Mason, P., and Pryce, G. (2011) *Controlling for transactions bias in regional house price indices.* Housing Studies, 26 (5). pp. 639-660. ISSN 0267-3037

http://eprints.gla.ac.uk/40379/

Deposited on: 4 January 2012

Controlling for Transactions Bias in Regional House Price Indices*

* *A revised version of his article has since been published in Housing Studies: Mason, P. & Pryce, G. (2011) Controlling for Transactions Bias in Regional House Price Indices. Housing Studies, 26 (5) 639-660*

Phil Mason
Department of Urban Studies
University of Glasgow
Glasgow, Scotland
G12 8RS
Tel: +44 141 330 5048
(p.mason@lbss.gla.ac.uk)

and

Gwilym Pryce (**)
Department of Urban Studies
University of Glasgow
Glasgow, Scotland
G12 8RS
Tel: +44 141 330 4399
(g.pryce@lbss.gla.ac.uk)

4th May 2010

Working Paper: Please do not quote without permission.

ABSTRACT Transactions bias arises when properties that trade are not a random sample of the total housing stock. Price indices are susceptible because they are typically based on transactions data. Existing approaches to this problem rely 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—the UK included—no such data exist and there is little prospect of correcting for transactions bias in any of the regularly updated mainstream house price indices. This paper suggests a possible alternative approach, using information at postcode sector level and Fractional Probit Regression to correct for transactions bias in hedonic price indices based on one and a half million house sales from 1996 to 2004, distributed across 1200 postcode sectors in the South East of England.

(**) Corresponding author

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 to Patric Hendershott, Don Haurin, Jesse Abraham and other participants at the Conference in Honour of Patric Hendershott (Columbus Ohio, July 2006). Thanks also to two anonymous referees for useful comments and to Experian for the generous provision of postcode level data. Neither DCLG nor Experian necessarily agree with our findings. Any remaining errors are our own.

* *A revised version of his article has since been published in Housing Studies: Mason, P. & Pryce, G. (2011) Controlling for Transactions Bias in Regional House Price Indices. Housing Studies, 26 (5) 639-660*

* *A revised version of his article has since been published in Housing Studies: Mason, P. & Pryce, G. (2011) Controlling for Transactions Bias in Regional House Price Indices. Housing Studies, 26 (5) 639-660*

Introduction

House price indices are typically computed on the basis of the sale price of properties traded in a given period. If one is interested in changes in selling price of traded properties, then these indices will offer suitable measurement, provided one uses an appropriate method for computing price change and controls for dwelling heterogeneity.[1] However, if one is attempting to compute changes in the value of houses in the *entire stock* of dwellings, then indices based on traded dwellings may be subject to transactions bias because properties that trade in a given period may not be typical of all dwellings. Sample-selection bias becomes problematic when there is systematic tendency for certain property types and locations (such as low-density properties) to trade less frequently than others, and when properties that are less likely to trade have a different rate of price appreciation.

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; Hwang and Quigley 2004) or in hedonic indices (Gatzlaff and Haurin 1998) have been published in leading real estate journals and are cited frequently.

[1] Variation in the type of house coming on the market can distort the computation of the average price and so "hedonic" estimation methods (see Malpezzi 2003) have been developed as a way of estimating the value of a standardised unit of housing. We do not attempt to address all the problems associated with hedonic house price measurement. Issues not considered here include the effect of variation in selling times and the role of liquidity in changing the interpretation of sale prices in a given period (see Leung, Lau and Leong, 2002; Fisher *et al.*, 2003; Clayton, Miller and Peng, 2010; and Levin and Pryce, 2009) and various problems associated with hedonic price indices (Case *et al*. 2003; Hill and Melser 2008).

Yet, there remain two major problems with the application of these methods. First, the techniques used have not been adopted in the mainstream hedonic literature, nor have they changed the way that institutions calculate house price indices. Given that volumes of hedonic indices are published each year and that GHHQ (Gatzlaff, Haurin, Hwang and Quigley) have established the need for selection bias correction, why is there a discrepancy between the demand for and supply of corrected indices? Unfortunately, the methods used by GHHQ require data that 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 all those that do not.

A second problem is that existing approaches tend to allow temporal variation either in the determination of the probability of sale, or in the effect of that probability on the price equation, but not both. In reality, there is reason to expect that both will vary because changes in the probability that certain types of property come on the market at different times of the year and different phases of the economic cycle, driven by a the complex interaction of factors that affect the chances of certain types of property coming onto the market

Our goal in this paper is to address both these problems. We attempt to develop a method that: (i) can be applied to data that are readily available in countries such as 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. We also permit both the effect of the estimated sample-selection bias and its determination to vary over time.

The paper is structured as follows. In section 1 we summarise the existing literature and section 2 we explain in more detail the problems with the methods used. In section 3, we describe our proposed method for dealing with transactions bias. In sections 4 and 5 we describe our data and present our results. We conclude the paper with a brief summary.

1. Existing Approaches to Transactions Bias

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 and Hwang and Quigley, 2004). The standard approach applied widely in the wider economics literature and based on the Heckman (1979) two-step selection model, treats sample-selection bias as an omitted variable problem and corrects for it by introducing a term that captures the effect of observations being systematically excluded from the sample. So, if the hedonic price regression being estimated is given by [1], a second regression [2] needs to be estimated to compute the correction term used as an explanatory variable in [1]:

$$y_{1i} = X_{1i}\beta_1 + U_{1i} \qquad [1]$$

$$y_{2i} = X_{2i}\beta_2 + U_{2i} \qquad [2]$$

where y_{2i} is unobserved, and y_{1i} is only observed when $y_{2i} > 0$, the error term U_{2i} is assumed to be normally distributed, allowing [2] to be estimated using probit regression. From the probit regression, one can derive λ_i, the inverse Mills' ratio (often abbreviated to "Mills' ratio"), defined as the ratio of the standard normal density function to the cumulative density function. The procedure has become a standard way of dealing with censored data and is described in most intermediate econometrics textbooks.

There have been various advances on this in the wider economics literature, such as the development of procedures that permit correction for selection bias where the error term in the selection equation is not normally distributed. Olsen's (1980) linear selection model, for example, allows U_{2i} to be uniformly distributed, while Lee's (1983) results allow U_{2i} to have a logistic distribution, derived either from a binary logit or multinomial logit estimation of [2]. To our knowledge, the Lee and Olsen results have not been widely applied in the real estate or urban economics disciplines. The exception is Shroder (2001) who considers the rental real estate investment decision, where y_{1i} is the level of landlord investment and y_{2i} indicates whether or not the household is a landlord. Using Lee's results, Shroder is able to derive consistent estimates of β_1 by including the predicted probability of selection as an additional variable in the OLS estimation of equation [1]. We are not aware, however, of any attempt to create a selection equation where the dependent variable is not binary, which is the situation we are faced with here (see section 3 below).

In terms of correcting sample-selection bias in house price regressions, we should note that there is no agreed set of factors that determine the probability of sale. In the Hwang and Quigley's (2004) study, for example, the selection equation is determined entirely by property attributes and does not include economic and demographic drivers which Gatzlaff and Haurin (1998) find to be important. A notable omission from these papers are equity and loss aversion effects on the decision to sell (see Stein, 1995; Genesove and Mayer 1997, 2001). There is good reason for this (and for their omission in our analysis below). Estimating equity requires one to estimate house value but this in turn requires one to correct for transactions bias. By including local economic and demographic drivers, Gatzlaff and Haurin (1998) are likely to have gone some way to controlling for such effects and this is the approach we use here.

2. Problems with existing approaches

As noted in the introduction, there are two main drawbacks with the GHHQ research.

(a) Requires Data on the Entire Housing Stock

First, 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 entire housing stock at individual dwelling level, which makes it impossible for providers of published house price indices to correct for sample-selection bias.

For example, of the eight main providers of house price indices in the UK: Land Registry (LR), Department of Communities and Local Government (DCLG), Nationwide, HBOS/Halifax, Financial Times, Royal Institution of Chartered Surveyors (RICS), Hometrack, and Rightmove, none attempts to correct for transactions bias. The Land Registry/Registers of Scotland survey is based on the records of all property transactions registered and, as a measure of the value of *traded* properties, there is very unlikely to be any major sampling bias associated with this index. The same is not true, however, if one uses the LR data to measure of the value of the *entire stock* since it only includes properties that transact. The DCLG index is based on mortgage origination data from around fifty lenders, collected through the Survey of Mortgage Lenders. 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, but again there is no attempt to provide a correction. Similarly, the Nationwide and HBOS/Halifax indices are based on mortgage origination data from the loan book records of individual lenders. So, in addition to the bias that results from having no data on properties that do not transact, these indices do not contain information on cash purchases, or on mortgage transactions through other lenders, and the samples used are further potentially biased by variations in the market share across different areas and over time. Both indices use a form of hedonic adjustment to correct for variations in the type of properties traded (see Meen and Andrew, 1998, p. 10) but there is no correction for sample-selection bias.

The Financial Times house price index (also called AcadHPI) is a composite index computed by Acadametrics which combines the Nationwide, HBOS/Halifax and Office of the Deputy Prime Minister (ODPM) / Communities and Local Government (CLG) house price indices with Land Registry (LR) records. The FT approach is founded on the assumption that LR sample is the most comprehensive and least biased of all UK house price data sources. However, there is usually a time lag in the release of LR data, so Academetrics employ an "index of indices" forecasting model to 'account for transactions not yet reported to the Land Registry' (Academetrics 2008). Academetrics claim that their index has been "chosen by the Chicago Mercantile Exchange, the world's largest derivative exchange, for their proposed residential house price derivative which we expect will be launched once the current falls in house prices can be seen nearing an end." (Academetrics 2010). However, note that Academetrics do not appear to measure or correct for transactions bias.

RICS and Hometrack base their results on a survey of market agents, and the results are also potentially distorted by transactions bias because respondents may base their perceptions primarily on properties that trade. There may additional distortions associated with survey-based indices due, for example, to respondents having an incentive to "talk-up" the market when prices and transactions are falling, 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 on 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 on properties offered for sale are reported, and there is no correction for mix adjustment or sample-selection bias arising from such properties not being typical of the stock of all dwellings.

Consideration of the methods used in other countries to compute house price indices (as described on the respective websites) reveals a similar absence of measures and corrections for the sampling bias arising from non-transacting properties. This is the case for hedonic-based indices, such as the IAS360 (USA), EUROPACE HPX hedonic house price index

(Germany), Statistics Norway house price index (Norway), Nasdaq OMX Valueguard Housing Index (Sweden), Permanent TSB/ESRI house price index (Ireland), Indice de Precios de Vivienda (Spain), and one of the RP Data-Rismark Home Value Indices (Australia). Neither is there is any explicit mention of attempts to measure transactions bias in the methodologies of repeat-sales-based indices such as the Federal Housing Finance Agency house price index, the S&P and the FISERV Case-Shiller indices (USA), Teranet – National Bank HPI (Canada), the Woningwaarde Index Kadaster (Netherlands), and the Case-Shiller-based RP- Data-Rismark Home Value Index (Australia). Finally, there is no apparent correction for transactions bias in summary house price indices such as the EUROPACE HPX mean (Germany), Canadian Real Estate Association index (Canada), the Urban Land Price and National Wooden House Market Value Indices (Japan).

The above survey highlights the imperative across many (if not all) developed countries to find a way of correcting for transactions bias that does not rely on data being available on each and every dwelling in the housing stock. The need is made all the more apparent when one considers the role of house price indices in a wide spectrum of economic and policy decisions. They are central to the debate over demand and supply imbalances at the intra and inter regional level and the role of market signals in determining planning decisions (see, for example, Barker, 2003). They are used in the measurement of affordability (Meen *et al*. 2008) and wealth inequalities (Levin and Pryce 2010), and are crucial to understanding the macro economic relationships between housing equity, interest rates and consumer spending (Goodhart and Hofmann 2008). There is also growing interest in the use of house price indices as the basis of derivatives as the

Academetrics (2010) quote above indicates (note the importance of having a correct measure of house price volatility in computing risk-return trade offs in optimal portfolio calculations – something that may be fundamentally undermined by sample selection bias, as we discuss later). And the measurement of house prices may have had a significant role in how the value of bank assets were calculated following Basel II and the use of mark-to-market valuation which may have subsequently exacerbated the liquidity crisis of 2008-2009 (see Levin and Pryce 2010; Hemmer 2008, Sapra 2008). The meaning and reliability of the indices used in each of these respective fields is 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. Of course, for some applications, transactions bias is not relevant – estate agents and lenders, for example, may only be interested in the price trends of properties that actually sell. However, in other contexts, particularly the measurement of housing wealth, the potential for equity withdrawal and 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 need to find ways of measuring and correcting for transactions bias.

(b) Variation in Sample-selection Determination and Bias over Time

A second problem with existing approaches is the failure to allow both the determination of the probability of sale, and the hedonic Mills' ratio coefficient to vary over time. Gatzlaff and Haurin (1998), for example, allow the coefficient on the Mills' ratio to vary, but estimate a single probit on all years. Hwang and Quigley (2004), on the other hand, include time dummies in the probit equation, but do not appear to allow temporal variation of the Mills' ratio coefficient in the repeat-sales regressions.

This may be problematic. For example, if the prices of low-turnover dwellings rise relative to high-turnover properties, then one would expect the coefficient on the Mills' ratio to change over time. 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 tend to have a lower turnover of stock. And there may be periods when the value of low-density housing is likely to rise 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, which increases the supply of high-density housing relative to that of low-density housing. Thus, in certain circumstances, prices of low-density, low-turnover stock would rise in value at a faster rate than high-density, high-turnover dwellings.

It seems implausible, however, that properties that trade infrequently (and hence have a low probability of entering a database of transacted properties in every year), will have a permanently different rate of price appreciation from those that trade frequently (and hence have a higher probability of being traded in a given year). It is more likely that certain types and locations of houses will experience lower rates of price appreciation than the average for a period and then go through a catch-up phase. At least, this is the story one might infer from the findings of the house price convergence literature (Meen, 1999; Cook, 2003, 2009; Holmes and Grimes, 2008) and from the cycles in housing wealth inequality observed by Levin and Pryce (2010). As a result, one might expect both the determination of probability of sale and the hedonic coefficient on the sample-selection correction term to vary over time as particular areas—and particular house types—have bouts of increased/decreased sales volumes, and corresponding periods of divergence/convergence in the rates of price appreciation. So, while the selection effect is unlikely to cause adjusted and unadjusted price levels to diverge inexorably over long periods (it seems implausible to expect non-traded properties to rise at ever greater or lesser rates), one might expect selection bias to affect the short term rate of price increase—i.e. the volatility of prices.

These arguments are closely related to the notion of submarkets. 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 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 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 the values of dwellings have increased at a faster rate than in other subregions. The adjustment process could be less than smooth due to tipping points that arise in the volume of subregional transactions caused by 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 dam-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 loses the expensive low-density properties that 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).

Taken together, these arguments highlight the multifaceted nature of housing transactions and the difficulty in knowing *a priori* what the effect of a particular type of housing or rate of turnover will be on the direction of the sample selection bias. The situation is made more complex by the interaction of spatial, temporal and structural effects. Quality and type of construction of a dwelling along with other factors will determine the desirability of a neighbourhood; the history of planning decisions and economic development will affect the spatial clustering of property types across neighbourhood desirability; market cycles, local demographic trends, and information imperfections will shift selection patterns over time.

3. Proposed Econometric Solution

Whilst regularly updated data on each and every dwelling in the housing stock are not available in many countries (the UK included), it is often possible to access the total number of dwellings in an area (from the UK Postal Address File, for example). Provided the data on house price transactions include the postcode sector, it will be possible to compute the proportion of the housing stock that trades in each area in a given period. By combining this information with data on socio-economic variables that affect the number of properties selling, it is feasible, in principle, to estimate the probability of a property in a given postcode sector selling in a particular period.

If we use the proportion of sales in each postcode sector as our dependent variable, the probability of sale cannot be modelled using standard probit or logit because the dependent variable will not be dichotomous. Neither will OLS yield appropriate estimates because proportions are bounded at zero and one—OLS assumes the dependent variable to be unconstrained and so could predict outside of the feasible range. The solution proposed here is to use Fractional Probit Regression

(FPR) developed by Papke and Wooldridge (1996) to model situations when the dependent variable is continuous and bounded between zero and one, or in fact any situation when the dependent variable is continuous but restricted to an interval [*c*, *d*].[2] Whilst Fractional Logit Regression (FLR) has been applied in the housing/real estate literature to model mortgage debt as a proportion of house value (Hendershott and Pryce 2006) and the determination of estate agent idiom (Pryce and Oates 2008), we are not aware of any housing/real estate applications of Fractional Probit, or of attempts to use FPR nor FLR to estimate the selection equation.

An earlier solution to the problem of modelling variables bounded between zero and one had been 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, $\boldsymbol{\beta}$ 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 $E(y|\mathbf{x})$ either as a logistic function (Fractional Logit Regression), or as a probit function (Fractional Probit Regression), 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). A particularly attractive feature of FLR

[2] Where *c* and *d* do not equal 0 and 1 respectively, fractional probit estimation can be applied by transforming y_2 to ensure that it lies in the [0,1] range. Wooldridge (2002, p. 661) suggests the following simple transformation: $(y_2 - c)/(d - c)$.

and FPR from a practical point of view is that it can be easily estimated using standard software packages,

> "Interestingly, the robust standard errors … in the context of ordinary logit and probit are computed almost routinely by certain statistics and econometrics packages, such as STATA® and SST®. Unfortunately, the packages with which we are familiar automatically transform the dependent variable used in logit or probit into a binary variable before estimation, or do not allow non-binary variables at all (fall into the first category). With the minor change of allowing for fractional *y* in so-called binary response analysis, standard software packages could be used to estimate the parameters… and to perform asymptotically valid inference." (Papke and Wooldridge (1996), p.623).

Fortunately, STATA® have since made the recommended amendment as part of the "glm" command.[3]

Application of FPR to equation [2] opens up the possibility of correcting for sample-selection bias in situations where there is a lack of information at individual level on the whole population, but where there is information on groups of individuals for the whole population. For example, suppose [1] were a hedonic house price equation which we were attempting to estimate for the whole country or region. Suppose we have detailed, individual-level data for each dwelling that sells, and on the neighbourhood where the property is located. However, for the population of dwellings as a whole (including those that do not sell), we lack information at the individual level, and we do not know which property sells or does not sell in a given period. What we do have, however, is information on the proportion of properties in each neighbourhood that sell. So in principle, we could use Fractional Probit regression to model the proportion of properties that sell in a neighbourhood (equation [2]) in terms of neighbourhood characteristics and from this derive the inverse Mills'

[3] The command STATA version 10 is: `glm y2 x2, fam(bin) link(probit) robust`.

ratio, the correction term to be included in the estimation of the hedonic price equation.

This is the method proposed here to correct for sample-selection bias in hedonic house price indices. Consider the following pseudo-Heckman two-step estimation:

$$p = a_0 + a_1\, detached + a_2 semi + a_3 terraced + a_4\,\mathbf{N} + a_5\,\lambda \qquad [1]'$$

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

where:

p	=	ln(price)
s	=	proportion of properties in a particular postcode sector that trade in a given month
λ	=	Mills' ratio, derived from [2]'
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.

One important difference between the approach presented here and the standard Heckman method is the computation of standard errors in the hedonic regression. The usual Heckman computation of standard errors will almost certainly be incorrect because residuals are clustered within years (because many of the explanatory variables are annual rather than monthly) and within postcode sectors (many of the explanatory variables are at postcode sector level, including, of course, the selection term), whereas the dependent variable, house price, is measured at the individual dwelling level.

To address this we adopt a method for computing standard errors that allows for intragroup correlation, relaxing the usual requirement that the observations be independent.[4] This assumes that observations are independent across groups (clusters) but not necessarily within groups. Allowing for intra-group correlation of errors, when combined with the inclusion of neighbourhood variables in the price regression, had a major effect on the hedonic results, reducing the t-ratios considerably. On the basis of these corrected standard errors we refine the hedonic regression, keeping only variables that were consistently statistically significant. The standard errors in the Fractional Probit Regression were computed as specified by Papke and Wooldridge (1996, see summary above).

[4] by using the vce(cluster) option in STATA.

4. Data

In principle, the correction technique described above could be applied to any hedonic house price index provided one is able to source the postcode sector of each house transaction in the sample and then model the proportion that sell in each period using socio-economic drivers measured at that level. Given that the approach could be used in conjunction with many of the UK indices, why did we choose to use Land Registry data? Because the Land Registry data are the most comprehensive sample of house transactions in the UK, they are often viewed as the benchmark by which other UK house price data are judged (see, for example, the Academetrics 2008, 2010 approach to computing the FT house price index). We thought it apt, therefore, to test whether the LR data are characterised by sample selection bias. If so, then all other indices that are based on, or are a subsample of the LR sample (which presumably includes all indices based on mortgage origination data) are potentially subject to transactions bias.[5]

Whether the technique can be applied in other countries depends on knowing the number of houses in each postal area, census tract etc. Provided the number of addresses in each area (however defined) are known, and the data source on house transactions is geocoded to this area, it will be possible to compute the % all houses in each area that enter the researcher's dataset. The advent of mass marketing via junkmail means that it is likely that all residential addresses in a country are held somewhere. At what cost and

[5] We say "*potentially* subject to bias" rather than "inevitably" because it is conceivable (but unlikely) that mortgage lending data is a non-random selection from the LR sample, and that this non-random selection from LR data precisely cancels out any non-randomness intrinsic to the LR sample itself.

spatial level that data are available in each country we do not know but it is something that is likely to be known to indigenous housing researchers.

Data used in the Estimation

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 2004. 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 1.6 million housing transactions over the period 1996 to 2003. The first half of Table 1 lists descriptive statistics on the Land Registry dwelling-level variables, and the second half provides summary statistics on variables measured at postcode sector level.

The selection equation was estimated on 1,198 postcode sectors in the South East of England and variables that explain the probability of sale were collated for each year for each of these. Explanatory 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.

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 bureaucracy 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 screens 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 that of 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 were comprised only of private housing.[6]

[6] This seemed to be the best way to control for social housing, given that in the house price data set we use here we were unable to drop former social-sector dwellings from the sample because we did not know which of the properties that sold were previously owned by social landlords

Table 1 Descriptive Statistics

5 Results

The probability of sale in each postcode sector for each year was estimated by running separate selection equation regressions for each year. The dependent variable in each Fractional Probit 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 burglary, the average distance between dwellings, the proportion of dwellings that were built before 1920, the proportion of semi-detached housing, the percentage change in population over the preceding ten years and the proportion of the population over 65.

As a baseline, we first present the OLS results of these annual regressions (Table 2). On the whole, we were able to explain around a third of the variation in the dependent variable (the adjusted R^2 ranges between 0.308 in 2004 to 0.313 in 1996). This compares very well with the Gatzlaff and Haurin (1998) OLS estimates of the probability of a house selling, which explained less than 1% of the variation in the dependent variable (adjusted R^2 of just 0.003). It is difficult to ascertain how well this compares with Hwang and Quigley (2004) because they only present the probit results, which do not include an adjusted R^2 diagnostic. We can, however, compare significance levels on individual coefficients. Our FPR results are reported in Table 3 run on each year separately and each year's regression yields 4 or 5 variables with t-ratios greater than 2. Again, this compares favourably with GHHQ. In Gatzlaff and Haurin (1998) probit regression, for example, only one variable, age of dwelling, has a t-ratio greater than 2 (their full list of reported t-ratios are 0.3, 0.0, 2.9, 1.6, 0.8, 0.1, 0.9, 1.8 and 0.8). Hwang and Quigley (2004) report six out of nine variables in the probit regression with t-ratios greater than 2.

In our results, the most significant variable was the proportion of socially rented housing (t-ratios of around -13.5). We found that the greater the proportion of socially rented housing in an area, the lower the probability of sale in a given year. Better school performance was significant in all years and tended to raise the probability of sale in each year (perhaps good schools attract a steady inflow of parents seeking access for their children, and a steady outflow of households for whom access to good schooling is no longer of value because their children have left school).

Distance between dwellings also proved to be highly significant in most years (t-ratios of around –9) and to have a negative effect, which suggests that dwellings in low-density areas have a lower probability of trading in a given year, other things being equal. Violent crime, burglary and the proportion of dwellings built pre-1920 did not appear to be particularly significant. Increases in population raised the probability of sale and the effect was statistically significant in all years except 2003 and 2004. The impact of the proportion of households aged over 65 had a positive effect in most years, and the effect was only significantly different from zero in all years. Table 3 presents the FPR results which have a similar pattern of statistical significance to those reported in the OLS regressions (Table 2).

Table 2 Estimation of the Selection Equation: OLS

Table 3 Estimation of the Selection Equation: FPR

Table 4 presents the results of ln(price) regressions run on all years of the data, first without the *Mills' ratio* variable—the estimated correction term derived from the FPR—and then with. These are not the regressions used to compute the price index (instead we use the Fleming and Nellis 1984 method described below) but we present these regressions run on the entire dataset of one and a half million observations to facilitate comparison with GHHQ. Crucially, the Mills' ratio is negative and highly significant (t = -13.4 which compares favourably with the t = 5.4 value reported in Hwang and Quigley, 2004)[7]. This suggests that properties that trade, as reported in the LR data, are not a random subsample of the housing stock, a corollary of which is that most other UK house price indices (which are based on a subset of the LR data) are likely to be characterised by transactions bias.

Table 4 Hedonic Estimates: Regressions Run on All Years

Have properties which are less likely to trade increased in value at a different rate? We investigated this by considering whether the coefficient on the *Mills' ratio* variable changed over time. This could have been achieved by interacting the Mills' ratio with a series of time dummies (as in Gatzlaff and Haurin 1998). However, since one of our goals is to derive a measure of selection-adjusted house price inflation that can easily be updated, we avoided using the dummy variable approach (since the addition

[7] Gatzlaff and Haurin (1998) interact the Mills' ratio with quarterly dummies and do not report the individual significance of each interaction term. They do, however, report the results of an F-test for joint significance; they find F=6.89 compared with a 5% critical value of 1.39.

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). Instead, we adopt the Fleming and Nellis (1984) method (see below) 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 Mills' ratio).

The coefficients on the Mills' ratio from each of these monthly regressions are plotted in Figure 1 along with the 95% confidence intervals. The coefficient is statistically significant (represented by the confidence interval lying entirely above or entirely below 0) in all but four (March 1996, February 1997, September 2000, and October 2001) of the 108 months for which the index is estimated. We have argued that the selection effect may vary because of changes in the probability that properties of a certain type/location trade at different times of the year and different phases of the economic cycle, along with the complex interaction of factors that affect the chances of certain types of property coming onto the market. It can be seen from Figure 1 that in most years the coefficient on the Mills' ratio remains negative, but that there is significant variation from year to year (the upper confidence interval in some months falls below the lower confidence interval in others, and *vice versa*). There is also evidence that the coefficient temporarily changed sign (becoming significantly positive) in two of the months analysed (February 1996, September 2001).

Figure 1 Coefficient on Inverse Mills' Ratio in Hedonic Ln(Price) Equation (with 95% confidence interval)

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), which is useful for our purposes because we want to allow the coefficient on Mills' ratio to vary. Unlike running a single hedonic regression on the entire dataset as in Table 4, the Fleming and Nellis method also has the advantage that it can be readily updated with information on subsequent time periods without changing all previous parameters and index values).[8]

Exponentiated predicted values from each regression using the average set of characteristics from 1996 are presented in Table 5, and the index values for each month are plotted in Figure 2. The cumulative effect over the entire period appears to be that the unadjusted index tends to overstate the true rate of price inflation of the stock of private housing (we observe a 306% increase in the unadjusted index compared with a 261% increase in the adjusted index). This is not dissimilar to the Hwang and Quigley (2004) study which found that the unadjusted index yielded a cumulative price increase of around 370% whereas the Mills' ratio-adjusted index gave a cumulative increase of around 310%. Gatzlaff and Haurin

[8] The time-varying nature of this type of hedonic index calculation may be further justified by the instability of implicit prices suggested by stochastic dynamic general equilibrium theory—see Leung, Wong and Cheung (2007).

(1998) find the opposite effect – the unadjusted hedonic tends to underestimate the rate of price change – but the effect is very small, perhaps due to the weak explanatory power of their probit selection equation.

The main difference between our results and those of GHHQ is that the adjusted hedonic index varies much more from month to month than the unadjusted hedonic index (the coefficient of variation of monthly change = 3.6 and 1.1 for the adjusted and unadjusted indices respectively).[9] This may be due to the fact that we have allowed both the coefficients in the Fractional Probit selection equation and the coefficient on the Mills' ratio in the hedonic regressions to vary over time. If so, previous studies may have overlooked an important aspect of unadjusted series: that they underestimate the month to month volatility in house prices. Indeed, greater volatility, rather than long-term differences in the rates of change, is what one might expect from a selection-adjusted index, given the likely convergence over time of house price appreciation in different housing sectors. However, our time series is too short to draw firm conclusions — further investigation using a longer time series and simulated data would be required to verify whether the cause of this discrepancy with GHHQ is due to genuine volatility in the price of the stock of dwellings, or whether it is a characteristic of our data or method.

Table 5 Adjusted and Unadjusted House Price Indices (South East England 1996 to 2003)

[9] Although GHHQ do not report the coefficient of variation of monthly change in their indices, there is no obvious increase in month to month variation from the graphs they present.

Figure 2 Adjusted and Unadjusted Monthly Nominal Constant Quality Price Indices (South East England 1996 to 2004)

Conclusion

This paper has not solved all the issues associated with the computation of house price indices. We have not, for example, considered how changes in liquidity over the housing cycle affect the interpretation and measurement of house price indices (see Leung, Lau and Leong, 2002; Fisher *et al.*, 2003; Clayton, Miller and Peng, 2010; and Levin and Pryce, 2009); nor have we addressed the implications of hedonic methods for the price index problem raised by Hill and Melser (2008) or the broader set of issues associated with hedonic methods (Malpezzi, 2003; Case *et al*. 2003). Our objective has been focussed on establishing whether it is possible to develop a method for correcting transactions bias, a distortion that is widely acknowledged but generally overlooked—we are not aware of any published house price index across the world that either measures or corrects for the bias that arises from traded properties being a non-random sample of all properties in the stock of housing. We have argued that, while the selection effect is unlikely to cause adjusted and unadjusted price levels to diverge inexorably (it seems implausible to expect non-traded properties to rise at ever greater or lesser rates), one might expect selection bias to affect the short term rate of price increase—i.e. the volatility of prices.

Our approach has been to develop a method that could conceivably correct transactions bias in house price indices where attribute data on individual dwellings are not available for the population of dwellings, 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 Probit Regression was used to derive an estimate of the inverse Mills' ratio for inclusion as a correction term in the hedonic house price regression—similar to the traditional Heckman (1979) approach except that the selection equation explains the probability of sale in each area rather than the probability of sale of each individual dwelling. We found evidence that the inverse Mills' ratio had a statistically significant effect in a simple hedonic price equation, suggesting that sample-selection bias was indeed present. We also found evidence that the coefficient on this correction varied over time, suggesting that selection bias was not constant. Overall, the unadjusted index tended to overestimate the true rate of price appreciation of the stock of private housing (consistent with the findings of Hwang and Quigley (2004) which were based on dwelling-level probit regressions).

However, our results also revealed greater month-to-month volatility, which was not apparent in earlier studies (if anything, the adjusted series in Gatzlaff and Haurin 1998 and Hwang and Quigley 2004 look slightly smoother). While different rates of volatility, rather than differences in long term price appreciation, is what one might intuitively expect from a comparison of selection-adjusted and -unadjusted price indices, further investigation is needed to confirm whether the increase in month-to-month variation is peculiar to our data (one could, for example, construct a simulated population of houses and explore the conditions under which sample selection bias causes smoothing).

In principle, our approach could be adapted to correct for transactions bias in repeat sales indices (such as the Case-Schiller index). For example, there is no obvious reason why the FPR Mills' ratio could not be incorporated into repeat sales indices in much the same way that Gatzlaff and Haurin (1997) and Hwang and Quigley (2004) incorporate the standard binary-probit Mills' ratio into repeat sales and hybrid index estimates. Our method could also be used to correct indices based on subsamples of the traded stock, such as mortgage transactions data (e.g. Nationwide and Halifax in the UK) where the selection bias is potentially greater (because of the exclusion of cash purchases and transactions based on mortgages from other lenders). It is less obvious, however, how our method could be applied to survey/market sentiment based indices, such as those published by RICS and Hometrack, where there is no regression to which the FPR Mills' ratio can be added.

A further area for future research is the effect that sample selection has on the ability to predict changes to the value of the housing stock. Forecasting is likely to be made problematic by the potential for the selection process to change. Finally, we should note that there may be applications of our approach to sample selection problems other than house price index calculation. In principle, the FPR Mills' ratio could be useful whenever probit regressions on individual observations cannot be estimated but where Fractional Probit Regressions can be used to model selection probabilities based on proportions of the population within mutually exclusive areas or groups.[10]

[10] with the caveat that further work needs to be done to understand the econometric properties and sensitivities of our method.

References:

Academetrics (2008) FT House Price Index Nov 2008, www.academetrics.co.uk

Academetrics (2010) What we do, www.acadametrics.co.uk/acadHousePrices.php

Barker, K. (2003) Review of Housing Supply Securing our Future Housing Needs: Interim Report – Analysis. HM Treasury: London.

Case, B., Pollakowski, H.O., and Wachter, S.M. (2003) On Choosing Among House Price Index Methodologies, *Real Estate Economics* 19, 3, 286-307

Clayton, J., Miller, N. and Peng, L. (2010) Price-volume correlation in the housing market: causality and co-movements. *Journal of Real Estate Finance and Economics* 40(1) pp. 14-40.

Cook, S. (2003) The convergence of regional house prices in the UK, *Urban Studies*, 40(11) pp. 2285-2294.

Fisher, J., Gatzlaff, D., Geltner, D. and Haurin, D. (2003) Controlling for the impact of variable liquidity in commercial real estate price indices, *Real Estate Economics*, 31 pp. 269-303.

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

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

Genesove, D. and Mayer, J. (1997) Equity and time to sale in the real estate market, *American Economic Review* 87(3) pp. 255-269.

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

Goodhart, C. and Hofmann, B. (2008) House Prices, Money, Credit and the Macroeconomy", *Oxford Review of Economic Policy*, 24(1), 180-205.

Heckman, J. (1979) Sample selection bias as a specification error, *Econometrica* 47 pp. 153-161.

Hemmer, T. (2008) "Discussion of Marketing-to-market: Panacea or Pandora's box?", *Journal of Accounting Research*, 46, 461-466.

Hendershott, P. and Pryce, G. (2006) The Sensitivity of Homeowner Leverage to the Deductability of Home Mortgage Interest, *Journal of Urban Economics* 60(1): pp.50-68

Hill, R.J. and Melser, D. (2008) Hedonic imputation and the price index problem: an application to housing, *Economic Enquiry* 46 (4) pp. 593-609.

Holmes, M.J. and Grimes, A. (2008) Is there long-run convergence among regional house prices in the UK? *Urban Studies*, 45(8) pp. 1531-1544.

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) pp. 161-178.

Jones, C., Leishman, C. and Watkins, C. (2003) Structural change in a local urban housing market, *Environment and Planning A* 35(7) pp. 1315-1326.

Kim, J. H., Pagliara, F. and Preston, J. (2005) The intention to move and residential location choice behaviour', *Urban Studies* 42 (9) pp. 1621-1636.

Lee, Lung-Fei (1983). "Generalized Econometric Models With Selectivity." *Econometrica*, 51:507-513.

Leung, C. K. Y., Lau, G. C. K. and Leong, Y. C. F. (2002) Testing alternative theories of the property price-trading volume correlation, *Journal of Real Estate Research* 23 (3) pp. 253-63.

Leung, C. K. Y., Wong, S. K. and Cheung, P. W. Y. (2007) On the stability of the implicit prices of housing attributes: a dynamic theory and some evidence, *International Real Estate Review* 10(2), pp. 65-91.

Levin, E. and Pryce, G. (2009) A regional house price model of excess demand for housing, RICS: London.

Levin, E. and Pryce, G. (2010) The dynamics of spatial inequality in UK housing wealth (forthcoming).

Levin, E., Montagnoli, A., and Pryce, G. (2010) Mark-to-Market and House Asset valuation, University of Glasgow mimeo.

Malpezzi, S. (2003) Hedonic pricing models: a selective and applied review, In O'Sullivan, T. and Gibb, K. ed (2003) *Housing Economics and Public Policy*, Blackwell: Oxford.

Meen G (1999) Regional house prices and the ripple effect: a new interpretation", *Housing Studies* 14(6) pp. 733-753.

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.

Meen, G., Andrew, M., Ball, M. Goody, J., Kasparova, D., Pryce, G., Whitehead, C., Wood, G. (2008) *Recent Developments in the Communities and Local Government Affordability Model*, London: Department for Communities and Local Government.

Olsen, R.J. (1980) A Least Squares Correction for Selectivity Bias, *Econometrica*, 48, 7, 1815-1820.

Papke, L. E. and Wooldridge, J.M. (1996) Econometric methods for fractional response variables with an application to 401(k) plan participation rates, *Journal of Applied Econometrics* 11 pp. 619-632.

Pryce, G. and Oates, S. (2008) 'Rhetoric in the Language of Real Estate Marketing', *Housing Studies*, 23(2), pp.319-348.

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

Sapra, H. (2008) “Do accounting measurement regimes matter? A discussion of mark-to-market accounting and liquidity pricing”, Journal of Accounting and Economics, 45, 379-387.

Shroder, M. (2001) What Makes a Landlord? Ownership of Real Estate by US Households, *Urban Studies*, 38, 1069-1081.

Stein, J. 1995. Prices and trading volume in the housing market: a model with downpayment constraints. *Quarterly Journal of Economics* 110(2) pp. 379-406.

Wooldridge, J. M. (2002) *Econometric analysis of cross section and panel data*, Cambridge and London: MIT Press.

Table 1 **Descriptive Statistics**

Variables at Dwelling Level (n = 1,599,859, all years)	**Mean**	**Std.Dev.**
price	£146,817	£145,456
detached	26%	
semi	28%	
terraced	28%	
flat	18%	
Transacted in 1996	9%	
Transacted in 1997	10%	
Transacted in 1998	11%	
Transacted in 1999	12%	
Transacted in 2000	11%	
Transacted in 2001	12%	
Transacted in 2002	13%	
Transacted in 2003	11%	
Transacted in 2004	11%	
Variables at Postcode Sector Level (n = 1,198)		
Former social rented	12.1%	8.9%
Economically active	64.7%	6.7%
Education score	55.32	5.28
Violent Crime	0.9%	0.4%
Burgulary	0.5%	0.2%
Average Dist between dwellings	20.33	17.69
Proportion of Dwellings built pre 1920	24.5%	14.8%
Proportion of dwellings in PCS built 1920-45	19.4%	7.4%
Proportion of dwellings in PCS built 1946-1979	27.8%	10.2%
Proportion of dwellings in PCS built 1980+	28.3%	15.1%
Average height above sea level	58.05	40.12
Average size of dwelling	113.06	15.55
Population change in 10 years preceding 1996	5.6%	6.4%
Population change in 10 years preceding 1997	5.2%	5.8%
Population change in 10 years preceding 1998	5.3%	5.0%
Population change in 10 years preceding 1999	5.0%	4.7%
Population change in 10 years preceding 2000	5.6%	4.3%
Population change in 10 years preceding 2001	5.6%	4.2%
Population change in 10 years preceding 2002	5.6%	4.4%
Population change in 10 years preceding 2003	5.5%	4.2%
Population change in 10 years preceding 2004	5.6%	3.9%

Table 2 Estimation of the Selection Equation at Postcode Sector Level: OLS

	1996	1997	1998	1999	2000	2001	2002	2003	2004
Social rented	-0.029	-0.029	-0.028	-0.029	-0.029	-0.029	-0.028	-0.028	-0.02
	(-14.054)	(-14.013)	(-13.880)	(-13.774)	(-13.833)	(-13.932)	(-14.039)	(-13.964)	(-13.93
Economically active	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.006	0.00
	(1.218)	(1.274)	(1.285)	(1.228)	(1.163)	(1.239)	(1.315)	(1.498)	(1.55
Education score	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.000
	(3.198)	(3.014)	(2.742)	(2.774)	(2.814)	(2.934)	(3.006)	(2.818)	(2.70
Violent Crime	0.010	0.000	-0.006	-0.006	-0.005	-0.008	-0.009	-0.012	-0.01
	(0.213)	(-0.001)	(-0.135)	(-0.128)	(-0.107)	(-0.175)	(-0.190)	(-0.258)	(-0.30
Burglary	0.055	0.059	0.056	0.051	0.050	0.049	0.051	0.058	0.06
	(0.956)	(1.014)	(0.972)	(0.880)	(0.857)	(0.839)	(0.861)	(0.984)	(1.04
Dist. between dwells	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.000
	(-9.490)	(-9.356)	(-9.302)	(-9.312)	(-9.365)	(-9.289)	(-9.262)	(-9.103)	(-9.01
Dwellings pre 1920	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.003	-0.00
	(-1.668)	(-1.762)	(-1.824)	(-1.825)	(-1.793)	(-1.809)	(-1.820)	(-2.056)	(-2.19
Semi-detached	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.00
	(-0.598)	(-0.688)	(-0.737)	(-0.770)	(-0.762)	(-0.796)	(-0.863)	(-1.016)	(-1.11
Population change	0.007	0.006	0.007	0.008	0.009	0.008	0.007	0.005	0.00
	(2.324)	(2.073)	(1.961)	(2.153)	(2.176)	(1.948)	(1.783)	(1.052)	(0.66
Population over 65	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.01
	(4.718)	(4.745)	(4.756)	(4.790)	(4.755)	(4.706)	(4.744)	(4.700)	(4.67
Constant	0.015	0.015	0.016	0.015	0.016	0.015	0.015	0.015	0.01
	(3.981)	(4.086)	(4.164)	(4.144)	(4.148)	(4.150)	(4.139)	(4.233)	(4.30
n	1,198	1,198	1,198	1,198	1,198	1,198	1,198	1,198	1,19
Adj R2	0.313	0.312	0.311	0.311	0.311	0.311	0.311	0.309	0.30

Dependent variable = proportion 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 3 Estimation of the Selection Equation at Postcode Sector Level: FPR

	1996	1997	1998	1999	2000	2001	2002	2003	2004
Former social rented	-0.617	-0.616	-0.616	-0.619	-0.619	-0.617	-0.616	-0.612	-0.609
	(-13.641)	(-13.589)	(-13.455)	(-13.384)	(-13.452)	(-13.529)	(-13.609)	(-13.499)	(-13.447
Economically active	0.135	0.139	0.141	0.138	0.133	0.138	0.142	0.155	0.160
	(1.630)	(1.680)	(1.684)	(1.631)	(1.562)	(1.633)	(1.710)	(1.883)	(1.940)
Education score	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
	(3.424)	(3.239)	(2.952)	(2.974)	(3.015)	(3.161)	(3.248)	(3.080)	(2.978)
Violent Crime	0.097	-0.105	-0.233	-0.226	-0.212	-0.272	-0.281	-0.333	-0.375
	(0.105)	(-0.116)	(-0.260)	(-0.252)	(-0.237)	(-0.303)	(-0.313)	(-0.370)	(-0.417
Burglary	1.123	1.195	1.149	1.037	1.023	1.002	1.026	1.151	1.221
	(1.008)	(1.070)	(1.031)	(0.933)	(0.919)	(0.894)	(0.911)	(1.024)	(1.087)
Dist between dwellings	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
	(-9.523)	(-9.396)	(-9.334)	(-9.348)	(-9.391)	(-9.328)	(-9.319)	(-9.163)	(-9.070
Dwellings pre 1920	-0.021	-0.024	-0.026	-0.026	-0.025	-0.025	-0.025	-0.031	-0.034
	(-0.879)	(-0.978)	(-1.045)	(-1.067)	(-1.026)	(-1.020)	(-1.028)	(-1.251)	(-1.396
Semi-detached	0.013	0.010	0.009	0.008	0.008	0.007	0.006	0.002	-0.001
	(0.481)	(0.389)	(0.345)	(0.292)	(0.305)	(0.280)	(0.208)	(0.067)	(-0.048
Population change	0.142	0.134	0.138	0.156	0.177	0.173	0.159	0.111	0.086
	(2.488)	(2.249)	(2.155)	(2.317)	(2.336)	(2.157)	(2.004)	(1.329)	(0.957)
Population over 65	0.362	0.364	0.365	0.366	0.366	0.362	0.365	0.362	0.361
	(4.621)	(4.654)	(4.673)	(4.709)	(4.684)	(4.627)	(4.671)	(4.628)	(4.606)
Constant	-2.1947	-2.1881	-2.1805	-2.1805	-2.1796	-2.1836	-2.1871	-2.1846	-2.1814
	(-28.418)	(-28.447)	(-28.285)	(-28.333)	(-28.333)	(-28.555)	(-28.671)	(-28.656)	(-28.558
No. variables with t >2	5	5	5	5	5	5	5	4	4
n	1198	1198	1198	1198	1198	1198	1198	1198	1198
ll	-100.37	-100.38	-100.38	-100.38	-100.38	-100.38	-100.38	-100.38	-100.3

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

Table 4 Hedonic Estimates: Regressions Run on All Years

	Without Correction Term	With Correction Term
House is detached	0.661	0.657
	(161.845)	(157.708)
House is semi-detached	0.329	0.334
	(92.712)	(92.037)
House is terraced	0.186	0.195
	(43.315)	(45.817)
Average size of dwelling in PCS	0.010	0.010
	(95.416)	(94.867)
Proportion of dwellings in PCS built 1920-45	-0.088	-0.069
	(-5.511)	(-4.486)
Proportion of dwellings in PCS built 1946-1979	-0.058	-0.038
	(-5.278)	(-3.557)
Proportion of dwellings in PCS built 1980+	0.064	0.076
	(6.049)	(7.347)
Average height above sea level in PCS	0.001	0.001
	(8.488)	(9.895)
Average distance between dwellings in PCS	4.346	4.352
	(21.972)	(21.846)
Education Score in PCS	0.014	0.013
	(22.750)	(21.801)
Violent Crime in PCS	-3.045	-1.880
	(-2.740)	(-1.655)
mills	-	-0.848
	-	(-13.381)
Constant	8.984	11.783
	(214.696)	(54.669)
+ month dummies		
n	1,500,887	1,498,965
Adjusted R^2	0.649	0.652

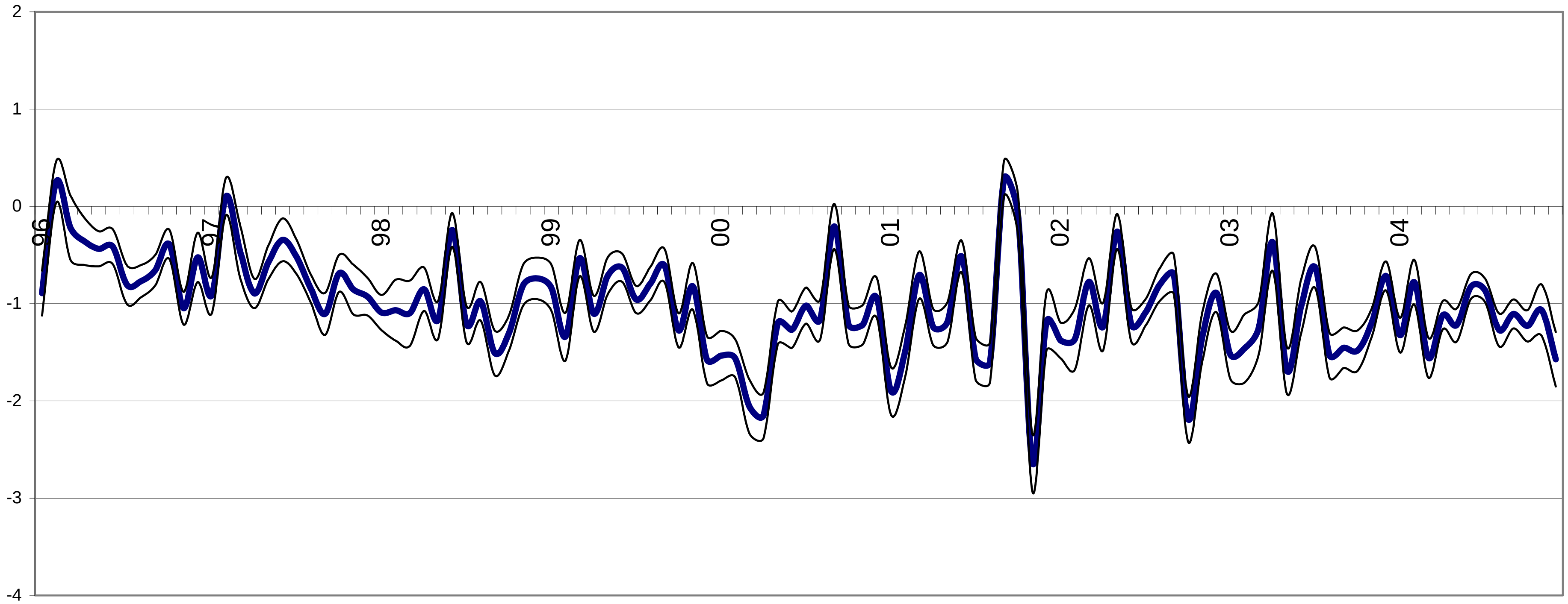
Figures in brackets are t-ratios based on standard errors that allow for intragroup correlation.
PCS = Post code sector

Table 5 Adjusted and Unadjusted House Price Levels in South East of England

Year	Month	Unadjusted	Adjusted
1996	Jan	£ 59,374	£ 78,077
	Feb	£ 59,120	£ 77,524
	Mar	£ 58,754	£ 73,010
	Apr	£ 60,501	£ 74,083
	May	£ 60,914	£ 71,707
	Jun	£ 62,020	£ 74,566
	Jul	£ 63,438	£ 72,067
	Aug	£ 63,737	£ 70,703
	Sep	£ 63,330	£ 73,354
	Oct	£ 63,299	£ 72,124
	Nov	£ 63,349	£ 71,542
	Dec	£ 64,455	£ 74,559
1997	Jan	£ 64,629	£ 79,973
	Feb	£ 64,410	£ 80,406
	Mar	£ 65,678	£ 79,646
	Apr	£ 66,978	£ 79,090
	May	£ 68,160	£ 78,209
	Jun	£ 68,660	£ 78,869
	Jul	£ 70,800	£ 77,889
	Aug	£ 72,658	£ 81,344
	Sep	£ 72,325	£ 82,631
	Oct	£ 72,978	£ 82,524
	Nov	£ 73,418	£ 85,626
	Dec	£ 74,893	£ 85,489
1998	Jan	£ 74,602	£ 91,578
	Feb	£ 74,349	£ 92,526
	Mar	£ 75,563	£ 89,792
	Apr	£ 77,810	£ 90,377
	May	£ 78,521	£ 89,612
	Jun	£ 79,755	£ 89,335
	Jul	£ 81,220	£ 87,461
	Aug	£ 81,995	£ 91,569
	Sep	£ 81,497	£ 92,304
	Oct	£ 81,318	£ 92,350
	Nov	£ 80,710	£ 92,981
	Dec	£ 82,028	£ 94,414
1999	Jan	£ 80,043	£ 99,882
	Feb	£ 81,367	£ 101,686
	Mar	£ 82,374	£ 95,918
	Apr	£ 84,189	£ 96,293
	May	£ 85,325	£ 98,375
	Jun	£ 87,467	£ 96,091
	Jul	£ 88,745	£ 93,634
	Aug	£ 89,717	£ 96,563
	Sep	£ 92,044	£ 101,078
	Oct	£ 92,394	£ 100,725
	Nov	£ 93,519	£ 103,182
	Dec	£ 96,367	£ 107,780
2000	Jan	£ 96,937	£ 116,838
	Feb	£ 96,889	£ 115,191
	Mar	£ 99,700	£ 111,307
	Apr	£ 102,932	£ 116,885
	May	£ 104,498	£ 118,332
	Jun	£ 107,626	£ 118,580

	Jul	£ 108,925	£ 121,000
	Aug	£ 110,262	£ 123,704
	Sep	£ 109,471	£ 126,493
	Oct	£ 108,166	£ 126,190
	Nov	£ 109,614	£ 127,854
	Dec	£ 111,382	£ 127,382
2001	Jan	£ 110,693	£ 138,047
	Feb	£ 111,004	£ 137,454
	Mar	£ 112,131	£ 131,065
	Apr	£ 115,985	£ 133,192
	May	£ 117,233	£ 130,698
	Jun	£ 119,506	£ 130,986
	Jul	£ 121,263	£ 129,333
	Aug	£ 122,617	£ 129,113
	Sep	£ 123,330	£ 139,072
	Oct	£ 122,887	£ 136,714
	Nov	£ 122,813	£ 137,881
	Dec	£ 125,872	£ 146,627
2002	Jan	£ 125,721	£ 152,962
	Feb	£ 125,269	£ 151,047
	Mar	£ 128,222	£ 146,392
	Apr	£ 131,114	£ 148,249
	May	£ 135,391	£ 142,550
	Jun	£ 138,790	£ 156,365
	Jul	£ 142,318	£ 148,278
	Aug	£ 146,462	£ 153,406
	Sep	£ 147,432	£ 164,513
	Oct	£ 148,898	£ 164,991
	Nov	£ 152,408	£ 168,830
	Dec	£ 153,940	£ 175,878
2003	Jan	£ 153,956	£ 182,197
	Feb	£ 153,586	£ 186,623
	Mar	£ 154,929	£ 190,564
	Apr	£ 157,900	£ 186,791
	May	£ 158,708	£ 184,065
	Jun	£ 159,870	£ 187,615
	Jul	£ 162,888	£ 181,452
	Aug	£ 164,871	£ 181,652
	Sep	£ 164,377	£ 181,579
	Oct	£ 164,987	£ 177,060
	Nov	£ 166,229	£ 181,035
	Dec	£ 166,861	£ 182,562
2004	Jan	£ 167,597	£ 209,472
	Feb	£ 168,623	£ 210,859
	Mar	£ 169,417	£ 197,589
	Apr	£ 173,134	£ 198,898
	May	£ 176,155	£ 203,505
	Jun	£ 177,747	£ 195,897
	Jul	£ 181,828	£ 192,114
	Aug	£ 183,124	£ 196,447
	Sep	£ 184,687	£ 202,814
	Oct	£ 181,556	£ 198,223
	Nov	£ 183,491	£ 203,139
	Dec	£ 181,883	£ 203,831

Figure 1 Coefficient on Inverse Mills Ratio in Hedonic Ln(Price) Equation (with 95% confidence interval)



Note: a table of the values plotted here can be obtained upon request from the authors.

Figure 2 Adjusted and Unadjusted Monthly Nominal Constant Quality Price Indices (South East England 1996 to 2004)