (we actually use the probability of selection, and then deduct predicted values from one) and use the predicted values to derive an instrument for *pnonselect*.

# Data

Our investigation is based on the analysis of data on postcode sectors and individual dwelling transactions in the South East of England over the period 1996 to 2003. Our results (particularly for the price equation) are based on very large samples and are drawn from the integration of different sources of spatial data (including Mosaic, Hometrack, Land Registry and The Ordinance Survey).

The primary data source was the Land Registry house price database supplied by the Department of Community and Local Government. This contained basic price, date and attribute information (detached, terraced, semi-detached, flats) for 1.4 million housing transactions over the period 1996 to 2003. Year dummies were created with the prefix "year_". The first half of Table 2 lists the mean and standard deviation of each Land Registry variable.

The selection equation was estimated at postcode sector level. There are 1,241 postcode sectors in the South East and variables that explain the probability of sale were collated for each year for each postcode sector. Variables include the incidence of crime, the proportion of social renting, the average education score, the average distance between dwellings (computed by Hometrack from Ordinance Survey Master Map data), the proportion of semi-detached dwellings (supplied by Mosaic), the change in population over the preceding ten years (local authority and Census estimates), and the proportion of the population over 65 (Mosaic). The proportion of dwellings that sell in any one year was calculated by dividing the total number of address points in each postcode sector by the total number of house transactions in that postcode sector.

**Table 2 Descriptives**

| Variables at Dwelling level: | n | Mean | Std.Dev. |
|---|---|---|---|
| price | 1,418,153 | £ 137,160 | £ 141,427 |
| detached | 1,418,153 | 26% | 0.437 |
| flat | 1,418,153 | 18% | 0.382 |
| semi | 1,418,153 | 27% | 0.447 |
| terraced | 1,418,153 | 28% | 0.450 |
| year_1996 | 1,418,153 | 10% | 0.296 |
| year_1997 | 1,418,153 | 11% | 0.317 |
| year_1998 | 1,418,153 | 12% | 0.324 |
| year_1999 | 1,418,153 | 14% | 0.345 |
| year_2000 | 1,418,153 | 13% | 0.331 |
| year_2001 | 1,418,153 | 13% | 0.341 |
| year_2002 | 1,418,153 | 15% | 0.355 |
| year_2003 | 1,418,153 | 12% | 0.330 |
| **Variables at postcode sector level:** | | | |
| Social rented | 1,241 | 12% | 0.090 |
| Economically active | 1,241 | 65% | 0.068 |
| Average Education score | 1,241 | 55.3 | 5.272 |
| Violent Crime | 1,241 | 0.9% | 0.004 |
| Burgulary | 1,241 | 0.5% | 0.002 |
| Distance between dwellings | 1,241 | 20.6 | 18.927 |
| Dwellings pre 1920 | 1,241 | 24% | 0.149 |
| Semi detached | 1,241 | 25% | 0.102 |
| Population change | 1,241 | 6% | 0.039 |
| Population over 65 | 1,241 | 17% | 0.060 |
| Proportion of stock that trades in a given year | 1,241 | 3% | 0.005 |

## *Results:*

The probability of sale in each postcode sector for each year was estimated by running separate regressions in each year. The dependent variable in each regression was the proportion of the housing stock that sold in that year. Explanatory variables included the proportion of socially rented dwellings, the proportion of economically active households, the average education score, the incidence of violent crime and burgulary, the average distance between dwellings, the proportion of dwellings that were built before 1920, the proportion of semi-detached housing, the % change in population over the preceding ten years and the proportion of the population over 65.

As a baseline, we present the OLS results of these annual regressions in Table 3. On the whole, we were able to explain around a quarter of the variation in the dependent variable (the adjusted $R^2$ ranges between 0.232 in 2000 to 0.313 in 1996). The FLR results are reported in Table 4. The most significant variable was the proportion of socially rented housing (t-ratios varied between –9 and –16). We found that the greater the proportion of social rented housing in an area, the lower the probability of sale. Better school performance tended to raise the probability of sale, though the effect was not significantly different from zero in four of the years. Distance between dwellings also proved to be highly significant in most years (t-ratios varied between –1 and –9) and to have negative effect on the probability of sale. The sign and significance of violent crime and burgulary varied between years, (with one tending to be lower if the other is higher, presumably due to multicolinearity), as did the impact of the proportion of dwellings built pre-1920. Increases in population raised the probability of sale and the effect was statistically significant in all years except 1998. The impact of the proportion of households aged over 65 also had a positive effect in most years, though the effect was only significantly different from zero in years 1996-1999.

Table 5 presents the results of ln(price) regressions run on all years of the data, first without the *pnonselect* variable – the estimated probability of non-selection – and then with. The variable is highly significant (t-ratio = 45) and positive. This suggests that properties that do not trade are not a random subsample of the housing stock, but tend, in fact, to be more expensive. We are also interested, however, in whether properties that are less likely to trade have increased in value at a different rate to those that frequently trade. Does the coefficient on the *pnonselect* variable change over time and is the variation sufficient to impact on price index calculation?

Since one of our goals is to derive a measure of selection-adjusted house price inflation that can easily be updated, we avoid using the dummy variable approach (since the addition of more recent data would cause all parameters to change and all previous values of the index would need to be updated each time another year of data is included). As a result we adopt a method commonly adopted in the UK in which a separate hedonic regression is applied to each period. We have a very large number of observations and this means that there are sufficient degrees of freedom to run a separate regression on each month (i.e. regression of ln(price) on *detached*, semi, terraced, and *pnonselect*). The coefficients on the probability of *pnonselect* from each of these monthly regressions are listed in Table 6 and plotted in Figure 1 along with the 95% confidence interval. It can be seen that in most years the coefficient remains positive, but that there is significant variation from year to year and evidence that the coefficient temporarily changed sign during 1999. Note that the steep slope of the line at the end of particular years is likely to be due in part to the fact that the probability of non-selection was calculated on an annual basis – the true adjustment in the coefficient over time is probably somewhat smoother.

Predicted values from the monthly ln(price) regressions were used to derive adjusted and unadjusted house price indices using the following adaptation of the Fleming and Nellis (1984) method – the approach used to construct the Halifax house price index (see Meen and Andrew, 1998, p. 10):

$$I_t = \frac{\sum \exp\left(\beta_{j,t} X_{j,1996}\right)}{\sum \exp\left(\beta_{j,1996} X_{j,1996}\right)}.$$

Advantages with this approach are that it incorporates the possibility that 'implicit prices may change over time' (Meen and Andrew, 1998, p. 10) and can also readily be updated with information on subsequent time periods without changing all previous parameters and index values. Exponentiated predicted values from each regression using the average set of characteristics from 1996 are presented in Table 7 along with the index values for each month. The cumulative effect over the entire period appears to be that the unadjusted index tends to understate the true rate of price inflation of the stock of private housing (280% increase in the unadjusted index compared with the 302% increase in the adjusted index).

## Summary

This paper has attempted to establish a method for correcting transactions bias in house price indices that could realistically be applied to all regions of the UK and also to other countries where attribute data on individual dwellings are not available for the whole stock of housing, but where information exists at neighbourhood level on factors that influence the probability of sale (factors such as crime, population change, tenure, school performance and density). Fractional Logit Regression was used to derive an instrument for the probability of non-selection. This technique is generally applicable to situations where the dependent variable is bounded between zero and one – and hence appropriate for modelling the variable of interest here: the proportion of properties that sold in each postcode sector in each year. We found evidence that our estimated probability of non-selection had a statistically significant effect in a simple hedonic price equation. We also found that the coefficient tended to vary over time, even changing sign in 1999. Overall, the unadjusted index tended to underestimate the true rate of price appreciation of the stock of private housing.

## Table 3 Estimation of the Selection Equation: OLS

| | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 |
|---|---|---|---|---|---|---|---|---|
| Social rented | -2.850 | -2.787 | -2.726 | -2.855 | -3.173 | -2.999 | -3.323 | -2.830 |
| | (-14.054) | (-15.275) | (-12.219) | (-12.152) | (-7.817) | (-17.821) | (-9.109) | (-7.662) |
| Economically active | 0.482 | 0.233 | 0.678 | 0.619 | -1.082 | 0.379 | -0.778 | -0.457 |
| | (1.218) | (0.746) | (1.979) | (1.130) | (-0.970) | (1.123) | (-0.770) | (-0.445) |
| Education score | 0.011 | 0.004 | 0.005 | 0.011 | 0.014 | 0.003 | 0.011 | 0.004 |
| | (3.198) | (1.185) | (1.552) | (2.589) | (3.209) | (0.835) | (2.679) | (1.081) |
| Violent Crime | 1.010 | 6.508 | 1.390 | -6.714 | 6.922 | 5.527 | 3.671 | -0.078 |
| | (0.213) | (1.378) | (0.319) | (-1.298) | (1.065) | (1.046) | (0.617) | (-0.013) |
| Burgulary | 5.532 | -1.966 | -3.943 | 1.406 | 2.344 | -3.851 | -16.486 | -12.154 |
| | (0.956) | (-0.361) | (-0.660) | (0.227) | (0.279) | (-0.640) | (-2.220) | (-1.584) |
| Dist. between dwells | -0.010 | -0.010 | -0.007 | -0.006 | -0.009 | -0.003 | -0.003 | -0.005 |
| | (-9.490) | (-8.392) | (-6.433) | (-4.938) | (-6.566) | (-1.249) | (-1.391) | (-5.073) |
| Dwellings pre 1920 | -0.214 | -0.111 | 0.198 | 0.400 | 0.504 | -0.192 | 0.080 | 0.140 |
| | (-1.668) | (-0.840) | (1.534) | (2.640) | (2.855) | (-1.102) | (0.459) | (0.885) |
| Semi-detached | -0.080 | -0.138 | -0.454 | -0.394 | -0.458 | -0.533 | -0.615 | -0.676 |
| | (-0.598) | (-0.926) | (-2.799) | (-2.193) | (-2.401) | (-3.752) | (-3.731) | (-3.961) |
| Population change | 0.687 | 0.708 | 0.562 | 1.914 | 2.997 | 1.750 | 2.059 | 1.986 |
| | (2.324) | (2.116) | (1.671) | (4.888) | (5.689) | (4.218) | (4.583) | (5.338) |
| Population over 65 | 1.886 | 1.553 | 0.853 | 2.443 | 0.456 | 0.443 | -0.053 | -0.213 |
| | (4.718) | (5.127) | (2.594) | (5.204) | (0.508) | (1.422) | (-0.064) | (-0.246) |
| Constant | 1.476 | 2.246 | 2.135 | 1.669 | 2.652 | 2.639 | 3.312 | 3.276 |
| | (3.981) | (7.246) | (6.448) | (3.441) | (2.830) | (8.227) | (3.854) | (3.674) |
| n | 1,198 | 1,198 | 1,205 | 1,241 | 1,263 | 1,267 | 1,280 | 1,280 |
| Adj R2 | 0.313 | 0.303 | 0.260 | 0.272 | 0.232 | 0.260 | 0.236 | 0.237 |

Dependent variable = % of the total housing stock that trades in a given year
Figures in brackets are t-ratios based on Mackinnon and White (1985) HC2 standard errors.

## Table 4 Estimation of the Selection Equation: FLR

| | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 |
|---|---|---|---|---|---|---|---|---|
| Social rented | -1.518 | -1.324 | -1.269 | -1.274 | -1.428 | -1.289 | -1.336 | -1.216 |
| | (-13.576) | (-14.578) | (-10.970) | (-11.924) | (-9.091) | (-15.963) | (-10.221) | (-8.663) |
| Economically active | 0.336 | 0.143 | 0.345 | 0.321 | -0.378 | 0.212 | -0.226 | -0.135 |
| | (1.666) | (1.022) | (2.327) | (1.475) | (-0.920) | (1.551) | (-0.661) | (-0.356) |
| Education score | 0.006 | 0.002 | 0.002 | 0.005 | 0.006 | 0.001 | 0.004 | 0.002 |
| | (3.439) | (1.332) | (1.737) | (2.763) | (3.353) | (1.036) | (2.894) | (1.275) |
| Violent Crime | 0.114 | 2.392 | 0.368 | -2.891 | 2.572 | 2.091 | 1.287 | -0.116 |
| | (0.051) | (1.164) | (0.203) | (-1.437) | (0.984) | (1.020) | (0.593) | (-0.050) |
| Burgulary | 2.840 | -0.785 | -1.399 | 0.577 | 1.141 | -1.277 | -5.981 | -4.737 |
| | (1.043) | (-0.330) | (-0.558) | (0.237) | (0.344) | (-0.543) | (-2.234) | (-1.595) |
| Dist. between dwells | -0.005 | -0.005 | -0.003 | -0.003 | -0.004 | -0.001 | -0.001 | -0.002 |
| | (-9.514) | (-8.444) | (-6.561) | (-5.165) | (-6.521) | (-1.403) | (-1.585) | (-5.293) |
| Dwellings pre 1920 | -0.048 | -0.018 | 0.113 | 0.190 | 0.225 | -0.041 | 0.061 | 0.080 |
| | (-0.815) | (-0.315) | (2.149) | (3.288) | (3.289) | (-0.607) | (1.003) | (1.361) |
| Semi-detached | 0.035 | -0.016 | -0.143 | -0.105 | -0.149 | -0.169 | -0.188 | -0.240 |
| | (0.556) | (-0.235) | (-2.040) | (-1.452) | (-1.876) | (-2.948) | (-3.122) | (-3.583) |
| Population change | 0.346 | 0.309 | 0.254 | 0.761 | 1.253 | 0.719 | 0.788 | 0.821 |
| | (2.454) | (2.098) | (1.787) | (5.048) | (6.085) | (4.438) | (4.885) | (5.684) |
| Population over 65 | 0.883 | 0.637 | 0.369 | 0.921 | 0.191 | 0.204 | -0.003 | -0.070 |
| | (4.675) | (4.999) | (2.711) | (5.099) | (0.581) | (1.692) | (-0.009) | (-0.222) |
| Constant | -4.244 | -3.808 | -3.879 | -4.053 | -3.661 | -3.680 | -3.421 | -3.407 |
| | (-22.490) | (-27.414) | (-27.079) | (-20.876) | (-10.521) | (-28.568) | (-11.725) | (-10.373) |
| n | 1,198 | 1,198 | 1,205 | 1,241 | 1,263 | 1,267 | 1,280 | 1,280 |
| ll | -100.4 | -106.7 | -111.4 | -120.6 | -117.0 | -122.5 | -130.0 | -122.3 |

Dependent variable = proportion of the total housing stock that trades in a given year.
T-ratios, presented in parentheses, are based on Papke and Wooldridge (1996) robust standard errors.

**Table 5 Hedonic Estimates**

| | Without Correction Term | With Correction Term |
|---|---|---|
| *detached* | 0.989 | 0.981 |
| | (747.217) | (737.697) |
| semi | 0.448 | 0.439 |
| | (374.005) | (361.958) |
| terraced | 0.212 | 0.206 |
| | (176.953) | (170.707) |
| year_1997 | 0.113 | 0.131 |
| | (67.478) | (76.119) |
| year_1998 | 0.248 | 0.274 |
| | (145.890) | (153.617) |
| year_1999 | 0.351 | 0.399 |
| | (215.793) | (208.406) |
| year_2000 | 0.528 | 0.563 |
| | (311.244) | (310.389) |
| year_2001 | 0.655 | 0.701 |
| | (397.612) | (364.859) |
| year_2002 | 0.819 | 0.889 |
| | (513.942) | (410.443) |
| year_2003 | 0.955 | 0.993 |
| | (600.901) | (564.245) |
| *pnonselect* | - | 11.040 |
| | - | (45.456) |
| _cons | 10.678 | -0.061 |
| | (6712.393) | (-0.257) |
| N | 1,418,153 | 1,418,153 |
| r2_a | 0.510 | 0.511 |

Figures in brackets are t-ratios based on Mackinnon and White (1985) HC2 standard errors.

**Table 6 Coefficients from Individual Monthly Selection-Adjusted Hedonic ln(price) Regressions**

| Year | Month | b | se | CI95_L | CI95_U | Adj. R2 |
|---|---|---|---|---|---|---|
| 1996 | jan | 12.61 | 2.77 | 7.18 | 18.04 | 0.36 |
| | feb | 11.01 | 2.64 | 5.82 | 16.19 | 0.35 |
| | mar | 10.51 | 2.34 | 5.92 | 15.10 | 0.35 |
| | apr | 13.12 | 2.36 | 8.49 | 17.75 | 0.35 |
| | may | 10.95 | 2.25 | 6.54 | 15.36 | 0.34 |
| | jun | 14.84 | 2.10 | 10.72 | 18.96 | 0.37 |
| | jul | 18.47 | 1.96 | 14.62 | 22.31 | 0.37 |
| | aug | 18.99 | 1.92 | 15.22 | 22.76 | 0.36 |
| | sep | 11.66 | 2.03 | 7.70 | 15.63 | 0.38 |
| | oct | 12.15 | 2.04 | 8.14 | 16.15 | 0.35 |
| | nov | 14.74 | 1.87 | 11.07 | 18.41 | 0.37 |
| | dec | 14.70 | 1.94 | 10.90 | 18.49 | 0.38 |
| 1997 | jan | 44.28 | 2.67 | 39.04 | 49.52 | 0.38 |
| | feb | 40.34 | 2.45 | 35.54 | 45.14 | 0.40 |
| | mar | 42.62 | 2.34 | 38.03 | 47.21 | 0.38 |
| | apr | 43.23 | 2.28 | 38.76 | 47.70 | 0.40 |
| | may | 40.71 | 2.07 | 36.66 | 44.77 | 0.39 |
| | jun | 43.00 | 2.05 | 38.98 | 47.03 | 0.41 |
| | jul | 47.61 | 1.97 | 43.75 | 51.48 | 0.42 |
| | aug | 44.32 | 2.01 | 40.37 | 48.26 | 0.43 |
| | sep | 43.80 | 2.14 | 39.61 | 48.00 | 0.40 |
| | oct | 47.11 | 2.10 | 42.99 | 51.23 | 0.40 |
| | nov | 39.25 | 2.24 | 34.87 | 43.63 | 0.40 |
| | dec | 49.81 | 2.18 | 45.54 | 54.07 | 0.39 |
| 1998 | jan | 7.77 | 3.87 | 0.18 | 15.36 | 0.40 |
| | feb | 0.90 | 3.89 | -6.73 | 8.53 | 0.39 |
| | mar | 2.73 | 3.68 | -4.49 | 9.95 | 0.36 |
| | apr | 8.86 | 3.43 | 2.13 | 15.59 | 0.38 |
| | may | 10.71 | 3.16 | 4.51 | 16.90 | 0.40 |
| | jun | 6.56 | 3.17 | 0.36 | 12.77 | 0.38 |
| | jul | 4.42 | 3.01 | -1.48 | 10.33 | 0.39 |
| | aug | 8.57 | 3.24 | 2.22 | 14.92 | 0.39 |
| | sep | 3.44 | 3.37 | -3.17 | 10.06 | 0.37 |
| | oct | 13.35 | 3.31 | 6.86 | 19.85 | 0.38 |
| | nov | 13.68 | 3.41 | 7.00 | 20.37 | 0.37 |
| | dec | 12.60 | 3.43 | 5.87 | 19.32 | 0.37 |
| 1999 | jan | -23.83 | 2.22 | -28.18 | -19.48 | 0.39 |
| | feb | -18.87 | 2.14 | -23.07 | -14.67 | 0.40 |
| | mar | -11.15 | 1.85 | -14.76 | -7.53 | 0.40 |
| | apr | -22.92 | 1.77 | -26.38 | -19.45 | 0.41 |
| | may | -18.91 | 1.79 | -22.41 | -15.40 | 0.42 |
| | jun | -15.99 | 1.64 | -19.22 | -12.77 | 0.41 |
| | jul | -14.84 | 1.55 | -17.89 | -11.80 | 0.42 |
| | aug | -14.73 | 1.61 | -17.89 | -11.58 | 0.41 |
| | sep | -13.30 | 1.63 | -16.50 | -10.10 | 0.40 |
| | oct | -14.33 | 1.63 | -17.53 | -11.14 | 0.39 |
| | nov | -18.67 | 1.68 | -21.96 | -15.38 | 0.39 |
| | dec | -17.21 | 1.69 | -20.51 | -13.90 | 0.40 |

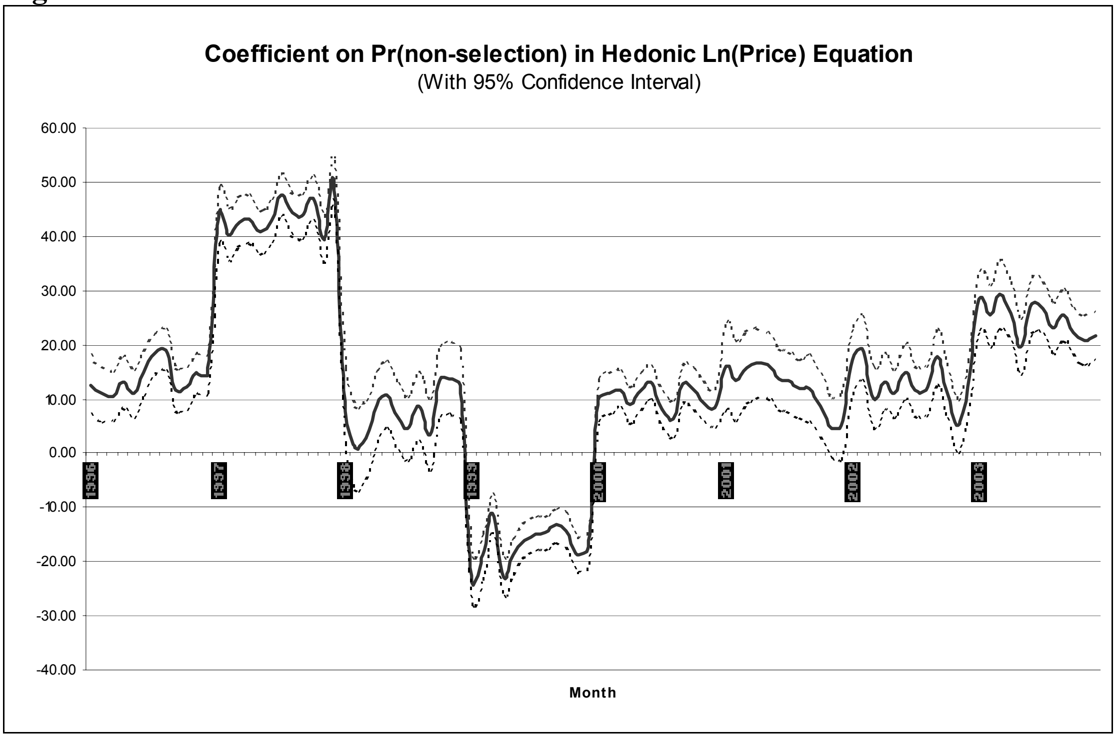
| | | | | | | |
|---|---|---|---|---|---|---|
| 2000 | jan | 9.99 | 2.02 | 6.03 | 13.95 | 0.38 |
| | feb | 10.97 | 1.94 | 7.17 | 14.77 | 0.38 |
| | mar | 11.73 | 1.75 | 8.31 | 15.16 | 0.37 |
| | apr | 8.87 | 1.77 | 5.40 | 12.35 | 0.39 |
| | may | 11.62 | 1.75 | 8.18 | 15.05 | 0.38 |
| | jun | 13.01 | 1.61 | 9.85 | 16.17 | 0.39 |
| | jul | 8.01 | 1.76 | 4.57 | 11.45 | 0.38 |
| | aug | 6.35 | 1.75 | 2.93 | 9.78 | 0.39 |
| | sep | 12.80 | 1.84 | 9.20 | 16.41 | 0.39 |
| | oct | 11.57 | 1.92 | 7.81 | 15.32 | 0.39 |
| | nov | 9.33 | 1.87 | 5.65 | 13.00 | 0.38 |
| | dec | 8.42 | 1.76 | 4.98 | 11.87 | 0.37 |
| 2001 | jan | 16.18 | 4.08 | 8.18 | 24.18 | 0.38 |
| | feb | 13.36 | 3.81 | 5.89 | 20.84 | 0.37 |
| | mar | 15.88 | 3.35 | 9.32 | 22.43 | 0.36 |
| | apr | 16.66 | 3.25 | 10.29 | 23.02 | 0.37 |
| | may | 16.22 | 3.02 | 10.29 | 22.15 | 0.37 |
| | jun | 13.63 | 2.89 | 7.96 | 19.30 | 0.35 |
| | jul | 13.30 | 2.79 | 7.83 | 18.78 | 0.38 |
| | aug | 11.89 | 2.70 | 6.59 | 17.18 | 0.37 |
| | sep | 11.94 | 3.15 | 5.76 | 18.13 | 0.35 |
| | oct | 8.63 | 2.98 | 2.78 | 14.47 | 0.36 |
| | nov | 4.56 | 2.92 | -1.15 | 10.28 | 0.35 |
| | dec | 5.18 | 3.21 | -1.11 | 11.46 | 0.35 |
| 2002 | jan | 17.21 | 3.17 | 11.00 | 23.42 | 0.37 |
| | feb | 19.29 | 3.08 | 13.26 | 25.32 | 0.34 |
| | mar | 10.14 | 2.81 | 4.63 | 15.65 | 0.34 |
| | apr | 13.24 | 2.63 | 8.08 | 18.39 | 0.36 |
| | may | 10.91 | 2.26 | 6.47 | 15.34 | 0.36 |
| | jun | 14.97 | 2.60 | 9.87 | 20.08 | 0.36 |
| | jul | 11.47 | 2.27 | 7.03 | 15.92 | 0.39 |
| | aug | 11.52 | 2.28 | 7.04 | 15.99 | 0.37 |
| | sep | 17.81 | 2.64 | 12.64 | 22.97 | 0.35 |
| | oct | 11.29 | 2.52 | 6.36 | 16.23 | 0.38 |
| | nov | 5.11 | 2.48 | 0.25 | 9.97 | 0.38 |
| | dec | 12.26 | 2.58 | 7.21 | 17.31 | 0.37 |
| 2003 | jan | 28.38 | 2.81 | 22.88 | 33.89 | 0.39 |
| | feb | 25.37 | 2.94 | 19.61 | 31.13 | 0.38 |
| | mar | 29.36 | 3.16 | 23.17 | 35.55 | 0.38 |
| | apr | 25.89 | 2.74 | 20.52 | 31.26 | 0.39 |
| | may | 19.67 | 2.67 | 14.43 | 24.91 | 0.39 |
| | jun | 27.45 | 2.68 | 22.21 | 32.70 | 0.38 |
| | jul | 26.68 | 2.41 | 21.96 | 31.40 | 0.41 |
| | aug | 23.16 | 2.40 | 18.45 | 27.87 | 0.41 |
| | sep | 25.49 | 2.42 | 20.75 | 30.24 | 0.40 |
| | oct | 22.34 | 2.21 | 18.02 | 26.66 | 0.42 |
| | nov | 20.87 | 2.29 | 16.39 | 25.36 | 0.42 |
| | dec | 21.61 | 2.27 | 17.16 | 26.06 | 0.40 |
| **Average:** | | **13.54** | **2.47** | **8.70** | **18.39** | **0.38** |

## Table 7 Adusted and Unadjusted House Price Indices

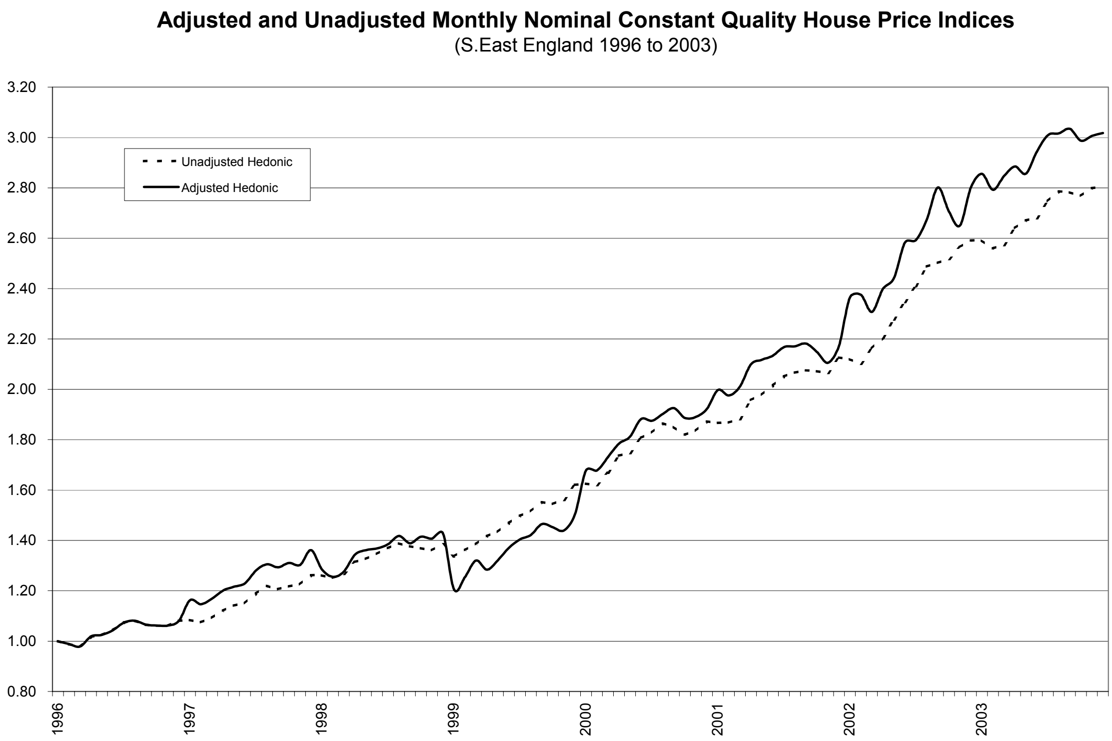
| Year | Month | Unadjusted | Adjusted | Unadjusted Index | Adjusted Index |
|---|---|---|---|---|---|
| 1996 | jan | £ 65,368 | £ 65,336 | 1.00 | 1.00 |
| | feb | £ 64,573 | £ 64,594 | 0.99 | 0.99 |
| | mar | £ 63,898 | £ 63,911 | 0.98 | 0.98 |
| | apr | £ 66,509 | £ 66,546 | 1.02 | 1.02 |
| | may | £ 66,974 | £ 66,971 | 1.02 | 1.03 |
| | jun | £ 68,208 | £ 68,181 | 1.04 | 1.04 |
| | jul | £ 70,213 | £ 70,165 | 1.07 | 1.07 |
| | aug | £ 70,650 | £ 70,615 | 1.08 | 1.08 |
| | sep | £ 69,604 | £ 69,611 | 1.06 | 1.07 |
| | oct | £ 69,352 | £ 69,382 | 1.06 | 1.06 |
| | nov | £ 69,366 | £ 69,385 | 1.06 | 1.06 |
| | dec | £ 70,630 | £ 70,621 | 1.08 | 1.08 |
| 1997 | jan | £ 70,830 | £ 75,952 | 1.08 | 1.16 |
| | feb | £ 70,272 | £ 74,906 | 1.08 | 1.15 |
| | mar | £ 71,409 | £ 76,352 | 1.09 | 1.17 |
| | apr | £ 73,262 | £ 78,457 | 1.12 | 1.20 |
| | may | £ 74,581 | £ 79,463 | 1.14 | 1.22 |
| | jun | £ 75,201 | £ 80,329 | 1.15 | 1.23 |
| | jul | £ 77,707 | £ 83,653 | 1.19 | 1.28 |
| | aug | £ 79,710 | £ 85,290 | 1.22 | 1.31 |
| | sep | £ 78,876 | £ 84,458 | 1.21 | 1.29 |
| | oct | £ 79,551 | £ 85,642 | 1.22 | 1.31 |
| | nov | £ 80,094 | £ 85,087 | 1.23 | 1.30 |
| | dec | £ 82,470 | £ 88,942 | 1.26 | 1.36 |
| 1998 | jan | £ 82,426 | £ 83,973 | 1.26 | 1.29 |
| | feb | £ 81,820 | £ 81,996 | 1.25 | 1.25 |
| | mar | £ 82,835 | £ 83,382 | 1.27 | 1.28 |
| | apr | £ 85,897 | £ 87,760 | 1.31 | 1.34 |
| | may | £ 86,697 | £ 88,961 | 1.33 | 1.36 |
| | jun | £ 87,976 | £ 89,370 | 1.35 | 1.37 |
| | jul | £ 89,497 | £ 90,455 | 1.37 | 1.38 |
| | aug | £ 90,739 | £ 92,609 | 1.39 | 1.42 |
| | sep | £ 90,001 | £ 90,737 | 1.38 | 1.39 |
| | oct | £ 89,511 | £ 92,409 | 1.37 | 1.41 |
| | nov | £ 88,962 | £ 91,942 | 1.36 | 1.41 |
| | dec | £ 90,506 | £ 93,291 | 1.38 | 1.43 |
| 1999 | jan | £ 87,380 | £ 78,774 | 1.34 | 1.21 |
| | feb | £ 88,998 | £ 81,960 | 1.36 | 1.25 |
| | mar | £ 90,557 | £ 86,247 | 1.39 | 1.32 |
| | apr | £ 92,634 | £ 83,853 | 1.42 | 1.28 |
| | may | £ 93,749 | £ 86,308 | 1.43 | 1.32 |
| | jun | £ 96,012 | £ 89,541 | 1.47 | 1.37 |
| | jul | £ 97,882 | £ 91,725 | 1.50 | 1.40 |
| | aug | £ 99,080 | £ 92,884 | 1.52 | 1.42 |
| | sep | £ 101,490 | £ 95,704 | 1.55 | 1.46 |
| | oct | £ 100,985 | £ 94,841 | 1.54 | 1.45 |
| | nov | £ 102,053 | £ 94,051 | 1.56 | 1.44 |
| | dec | £ 105,935 | £ 98,229 | 1.62 | 1.50 |
| 2000 | jan | £ 106,232 | £ 109,599 | 1.63 | 1.68 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | feb | £ | 105,927 | £ | 109,615 | 1.62 | 1.68 |
| | mar | £ | 109,091 | £ | 113,077 | 1.67 | 1.73 |
| | apr | £ | 113,481 | £ | 116,585 | 1.74 | 1.78 |
| | may | £ | 114,232 | £ | 118,383 | 1.75 | 1.81 |
| | jun | £ | 118,131 | £ | 122,928 | 1.81 | 1.88 |
| | jul | £ | 119,523 | £ | 122,512 | 1.83 | 1.88 |
| | aug | £ | 121,904 | £ | 124,319 | 1.86 | 1.90 |
| | sep | £ | 120,991 | £ | 125,779 | 1.85 | 1.93 |
| | oct | £ | 118,926 | £ | 123,241 | 1.82 | 1.89 |
| | nov | £ | 119,974 | £ | 123,492 | 1.84 | 1.89 |
| | dec | £ | 122,428 | £ | 125,597 | 1.87 | 1.92 |
| 2001 | jan | £ | 122,026 | £ | 130,493 | 1.87 | 2.00 |
| | feb | £ | 122,104 | £ | 129,079 | 1.87 | 1.98 |
| | mar | £ | 122,980 | £ | 131,369 | 1.88 | 2.01 |
| | apr | £ | 127,944 | £ | 137,120 | 1.96 | 2.10 |
| | may | £ | 129,323 | £ | 138,351 | 1.98 | 2.12 |
| | jun | £ | 131,730 | £ | 139,411 | 2.02 | 2.13 |
| | jul | £ | 134,072 | £ | 141,626 | 2.05 | 2.17 |
| | aug | £ | 135,027 | £ | 141,817 | 2.07 | 2.17 |
| | sep | £ | 135,681 | £ | 142,539 | 2.08 | 2.18 |
| | oct | £ | 135,425 | £ | 140,327 | 2.07 | 2.15 |
| | nov | £ | 135,013 | £ | 137,583 | 2.07 | 2.11 |
| | dec | £ | 138,969 | £ | 141,961 | 2.13 | 2.17 |
| 2002 | jan | £ | 138,554 | £ | 154,473 | 2.12 | 2.36 |
| | feb | £ | 137,363 | £ | 155,213 | 2.10 | 2.38 |
| | mar | £ | 141,345 | £ | 150,771 | 2.16 | 2.31 |
| | apr | £ | 144,115 | £ | 156,709 | 2.20 | 2.40 |
| | may | £ | 148,892 | £ | 159,511 | 2.28 | 2.44 |
| | jun | £ | 153,450 | £ | 168,711 | 2.35 | 2.58 |
| | jul | £ | 157,519 | £ | 169,386 | 2.41 | 2.59 |
| | aug | £ | 162,589 | £ | 174,826 | 2.49 | 2.68 |
| | sep | £ | 163,543 | £ | 183,048 | 2.50 | 2.80 |
| | oct | £ | 164,600 | £ | 176,795 | 2.52 | 2.71 |
| | nov | £ | 167,666 | £ | 173,161 | 2.56 | 2.65 |
| | dec | £ | 169,379 | £ | 183,099 | 2.59 | 2.80 |
| 2003 | jan | £ | 169,377 | £ | 186,582 | 2.59 | 2.86 |
| | feb | £ | 167,297 | £ | 182,446 | 2.56 | 2.79 |
| | mar | £ | 168,185 | £ | 185,966 | 2.57 | 2.85 |
| | apr | £ | 172,602 | £ | 188,506 | 2.64 | 2.89 |
| | may | £ | 174,553 | £ | 186,620 | 2.67 | 2.86 |
| | jun | £ | 175,188 | £ | 192,323 | 2.68 | 2.94 |
| | jul | £ | 179,590 | £ | 196,619 | 2.75 | 3.01 |
| | aug | £ | 182,144 | £ | 197,080 | 2.79 | 3.02 |
| | sep | £ | 181,883 | £ | 198,262 | 2.78 | 3.03 |
| | oct | £ | 180,961 | £ | 195,170 | 2.77 | 2.99 |
| | nov | £ | 182,957 | £ | 196,400 | 2.80 | 3.01 |
| | dec | £ | 183,251 | £ | 197,184 | 2.80 | 3.02 |

## Figure 1




## Figure 2



## References:


Fleming, M.C. and Nellis, J.G. (1984) 'The Halifax House Price Index Technical Details', published by the Halifax Building Society and cited in Meen and Andrew, 1998, p. 10.

Gatzlaff, D.H. and D. R. Haurin. 1994. Measuring changes in local house prices. *Journal of Urban Economics* 35: 221-244.

Gatzlaff, D.H. and D. R. Haurin. 1997. Sample selection bias and repeat-sales index estimates. *Journal of Real Estate Finance and Economics* 14: 33–50.

Gatzlaff, D.H. and D. R. Haurin. 1998. Sample selection and biases in local house value indices. *Journal of Urban Economics* 43: 199–222.

Genesove, D. and J. Mayer. 1997. Equity and Time to Sale in the Real Estate Market. *American Economic Review* 87(3): 255-269.

Genesove, D. and J. Mayer. 2001. Loss aversion and seller behavior: Evidence from the housing Market. *Quarterly Journal of Economics* 116(4): 1233-1260.

Grigsby, W. 1963. *Housing Markets and Public Policy.* University of Pennsylvania Press Philadelphia.

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

Hendershott, P. and Pryce, G. (2006) The Sensitivity of Homeowner Leverage to the Deductibility of Home Mortgage Interest, *Journal of Urban Economics* (forthcoming).

Hwang, M. and Quigley, J. (2004) Selectivity, Quality Adjustment and Mean Reversion in the Measurement of House Values, *Journal of Real Estate Finance and Economics*, 28:2/3, 161-178.

Jones, C., C. Leishman and C. Watkins. 2003. Structural Change in a Local Urban Housing Market. *Environment and Planning A* 35(7): 1315-1326.

Kim, J.H. Pagliara, F. and Preston, J. (2005) 'The Intention to Move and Residential Location Choice Behaviour', *Urban Studies*, 42, 9 1621-1636.

Meen, G. and Andrew, M. (1998) "*Modelling Regional House Prices: A Review of the Literature*", Report Prepared for the Department of the Environment, Transport and the Regions, London.

Pryce, G. (2004) *The Micro and Macro Effects of the Location of New Housing Supply*, ODPM, London.

Rosenthal, L. (1997) "Chain formation in the owner-occupied housing market", *Economic Journal*, 107, 3, 475-488.

Stein, J. 1995. Prices and Trading Volume in the Housing Market: A Model with Downpayment Constraints. *Quarterly Journal of Economics* 110(2): 379-406.