Office of the
Deputy Prime Minister
Creating sustainable communities

New Horizons Programme Which House Price? Finding the Right Measure of House Price Inflation for Housing Policy

Technical Report

Office of the
Deputy Prime Minister
Creating sustainable communities

New Horizons Programme Which House Price? Finding the Right Measure of House Price Inflation for Housing Policy

Technical Report

April 2006
Office of the Deputy Prime Minister: London

By Gwilym Pryce and Philip Mason

Although this report was commissioned by the Office, the findings and recommendations are those of the authors and DO NOT necessarily represent the views of the Office of the Deputy Prime Minister. This report will form part of our evidence base when tackling future issues and policies.

Office of the Deputy Prime Minister
Eland House
Bressenden Place
London
SW1E 5DU
Telephone: 020 7944 4400
Website: www.odpm.gov.uk

Further copies of this publication and alternative formats are available from:

ODPM Publications
PO Box 236
Wetherby
West Yorkshire
LS23 7NB
Tel: 08701 226 236
Fax: 08701 226 237
Textphone: 08701 207 405
Email: odpm@twoten.press.net

or online via the ODPM website: www.odpm.gov.uk

Printed in the UK on material containing no less than 75% post-consumer waste.

April 2006

Product Code 05 ASD 03771/a

Contents

1 Introduction

1.1 Context

House prices are notoriously difficult to predict. There is no better illustration of this than the recent failure of housing economists to accurately forecast the end of the UK housing boom. Housing pundits and analysts have been queuing up to predict when the market would turn, and although the housing market is now indeed less than buoyant in many parts of the country, this is no proof of their forecasting abilities. As a recent article in the Financial Times aptly observes,

> "Even a stopped clock gives the right time twice a day… We have been predicting the end of the housing boom for so long that, sooner or later, one of us must get it right".[1]

Indeed, the article goes on to document in detail the predictions of imminent doom in the housing market since the turn of the millennium. The most notable example is the very bold prediction in November 2002 by the high-profile economist Andrew Oswald:

> "I think we are about to go through the great housing crash of 2003 to 2005… I advise you to sell your house, and move into rented accommodation… Panic will then set in…"[2]

A website has even been established, dedicated to predicting the deflation of the housing bubble (www.housepricecrash.co.uk).

Now that the "crash" is here, it seems far less severe than many anticipated, with house prices still rising in certain parts of the country (such as the West of Scotland – see Pryce, 2005, Housing Market Commentary) and possible signs of recovery already on the horizon elsewhere.[3] Indeed, the most recent figures from the Land Registry (Table 1-1) demonstrate that prices have risen by more than the rate of general inflation (2.5%) in most regions over the past year.

1 Jim Pickard, Financial Times, Property Special, Sept 2005, pp. 6-13.

2 Quoted in Jim Pickard, Financial Times, Property Special, Sept 2005, p.9.

3 The most recent survey from the Nationwide (www.nationwide.co.uk/hpi/), for example, suggests a slight pick-up in house prices. Nationwide estimate house prices to have increase by 1.3 % in October alone, bringing the annual growth rate to 3.3%.

Table 1-1 Land Registry House Price Results

Region	Jul-Sept 2004	Jul-Sept 2005	Price increase %
Wales	£135,162	£145,188	7.42%
Yorkshire	£133,552	£141,188	6.01%
North	£123,606	£130,948	5.94%
North West	£133,878	£139,929	4.52%
Greater London	£287,470	£300,329	4.47%
South East	£227,991	£234,833	3.00%
East Anglia	£174,949	£180,053	2.92%
East Midlands	£151,405	£155,630	2.79%
West Midlands	£159,203	£161,076	1.18%
South West	£201,156	£202,249	0.54%
England & Wales	£187,971	£194,589	3.52%

Source: Land Registry

But do the house price indices published by Land Registry, ODPM and the main lenders, mean what we think they mean? Our interest, in this report, is not in whether we can accurately predict movements in future price trends – that is a question we already know the answer to, as the recent Financial Times survey quoted above clearly demonstrates. Instead, our enquiry considers a still more basic question: whether house price indices can even give us a reliable account of the past.

Time was when house price data were released but once a year, and then only for the whole country. We now have a plethora of different index providers for ever-smaller geographical areas, released with ever-greater frequency. But what do these indices really tell us? Are they all measuring the same thing, and do they truly reflect changes in the value of the underlying housing stock? If we cannot manage to measure accurately past movements in prices, how can we hope to predict future trajectories? Our concern, then, is with the reliability of house price indices and, more specifically, with the implications of possible bias in the samples used to compute indices.

1.2 Aim

This report considers five questions about the reliability of house price measurement:

1. Does it matter whether house prices are measured accurately or not?

2. Where does the sample come from?

3. What is the mix adjustment?

4. What about those properties not recently sold?

5. What can be done to correct for transactions bias?

Since questions 1, 2 and 3 have already been the subject of much attention, our focus will be on the final two questions which have not been studied in any great depth in the UK. Our two main goals are to identify whether there is any evidence that house sales in the UK are non-random, and whether non-randomness (if it exists) can be corrected for, given the limitations of available UK data.

These two tasks are not trivial. Indeed, as our investigation proceeded, we began to realise why no previous UK research existed on these topics! It transpires that methods developed elsewhere, most notably in the US, cannot be applied in the UK because of data inadequacies. New techniques would need to be developed to analyse these topics in any great depth. Necessity, however, is the mother of invention and these difficulties forced us to think of new ways of tackling the transactions-bias problem. We believe that the following pages contain a great deal of innovation that offers potentially new possibilities for house price adjustment, not only in those regions studied (Scotland and the South East) but also in the remaining UK regions and indeed in other countries).

1.3 Plan

The question of whether the reliability of house price measurement is of any great import is our logical starting point and is the subject of Chapter 2. We consider briefly the implications of reliable house price measurement for (1) the analysis of demand and supply imbalances; (2) the measurement of affordability and wealth inequalities; (3) the measurement of the impact of new supply; and (4) modelling the impact of house prices on consumer spending. We conclude that, because house prices affect so many fundamental factors in the economy, personal finance and the planning system, reliable measurement is paramount.

In Chapter 3 we examine the various existing measures of house price change in the UK. We consider in particular the differences in the samples used in these measures, and differences in the methods of calculating the index. A particularly important issue – one which has received a great deal of attention in the literature – is how to adjust for differences in the mix of properties coming onto the market in different time periods. The final section of the chapter introduces the topic of greatest interest to us: what about properties that rarely trade? All current indices are based on transactions data – information on price and dwelling type gleaned from properties that have recently sold. But is this data a truly random sample of the entire housing stock in the region of interest? We cite evidence from US literature suggesting that it is unlikely that traded properties are typical of all dwellings.

Chapter 3 considers approaches developed in the housing economics literature to measure and correct for transactions bias. We provide new insights into the theoretical weaknesses of these approaches and, most importantly, we demonstrate their very limited usefulness to the UK (and presumably to many other countries) because of the absence of up-to-date information on the entire housing stock.

Chapter 4 expands on one of the key theoretical weaknesses of the received wisdom identified in Chapter 3; that of duration dependence. In particular, we explore methods of measuring and correcting for duration dependence in the probability of sale that are not only a methodological improvement generally, but also open up the possibility of providing sample-selection correction to UK indices in future.

Existing studies of transactions bias have focussed either the impact of variation in economic/demographic forces that influence the probability of sale, or on the implications of property type on that probability. Both assume that property types are randomly distributed across space. In Chapter 5 we argue that this is a major omission. Spatial concentrations of particular property types, neighbourhood types and socio-economic factors, will conspire to cause non-randomness across space in the probability of sale. If this is the case, then the possibility arises for a method of correction that could be applied to all parts of the UK because detailed data now exist on the characteristics of areas down to the level of individual postcodes. Given the potential practical importance of being able to identify and measure systematic spatial variation in the probability of sale, we investigate the phenomenon in-depth, analysing and comparing a vast amount of data on the South East housing market. The result is an overwhelming body of evidence demonstrating that the probability of sale does indeed vary systematically across space. We then consider ways in which the variation in the probability of sale could be used to correct house price indices. Such corrections would, of course, be somewhat redundant if their impact is negligible. Transactions bias may well exist, but if it does not affect our estimation of house price inflation to any great degree, then there is no compelling case for changing existing measures to include the correction. The most important element of Chapter 5 (and possibly of the whole report) is the results of our initial estimation of the impact of selection bias. We compare unadjusted results with those from a simple correction procedure and find significant differences, not only in estimated house price levels, but also in house price trajectories.

Chapter 6 summarises the findings of the report and consolidates the case for introducing transactions-bias correction to the existing house price index measures in the UK. The report concludes with a series of recommendations.

2 The Importance of Reliable House Price Measurement

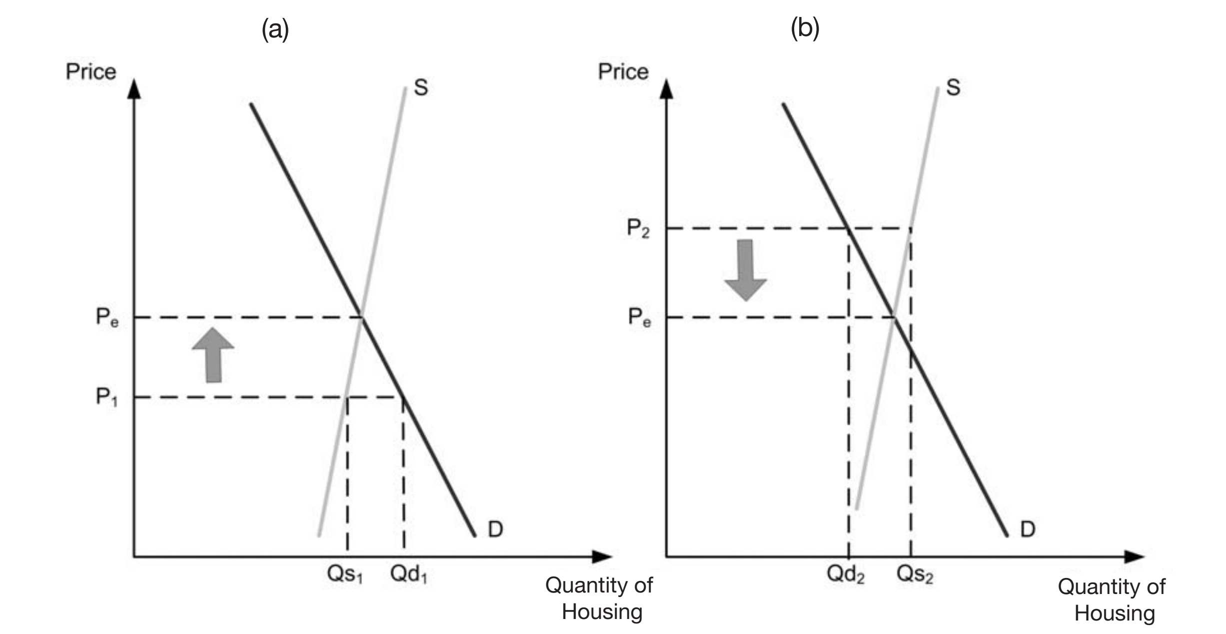
2.1 Introduction

Prices play a pre-eminent role in our understanding of any market. Housing is no exception. It is, however, more complex than most goods because each dwelling represents a unique mix of attributes and location. As a result, no one method of measuring house prices can capture all aspects of housing market dynamics. The challenge for housing economists and policy makers is finding the right measure of house price change for the task in hand. There are already many different indices available. A problem common to them all (in the UK at least), however, is that they are transactions-based: all the information used to calculate the indices is taken from records of recent sales. Dwellings sold infrequently are underrepresented in such indices, and this has the potential to cause bias. This is not the only problem that plagues house-price measurement but it is potentially one of the most crucial and it is certainly the one that has received the least research, at least in the context of the UK. Before we discuss these problems in more depth, we first need to justify the attention paid to house price measurement, and the potential introduction of new methods of calculation and bias-correction. In this chapter we look at a number of important implications of house prices that necessitate accurate measurement of the value of the housing stock.

2.2 Demand and Supply Imbalance

The role of house prices in facilitating housing market adjustment is depicted in the two diagrams of Figure 2-1 below. The relationship between housing supply and price is depicted by the upward sloping line denoted S – as prices rise, quantity supplied rises (although not by as much as we might expect or hope for – see Barker 2003, 2004; Pryce 2004). The relationship between housing demand and house prices is represented by the downward sloping line denoted D – as house prices rise, one would expect demand for housing to fall, all other things being equal. In panel (a), price is initially at P1. At this price, quantity demanded, Qd1, exceeds the supply of housing, Qs1. As a result, there is upward pressure on price – house prices rise until housing demand equals housing supply. So, rising house prices are indicative of excess demand. The converse is also true, as depicted in panel (b). Here prices are too high at P2, and quantity supplied, Qs2, exceeds quantity demanded, Qd2. There is downward pressure on prices and prices fall until demand equals supply once more.

Figure 2-1 Price Adjustment



Prices only change if there is disequilibrium – either excess demand, or excess supply. House price changes therefore provide a very important element to our understanding of demand and supply imbalances. Planners, for example, can analyse house price changes to identify areas of demand and supply imbalances. Indeed, one of the recommendations of the Barker review was that land allocations should be based on a better understanding of local housing market dynamics. In other words, full use should be made of the information incorporated in local house price series and other market indicators when making land planning decisions.

In practice, this amounts to comparing house price trajectories for different parts of a region, local authority or city in order to ascertain where greater supply is most needed. Crucial to this analysis is the assumption that differences in the rate of growth of available price indices reflect genuine differences in the appreciation of the underlying stock of housing between areas. Unfortunately, as we shall demonstrate in subsequent chapters, failure to adequately correct for property characteristics and/or failure to correct for the bias caused by the absence of untraded properties, can distort estimated price trajectories, and these distortions will not be the same for all areas, frustrating comparison of house price inflation across space.

2.3 Measuring Affordability and Wealth Inequalities

A possible unintended consequence of policy founded on transactions-based indices may be to exacerbate the price differential between desirable and undesirable properties, and hence, exacerbate housing wealth inequality. If standard indices implicitly encourage the construction of dwellings of a type that are frequently traded, then prices will be depressed in this sector relative to the larger, more desirable properties in the infrequently traded sector. As the recent Shelter report indicates, there are important long-term sociological implications of the growing housing gap that should not be ignored in the pursuit of short-term policy goals:

> "… children born this century will be starting life more financially unequal than has been the case since Victorian times… [T]he growing inequality in housing is marginalising a whole section of society... Those whose parents have housing wealth are more likely to be advantaged in childhood and to benefit from financial assistance, for example, in finding their own homes… For the children of the poor there will be large parts of the country to which they cannot consider moving in the future even if they should wish to. When they have problems in their lives, there will not be recourse to family wealth to bail them out, to help with a time when they cannot work or find work, to help pay their way through university…"

Whether or not one shares the political philosophy upon which the Shelter report is predicated, one cannot deny that there are potentially far reaching implications of using house price measures that potentially encourage policy decisions that inadvertently exacerbate housing wealth inequality. For example, if high density properties trade more frequently but have a lower rate of price appreciation, policies that encourage the construction of high density dwellings will ostensibly reduce overall price inflation because an outward shift of supply of high density properties will cause their price to fall, and since such properties trade more frequently, the boosted supply will have a disproportionately large downward effect on average price. However, there is nothing in this policy to suppress the values of existing low-density dwellings, and so they will continue to appreciate at least the same rate as before. The outcome in the long run will be an inevitable polarisation of housing wealth between those living in high-density accommodation, and those living in low-density dwellings.

It is of considerable importance, therefore, to ascertain whether there are indeed differences in the rate of sale, and/or of price appreciation, between different levels of density, or other neighbourhood/property characteristics. And if such differences do exist, there is an imperative to develop methods of correcting house price measures so that they accurately reflect changes in the value of the whole stock, not just those that frequently trade.

2.4 Measuring the Impact of New Supply

For some uses of house price information, sample selection bias does not matter. Estate agents, and to some extent mortgage lenders, for example, may be content with knowing only the price trends of properties that trade. After all, the revenues of firms in these industries are not typically affected by changes in the values of properties that rarely trade, with some notable exceptions (such as estate agents that specialise in a niche market that includes rarely traded properties). However, for certain policy decisions, particularly for those relating to the supply of new housing, transactions bias could result in significant distortions. This is because houses that sell infrequently tend to belong to different submarkets than those that sell frequently and have a different pattern of response to general economic conditions and also to new supply.

Following HM Treasury Barker Review of UK housing Supply (2004), it has become clear that one of the key requirements of a house price measurement is that it must be enable the production of reliable estimates of the impact of new supply. The Barker Review Interim Report (2003, p.58) estimated that the current number of new homes per annum would need to be doubled in order to "achieve the European trend rate", and “more than double to get real price stability”. Such estimates are founded on transactions-based indices, however, and we need to ask whether these are the appropriate measures for such analysis.

The argument applied above to the analysis of sub-regional differences in house price appreciation, applies also to macro indices. If, for example, newly constructed dwellings tend to be of a type that sell frequently, they will be in the same submarket as properties that repeatedly enter the official indices of house price inflation, and may appear to have the desired dampening effect on those indices. However, these new dwellings may have relatively little impact on the prices of infrequently sold properties, particularly if the latter are of a markedly different size and type. So the value of the overall housing stock may well be appreciating at a rate above the target rate of house price inflation, even though macro house price measures suggest the target has been met.

There is a need, therefore, to find a practical way of correcting for sample selection bias in UK macro house price indices. If the policy target is to reduce house price inflation, then policy makers have to be aware that current measures carry with them an intrinsic bias towards frequently traded properties. Does this computational bias reflect the desired policy bias? It may well do. For example, if it is true that luxury, low density properties trade less frequently and continue to appreciate at an unabated rate, this may be of little consequence if the entire focus of policy is only to ensure that properties at the lower end of the market remain affordable.

Such a policy priority has two major drawbacks. First, there is the exacerbation of wealth inequality noted above, and with it the implications for equality of opportunity and continued intergeneration polarisation the *haves* and *have-nots*. Second, length of stay may to some degree be measure of consumer satisfaction with property type and neighbourhood. For example, given the emotional and pecuniary upheaval associated with moving house, a family may only consider moving if they anticipate a significant improvement in living standards from doing so (other things being equal). A major reason for moving will therefore be dissatisfaction with current living conditions due to lack of space, unsuitable layout or due to neighbourhood problems. A policy that effects a major increase in the stock of frequently sold dwellings may achieve the superficial policy goal of reducing the rate of change of a transactions-based price index, but at the same time have resulted in an increase in the proportion of the total housing stock with which homeowners are generally dissatisfied.

In depressed areas, the goal of policy may be to actually raise house prices (or this may at least be perceived to be an indicator that successful regeneration has taken place). Pryce and Gibb (2003) have presented preliminary evidence that new construction adjacent to a deprived area can have a regeneration effect. This positive force for change may be all the more potent if the dwellings being constructed are of high quality and of the kind that encourage the formation of stable, long-stay communities (rather than high-density, high-turnover estates). Existing transactions based measures of house price appreciation may therefore underestimate the positive effects of regeneration in those areas if the index includes few long-stay properties.

2.5 Private Sector Investment Decisions

Understanding the role of frequency of sale is not just of relevance to policy makers, however. The bias it implies for house price indices has the potential to distort private sector investment decisions. The lack of an appropriately adjusted house price index will be a source of uncertainty for potential investors in either new construction or potential landlords. To make appropriate financial decisions, such investors need to be able to readily compare the performance of the housing sector with that of other tradable assets such as stocks and bonds and so lack of information in the housing asset market relative to other asset markets will further reduce the attractiveness of housing construction as a destination for investment funds. This applies to the macro decisions of institutional investors, but also to the small-scale decisions of thousands of smaller investors and landlords who also have to make rational choices about how best to use their funds.

And not least, it applies to individual households for whom property values play an increasingly important role in their planned provision to fund retirement. Since the success of pension provisions now rests so crucially on the expectations about future house price growth, price measures that either under or over estimate the true rate of house price inflation will affect the realism of those expectations and have potentially far reaching consequences.

2.6 Modelling the Impact on Consumer Spending

There is a close correlation between cycles of consumption growth and house price growth. Although much of this correlation is likely to be due to consumption and house prices being influenced by the same common factors (particularly those which cause revisions to households' expected lifetime income, such as productivity growth or tax changes – see Orazio et al, 2005; Aoki et al 2003), at least part of the correlation will be causal. For example, house price increases raise household wealth, which in turn raises desired expenditure. Also, house price growth "increases the collateral available to homeowners, reducing credit constraints and thereby facilitating higher consumption" (Orazio et al, 2005 op cit).

Note, however, that distinguishing between the "common factors" explanation and the "causal explanation" is perhaps a false dichotomy. As recent research by Giuliodori (2005) has demonstrated, one of the crucial roles of housing markets is the part it plays in the monetary transmission mechanism (MTM). Interest rates affect consumer spending directly, but they also affect house prices which in turn affects consumer spending:

> "The results indicate that in those countries (particularly the UK) with more competitive mortgage markets and more efficient housing systems, house prices appear to play an important role in the transmission of an interest rate shock to household consumer spending." (Giuliodori 2005, p. 539).

Untangling the relationship between house price appreciation, interest rates, productivity, consumption and saving is clearly of major importance in understanding how the macro economy functions. It is also crucial to our ability to construct statistical models that accurately predict future changes in aggregate demand in response to changes to housing market factors. Such models rest ultimately, however, on reliable measures of the variables involved. If our measure of house prices is an increasingly poor measure of the true value of the overall housing stock (because it only includes properties that trade or get re-mortgaged), then models of the macro economy will produce increasingly distorted interpretations of the relationship between the housing market and other aggregate variables.

2.7 Conclusion

In this chapter we have argued that reliable measurement of house price appreciation at the sub regional level is essential for the appropriate identification of demand and supply imbalances and of efficient land allocation. We have also argued that existing measures of house price inflation could potentially seriously underestimate the rate of wealth polarisation over the long run and hence disguise the potentially harmful effects on social inclusion and cohesion that this implies. Transaction-based price indices may also distort policy decisions, and give lead to false impressions of policy success, particularly with regard to the achievement of house price inflation targets. Reliable house price measurement is also needed if we are to accurately model the impact of new housing supply, or the relationships between the housing market and key macro variables.

Given the range and magnitude of the potential effects of bias in house price measurements, there is a very great imperative to establish whether such bias exists and whether it can be corrected.

3 What do Current Measures Mean?

3.1 Introduction

The number and frequency of house price indices has bourgeoned in recent years. This does not necessarily mean that we understand more about the value of the housing stock than we did ten years ago. Each index is based on a particular source of information, and each uses a particular method to calculate proportionate change in prices. In this chapter we ask three crucial questions about the construction of house price indices in the UK:

1. Where does the sample come from?

2. What is the mix adjustment?

3. What about properties that have not recently sold?

3.2 Where does the sample come from?

A detailed summary of the indices discussed below is presented as a table at the end of this chapter. In this section we draw on that information to answer the three questions posed above regarding the meaning and reliability of each method.

3.2.1 LAND REGISTRY/REGISTERS OF SCOTLAND

1. **Source of the sample:** The Land Registry survey comes out once every three months and is based on the records of property transactions registered over the period. Since it is a legal requirement for property transactions to be logged with Land Registry, the data used is potentially comprehensive (though certain transactions, such as repossessions and property transfers following a divorce, are deliberately 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.

2. **Mix adjustment:** Only very basic details on properties are recorded (particularly in Scotland) and so there is no mix adjustment.

3. **What about properties that do not trade?** As a measure of the value of the entire housing stock, this index is potentially biased because there is no correction for properties that do not trade.

3.2.2 ODPM/SML

1. **Source of the sample:** This survey is based on mortgage origination data from around fifty lenders, collected through the Survey of Mortgage Lenders. Until around two years ago, this survey was only a sample of 5% 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.

2. **Mix adjustment:** Mortgage origination data 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.

3. **What about properties that do not trade?** As a measure of the value of the entire housing stock, this index is potentially biased because there is no correction for properties that do not trade, or for properties traded without a mortgage.

3.2.3 NATIONWIDE AND HALIFAX

1. **Source of the sample:** Both surveys 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, nor 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.

2. **Mix adjustment:** Both indices use a form of hedonic adjustment to correct for variations in the type of properties traded over time.

3. **What about properties that do not trade?** There is no correction for properties that do not trade, or for properties traded without a mortgage, or for bias in the respective market shares of the two lenders.

3.2.4 ROYAL INSTITUTION OF CHARTERED SURVEYORS (RICS)

1. **Source of the sample:** This survey is based 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 a statistical 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).

2. **Mix adjustment:** No formal adjustment made.

3. **What about properties that do not trade?** Given the difficulty of ascertaining the different between price trajectories of traded vs untraded properties even when advanced statistical methods are applied to historical data, it seems unlikely that the current perceptions of respondents will be able to adequately capture movements in the value of the entire stock.

3.2.5 HOMETRACK

1. **Source of the sample:** Similar to the RICS survey, 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).

2. **Mix adjustment:** No formal adjustment made.

3. **What about properties that do not trade?** As with the RICS survey, this is really a measure of changes in the prices of traded properties, rather than of the entire housing stock.

3.2.6 RIGHTMOVE

1. **Source of the sample:** The sample is based 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 are reported, and it is possible that bias could emerge due to the difference between asking and selling prices is not constant either across areas or over time (see Pryce 2004).

2. **Mix adjustment:** No formal adjustment made.

3. **What about properties that do not trade?** No correction is made for properties that do not come onto the market.

3.2.7 FINANCIAL TIMES

1. **Source of the sample:** Compares 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.

2. **Mix adjustment:** the mix adjustment 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".

3. **What about properties that do not trade?** No correction is made for properties that do not come onto the market, and so like the Land Registry index, is potentially subject to transactions bias.

3.3 The Importance of Mix Adjustment

Particularly important when looking at price movements for a small area. Changes in the type of dwellings coming onto the market can have a big impact on average price. E.g. mansion comes onto the market pulls up the average price in that month for that area, even if prices have not really changed.

A particular problem with a regional and sub-regional measures of house price change is how to account for the fact that sales in some areas will be dominated by a type of property not common in others. For example,

> "transactions in the West End of Glasgow tend to be of traditional, stone tenements; whereas in other areas of the city, semi-detached houses or some other dwelling type may tend to dominate the sales figures. Comparing the headline average value of properties sold in such contrasting districts can therefore be misleading. This is borne out in the GSPC data. In 1999 quarter 1, the average price of a dwelling sold in the West End was £57,806, which was less than the average price of properties sold in South Lanarkshire (£59,197). However, if it were possible to compare prices of an identical dwelling in the two areas, anyone familiar with the localities would expect West End prices to be much higher." (Pryce, 2004, Housing Market Commentary No.1).

Changes occur also over time in the size and type of dwellings that come on the market, though this is less of a problem for the computation of regional or national indices where short fluctuations in type tend to be minimal when taken over a very large number of transactions. Indices based on smaller samples, however, are susceptible to distortions caused by one or two large or expensive properties entering the sample for a particular period, inflating the average price even if the value of all traded properties has remained unchanged.

3.3.1 HEDONICS

The heterogeneity of housing means that even if data were available on the whole housing stock, any estimate of house price change would need to control for differences in the attributes of dwellings. The hedonic method typically uses regression analysis to decompose the value of a house into its constituent parts, such as structural attributes and location factors. Estimated values of each component can then be used to derive the average value of a constant quality dwelling over a given time period. Although a theoretical basis for the hedonic approach has been well established (Rosen, 1974), a range of specification issues have emerged in the literature (see Linneman 1980; Halvorsen and Pollakowski 1981; Butler 1982 and Malpezzi 2004). Nonetheless, hedonic analysis is widely regarded as the best approach to correcting for variations in the mix of dwellings coming onto the market, and it forms the basis for the mix-adjustment procedures of both the Nationwide and Halifax indices.

3.3.2 REPEAT SALES

An alternative to the hedonic approach, is the repeat sales method. Although rarely used in the UK because of data limitations (Jones et al 2002 is a rare example), it is the most commonly used means of calculating price indices in the US. The method is simple. Prices on properties that sell more than once are compared between each sale to give a very accurate measurement of the price appreciation of those individual properties. The average of these price increases is then calculated for particular regions. Unfortunately, the repeat sales approach is very prone to sample selection bias because only properties that trade more than once enter the index.

3.4 Properties not recently sold

Although the hedonic approach used by Nationwide and Halifax will help minimise distortions that arise when an atypical selection of properties come onto the market in a particular period, it does not take into account distortions that arise from the fact that certain types of property, and/or properties in certain areas, rarely come onto the market at all, and that these properties may be appreciating at a different rate to the bulk of properties that enter the index more frequently.

3.4.1 IMPACT OF UNSAMPLED PROPERTIES ON HEDONICS

If properties that do not sell are on average similar to those that do, then hedonic estimation will be unbiased. If, however, properties that do not sell are different, then hedonic estimation may be biased. Particularly if the contribution of particular attributes to the overall price of a house is different for untraded properties.

In Figure 3-1, below we assume that high quality properties in desirable surroundings rarely enter the market and so are under-represented in the sample used to draw the regression line that is used to create a constant quality /mix-adjusted price index. For simplicity, we assume only one dwelling attribute – floor area. If the slope of the line = 100, then it means that for traded properties, an extra square metre adds £100 to the value of a house.

In Figure 3-2 we compare this regression line from the one that would be obtained if the values of all properties in the housing stock were included. We see that a regression line based on the entire stock would have a slope of £130. Hedonic predictions of the average price of a dwelling of a particular size will therefore underestimate the true average value if the data used comprises only traded properties.

The problem is compounded if untraded properties have a different rate of price appreciation. This is demonstrated in Figure 3-3 where difference between the regression slope based on transacted properties and the slope based on all properties widens over time. In such circumstances, the traditional uncorrected hedonic approach could seriously under or over predict the rate of increase in the value of the housing stock, particularly over a prolonged period.

Figure 3-1 Regression Line Based Only on Properties that Trade

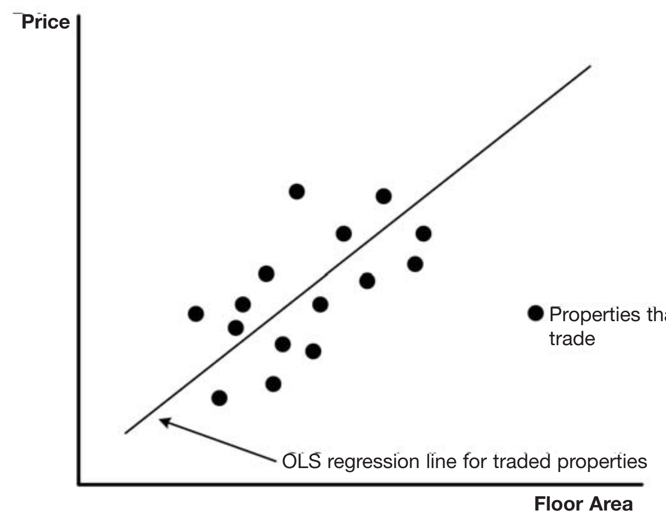


Figure 3-2 Regression Lines With and Without Properties that Do Not Trade

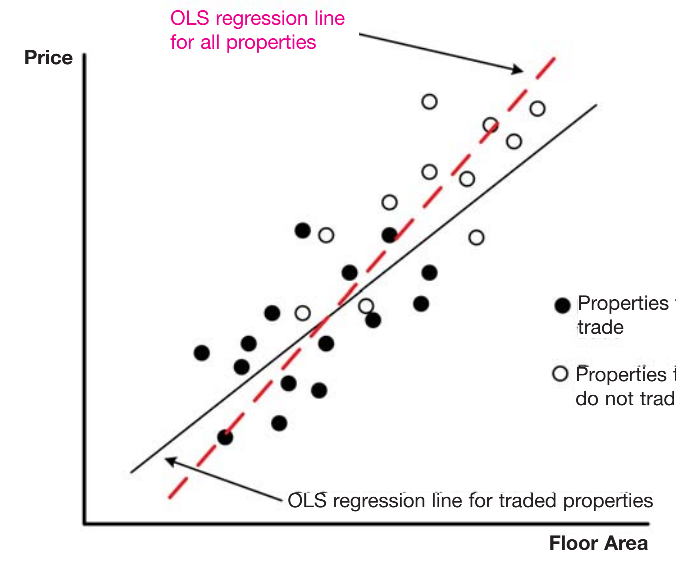
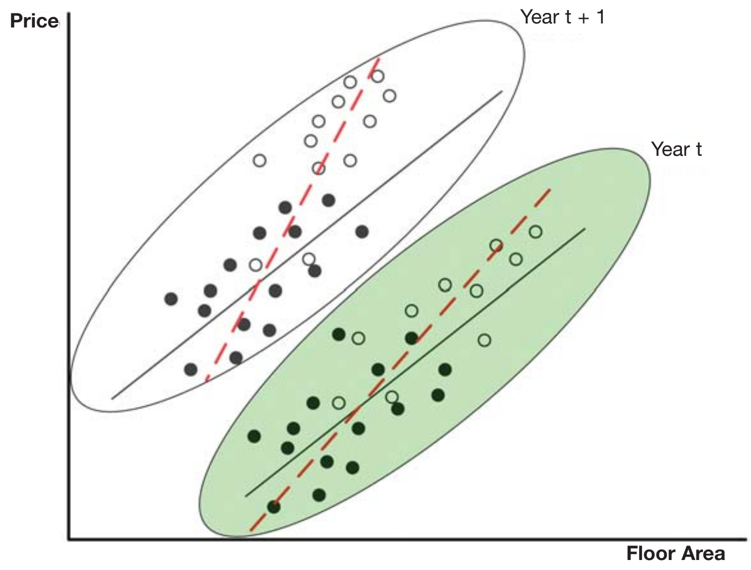


Figure 3-3 Hedonic Regression Bias when Untraded Properties have a Different Inflation Rate



One of the most important aspects of the discussion initiated in the previous chapter is how the choice of house price measure has implications for national and regional supply policy. If the target is to reduce house price inflation, then policy makers have to be aware that current measures carry with them an intrinsic bias towards frequently traded properties. Does this computational bias reflect the desired policy bias? It may be, for example, that there are differential house price inflation rates between different sectors of the housing market according to frequency of sale. If less frequently traded property is experiencing higher inflation than high turnover dwellings, then there may be a case for giving greater policy priority to alleviating demand pressures in the infrequently traded sector.

And, as noted earlier, policy implications are magnified if length of stay is to some degree a measure of consumer satisfaction with their property type and neighbourhood. A policy that effects a major increase in the stock of frequently sold dwellings may achieve the superficial policy goal of reducing the rate of change of a transactions-based price index, but at the same time have resulted in an increase in the proportion of the total housing stock with which homeowners are generally dissatisfied. House prices are also crucial to a wide range of personal and corporate financial decisions, not to mention the implications for the macro economy and the monetary transmission mechanism.

The absence of transactions bias correction from all current measures of house price change in the UK is therefore cause for concern. Our survey thus far provide a very strong motivation for investigating the extent of sample selection bias and to explore possible solutions.

3.5 Conclusion

In this chapter we have surveyed the current measures employed in the UK to gauge house price change. We found considerable differences in the sample bases used to compute the different indices that are currently published, indicating that in fact each index is in fact measuring something slightly different.

Two welcome developments in recent years are: 1. the more widespread use of mix adjustment methods, and 2. attempts to construct a combined index that makes use of the best aspects of the component measures. Unfortunately, neither of these advancements deal with the potentially serious problem of transactions bias.

Table 3-1 Detailed Summary of UK House Price Indices

	Land Registry	ODPM / SML	Nationwide	Halifax	RICS	Hometrack	Rightmove	FT
Precise name of index	Residential Property Price Report (not an index as such)	Survey of Mortgage Lenders All Lenders Index	–	Halifax House Price Index	–	Hometrack Index	Rightmove House Price Index	FT House Price Index MA
Measure	Sales (mortgage and cash)	House price info for properties that reached completion during month for which mortgage provided	Agreed price after survey and mortgage approval	Mortgage transactions at the time they are approved rather than when they are completed	Whether prices went up or down or stayed the same in the previous 3 months	Whether prices are going up or down, most frequently listed properties and prices achieved on recent sales	Current residential asking prices as they come on the market via Rightmove's member estate agents over the previous month	Sales (mortgage and cash)
Property classification variables	property type (detached, terraced, semi-detached, flat)	Property type, age of property, floor area, tenure, garage, central heating, number of bathrooms, number of bedrooms; old/new; FTB/PO.	Types of property (detached, semi-detached, terraced and flats); FTB/PO; 3 property ages (New, modern and older).	Type of property; age of the property; tenure; number & type of rooms; central heating; Number of garages/spaces; Garden; Land area if greater than one acre; Road charge liability.	3 property ages (New, modern and older) and property types	Property types (detached, semi-detached, terraced, flats/apartments)	Property types (Detached, Semi-Detached, Terraced, Flat, Unique, and House)	
Frequency	every 3 months	monthly since Sept 2003; every 3 months 1968-2002	monthly and every 3 months	monthly and every 3 months	monthly	monthly	monthly and annual	monthly and annual
Sample	Obligatory registration of (almost) all residential property sales.	mortgage origination data from c50 mortgage lenders, >=5% sample of completions from all lenders. >5% from some lenders since 2001. 1since mid-2003: ~25,000 completions per month; 1993-2002: 26,000-35,000 transactions/year. Represents c25% of all mortgage completions. Provision of data is voluntary and may be submitted late	in-house mortgage origination data sufficient sample size to produce representative house price series, based on number of new loans written i.e. amount of gross lending for house purchase. Property sales financed by mortgage lending only;	in-house data based on mortgage approvals	information from c.350 chartered surveyors information from survey of "Hometrack recommended" estate agents in towns in E&W with population >10,000.	50-60% of properties on the market at any one time, a sample of more than 120,000 – 150,000 new properties per month	Based on LR figures, so based on all sales in E&W	

	Land Registry	ODPM / SML	Nationwide	Halifax	RICS	Hometrack	Rightmove	FT
Excluded from sample	1. All commercial transactions; 2. Before Jan 2000 – All sales below £10,000 and over £1million. (On line data for Jan 2000 onwards includes details of these sales). 3. Transfer, conveyances, assignments or leases at a premium with nominal rent which are: 'Right to buy' sales at a discount, subject to a lease, subject to an existing mortgage, to effect the sale of a share in a property, by way of a gift, by way of exchange, under a Compulsory Purchase order, under a court order, to Trustees, Vesting Deeds, Transmissions or Assents, of more than one property, Leases for 21 years or less	remortgages, purchases by sitting tenants; outliers (>5 SD)	no cash transactions; remortgages and further advances; right to buy sales at discounted price; Floor size has to be within specified limits for give property type – e.g. detached house has to have >=400 sq. ft floor area	transactions that do not constitute a fully consistent body of data for the purpose of house price analysis are excluded from the Indices. These exclusions primarily cover property sales that are not for private occupation and those that are likely to have been sold at prices which may not represent 'free' or 'normal' market prices, for example, most council house sales, sales to sitting tenants, etc. Only mortgages to finance house purchase are included; remortgages and further advances are excluded.	property sales that are not for private occupation and those that are likely to have been sold at prices which may not represent 'free' or 'normal' market prices, for example, most council house sales, sales to sitting tenants, etc. Remortgages and further advances are excluded	n/k	Land, Commercial Property, Hotels, Mobile Homes, New homes, and properties which are more than three standard deviations above the average or below the average (mean) price	Presumably the same as those excluded from LR records
Extent of cover	England & Wales	UK	UK	UK	England, Scotland & Wales	England & Wales currently, rest of UK in the future	England & Wales	England & Wales
Minimum area of resolution	Postcode (although ~20% of properties not registered by pcode)	Major regions	UK (monthly), UK and 13 regions (quarterly)	UK (monthly), UK and 12 regions (quarterly)	Region	postcode (preferably from 2 estate agents within same postcode area)	postcode area	Postcode (since it is based on LR data)

	Land Registry	ODPM / SML	Nationwide	Halifax	RICS	Hometrack	Rightmove	FT
Methodology	Straightforward collation of sale prices	Follows recommendations of "House Price Working Group"; regression model; annual chain-linked Laspeyres-type index (like Retail Prices Index)	updated in 1993 following publication of 1991 census data, so more robust to lower sample sizes because it better identifies and tracks representative house price	multivariate regression	net-balance survey	various analytical methodologies and data elements e.g. comparable sales prices, home characteristics, historical property price appreciation etc.	calculate average asking price per postcode-property type combination	Compares 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. Assumes LR data is unbiased.
Seasonal adjustment	No	In the future (not enough time-series data so far)	yes	yes	yes	no	no	yes
Mix-adjustment	No	yes	yes, hedonically	yes, hedonically	No formal adjustment	No formal adjustment	No formal adjustment	yes
Other weighting or adjustments	No	expenditure weighting (=transaction weights applied to average cell prices)	transaction weighting	transaction weighting	property type and postcode area	timeliness (<> seasonality)		
Source of information	http://www.landreg.gov.uk/	http://www.odpm.gov.uk/index.asp?id=1156183	http://www.nationwide.co.uk/hpi/method_qs.htm	http://www.hbosplc.com/economy/index methodology.asp	www.rics.org and telephone conversation with Ryan Emmet	http://www.hometrack.co.uk/	http://www.rightmove.co.uk and, for example, http://www.rightmove.co.uk/pdf/p/hpi/HousePriceIndex17October2005.pdf and telephone conversation with Charlotte Wheeler (0845 4568439)	http://www.acadametrics.co.uk/ftHousePrices.php and http://news.ft.com/cms/s/1d089640-fb60-11d8-8ad5-00000e2511c8.html

	Land Registry	ODPM / SML	Nationwide	Halifax	RICS	Hometrack	Rightmove	FT
Notes	Most comprehensive, most "authoritative", but long time lag between completion and receipt of details by LR and publication; limited information for each purchase	No cash sales included (~25% of the market), although aim to include this in the future; appears 2 months in arrears; expenditure weighting gives London and SE greater index. Year-on-year inflation rate reflects increase in total house-buying expenditure over past year. Index confirms earlier signals (eg from Halifax & Nationwide)	indices are based on *mortgage approvals* – so reflect house prices at least one month ahead of completion. Processed immediately, no wait for receipt of data from other lenders. Consequently provides much earlier indicator of house price trends than ODPM index. Over long periods Halifax and Nationwide series follow similar patterns, as both use similar statistical techniques to produce prices – differences because representative property differ in make up from that of Halifax	indices are based on mortgage approvals – so reflect house prices at least one month ahead of completion. Processed immediately, no wait for receipt of data from other lenders. Consequently provides much earlier indicator of house price trends than ODPM index. Over long periods Halifax and Nationwide series follow similar patterns, as both use similar statistical techniques to produce prices – differences because representative property differ in make up from that of Nationwide		Established 1999. Claims to be the only independent property research and database company in the UK	owned by four of the UK's biggest estate agency chains – Halifax, Countrywide Assured, Royal & Sun Alliance and Connells. However, ownership and operation of all mission critical business functions are provided in-house. Established in 2000	Calculated by Acadametrics for FT
Transactions bias correction:	No	No	No	No	No	No	No	No

4 Measuring and Correcting Transactions Bias – Existing Approaches

4.1 Introduction

Our conclusion in the last chapter was that none of the existing measures of house price change currently published in the UK correct for sample selection bias. As we shall see in this chapter, there is very good reason for this: the approaches that have been developed elsewhere (largely in the US) require detailed and up to date information on the entire housing stock. Since this information is not readily available in the UK, these techniques are not readily applicable.

The situation may not, in fact, be as gloomy as it may first seem. The correction methods deployed in the US are not, as we shall demonstrate in this chapter, not without their drawbacks, even when the required data are available. Before presenting our critique we shall set out the theoretical model that underpins the US techniques. We then identify a number of possible shortcomings in this model. These weaknesses point to alternative correction methods which may, ironically (and indeed, somewhat fortuitously), be more amenable to estimation in the UK data context than the methods currently being employed.

4.2 Heckman Correction

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[4] of a property coming onto the market. This estimated probability[4] can then be entered into the sale price equation to correct for sample selection bias. More specifically, let S_{it}* be the latent variable that drives the decision of household i of whether or not to sell in period *t:*

$$S_{it}^* = P_{it}^O - P_{it}^R \qquad [1]$$

4 More precisely, the inverse Mills ratio is calculated.

where P_{it}^{O} is the offer price and P_{it}^{R} is the sellers reservation price. Offer prices are determined as follows,

$$P_{it}^{O} = V_{it} + e_{it}^{O} \qquad [2]$$

where e_{it}^{O} refers to the seller's preferences/information (assumed to be random normally distributed) and V_{it} is the market value of the house, which is a function of attributes and location,

$$V_{it} = \Sigma\alpha X^{H}_{ijt} + \Sigma\beta\ D_t \qquad [3]$$

where X^{H}_{ijt} is a vector of j property and neighbourhood characteristics of the ith property at time t, and D_t is a dummy variable equalling 1 in the period that V_{it} is observed (t = 1 is omitted).

The seller's reservation price, P_{it}^{R}, depends on the opportunity cost of waiting for a better offer, and on the seller's knowledge of the determination of the value of the property, and the distribution of potential offers:

$$P_{it}^{R} = V_{it} + e^{R}_{it} \qquad [4]$$

It follows from [1], [2] and [4] that the latent variable determining whether the owner sells or not is given by the difference between the random two random variables associated with buyers and sellers respectively,

$$S_{it}^{*} = P_{it}^{O} - P_{it}^{R} = e_{it}^{O} - e_{it}^{R} \qquad [1]$$

Gatzlaff and Haurin (1998) argue that these two error terms will determined by personal attributes x^p, and local or national macro economic factors z_t:

$$e_{it}^{O} - e_{it}^{R} = f(x^p, z_t) \qquad [5]$$

Although S_{it}^{*} is not observable, we can observe the outcome S_{it} of the household decision of whether or not to sell, where S_{it} = 1 if the household decides to sell, and = 0 otherwise. These observed values can be explained using a probit regression with S_{it}^{*} as the dependent variable, and x^p, z_t as the explanatory variables. This estimated regression can be used to derive, λ_{it}, the hazard of non-selection ('inverse Mills ratio')

$$\lambda_{it} = \lambda_{it}(x^p, z_t) \qquad [6]$$

Applying Heckman's (1979) result to hedonic estimation, it can be shown that the omitted variable bias associated with having a non-random sample in the OLS estimation can be overcome by including λ_{it} as an explanatory variable. So in,

$$P_{it} = \Sigma\alpha X^{H}_{ijt} + \Sigma\beta\ D_t + \gamma\ \lambda_{it} + u_{it} \qquad [7]$$

α and β are unbiased, where P_{it} is the observed selling price on properties that actually sell.

4.3 Problems with Heckman Correction

4.3.1 PROBLEM 1: DURATION DEPENDENCE

There is, however, a crucial missing ingredient in Gatzlaff and Haurin's (1998) specification of the seller's reservation price in [4], the current length of stay, L_{it}. If L_{it} is included,

$$P_{it}^{R} = V_{it} + L_{it} + e^{R}_{it} \qquad [4]\#$$

then,

$$e_{it}^{O} - e_{it}^{R} = f(x^{p}, z_{t}, L_{it}) \qquad [5]\#$$

and,

$$\lambda_{it} = \lambda_{it}(x^{p}, z_{t}, L_{it}) \qquad [6]\#$$

which suggests that the estimation of the hazard of non-selection also has to include the length of stay. Why might we expect duration dependency? There are nine distinct reasons

1. **Unpacking:** households cannot consume the durable goods they possess while they are still in boxes!

2. **Customisation:** properties often bought for their potential, rather than the actual condition/style of the house at the point of purchase. So the optimal consumption of housing will not be achieved until several months, if not years, of decoration, renovation and improvement have taken place.

3. **Social Capital:** it takes time to establish good relationships with neighbours so it is worthwhile staying put until those connections are well developed.

4. **Schooling:** frequent moves disrupt the human capital accumulation of children – in other words, parents have an incentive not to move to often if they want to optimise the educational experience of their children.

5. **Employment:** frequent moves may be perceived as a negative signal to future employers. It may, for example, be interpreted as being indicative that the person has little commitment to work, or fails to get along with colleagues.

6. **Equity:** Stein/Genesove and Mayer have argued that people only move when they have accumulated sufficient equity in their homes to pay for the cost of relocating.

7. **Marketing Time:** the time taken to sell a property may vary over time or across regions. This will affect length of stay because:

   length of stay = time-off-the-market + time-on-the-market

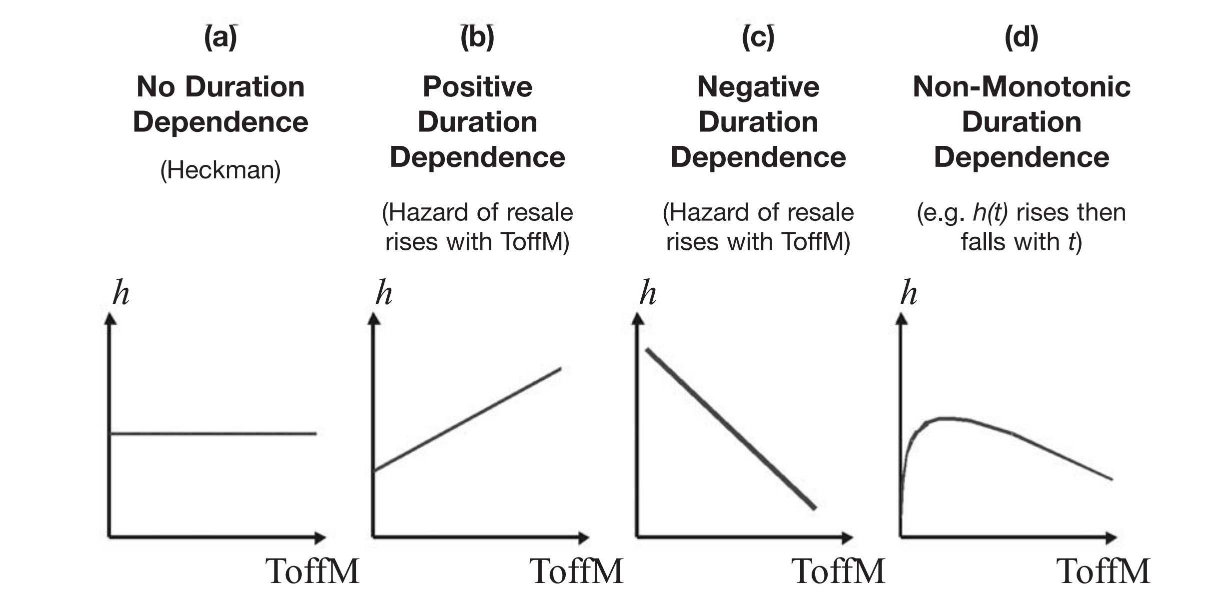
8. **Consumers seek to minimize transaction costs over their lifetime:** other things being equal, the sum of moving costs over a person's lifetime will be lower if they make fewer moves.

These factors combine to make the probability of sale vary the longer someone is living in a property. This is called *duration dependence*, and if it is present, the most appropriate way of estimating the probability that a property will come onto the market is to use time-to-event estimation techniques such as Cox Proportional-Hazard estimation and Log-logistic regression. Applying duration analysis to the estimation of λ_{it} would therefore overcome an important potential weakness in the Gatzlaff and Haurin (1998) analysis. By applying probit rather than a duration model approach that allows for duration dependence, Gatzlaff and Haurin implicitly assume that there is no "duration dependence" in the process by which properties come onto the market. That is, the longer a property remains off the market does not in any way affect the probability of it entering the market in the next period. This is akin to saying that a person who moves into a property is equally likely to put the house on the market the following day as he is in ten years time. However, given the emotional upheaval and transactions costs associated with moving, it is highly unlikely that this will be the case.

In order to choose the appropriate form of sample selection correction, we need to ascertain whether there exists some degree of duration dependence. If there is zero duration dependence then the probability of a property coming onto the market estimated can reliably estimated from a simple probit regression (as in the Heckman two step approach adopted by Gatzlaff and Haurin). If this were the case, we would expect the "hazard" of a property coming onto the market to remain unaffected by how long it had been off the market. *Hazard* is a statistical term which for the purposes of the current discussion can be interpreted the same as "probability" (the only difference is that hazards vary from zero to infinity, whereas probabilities vary from zero to one).

If the hazard of sale does not change the longer someone stays in a property, then the hazard curve will be a horizontal line as depicted in panel (a) of Figure 1. If, however, the hazard of a property coming onto the market increases the longer remains off the market, then we would expect the hazard curve to be upward sloping, as represented in panel (b) of Figure 4-1 below. The converse is depicted in panel (c). Finally, the hazard curve may be non-linear, rising (falling) initially, reaching a zenith (trough) and then declining (increasing) or undulating in a regular or irregular pattern. Panel (d) depicts a simple hill-like hazard function.

Figure 4-1 Hazard Function Types for Different Levels of Duration Dependence



4.3.2 PROBLEM 2: NON-NORMALITY IN THE PROBABILITY OF SELECTION

Research on the Heckman two-step model has revealed that its results are highly sensitive to the assumption of normality (see Greene 2003; Goldenberger 1983). This means that if the probability of a property trading in a given period is indeed normally distributed, then the Heckman approach is likely to work very well. Unfortunately, if the probability is in fact non-normal, the Heckman correction may itself be a source of estimation bias.

This is a particular concern given the above discussion since the existence of duration dependence is likely to imply that the probability of sale is almost certainly not normal but in fact a highly skewed distribution. Such non-normality is likely to be further exacerbated by the spatial variation and spatial spillover effects that constitute our next two criticisms of the Heckman correction approach.

4.3.3 PROBLEM 3: SPATIAL VARIATION

We have so far assumed that the proportion of properties that sell varies randomly across space. In some periods, more properties will sell in area A than in area B, but in subsequent periods, the reverse may be true. It is assumed that in the long run, the proportion of properties that sell has no systematic spatial pattern. But is this likely to be true? The Hwang and Quigley (2004) application of the Heckman approach models the probability of sale as a function of dwelling characteristics (living area, utility area, lot size, tiled bath, sauna, detached, laundry, winter quality walls, slate/ copper roof) whereas the Gatzlaff and Haurin (1998) application models the probability of sale as a function of macro-economic indicators, property characteristics, and occupant characteristics. (Note that the Gatzalff and Haurin data is "limited to single-family detached homes with between 600 and 6000 square feet of living area and less than five acres of land" op cit, p.209, which presumably precludes the potentially important variation in frequency of sale and house-price inflation due to variation in property type).

But the proportion of properties that sell in any given period may also vary according to neighbourhood characteristics. Indeed, this may well be the most potent single determinant because other factors (such as household life-cycle stages, economic factors, and duration dependence) are all likely to be clustered across space. Economic and social deprivation are highly spatially (hence the development of deprivation indices in both Scotland and England and Wales, and location-targeted area regeneration initiatives). Dwelling types have a strong spatial dimension due to the lumpy and non-random nature of land release and the different building styles of particular periods and of particular developers. Finally, household types are likely to non-randomly distributed across space due to the above (spatial patterns economic indicators and dwelling types) and due the preferences of households to locate adjacent to other households with similar characteristics.

A particularly important potential spatial pattern is the impact of housing density on the probability of sale. Assume, for sake of argument, that households generally have a preference for low-density living, and that this preference increases in the latter stages of the life-cycle. One conception of the housing consumption decision is that people keep moving up the housing ladder until they find a property that most closely matches their preferred type and location. Once achieved, they remain there and only move when there is a significant change to their family or employment circumstance (such as children leaving home, divorce or job relocation). The corollary of these two assumptions is that duration of stay will be much longer in low density areas and therefore the proportion of dwellings that trade in any given period in such areas will be significantly lower than in high-density areas.

4.3.4 PROBLEM 4: SPATIAL SPILLOVERS

Not only are there likely to be concentrations of probability of sale in particular areas, but also spatial interactions with neighbouring areas in the determination of those probabilities. There are several reasons for this.

(1) **Relationships and family ties:** First, one of the motives for household A locating and remaining in a particular area is the proximity of that location to friends and relatives. If those friends and relatives decide to move, this will obviously affect the probability of household A also putting their house up for sale. The complex network of relationships that makes up local bedrock of social capital implies complex chains on transaction contingency.

(2) **Changes to local amenities:** The location of new amenities (such as the relocation of a school; the closure of a train station) rarely affects a single neighbourhood alone. Other localities will also be affected, albeit to diminishing degrees. The proportion of sales in each neighbourhood will therefore be correlated with the incidence of sale in neighbouring areas.

(3) **Crime and the Fear of Crime:** Some of the most important determinants of the decision to move are the most difficult to calibrate. The arrival of a 'problem family' in a neighbourhood can have a massive effect on the preference for relocation of neighbouring families, and rising crime rates in one area can profoundly shape the fear of crime in surrounding neighbourhoods, if not the actual risk of crime has remained unchanged. The dissemination across space of perceptions about crime and other social ills leads, at least in principle, to another reason to expect spatial correlation in the turnover of the housing stock.

(4) **Many moves are local:** In many areas, a large proportion of moves are local. People seek to move upmarket, or downsize, without losing connections to friends, relations, schools and other local amenities. So if someone puts their house on the market in area A, that gives another household in adjacent area B the opportunity to relocate. If a household in area B does indeed move, that in turn offers an opportunity to households in proximity to area B to also relocate.

This phenomenon arises because the vast majority of housing transactions are second hand. Therefore, it is only possible to move into one's house of choice if the current owner of that house decides to offer it for sale, which invariably means they in turn need to move out and find accommodation elsewhere (which is often a relatively local move). Housing chains (both latent and transactional) therefore present a very strong a priori reason to expect spatial spill-overs in the incidence of sale.

(5) **Equity gains by submarket:** US research (Genesove & Mayer; Stein and others) suggests that the probability of moving is a function of equity. Households are averse to moving if it entails making a net loss. They therefore tend only to move, given the choice, when they have accumulated enough equity in their dwelling to cover transactions costs. Combine this principle with the Law of One Price which suggests that properties in the same or similar submarkets likely to appreciate at the same rate (see Jones et al), and the corollary is that the probability of sale will also be correlated across similar submarkets. Since submarkets can be aspatial (Goodman and Thibeadeu; Bourassa), distance alone will not capture the spatial correlation – some measure of the similarity of neighbourhoods would also be needed.

However, there has, as yet, been no attempt to incorporate spatial dependency into existing models of transaction bias.

4.3.5 PROBLEM 5: DATA REQUIREMENTS

In addition to the duration dependency problem, the applicability of Gatzlaff and Haurin (1998) approach to the UK context is limited since comprehensive data on unsold properties are rarely available. The Heckman approach assumes that you have extensive information on the characteristics of properties that have not sold. Even if the above problems in the Heckman approach did not exist, it would be of little use to UK housing analysts because of data adequacy problems.

5 Dealing with Duration Dependence

5.1 Introduction

Our initial intention in exploring the phenomenon of duration dependence was to identify duration of stay from repeat sales of Land Registry data for the South East over a prolonged period (such as ten to fifteen years). It quickly became apparent, however, that this would not be feasible since Land Registry data for England and Wales has only recently been supplied with sufficient property type and location information to make the identification of repeat sales possible. As a result, we resorted to extending the example used in the Pryce (2004) ODPM report which was based on SASINES/Registers of Scotland data which have long been made available with detailed location information.

5.1.1 EVIDENCE OF TRANSACTIONS BIAS: THE EVIDENCE SO FAR

To illustrate the kind of biases endemic in existing price indices, Pryce (2004) considered the number of properties in each West of Scotland local authority that sold either once, twice, three times, four times or five or more times in the 1991 to 2000 period. He found evidence of considerable 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 next question to consider is whether there are different patterns of house price values for different rates of turnover. If so, is possible that house price indices are being distorted by the systematic variation frequency of sale. Other than Pryce's (2004) preliminary analysis, there is very little work to draw on to other than studies based on data from the US (for example: Gatzlaff and Haurin 1994, 1997, 1998; Fisher et al 2003) or Sweden (Hwang and Quigley 2004). Certainly, the international studies suggest that the frequency of sale is not independent of property type. Hwang and Quigley, for example, compute the average characteristics of dwellings for dwellings that sell only once, and compare these averages to dwellings that sell twice, three times, four times etc. Their results (reproduced below in Table 5-1) clearly demonstrate a relationship between frequency of sale an dwelling attributes. A cursory reading of the data suggests that larger, detached properties generally tend to sell less frequently, but upon closer inspection, the figures suggest that the relationship is more complex. This is more clearly seen if we plot the data as a bar chart, as in Figure 5-1 which shows a non-linear relationship between lot size and frequency of sale. The same is true of the relationship between the proportion of detached properties for each category of sale frequency (Figure 5-2) and, to a lesser extent, the average living area (Figure 5-3).

Table 5-1 Average Characteristics of Stockholm Dwellings as a Function of Sales Frequency

Number of sales	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Number of Dwellings	59,550	20,202	5,839	1,296	239	26	2
Price (000s SEK)	980.62	956.49	949.04	935.31	977.54	1006.20	1087.81
Living area (m^2)	124.24	123.46	122.12	118.68	120.46	114.00	119.50
Utility area (m^2)	40.49	38.32	37.69	37.39	38.44	44.21	18.00
Lot size (m^2)	1,117	912	803	839	824	1,039	457
Tiled bath (yes = 1)	0.1699	0.1670	0.1573	0.1588	0.1431	0.1667	0.0000
Sauna (yes = 1)	0.2152	0.2244	0.2242	0.2253	0.2243	0.2692	0.4286
Detached (yes = 1)	0.6653	0.6137	0.5833	0.5874	0.6033	0.6923	0.2143
Laundry (yes = 1)	0.7555	0.7798	0.7799	0.7510	0.7824	0.7115	0.9286
Winter quality walls (yes = 1)	0.8170	0.8277	0.8254	0.8490	0.8285	0.9615	0.5000
Slate/copper roofs (yes = 1)	0.7637	0.7391	0.7493	0.7486	0.7674	0.8141	0.3571

Adapted from Table 2 of Hwang and Quigley (2004)

Figure 5-1 Lot size and Frequency of Sale: Evidence from Stockholm

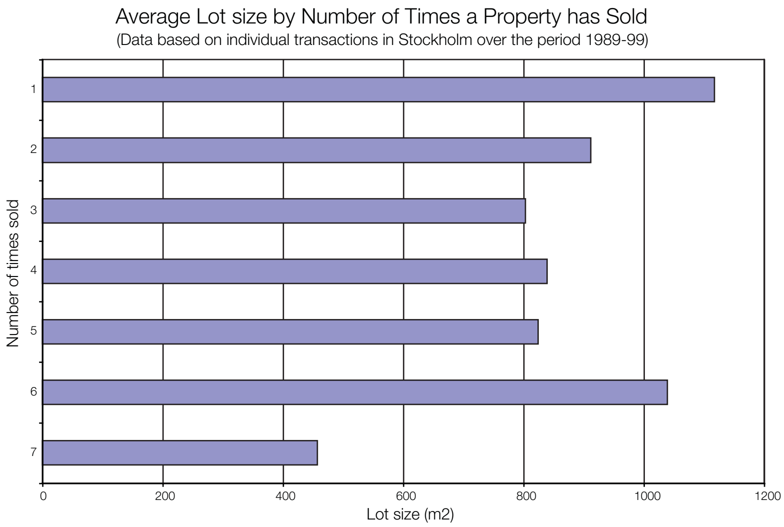


Figure 5-2 Proportion of Detached Properties by Frequency of Sale: Evidence from Stockholm

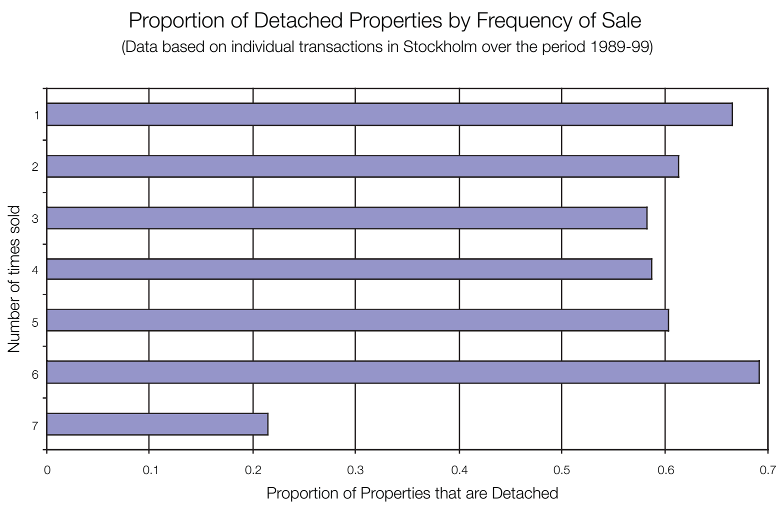
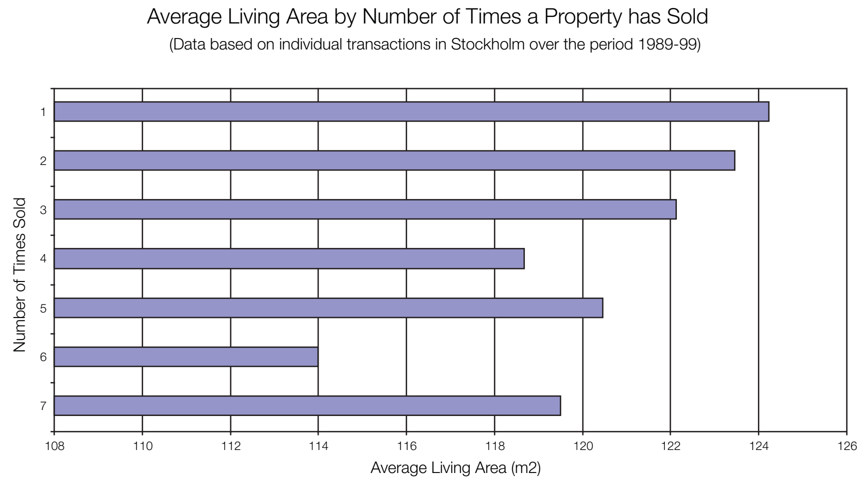


Figure 5-3 Average Living Area by Frequency of Sale: Evidence from Stockholm



Pryce (2004) did not provide a breakdown of frequency of sale by property type because the SASINES data on which the analysis was based did not provide information on dwelling characteristics. He did, however, find that mean house price tended to be lower for properties which frequently sell. A notable exception was the City of Glasgow which is a very heterogeneous area and likely to be biased by the West End which is a generally considered a separate submarket (it is a high value area with high turnover). Further analysis would be needed to ascertain whether the effect was caused by the heterogeneity of the City of Glasgow (we would like to know, for example, whether *within* the West End, more frequently traded properties tend to be of lower value).

Nevertheless, the preliminary results provided enough evidence to believe that house prices varied systematically by frequency of sale and that grouping all properties together without accounting for this non-randomness was likely to result in house price indices giving a biased picture of the level of prices at a given point in time. There was also evidence to support the argument that in many areas, properties that remain off the market for long periods yield higher yields of "satisfaction" (whether due to location, size or quality) as they tend to sell for a higher price than frequently sold dwellings. This might reflect simple lifecycle patterns or it might also be the result of information asymmetries in the housing market (buyers know less than sellers about the true quality of the dwelling and the desirability of its location) and this can result in the stock of dwellings for sale at a given point in time being characterised by a disproportionate number of poor quality properties (a process called "adverse selection – see Akerlof's 1963 seminal theoretical paper on the "Market for Lemons").

Perhaps the most telling aspect to the work done by Pryce (2004) was the analysis of rates of house price change across the different rates of property turnover, the results of which are presented below in Table 5-2. It can be seen that the increase in prices tended to be greater for properties that sold only once over the 1990-2000 period. The implications of these findings are profound: they suggest that using the change in average of all properties would systematically underestimate the true growth in the value of the entire housing stock. Note that if a repeat sales index were used, the bias would be even greater since not only would untraded properties be omitted, but even properties that sold once would be excluded. In some areas the difference was found to be very substantial. For example, the percentage increase in average prices in

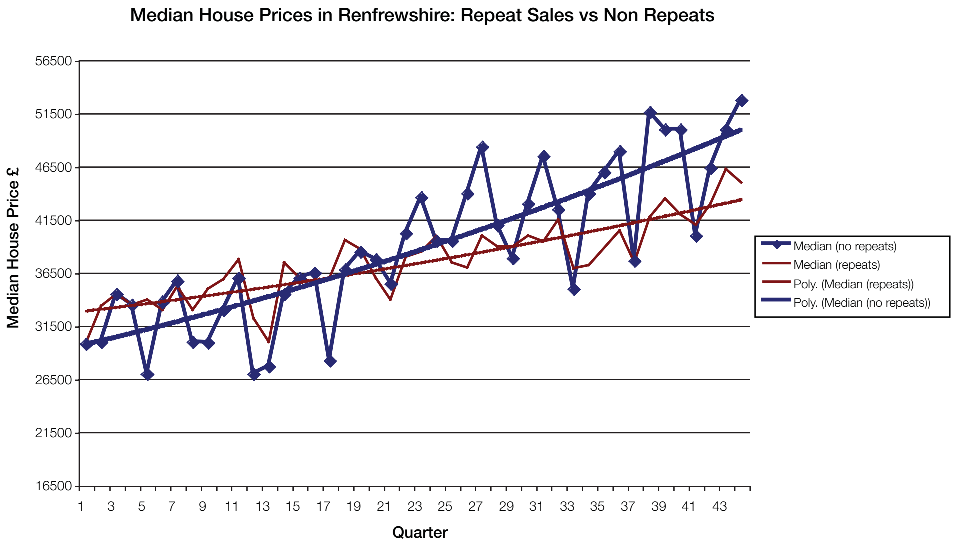
Renfrewshire over the 1990 to 2000 period for properties that sold once was double that of either properties that sold twice or three times (44.5% compared with 22% and 22.5%), and four times that of properties sold four times (44.5% compared with 10%). The graphs in Figure 5-4 illustrate how properties that sell frequently follow a different price trajectory to those that sell only once, and how the impact of this effect is cumulative over time. The corollary is that the longer the time period for which price indices are being used, the greater the bias that may be inherent in those indices unless there are mitigating factors not evident in the particular time period and region considered in the Pryce (2004) analysis.

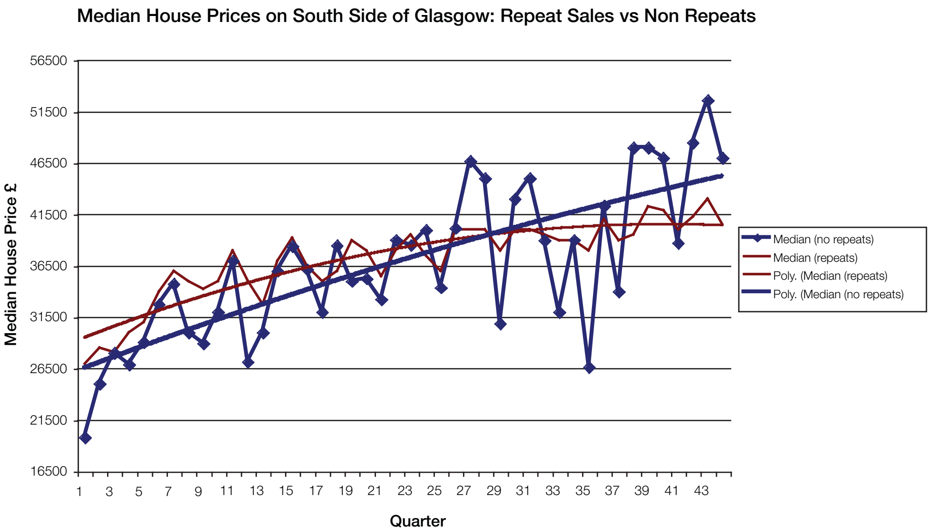
Table 5-2 % Change in Average House Price by No. Times sold in 1991-2000 Period

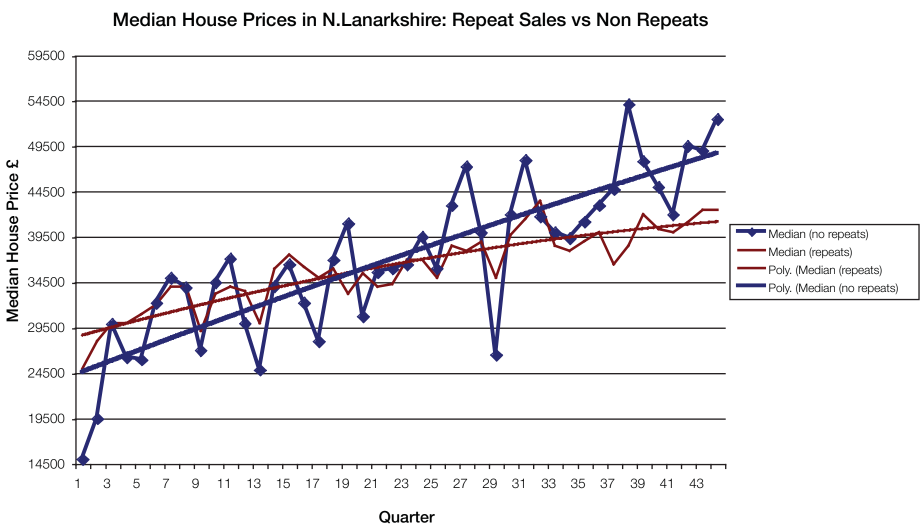
Local Authority		Number of Times Sold in the Period 1991-2000					
		1	2	3	4	5+	All
City of Glasgow	Mean	46.7%	51.8%	46.8%	41.8%	30.5%	48.1%
	n	(9192)	(4138)	(1492)	(343)	(103)	(15268)
E. Renfrewshire	Mean	50.8%	34.5%	40.6%	21.0%	40.2%	46.5%
	n	(1235)	(579)	(173)	(47)	(15)	(2132)
N. Ayrshire	Mean	40.7%	24.2%	29.9%	11.4%	90.1%	35.3%
	n	(2165)	(755)	(236)	(39)	(4)	(3316)
Renfrewshire	Mean	44.5%	22.0%	22.5%	10.7%	-1.1%	35.7%
	n	(2777)	(1159)	(363)	(75)	(16)	(4509)
E. Ayrshire	Mean	48.3%	28.5%	24.2%	2.2%	33.3%	41.3%
	n	(1688)	(634)	(189)	(43)	(3)	(2568)
E. Dunbartonshire	Mean	38.3%	36.7%	42.1%	29.7%	41.3%	38.4%
	n	(1403)	(718)	(201)	(47)	(6)	(2378)
S. Lanarkshire	Mean	50.1%	43.4%	22.3%	28.4%	15.1%	44.9%
	n	(4018)	(1999)	(782)	(200)	(86)	(7085)
Argyll & Bute	Mean	62.6%	40.4%	39.3%	102.4%	89.2%	58.1%
	n	(1330)	(320)	(64)	(11)	(7)	(1732)
S. Ayrshire	Mean	34.6%	26.4%	25.5%	22.2%	6.0%	31.6%
	n	(1783)	(703)	(210)	(44)	(7)	(2756)
N. Lanarkshire	Mean	49.5%	34.0%	26.5%	18.3%	2.0%	42.3%
	n	(3807)	(1583)	(587)	(142)	(31)	(6150)
W. Dunbartonshire	Mean	43.6%	29.2%	27.7%	11.2%	19.3%	37.6%
	n	(1294)	(558)	(175)	(42)	(5)	(2074)
Inverclyde	Mean	36.4%	39.8%	40.8%	18.3%	-37.5%	36.4%
	n	(1296)	(416)	(123)	(28)	(7)	(1870)

Source: adapted from Pryce 2004

Figure 5-4 Comparison of Repeats and Non-Repeats For Different Submarkets





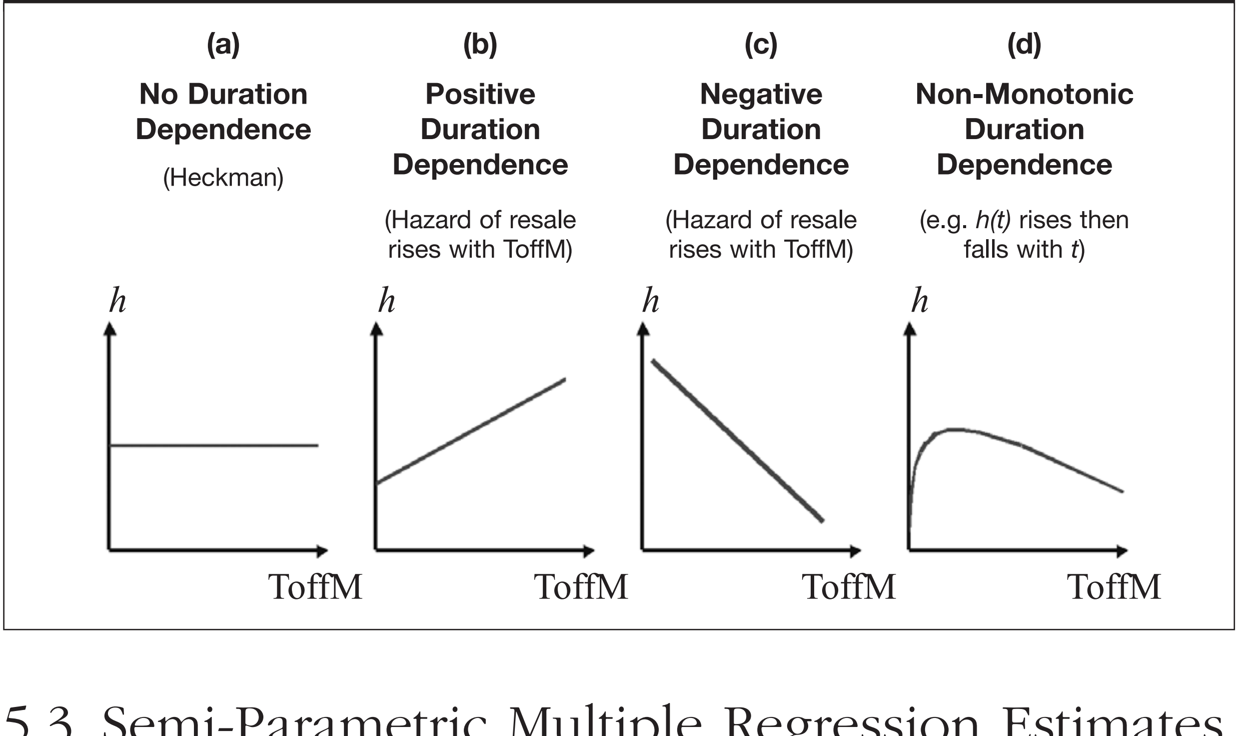


5.2 Econometric Model of Duration Dependence

Our discussion in the previous chapter led us to the conclusion that, in order to choose the appropriate form of sample selection correction, we would need to ascertain whether there exists some degree of duration dependence, and if present, our correction mechanism would need to account for this. Our goal in this section, therefore, is to extend the work of Pryce (2004) by developing a formal econometric model of duration dependence.

If we find zero duration dependence then the probability of a property coming onto the market estimated can reliably estimated from a simple probit regression (as in the Heckman two step approach adopted by Gatzlaff and Haurin). If this were the case, we would expect the hazard of a property coming onto the market to remain unaffected by how long it had been off the market. This would result in a horizontal hazard function as depicted in panel (a) of Figure 5-5. If, however, the hazard of a property coming onto the market increases the longer remains off the market, then we would expect the hazard curve to be upward sloping, as represented in panel (b) of Figure 1. The converse is depicted in panel (c). Finally, the hazard curve may be non-linear, rising (falling) initially, reaching a zenith (trough) and then declining (increasing) or undulating in a regular or irregular pattern. Panel (d) depicts a simple hill-like hazard function. We could not, in scenarios (b), (c) or (d), use a simple probit based analysis (as in the Heckman correction method) which assumes zero duration dependence and normally distributed probability of sale.

Figure 5-5 Hazard Function Types for Different Levels of Duration Dependence



5.3 Semi-Parametric Multiple Regression Estimates

An alternative approach to the Heckman correction method would be to make use of the duration of stay information that could potentially be gleaned from Land Registry data. Since it is only in recent years that the full address details of transacted properties have been included in the English and Welsh Land Registry data releases, this approach, while feasible in principle, will not be of practical use for some time to come.

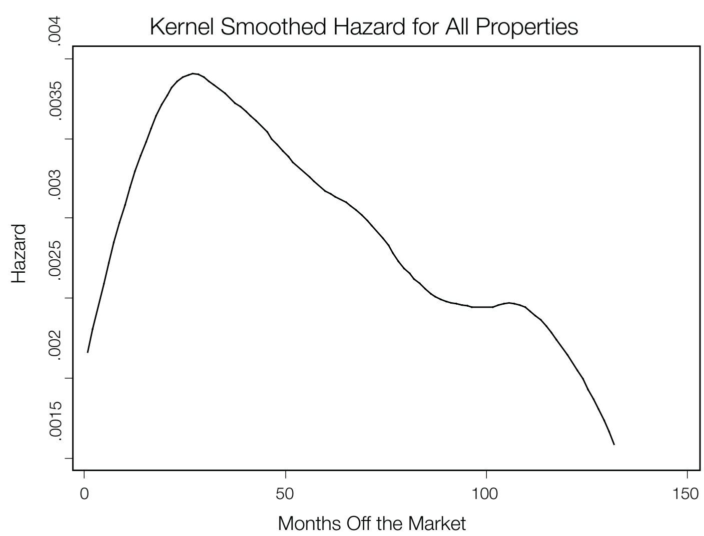
Scottish Land Registry data ("SASINES") has, however, recorded the full address of each sale for more than a hundred years, and digitised for the last fifteen years or so. This should be sufficient for the kind of analysis we have in mind. If fifteen to twenty years of SASINES data can be compiled for a region, while information on properties that do not sell at all would not be available, we would be able to examine the nature of any property that sells at least once during that period. Duration modelling techniques could then be applied to explain the length of time the property remains off the market using techniques that control for "censored" observations – properties that sold once but currently remain off the market. Since duration until resale varies with market buoyancy, this bias could be controlled for by predicting the hazard rate for each property for a set of "controlled" market conditions. This hazard rate could then be entered into the house price equation to control for sample selection bias.

To some extent the distinction between properties that sell and those that do not is a false dichotomy, or at least an incomplete one. The real issue is frequency of sale. A window of ten years of all property transactions, for example, will not include all properties since some will not sell at all. However, it will most probably include all types of properties. Even though a there may exist a type of property that sells once in twenty years, provided this class of properties is of reasonable size, it is likely that a number of these properties will trade within the ten year window. So the ten year window should give a random sample of all levels of frequency of sale. Application of censored duration techniques should adequately control for those properties that sell only once in this period.

5.3.1 AN INITIAL ESTIMATE

We computed the duration of stay of over quarter of a million transactions in the 1990-2000 period in the West of Scotland. From this we were able to derive the kernel-smoothed hazard curve drawn in Figure 5-6. Although the curve is constructed assuming no variation in hazard rates between properties of different types or location, it gives us a clear indication that the hazard function is likely to be both non-linear and non-monotonic, both of which are indicative of a high degree of duration dependence.

Figure 5-6



5.3.2 CONTROLLING FOR AREA TYPE, DEMOGRAPHICS AND ECONOMIC FACTORS

To arrive at a more robust confirmation of the existence of duration dependency, we need to control for other factors. For example, a general market decline in turnover over the period or the collapse of a high turnover area in the second half of the 1990s might be the cause of the fall-off in the hazard function. Cross-sectional variation might occur because certain areas are less attractive and/or have thinner markets, or because some areas attract an older population. So in the regressions that follow we include:

- the deprivation score of an area and,
- the proportion of the population that are over 54.
- Population changes and,
- cycles in unemployment are both factors that might vary both over time and across space.
- Also the number of newbuild is a potentially important determinant of the probability of a property coming onto the market.
- Also, the distance travelled by buyers might be a factor given that areas that predominantly attract buyers from well outside the area will be characterised by households that have few local ties to discourage resale.

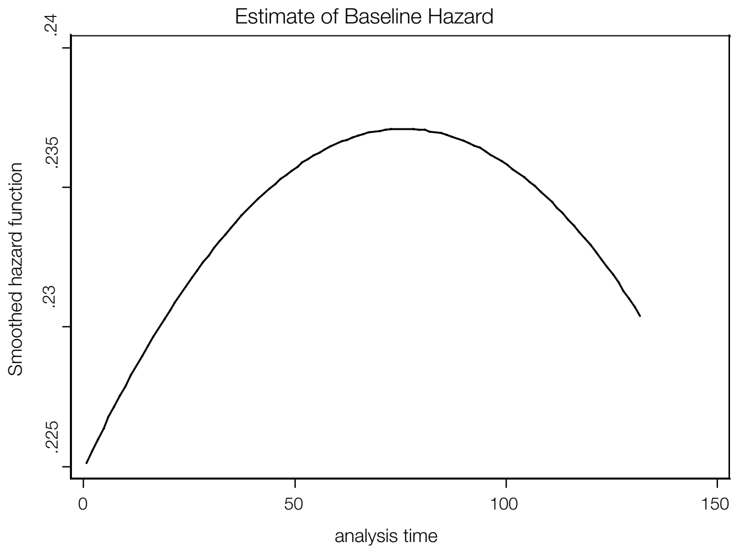
The following is a proportional hazards Cox regression that takes into account the possibility of "repeat failure" (i.e. a property repeatedly coming onto the market). It can be seen that higher deprivation scores are associated with lower hazard rates, and that an increase in the claimant count has a similar (but much greater) effect. Also, the more newbuild in an area, the less likely a property is to come onto the market. Factors that increase the hazard of entering the market include the distance travelled by the buyer and positive changes in the population of a district.

Table 5-3 Cox Proportional Hazard Model of the Hazard of Sale

	Slope Coefficient	t	Sig	95% Confidence Interval for the Slope Coefficient	
				Lower bound	Upper Bound
deprivation	0.995	-3.260	0.001	0.992	0.998
Dunemp	0.431	-18.480	0.000	0.394	0.471
buyer_distance	1.039	16.070	0.000	1.034	1.043
newbuild	0.999	-5.920	0.000	0.999	0.999
prop_gt54	0.016	-15.930	0.000	0.009	0.026
Δpop	825.791	5.590	0.000	78.277	8711.80
n	286,575				
Ll	-596064.56				
Chi Sq	317750.24				
Chi Sq Sig.	0.0000				

Estimating this regression also allows us to derive the *baseline hazard* (a way of examining the shape of the hazard function holding other factors constant), plotted in Figure 4 below. This clearly shows a concave and highly duration dependent hazard function.

Figure 5-7 Semi-Parametric Estimate of the Baseline Hazard



Although this curve is estimated using non-parametric kernel density estimation, it shows a remarkably regular quadratic shape. Running a quadratic OLS regression of the estimated baseline hazard against analysis time (time off the market) confirms this as the following results show (Adj R-squared = 0.6521):

Table 5-4 Quadratic OLS Regression of Baseline Hazard on Time Off the Market

	Coef.	t	Sig	[95% Confidence	Interval]
Analysis time	.0275415	298.72	0.000000	.0273608	.0277222
Analysis time^2	-.0002018	-230.70	0.000000	-.0002035	-.0002001
Constant	.2300911	122.41	0.000000	.2264069	.2337752

For the hazard of entering the market to be non-duration dependent, the coefficients on both the linear and quadratic analysis time variable would have to equal zero and the t tests report that we can reject both these null hypotheses with infinitesimal chance of false rejection (note how narrow the confidence intervals are for both coefficients).

5.4 Parametric Estimation

Further confirmation of the duration dependency of the hazard function is found from considering parametric estimation. Gompertz, Weibull and Log-logistic regressions were run with the results and interpretation summarised below (regression tables follow the summary). The overwhelming evidence for duration dependence presented here precludes simple probit estimation of the probability of a property entering the market since such an approach has to assume that the length of time a property has been off the market has no bearing on whether it will re-enter the market in the next period.

This means that a property that has just been purchased has as much chance of entering the market again as a property that has been off the market for several years. Clearly, this is not the case and so a more robust approach to correcting the sample selectivity bias intrinsic to transactions based hedonic estimation is needed. That is, the estimation of the hazard of selection (or non-selection) needs to allow for duration dependency.

Table 5-5 Summary of Test Results for Duration Dependence

Gompertz Distribution:

if gamma > 0 then positive duration dependence;
if gamma = zero then no duration dependence;
if gamma < 0 then negative duration dependence.

Estimated value of gamma:

CI 95% (.3046027, .3086786) Clearly greater than zero **=> duration dependence**

Weibull Distribution:

if p > 1 then positive duration dependence;
if p = one then no duration dependence;
if p < 0 then negative duration dependence.

Estimated value of p:

CI 95% 4.735841, 4.800477) Clearly greater than one **=> duration dependence**

Log-logistic Distribution:

if	gamma = 1	then the hazard is monotonic and negative duration dependence
if	0.5 < gamma < 1	then the hazard rises steeply but declines shallowly indicating highly positive duration dependence at the outset, gradually becoming slightly negative duration dependent.
if	gamma < 0.25	then the hazard initially rises but declines steeply indicating gradually increasing duration dependence, which at some point rapidly becomes highly negatively duration dependent.

Estimated value of gamma:

CI 95% (.183254 .1857785) Clearly less than one **=> non-monotonic duration dependence**

All of these results confirm duration dependency and the log-logistic estimation confirms the non-monotonic nature of the hazard function (the baseline hazard from which is plotted in Figure 5-8 below). The full set of regression results from which Table 5-5 are presented in the tables that follow.

Figure 5-8

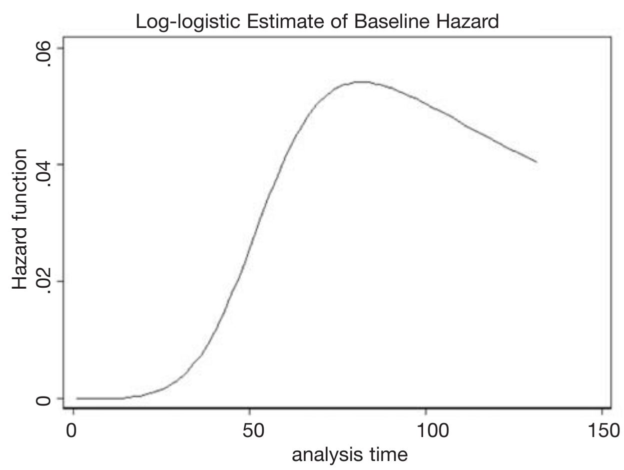


Table 5-6 Gompertz regression – log relative-hazard form

	Haz. Ratio	z	P>lzl	[95% Confidence	Interval]
deprivation	0.995	-3.400	0.001	0.992	0.998
Δ _unemployed	0.371	-21.940	0.000	0.340	0.405
buyer_dist	1.039	20.000	0.000	1.035	1.043
newbuild	0.999	-6.430	0.000	0.999	0.999
prop_gt54	0.015	-16.060	0.000	0.009	0.025
Δ pop	1343.01	5.99	0.000	127.22	14177.1
yr1991_d	0.025	-178.990	0.000	0.024	0.026
yr1992_d	0.001	-262.630	0.000	0.001	0.001
yr1993_d	0.000	-290.310	0.000	0.000	0.000
yr1994_d	0.000	-299.850	0.000	0.000	0.000
yr1995_d	0.000	-303.210	0.000	0.000	0.000
yr1996_d	0.000	-301.730	0.000	0.000	0.000
yr1997_d	0.000	-305.540	0.000	0.000	0.000
yr1998_d	0.000	-302.440	0.000	0.000	0.000
yr1999_d	0.000	-303.880	0.000	0.000	0.000
yr2000_d	0.000	-298.450	0.000	0.000	0.000
gamma	0.307	294.910	0.000	0.305	0.309
n	286,575				
LR chi2(16)	324595.53				
Log likelihood	-51570.394				
Prob > chi2	0.000				

Table 5-7 Weibull regression – log relative-hazard form

	Haz.Ratio	z	P>lzl	[95% Confidence	Interval]
deprivtn	0.996	-2.540	0.011	0.993	0.999
Δ_unemployed	0.173	-38.670	0.000	0.158	0.189
buyer_distance	1.037	18.910	0.000	1.033	1.041
newbuild	1.001	4.600	0.000	1.000	1.001
prop_gt54	0.016	-15.920	0.000	0.009	0.026
Δ_pop	753.047	5.510	0.000	71.237	7960.508
yr1991_d	0.026	-184.960	0.000	0.025	0.027
yr1992_d	0.002	-252.590	0.000	0.002	0.002
yr1993_d	0.000	-281.780	0.000	0.000	0.000
yr1994_d	0.000	-297.480	0.000	0.000	0.000
yr1995_d	0.000	-299.920	0.000	0.000	0.000
yr1996_d	0.000	-293.730	0.000	0.000	0.000
yr1997_d	0.000	-298.660	0.000	0.000	0.000
yr1998_d	0.000	-285.440	0.000	0.000	0.000
yr1999_d	0.000	-278.780	0.000	0.000	0.000
yr2000_d	0.000	-244.400	0.000	0.000	0.000
/ln_p	1.562	451.660	0.000	1.555	1.569
p	4.768			4.736	4.800
1/p	0.210			0.208	0.211
n	286,575				
LR chi2 (16)	249,620				
Log likelihood	-89,725				
Prob > chi2	0.000				

Table 5-8 Log-logistic regression -- accelerated failure-time form					
	Coef.	z	P>lzl	95% Confidence	Interval]
deprivtn	0.001	2.500	0.013	0.000	0.002
Δ_unemployed	0.383	38.320	0.000	0.363	0.402
buyer_dist	-0.008	-18.720	0.000	-0.009	-0.007
newbuild	0.000	-1.380	0.169	0.000	0.000
prop_gt54	0.893	15.230	0.000	0.778	1.007
Δ_pop	-1.367	-5.050	0.000	-1.898	-0.836
yr1991_d	0.876	214.250	0.000	0.868	0.884
yr1992_d	1.391	353.370	0.000	1.383	1.398
yr1993_d	1.787	485.020	0.000	1.779	1.794
yr1994_d	2.110	575.080	0.000	2.103	2.117
yr1995_d	2.388	580.320	0.000	2.380	2.396
yr1996_d	2.529	559.910	0.000	2.520	2.538
yr1997_d	2.812	559.160	0.000	2.803	2.822
yr1998_d	2.961	494.670	0.000	2.949	2.973
yr1999_d	3.121	460.830	0.000	3.108	3.135
yr2000_d	3.402	346.370	0.000	3.383	3.421
Constant	2.045	130.170	0.000	2.015	2.076
/ln_gam	-1.690	-484.190	0.000	-1.697	-1.683
gamma	0.185			0.183	0.186
n	286,575				
LR chi2(16)	247,246				
Loglikelihood	-90,460				
Prob>chi2	0.000				

5.5 Omitted Variable Bias in Hedonic Estimation

To illustrate how the estimated hazard of resale could be incorporated into submarket hedonic regressions the following regresses the log of selling price against distance to the nearest central business district (in this case, Glasgow city centre), the estimated hazard of sale for each property transaction, the interaction of this hazard with year dummies and stand alone year dummies.

Looking at the results it can be seen that the stand alone hazard variable has statistically significant positive effect on house prices in the West End submarket (interpreted as capturing the cyclical variation in market buoyancy), though when interacted with year dummies, it can be seen that the house price is generally lower for higher the hazard rates (i.e. for a given period, cross sectional variation in the hazard will reflect quality differences in properties/neighbourhoods and so higher hazards of resale will correspond to lower quality) but the interaction terms are not individually statistically significant.

The next stage in such an analysis would be to bootstrap these hedonic regressions since it is not clear whether Heckman's analytical formulae for computing the appropriate standard errors still applies when the hazard of selection is used (rather than the hazard of non-selection) and when other violations apply (see Greene 2003).

Table 5-9 Hedonic House Price Regression for the West End of Glasgow: SASINES

	Coef.	t	P>|t|	[95% Confidence	Interval]
distance to Glasgow centre (log)	-0.120	-5.050	0.000	-0.166	-0.073
Hazard of sale	2.520	2.140	0.032	0.216	4.825
Hazard of sale * 1991 dummy	-2.910	-2.020	0.043	-5.732	-0.089
Hazard of sale * 1992 dummy	-1.111	-0.780	0.435	-3.900	1.679
Hazard of sale * 1993 dummy	-1.985	-1.510	0.130	-4.555	0.584
Hazard of sale * 1994 dummy	-1.951	-1.190	0.235	-5.169	1.268
Hazard of sale * 1995 dummy	-1.509	-1.170	0.244	-4.047	1.029
Hazard of sale * 1996 dummy	-2.477	-1.380	0.168	-6.000	1.047
Hazard of sale * 1997 dummy	-2.025	-1.490	0.135	-4.683	0.634
Hazard of sale * 1998 dummy	-1.681	-1.320	0.187	-4.180	0.819
Hazard of sale * 1999 dummy	0.105	0.070	0.947	-2.998	3.208
Hazard of sale * 2000 dummy	-2.390	-0.720	0.472	-8.899	4.119
1991 dummy	-0.010	-0.060	0.951	-0.318	0.299
1992 dummy	0.164	1.070	0.283	-0.136	0.464
1993 dummy	0.167	1.140	0.256	-0.121	0.454
1994 dummy	0.013	0.070	0.943	-0.349	0.375
1995 dummy	0.117	0.760	0.448	-0.185	0.419
1996 dummy	0.148	0.650	0.515	-0.297	0.593
1997 dummy	0.214	1.210	0.226	-0.132	0.561
1998 dummy	0.281	1.700	0.088	-0.042	0.605
1999 dummy	0.741	3.550	0.000	0.332	1.150
2000 dummy	0.465	2.000	0.046	0.008	0.922
Constant	11.520	44.570	0.000	11.013	12.027
Number of obs	2,539				
F(22, 2516)	7.35				
Prob > F	0.000				
Adj R^2	0.0522				

Dependent variable = ln (Selling Price)

The findings for the West End are generally confirmed for when the regression is run on the whole of Strathclyde (Table 5-10) using area dummies to capture spatial variation. Neither of these regressions include dwelling attribute variables (hence the very low adjusted R^2) since the SASINES data do not record such information. However, the hazard regressions could be used to predict hazard rates for other data sets and result in a more robust estimation of constant quality price.

Table 5-10 Hedonic House Price Regression for the Whole of Strathclyde: SASINES

	Coef.	t	P>|t|	[95% Confidence	Interval]
distance to Glasgow centre (log)	-0.110	-12.650	0.000	-0.127	-0.093
Hazard of sale	1.411	3.520	0.000	0.626	2.196
Hazard of sale * 1991 dummy	-0.232	-0.480	0.629	-1.172	0.709
Hazard of sale * 1992 dummy	0.200	0.420	0.678	-0.743	1.142
Hazard of sale * 1993 dummy	-0.654	-1.460	0.144	-1.532	0.224
Hazard of sale * 1994 dummy	0.183	0.320	0.749	-0.940	1.307
Hazard of sale * 1995 dummy	-0.873	-1.960	0.050	-1.745	-0.001
Hazard of sale * 1996 dummy	-0.630	-1.060	0.288	-1.791	0.532
Hazard of sale * 1997 dummy	-1.589	-3.620	0.000	-2.449	-0.729
Hazard of sale * 1998 dummy	-0.962	-2.270	0.023	-1.794	-0.131
Hazard of sale * 1999 dummy	-0.529	-0.910	0.363	-1.670	0.611
Hazard of sale * 2000 dummy	0.602	0.420	0.677	-2.232	3.436
1991 dummy	0.179	3.360	0.001	0.074	0.283
1992 dummy	0.188	3.630	0.000	0.087	0.290
1993 dummy	0.158	3.160	0.002	0.060	0.257
1994 dummy	0.182	2.830	0.005	0.056	0.309
1995 dummy	0.089	1.650	0.099	-0.017	0.196
1996 dummy	0.198	2.540	0.011	0.045	0.352
1997 dummy	0.119	2.110	0.035	0.008	0.231
1998 dummy	0.302	5.450	0.000	0.193	0.410
1999 dummy	0.449	5.690	0.000	0.295	0.604
2000 dummy	0.496	4.990	0.000	0.301	0.691
+Area Dummies					
Constant	11.348	116.80	0.000	11.157	11.538
n	197,46				
$AdjR^2$	0.036				

5.6 Conclusion

This chapter has presented the results of an exploratory and pioneering study of sample selection correction that (a) can be applied to UK data; and (b) test for and capture the effects of duration dependence. To our knowledge, this is the first time such analysis has been attempted in any country. We extended Pryce's (2004) work on Registers of Scotland/SASINES data on the West of Scotland to by applying an extensive series of tests and estimation methods to ascertain whether there was duration dependence in the probability of a property selling. Our results offered overwhelming evidence that the probability of sale is indeed characterised by duration dependence. We also attempted to test whether the predicted hazard (probability) of sale was a significant variable when entered as a correction term in a hedonic price regression. Preliminary results using Registers of Scotland sales data indicated that this was indeed the case but since this data did not include dwelling attributes the overall explanatory power of these regressions was weak. Nevertheless, the predicted probabilities of sale could in principle be applied to hedonic regressions based on more detailed lender or estate agent data.

6 Dealing with Spatial Variation and Spatial Spillovers

6.1 Introduction

In this chapter we investigate whether there exists evidence of systematic variation in the probability sale across space. No doubt, the incidence of sale will vary across neighbourhoods in a given quarter simply because of random variation in the timing of properties coming on and off the market. The question is whether there is also a systematic component to that variation – whether the incidence of sale and the type of neighbourhood are at all correlated.

Our investigation is based on the analysis of the proportion of sales at postcode and postcode sector levels in the South East of England over the period 1996 to 2004. As such, our results are based on very large samples and are drawn from an unprecedented integration of different sources of spatial data from a range of sources (including Mosaic, Hometrack, Land Registry, ODPM and The Ordinance Survey). The combined data allows us to categorise localities down to postcode level according to the following types of variable:

(1) **Average Density:** measured in terms of average distance to nearest neighbour for each and every postcode (calculations were done by Hometrack based on Ordinance Survey data);

(2) **Elevation:** measured by the height above sea level of the cancroids of each postcode unit in the South East (analysis provided by the Department of Geography and Geomatics, University of Glasgow, based on Ordinance Survey data);

(3) **Typical dwelling size:** measured in terms typical plot footprint area (data supplied by Hometrack based on Ordinance Survey data); estimated typical total floor area (two measures were used: one from Hometrack and one from Mosaic, both based on a combination of actual measurements and interpolation), and the typical number of bedrooms in each and every postcode (data provided by Mosaic);

(4) **Typical dwelling type:** measured in terms of the proportion of properties in each and every postcode that are detached, the proportion that are semi-detached, the proportion that are terraced, the proportion that are Purpose Built (PB) flats, the proportion of Non-Purpose Built (NPB) flats, and the proportion of bungalows (data provided by Mosaic);

(5) **Typical dwelling age:** measured in terms of the proportion of dwellings built during particular periods (data provided by Mosaic);

The chapter is structured as follows: first we analyse the bivariate relationships between the neighbourhood characteristics listed above and the proportion of properties that sell each year in each postcode sector. Second, we consider how the spatial spillover effect described in chapter 5 might be captured using a single, continuous measure. Third, we use regression analysis to consider the relationship between neighbourhood characteristics and the probability of sale. Regression estimation is useful because it allows us to consider associations between variables in a multi-determinant setting – that is, one that allows us to estimate the relationship between our dependent variable (probability of sale) and each explanatory variable (neighbourhood characteristics) holding all other explanatory variables constant. A variety of regressions are run, both on the whole of the South East and also for each of the component counties that make up the region. In addition to the regressions on data aggregated to postcode sector level, we conduct a complementary analysis at postcode unit level. The fourth and final step in our investigation attempts to estimate whether spatial variation in the probability of sale has any significant impact on house price indices.

6.2 Spatial Variation: Bivariate Analysis of the Probability of Sale

In Chapter 5, we presented the results of Hwang and Quigley's (2004) analysis of Stockholm housing transactions data (see Table 5-1 and subsequent graphs). They 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 their table of descriptive statistics how the relationship between probability of sale and dwelling attributes varies over time (their probit estimation includes time dummies to capture this effect, but they do not report the coefficients for these time effects; also, it appears that they do not include time interaction effects to capture the possible changes in slope coefficients over time).

Since our graphical analysis of their data suggested potential non-linearities in the relationship between frequency of sale and dwelling characteristics (see Figure 5-1 and Figure 5-2), we felt that a similar graphical analysis of UK data would be the best way to present the bivariate aspect of our investigation. Because we are also interested in how relationships may change over time, we constructed graphs for each separate year of our data (1996 to 2004). We do not have frequency of sale partly because we are unable to identify repeat sales, and partly because in this chapter we are interested in the neighbourhood, rather individual dwelling, effects on probability of sale. Instead, we plot neighbourhood characteristics against decile of stock turnover, S_{it} where,

S_{it} = total number of sales in a postcode sector i in a given year t divided by the total number of residential dwellings in that postcode sector.

The results, presented in the Appendix (available separately on request), offer the most comprehensive description of the relationship between the proportion of properties that sell and neighbourhood characteristics yet to be published and represents a very substantial amount of work. Also, we believe it is the first time links between the proportion of dwellings that sell and spatially specific factors such as density and elevation have been analysed in any detail. We shall now summarise the bivariate results suggested by the graphs in the Appendix and use these findings to inform our multiple regression models developed in subsequent sections.

6.2.1 DENSITY

The bar charts in Figure A-1 plot the average distance to the nearest neighbour for each decile of the proportion of the housing stock that sells in a given year for each postcode sector in the South East. The figures are robust in that they are based on large samples – there are more than a thousand postcode sectors in the South East. The average distance (measured in metres) between each dwelling and the nearest neighbour for each postcode was aggregated for the purposes of the bar charts to postcode sector level.

We anticipated that areas with a smaller proportion of sales in any given year (i.e. those in the lowest deciles of S_{it}, where S_{it} is simply the proportion of the housing stock in area i that sells in year t) would tend to be of low density. What we did not anticipate was the magnitude of this effect. As the graphs in Figure A-1 show, areas in the lowest decile of S_i have an average distance to nearest neighbour of over 50m. This compares to an average distance to nearest neighbour of less than 20m in most of the remaining deciles of S_{it}. Clearly, the relationship is non-linear – the difference between the first and second deciles is far greater than the distance between the remaining deciles. There was some variation over time in this relationship, with the contrast between the first and second deciles increasing considerably between 1996 and 1999, but the overall effect was unchanged.

6.2.2 ELEVATION

How might height above sea level affect the probability of sale? Our hypothesis is that that height above sea level may be a proxy for quality of location and possibly also of construction quality. Anecdotally, planning decisions during and after the industrial revolution tended to locate the most desirable new build in areas above and upwind from the severest pollution. There may also be some concern about the role of flooding which makes elevated neighbourhoods more secure places to live and hence where households will be happy to live for prolonged periods. In contrast, households may contemplate living in low level areas for short periods, but prefer to move to more elevated locations when the opportunity arises to avoid the risk of flooding (and indeed to avoid finding themselves in the possession of a house that they can no longer sell because of previously unforeseen susceptibility to flood damage).

Height above sea level might also offer good views, further enhancing the location effect, though we reasoned that *variation* in height within a postcode sector might be a better reflection of the potential for views. On the other hand, variation within a postcode sector of height above sea level suggests that there are valleys as well as hill tops, and the high demand for elevated locations may be offset by low demand for low level locations.

The bivariate results in Figure A-2 generally support the hypothesis that postcode sectors which are generally elevated have lower rates of stock turnover. For most years, the overall relationship between elevation and the proportion of the stock that trades is negative: the first decile of S_{it} generally has greater elevation than the other deciles.

However, the relationship is also clearly non-linear with a peak in elevation also occurring between the sixth and ninth deciles in a number of the years investigated.

The relationship between the *variation* in elevation and the proportion of stock that trades is somewhat ambiguous (Figure A-3), perhaps for the reasons given above. Overall, the neighbourhoods with greater undulation tend to be in the highest decile of S_{it}.

6.2.3 SIZE

We have three measures of size: (i) average footprint area; (ii) estimated total floor area; and (iii) dominant number of bedrooms. Our analysis of these measures is presented in Figures A-4, A-5 and A-6 respectively.

We believe it likely that the first of these, average footprint area, is measured with greatest precision since the information (provided by Hometrack) is derived from calculations taken from the Ordinance Survey Master Map system, unlike the other measures which potentially include a considerable degree of interpolation and approximation. We have two measures of the total floor area of dwellings (one provided by Hometrack and the other by Mosaic), but we expect that for both these measures, precise data were not available for a significant number of properties and so approximations were used based on combining information from other measures such as footprint size, type of dwelling, and information from estate agents.

The third measure of size is perhaps most imprecise: number of bedrooms. The source of imprecision is obvious in that a dwelling may have many bedrooms, but if each bedroom is small, the overall size of the property may not be that great. The imprecision is likely to have been exacerbated by the need to interpolate for dwellings for which no information was available.

And, indeed, our expectations proved well founded: the effect of size on trading proportions is most pronounced for the footprint measurement. Properties in the lowest decile of S_{it} have nearly three times the average footprint area, producing a pattern in the bar charts not dissimilar to that of the density measure, which is obviously closely related.

As for the other two measures of size, average total floor area and the dominant number of bedrooms, we found very little variation in these variables across deciles of S_{it}, which suggests that there is a weak or non-existent relationship with stock turnover.

6.2.4 TYPE

We look at six measure of dwelling type for each postcode sector: (1) proportion of dwellings that are bungalows; (2) proportion of dwellings that are detached; (3) proportion of dwellings that are semi-detached; (4) proportion of dwellings that are terraced; (5) proportion of dwellings that are non-purpose built (NPB) flats; and (6) the proportion of dwellings that are purpose built (PB) flats.

Results for the proportion of bungalows (Figure A-7) and for the proportion of detached dwellings (Figure A-8) were surprising. Areas in the highest decile of stock turnover tended also to have the greatest proportion of detached dwellings and also the greatest proportion of bungalows. There is a degree of non-linearity – in most years, a peak in the % of detached in the first decile. Indeed, in all years bar 2003, the proportion of properties is higher in the first decile of Sit than the second or third decile of S_{it}.

As for semi-detached properties (Figure A-9), the relationship appears to be concave: areas in the first decile of stock turnover have the lowest proportions of semi-detached properties, as do those in areas in the tenth decile of stock turnover. Areas with the greatest proportion of semi-detached properties tend to be in the second to fifth deciles of S_{it}.

The results for terraced properties are presented in Figure A-10. Interestingly, although the lowest turnover areas (decile 1) do not have particularly high rates of terraced dwellings, areas in deciles 2, 3, and 4 typically have the highest proportions of terraced dwellings.

We expected the relationship between turnover and area type to be the least ambiguous for flats. We reasoned that, because of life-cycle factors, people tend only to ive in flats for particular periods in their lives (most notably when they are young, single, without children, and/or as first time buyers) or as second homes. These aspects of flat purchase all suggest relatively low emotional attachment and relatively high turnover rates. And indeed, this was borne out in the results for non-purpose built flats (Figure A-11) but not for purpose built flats (Figure A-12). The anomaly of the latter may be explained by one of two factors. First, older dwellings that have been converted to flats may tend to be the least amenable to disabled access, and so older and less mobile households who choose to live in a flat will prefer to purchase a purpose-built dwelling. These are also the types of household least likely to move frequently. The second possibility is that purpose built flats were built in periods when construction quality is low and/or in relatively deprived areas. The first theory explains low turnover in terms of low levels of supply (occupants rarely want to move out), whereas the second theory explains low turnover in terms of low demand (few want to move in).

Similar complexities arise when we consider the turnover of stock according to date of construction (Figures A-14 to A-17). We anticipated finding lower rates of turnover of older stock which we assumed to be characterised by larger proportions and more substantial construction standards. Clearly, however, the reality is not straightforward. Areas in the second decile of turnover have very high proportions of pre-1920 dwellings. However, this is not true of areas in the first decile, which actually tend to have high proportions of dwellings built since 1979.

6.2.5 A COMPLEX PICTURE

Taken together, these results highlight important aspects of the nature of housing transactions. Clearly, it is not only the quality and type of construction of a dwelling that affects satisfaction with a neighbourhood – 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, information imperfections and processes of gentrification.

For example, if detached bungalows increase in value over a prolonged period at a rate that exceeds that of other property types, that difference in appreciation may not be widely known because house price information may only be presented in the form of averages for all property types. When owners do eventually become aware of the accelerated appreciation of their houses, there may be a rush of detached bungalows being traded by households keen to access their accumulated equity, purchased by investors newly aware of the favourable long term prospects of this class of assets. Such variation in the types of property may be reinforced by processes of gentrification: clusters of detached bungalows 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. Note also that refusal to sell will always dominate – even the highest levels of demand only become realised if the current occupants are willing to sell (a buyer cannot *make* someone sell). So one would expect the lowest rates of turnover to 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 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 possible set of scenarios is summarised in Table 6-1, though the permutations considered are far from exhaustive.

Table 6-1 Turnover Rate Market Scenarios

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

* 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).

6.3 Measuring Spatial Interactions

We hypothesised in Chapter 4 that, not only are there likely to be concentrations of probability of sale in particular areas, but that there would also be spatial interactions with neighbouring areas. We presented several reasons for this, under the following headings (see section 4.3.4):

(1) Relationships and family ties

(2) Changes to local amenities

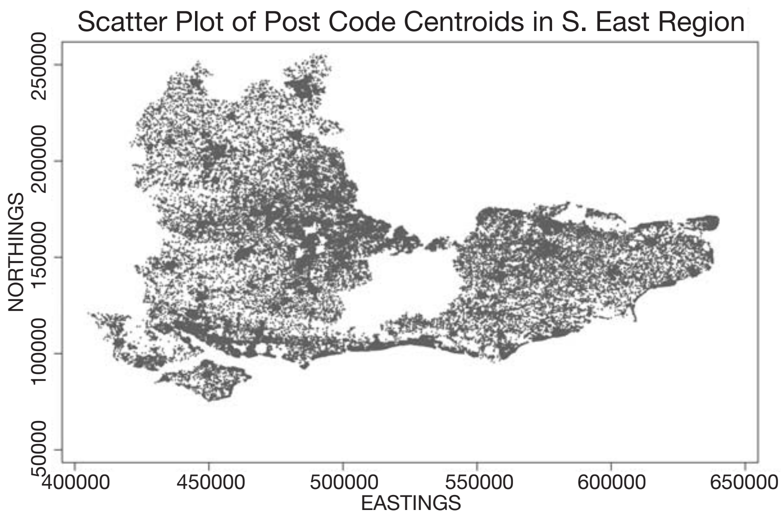
(3) Crime and the Fear of Crime

(4) Many moves are local

(5) Equity gains by submarket

To verify the spillover hypothesis we need to construct a measure of spatial interaction that will capture these effects. A standard method deployed in the spatial econometrics literature is to apply a "distance decay" function, which applies a greater weight to nearby neighbourhoods than to those located further away. This is the approach we take here. The proportions of dwellings that sell in the closest fifty neighbourhoods were weighted by their distance to the neighbourhood of interest. This was a major undertaking since there are over 190,000 postcode sector units in the South East. We had to calculate the distance from the centroid of every postcode unit (see Figure 6-1) to the centroid of every other postcode unit (entailing more than thirty six billion calculations). Each distance was ordered for each postcode to identify the fifty nearest neighbouring postcodes. The final selection of fifty distances was used to compute a weighted sum of neighbouring stock turnover proportions.

Figure 6-1 Spatial Distribution of Postcode Cancroids in S. East



(Source: Ordnance Survey)

6.4 Regression Results

It is clear from the literature (particularly Gatzlaff and Haurin, 1998) that it is not only property characteristics that will determine the rate of turnover of the housing stock. Socio-economic, tenure, and demographic factors will also be important. We attempt to control for these effects using the Mosaic categorization of postcodes (and hence postcode sectors). The Mosaic system, provided by Experian Business Strategies (www.business-strategies.co.uk) classifies areas according to tenure, socio-demographics and spending patterns. From these categories we created a comprehensive series of binary variables (called "dummies") to represent each identified type of area. We then included these dummies in our regressions as control variables, to account for the wide range of factors not fully reflected in the variables that are the subject of our current focus (density, elevation, type). We have not presented the coefficients for each of these dummies since they would only have meaning to someone who is fully familiar with the Mosaic classification schedule, but details can be provided by the authors on request (only statistically significant dummies were retained in the regressions).

6.4.1 LOG-LOG MODEL

The first set of regression models presented below (Table 6-2) are based on log-log functional forms. That is, the dependent variable and key continuous explanatory variables (the spatial spillover variable, density, footprint area, elevation, % in each age category, and % in each house type category) are converted to natural log scaling. This allows us to interpret the coefficients as elasticities, which is particularly useful because elasticities are independent of the scaling of the different variables in the regression. An elasticity tells us the percentage change in the dependent variable (S_{it}) in response to a percentage change in an explanatory variable, holding all other factors constant.

Low-density

Looking at the first regression of Table 6-1 (labelled lnOLS1996, where OLS = ordinary least squares – the simplest form of regression estimator, and ln = natural log), we can see that in 1996, the coefficient on our density variable (average distance to nearest dwelling) equals –0.16. This means that a ten per cent rise in average footprint area would result in a fall in the percent of properties in an area that trade by 1.6%. This coefficient varied somewhat over the following years (-0.14 in 1997; -0.08 in 1998; -0.16 in 1999 and 2004; -0.10 in 2000, 2001, 2003) but remained negative throughout, confirming our hypothesis that probability of sale is lower in low density area. In 2000, the variable was not significant (t ratio was less than 2) so it was removed from the regression.

Elevation

The height above sea level also had the anticipated negative effect on stock turnover, though the size of the effect was less pronounced: a ten per cent increase in elevation causes the rate of stock turnover to fall by between 0.2% and 0.4% (the elasticity varies between –0.02 and –0.04). Note that the effect of elevation was not statistically significant from zero in years 1996, 1998, 2001 and 2003, and hence the variable was removed from the regressions in those years.

Size

We tried all three size variables but only footprint area proved significant, and so it is only this variable that is reported in the regressions. The footprint area effect on turnover proved to be statistically different from zero in all nine years with coefficients ranging from –0.07 to –0.12, which means that a ten per cent rise in average footprint area will cause a fall in Sit of between 0.7% and 1.2%.

Age

The only consistently significant age variable in the log-log regressions was the pre-1990 category, which had a positive effect on stock turnover (a ten percent rise in the average proportion of dwellings that were built before 1920 results in a rise in Sit of between 0.6% and 1.7%).

Type

Similarly, the only consistently significant type measure was the detached variable. A ten percent increase in the average proportion of houses in a postcode sector that are detached causes the proportion of dwellings that trade to rise between 1.3% and 2.0%. The importance of this variable is increased all the more by the fact that it varies across postcode sectors more than any of the other variables in the regression (coefficient of variation, CoV, equals 0.4).

Spatial Spillover

As anticipated, the effect on turnover of the rates of turnover of neighbouring postcodes has a positive effect, with elasticities ranging from 0.06 to 0.33.

Table 6-2 Results from OLS Log-Log Model of % Dwellings that Sell (Postcode Sector Level)

Variable	CoV	OLS1996	OLS1997	OLS1998	OLS1999	OLS2000	OLS2001	OLS2002	OLS2003	OLS2004
Constant	–	2.70	2.76	2.85	2.12	5.32	2.55	1.88	1.80	2.12
	–	(10.07)	(9.00)	(7.86)	(11.07)	(11.10)	(7.06)	(12.96)	(10.85)	(11.07)
aSWc2to5_p1_ln	0.004	0.06	0.07	0.11	–	0.33	0.06	–	–	–
	–	(2.88)	(2.70)	(3.09)	–	(8.73)	(2.24)	–	–	–
aveFPA_ln	0.008	-0.07	-0.07	-0.08	-0.07	-0.12	-0.09	-0.07	-0.10	-0.07
	–	(-2.51)	(-2.70)	(-2.00)	(-2.00)	(-3.11)	(-2.49)	(-2.94)	(-3.76)	(-2.00)
aveNNm2_ln	0.089	-0.16	-0.14	-0.08	-0.16	–	-0.10	-0.10	-0.10	-0.16
	–	(-6.68)	(-6.42)	(-2.16)	(-6.68)	–	(-4.92)	(-6.24)	(-5.38)	(-6.68)
ht_elevation_ln	0.070	–	-0.02	–	-0.03	-0.04	–	-0.02	–	-0.03
	–	–	(-3.00)	–	(-2.14)	(-4.15)	–	(-2.52)	–	(-2.14)
MPPD_p20_45_ln	0.135	0.07	–	-0.10	–	-0.11	–	–	–	–
	–	(2.42)	–	(-3.82)	–	(-4.25)	–	–	–	–
MPPD_p46_79_ln	0.218	–	–	–	–	–	-0.10	-0.10	-0.07	–
	–	–	–	–	–	–	(-7.46)	(-9.08)	(-4.56)	–
MPPD_pre1920_ln	0.116	0.08	0.09	0.13	0.09	0.17	0.07	0.07	0.06	0.09
	–	(3.91)	(3.57)	(5.04)	(5.13)	(7.42)	(5.67)	(6.14)	(5.34)	(5.13)
MPPD_prop_bung_ln	0.262	–	–	–	–	–	0.04	–	–	–
	–	–	–	–	–	–	(3.13)	–	–	–
MPPD_prop_det_ln	0.404	0.19	0.15	0.14	0.20	0.18	0.12	0.14	0.13	0.20
	–	(9.77)	(10.22)	(6.86)	(11.03)	(9.92)	(9.06)	(10.71)	(7.45)	(11.03)
MPPD_prop_sdet_ln	0.123	–	–	–	-0.08	–	–	–	-0.05	-0.08
	–	–	–	–	(-3.93)	–	–	–	(-2.06)	(-3.93)
+ Mosaic categories										
N		1,182	1,181	1,188	1,219	1,241	1,248	1,258	1,257	1,219
r2_a		0.378	0.276	0.217	0.200	0.222	0.172	0.274	0.262	0.200
ll		-10.42	-24.74	-281.34	-361.49	-519.34	-148.36	278.46	248.49	-361.49
aic		50.84	87.47	594.68	746.98	1060.69	324.72	-534.93	-460.98	746.98

t-values based on HC2 standard errors

Table 6-3 OLS Results

Variable	OLS1996	OLS1997	OLS1998	OLS1999	OLS2000	OLS2001	OLS2002	OLS2003	OLS2004
Constant	-0.065	-0.066	0.968	0.550	0.760	1.329	2.368	2.182	0.550
	(-0.312)	(-0.235)	(5.307)	(2.628)	(2.821)	(4.600)	(47.524)	(29.033)	(2.628)
aSWc2to5_p1	10,700	11,400	10,600	11,000	7,922	3,416	–	1,609	11,000
	(8.526)	(9.023)	(9.852)	(9.999)	(7.547)	(4.516)	–	(2.026)	(9.999)
aveNNm2	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.004	-0.004	-0.003
	(-3.979)	(-3.625)	(-3.572)	(-3.341)	(-2.773)	(-3.792)	(-3.625)	(-2.207)	(-3.341)
ht_elevation	–	-0.002	-0.001	–	-0.001	–	–	–	–
	–	(-3.136)	(-2.057)	–	(-2.200)	–	–	–	–
MPPD_of	2.422	3.423	–	–	1.815	2.371	1.949	0.939	–
	(5.023)	(5.493)	–	–	(2.479)	(3.943)	(5.575)	(2.847)	–
MPPD_p20_45	2.180	–	–	–	–	–	–	–	–
	(4.770)	–	–	–	–	–	–	–	–
MPPD_p46_79	0.975	1.391	1.204	1.188	–	–	–	–	1.188
	(3.396)	(4.970)	(3.505)	(3.668)	–	–	–	–	(3.668)
MPPD_pre1920	–	–	0.990	1.509	0.531	–	–	–	1.509
	–	–	(4.255)	(6.215)	(2.237)	–	–	–	(6.215)
MPPD_prop_bung	1.544	1.574	–	0.927	1.503	1.139	–	–	0.927
	(3.927)	(3.229)	–	(2.081)	(3.328)	(2.591)	–	–	(2.081)
MPPD_prop_det	2.081	2.243	1.430	1.760	2.255	1.741	1.576	1.102	1.760
	(7.184)	(7.300)	(8.061)	(9.573)	(7.136)	(5.507)	(8.794)	(7.478)	(9.573)
MPPD_prop_sdet	–	1.278	–	0.839	0.829	0.946	–	–	0.839
	–	(4.093)	–	(3.106)	(2.595)	(2.915)	–	–	(3.106)
MPPD_prop_terr	0.918	1.281	–	–	1.099	0.929	–	–	–
	(3.212)	(3.822)	–	–	(2.934)	(2.465)	–	–	–
+ Mosaic categories									
N	1,307	1,306	1,307	1,307	1,307	1,307	1,307	1,306	1,307
r2_a	0.315	0.3	0.268	0.275	0.24	0.267	0.279	0.27	0.275

t-values based on HC2 standard errors

6.4.2 OLS RESULTS

We repeated the analysis without taking logs (Table 6-3). The magnitudes of the coefficients are more difficult to interpret, but the signs on the coefficients (+ or -) have broadly the same meaning, and generally confirm the findings of the log-log model. The goodness of fit measure ("Adjusted R^2") was generally better for the linear model than the OLS model, and the t-ratios on variables were typically higher. As such, the log-log functional form, although having the advantage of being easy to interpret, is unlikely to be the most appropriate way of modelling the data, particularly if our goal were to use it to predict the probability of sale as a means of correcting house price indices (as in the Heckman model).

Table 6-4 Cubic OLS Regression Results

Variable	CbOLS1996	CbOLS1997	CbOLS1998	CbOLS1999	CbOLS2000	CbOLS2001	CbOLS2002	CbOLS2003	CbOLS2004
Constant	-0.090	-0.433	0.141	0.099	0.775	1.758	2.319	1.632	0.099
	(-0.410)	(-1.444)	(0.557)	(0.372)	(2.884)	(5.856)	(25.313)	(11.982)	(0.372)
aSWc2to5_p1	11,900	12,300	11,700	10,900	7,539	2,566	1,504	2,371	10,900
	(8.877)	(8.180)	(8.362)	(8.420)	(6.780)	(2.805)	(2.016)	(3.446)	(8.420)
aveFPA	–	–	–	–	–	-0.001	-0.001	–	–
	–	–	–	–	–	(-2.517)	(-2.272)	–	–
aveFPAx2	–	–	–	–	–	0.000	0.000	–	–
	–	–	–	–	–	(2.157)	(2.053)	–	–
aveNNm2	-0.014	-0.013	-0.014	-0.015	-0.016	-0.009	-0.013	-0.009	-0.015
	(-5.366)	(-4.211)	(-5.597)	(-4.895)	(-6.246)	(-3.285)	(-4.350)	(-5.068)	(-4.895)
aveNNm2x2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(4.493)	(3.710)	(4.541)	(3.802)	(5.212)	(3.031)	(3.247)	(4.020)	(3.802)
aveNNm2x3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	–	0.000
	(-4.056)	(-3.538)	(-4.047)	(-3.270)	(-4.721)	(-3.017)	(-2.753)	–	(-3.270)
ht_elevation	-0.006	-0.007	-0.005	-0.006	-0.005	-0.009	-0.006	–	-0.006
	(-3.341)	(-4.568)	(-2.740)	(-3.755)	(-3.125)	(-3.017)	(-3.475)	–	(-3.755)
ht_elevationx2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	–	0.000
	(2.936)	(3.969)	(2.272)	(3.595)	(2.846)	(2.731)	(3.041)	–	(3.595)
ht_elevationx3	–	–	–	–	–	0.000	–	–	–
	–	–	–	–	–	(-2.340)	–	–	–
MPPD_of	–	6.032	–	–	–	1.464	–	–	–
	–	(2.849)	–	–	–	(2.883)	–	–	–
MPPD_ofx2	–	-28.861	–	–	–	–	3.345	–	-
	–	(-2.526)	–	–	–	–	(2.251)	–	–
MPPD_ofx3	–	47.613	–	–	–	–	–	8.527	–
	–	(2.651)	–	–	–	–	–	(2.562)	–
MPPD_p20_45	–	-4.334	-1.490	-8.738	–	–	–	–	-8.738
	–	(-2.782)	(-3.397)	(-2.355)	–	–	–	–	(-2.355)
MPPD_p20_45x2	–	6.456	–	29.790	–	–	–	–	29.790
	–	(2.228)	–	(2.091)	–	–	–	–	(2.091)
MPPD_p20_45x3	–	–	–	(-36.341)	–	–	–	–	(-6.341)
	–	–	–	(-2.110)	–	–	–	–	(-2.110)
MPPD_p46_79	3.524	5.026	4.177	5.554	–	–	–	–	5.554
	(5.837)	(5.916)	(5.405)	(5.861)	–	–	–	–	(5.861)
MPPD_p46_79x2	–	–	–	–	5.253	-1.502	–	–	–
	–	–	–	–	(2.274)	(-3.930)	–	–	–
MPPD_p46_79x3	-8.362	-10.633	-10.749	-13.968	-9.704	–	–	–	-13.968
	(-4.685)	(-4.898)	(-5.012)	(-6.085)	(-2.262)	–	–	–	(-6.085)
MPPD_pre1920	1.343	–	–	2.537	–	–	–	–	2.537
	(5.562)	–	–	(6.702)	–	–	–	–	(6.702)

Table 6-4 Cubic OLS Regression Results – *continued*

Variable	CbOLS1996	CbOLS1997	CbOLS1998	CbOLS1999	CbOLS2000	CbOLS2001	CbOLS2002	CbOLS2003	CbOLS2004
MPPD_pre1920x2	–	7.020	8.383	–	7.671	–	5.453	5.389	–
	–	(5.041)	(7.123)	–	(6.054)	–	(5.595)	(5.724)	–
MPPD_pre1920x3	–	-8.257	-9.628	–	-8.444	–	-7.500	-7.648	–
	–	(-4.405)	(-5.640)	–	(-4.650)	–	(-4.835)	(-4.596)	–
MPPD_prop_bung	–	–	–	–	–	1.483	–	–	–
	–	–	–	–	–	(3.715)	–	–	–
MPPD_prop_bungx3	18.084	19.825	18.168	16.937	18.451	–	–	10.911	16.937
	(5.746)	(4.779)	(5.969)	(4.519)	(5.255)	–	–	(3.212)	(4.519)
MPPD_prop_det	1.731	4.600	1.544	3.072	2.954	1.459	2.726	3.089	3.072
	(7.969)	(4.105)	(6.706)	(5.458)	(8.179)	(4.322)	(10.029)	(7.078)	(5.458)
MPPD_prop_detx2	–	-8.179	–	-1.863	–	–	–	-2.353	-1.863
	–	(-2.399)	–	(-2.283)	–	–	–	(-4.035)	(-2.283)
MPPD_prop_detx3	–	7.001	–	–	-1.515	–	-2.458	–	–
	–	(2.147)	–	–	(-2.570)	–	(-4.181)	–	–
MPPD_prop_sdet	3.287	3.532	5.978	6.323	0.566	4.003	–	–	6.323
	(3.792)	(3.714)	(3.590)	(3.599)	(2.052)	(2.517)	–	–	(3.599)
MPPD_prop_sdetx2	-4.622	-3.842	-15.979	-16.935	–	-11.036	–	–	-16.935
	(-2.979)	(-2.561)	(-2.954)	(-3.221)	–	(-2.297)	–	–	(-3.221)
MPPD_prop_sdetx3	–	–	14.365	16.005	–	9.350	–	–	16.005
	–	–	(2.723)	(3.245)	–	(2.246)	–	–	(3.245)
MPPD_prop_terr	–	0.828	–	–	1.705	1.754	–	–	–
	–	(2.433)	–	–	(2.810)	(2.077)	–	–	–
MPPD_prop_terrx2	–	–	–	–	–	-2.416	–	5.086	–
	–	–	–	–	–	(-2.164)	–	(2.491)	–
MPPD_prop_terrx3	–	–	–	–	-3.015	–	–	-7.661	–
	–	–	–	–	(-2.130)	–	–	(-2.304)	–
+ Mosaic categories									
N	1,307	1,306	1,307	1,307	1,307	1,307	1,307	1,306	1,307
r2_a	0.356	0.363	0.32	0.332	0.300	0.321	0.341	0.319	0.332

t-values based on HC2 standard errors

6.4.3 CUBIC OLS MODEL

The non-linearities evident in the bar-char graphs presented in the separate Appendix suggest that we should in fact attempt more exotic transformations of the explanatory variables if we are to create a well-fitting model (particularly important if our goal was prediction). In Table 6-4 we present the results of regressions in which each continuous explanatory variable was subject to linear, quadratic, and cubic transformations, and retained in the regression if the coefficients proved significantly different from zero. This process generally improved the model fit (the Adjusted R2 figures were in the range 0.300 to 0.363). Such a regression is difficult to interpret (even more so than the linear OLS model) and would really be used for prediction purposes only, so we shall add no further comment other than to say that the results present fairly irrefutable evidence of the existence of non-linearities.

Table 6-5 Regressions on % Sold Over Entire Period at the Level of Individual Postcodes

	Coefficient	t	lower	upper
Constant	5.854	31.690	5.49170	6.21587
aSWc2to50_p1	104.656	12.700	88.49872	120.81300
aveFPA	-0.0003	-2.030	-0.00055	-0.00001
aveNNm2	-0.010	-23.720	-0.01099	-0.00931
ht_elevation	-0.003	-7.930	-0.00376	-0.00227
MPPD_of	0.401	2.270	0.05409	0.74883
MPPD_p20_45	-0.472	-5.090	-0.65311	-0.29020
MPPD_p46_79	-1.203	-16.410	-1.34678	-1.05932
MPPD_pr~_det	0.912	11.670	0.75836	1.06466
MPPD_prop_~r	0.682	9.360	0.53914	0.82467
+ Mosaic categories				
n	193,282			
Adj. R^2	0.156			

t-values based on HC2 standard errors

6.4.4 REGRESSIONS RUN ON INDIVIDUAL POSTCODE UNITS

We ran the simple, untransformed regression also at postcode unit level. The disadvantage of this approach is that, at such a small spatial scale, each unit represents only eight houses or so. Consequently, the rate of stock turnover measure (our dependent variable) becomes rather 'lumpy' at the lower end. To illustrate, consider a postcode with eight dwellings where no properties sell. The rate of turnover, S_{it}, equals $0 \div 8 = 0\%$. Now consider another postcode which also has eight dwellings, but where one property sells. The rate of turnover for this postcode would be $1 \div 8 = 12.5\%$. To overcome this granularity problem, we run a regression on the entire 9 year period, which results in a much more fine grained dependent variable. We assume a property can potentially sell once a year, so the dependent variable is computed as the number of properties that sell in a postcode, divided by 9 times the total number of properties in the postcode. So if a postcode has only one property sale in the entire period, and there are in total eight dwellings in that postcode, S_{it} is computed as 1/(9x8) = 1.39%.

These complications are more than outweighed by two very substantial advantages to running regressions at postcode level for the entire period. First, the increase in sample size is almost twenty fold. There are nearly 200,000 post code units in the South East, and this allows us to derive very robust estimates of the effects of the proposed explanatory variables on the probability of sale. Second, there is the issue of duration dependence. If we run the regressions on individual years, we have to be aware that the proportion of dwellings that sold in a postcode in the previous year affects the proportion of properties that will sell in the current year. This is because, households who only moved-in last year, are highly unlikely to sell this year (for the reasons set out in Chapter 4, section 4.3.1). Taking the whole period together, we would expect short run duration dependency effects to cancel each other out. The results presented in Table 6-5 are therefore in many ways the most robust confirmation we can offer that there are indeed significant density, elevation, house type and spatial spillover effects. The evidence for the density and spillover effects is particularly overwhelming. The t-ratios of -23.72 and 12.70 respectively represent an infinitesimally small probability that the estimated effects are spurious. Elevation also has a very large t-ratio (t = -7.93), as does the proportion of properties built between 1946 and 1979 (t = -16.41), the proportion built between 1920 and 1945 (t = -5.1) the proportion detached (t = 11.67), and the proportion terraced (t = 9.360).

6.5 Impact on House Price Indices

Taken together, these results provide a very substantial body of evidence in support of the hypothesis that stock turnover varies systematically across space. Perhaps most interesting of all, is the finding that the proportion of dwellings that sell is dependent on construction density. This means that house price indices, as they are currently measured, are likely to have a rather complex relationship with housing and planning policy. Recent emphasis on encouraging high density new construction, will over time feed through into significant increases in the proportion of the stock that are high density, which is likely to suppress prices in the high-density sector, but not necessarily so in the low-density sector. As such the gulf in price appreciation is likely to widen, with property in low-density areas becoming ever more valuable relative to their high-density equivalents. But since our analysis has shown that properties in low density areas are significantly less likely to enter the market, and hence enter the computations of price index algorithms, the growing appreciation of low density areas will not be reflected in the estimates of price inflation. In short, house price inflation will be reported as stabilising, when in actual fact, the true average value of the housing stock may well be rising, and the variance in house price inflation may be increasing.

Whether or not this hypothetical scenario bears any relation to reality depends critically on whether low-density areas have different rates of inflation. If there is indeed evidence that density and house price inflation are not independent even before the impact of a policy to promote high-density construction has taken effect, then there is a very real basis for believing that existing house price measures not only mis-represent true inflation rates, but will do so to an increasing degree as the cumulative impact of planning policy brings about further change in the composition of the housing stock.

Our goal in this the final stage of our analysis is to explore whether any such evidence exists. That is, whether there is any empirical basis for believing that inflation rates are higher (or indeed lower) for properties located in low density-areas, than for properties located in high-density areas.

We employ a simple test to establish or reject our case. First we regress selling price on attributes and year dummies without any sample-selection-bias correction term. We then compare the predicted constant quality price series derived from this regression with that derived from an adjusted version of the regression. Since the Land Registry data is used as a benchmark for other indices (see the description of the FT index, for example, in Chapter 3), it makes sense to test whether there is any bias in this, the most comprehensive of housing samples. If it can be established that bias does indeed exist in the Land Registry data, then it is logical to conclude that the bias in other indices will be at least as great as that found in the Land Registry data. It will also imply that attempts to use the Land Registry data to correct errors in less comprehensive data sources (as in the FT approach) will be futile.

A further advantage of using the Land Registry data is that it provides us with a truly enormous sample (more than 1.5 million transactions between 1996 and 2004), which means that our results are very unlikely to subject to random sampling variation error – a source of error that frustrates the interpretation of results based on small samples – and any miscomputation of the standard errors (used to derive the t-ratios) will be less of a concern because the t-ratios on any genuine effects will be so large anyway.

6.5.1 HEDONIC REGRESSION RESULTS

Results of the unadjusted regressions are presented in Table 6-6 for the South East as a whole (first column of coefficients) and then for each of the nine subsidiary counties. The explanatory variables are limited to those provided in the Land Registry data: type of dwelling (detached, semi-detached, terraced, bungalow, flat or other), and year of sale. All variables are highly significant.

Results of the adjusted regressions are presented in Table 6-7. The adjustment we have introduced is very simple. We simply introduce, as an explanatory variable, the probability that a property will not enter the Land Registry sample (this probability is computed as one minus the dependent variable in the stock turnover regression reported in Table 6-5). We also interact this probability with the time dummies to pick up any variation in the effect on selling price of the probability of sale over time. The results show that the probability of non-selection variable is highly significant in the overall sample ($t = -7.763$), and also highly significant in most of the county regressions. Also, most of the interactive terms are highly significant (particularly the 2004 interaction variable).

6.5.2 IMPACT OF SAMPLE SELECTION BIAS ON HOUSE PRICE INFLATION ESTIMATES

We constructed constant quality house price series were from the unadjusted hedonic regression and compared these series with their adjusted counterparts. The results of this process are presented in Table 6-8. Column (3) of the table lists the unadjusted constant quality house price averages for each year, and column (6) presents the bias-adjusted averages. Figures in columns (4) and (7) refer to the annual percentage change in each series respectively. Columns (5) and (8) list the cumulative percentage change since 1996.

The final two columns directly compare the unadjusted and adjusted results. Column (9) presents the simple percentage point difference between the adjusted and the unadjusted annual inflation rates. For example, in 2000 the unadjusted annual inflation rate in Berkshire was 26.4%, whereas the adjusted inflation rate was 35.1%. The absolute difference between these two rates, 26.4 – 35.1, equals –8.6 percentage points. This means that unadjusted inflation estimate underestimates the actual inflation of the entire housing stock by 8.6 percentage points (see column (9)). On average, the unadjusted regressions understate the level of annual house price inflation by between 0.2 percentage points and 5.7% percentage points.

The final column – column (10) – tries to summarise the difference between the *cumulative* price inflation estimate of the unadjusted regressions as compared with the adjusted regressions. Our measure is simple: we compute the difference between the two final cumulative change figures (i.e. total cumulative % change in house prices between 1996 and 2004) and then present this difference as a proportion of the unadjusted cumulative change estimate. We find that our summary measure varies considerably, from a trivial positive value 0.4% to a very substantial negative value of –55.6%.

There are two important implications of this finding. First, in some areas the extent to which unadjusted hedonic based index measures underestimate cumulative house price inflation can be very substantial indeed, particularly over prolonged periods. Second, the fact that sample selection bias varies so much between counties suggests that comparisons of unadjusted price inflation series between counties is potentially highly misleading.

To illustrate this second point, compare the unadjusted cumulative inflation results for Oxfordshire (231%) with those of Surrey (262%). On this basis, the rate of inflation in the two areas is not that dissimilar. Now compare the inflation results derived from the transactions-bias corrected procedure. For Oxfordshire the adjusted rate of 230% is almost identical to the adjusted estimate. In contrast, the adjusted figure for Surrey (407%) is massively greater than the unadjusted value, indicating that in reality, the appreciation of the value of the housing stock in these two counties has been very different. Clearly, the implications for planning decisions and housing policy are profound. Based on the unadjusted estimates, we might conclude that both areas need a similar proportionate increase in new build to ameliorate house price inflation. Using the adjusted series we arrive at the very opposite conclusion: Surrey needs a far more radical boost to housing supply to restore price stability.[5]

Further research needs to be done into whether such large discrepancies occur between areas at smaller spatial scales (such as local authorities), but we can see no obvious *a priori* reason why this should not be the case. Undoubtedly, the hedonic regressions used to produce these results could be improved, but the results are of such a high order of magnitude, it is unlikely that improvements in functional form will remove the overwhelming implications of these results (indeed, they may actually exacerbate them). Potentially, datasets that have more attribute and neighbourhood information could yield hedonics which capture much of the sample selection probability variable, and mitigate the impact of having a biased sample. Against this, however, one has to bear in mind that the datasets which tend to have detailed attribute and neighbourhood information, tend to have an even less representative sample of dwellings (such as data from lenders or estate agents).

Put more positively, however, the results of this study indicate that it is feasible to construct an effective sample selection measure from existing data, not only for the South East, but potentially for all UK regions. This correction term could easily be incorporated into the main measures of house price inflation currently published.

5 In actual fact, the relationship between new construction and the stability of house prices, partly because of the impact of submarkets (Pryce 2004), and partly because of the complexities of the macro-economic adjustment mechanism (Meen 2005).

6.6 Conclusion

In this chapter we have attempted to investigate whether there exists evidence of systematic variation in the probability sale across space and, in particular, whether the incidence of sale and the type of neighbourhood are at all correlated. We conducted our analysis at both postcode and postcode sector levels, and used both bivariate and multiple regression techniques. Our results suggested that there was overwhelming support for the hypothesis that the probability of sale varies across space. We found that a significantly smaller proportion of dwellings in low-density areas trade than in high-density areas. Height above sea level was also found to have a negative effect on the probability of sale, and to a lesser extent, size and type of property.

The final step in our investigation attempted to estimate whether spatial variation in the probability of sale had any significant impact on house price indices. Again our results were unambiguous. Based on very large samples, we produced results which indicated very large sample selection bias in unadjusted house price inflation estimates. We also found that sample selection bias varies greatly between counties which means that comparisons of unadjusted price inflation series between counties is potentially highly misleading.

The results of this study also demonstrate that it is feasible to construct an effective sample selection measure from existing data, not only for the South East, but potentially for all UK regions. This correction term could easily be incorporated into the main measures of house price inflation currently published.

Table 6-6 Unadjusted Hedonic Regressions for Each County in S. East

Variable	OLSALL1U~j	OLScount~j	OLScount~j	OLScount~j	OLScount~j	OLScount~j	OLScount~j	OLScount~j	OLScount~j	OLScount~j
type_1	80,419	88,287	103,838	32,958	58,671	40,280	74,556	65,752	113,986	58,848
	(70.672)	(37.346)	(26.948)	(13.197)	(12.623)	(9.444)	(59.826)	(15.257)	(28.303)	(21.57)
type_2	-76,589	-83,966	-77,839	-85,038	-79,519	-43,903	-58,569	-82,062	-115,400	-90,249
	(-66.833)	(-35.265)	(-19.863)	(-34.156)	(-17.012)	(-10.223)	(-45.89)	(-18.68)	(-28.521)	(-32.955)
type_3	-28,396	-33,860	-31,012	-41,017	-32,731	-14,083	-14,291	-44,206	-57,408	-33,140
	(-24.983)	(-14.372)	(-8.051)	(-16.419)	(-7.038)	(-3.296)	(-11.555)	(-10.294)	(-14.265)	(-12.14)
type_4	-55,932	-61,412	-59,688	-51,792	-56,975	-27,341	-41,979	-57,327	-82,739	-57,740
	(-49.216)	(-26.061)	(-15.441)	(-20.77)	(-12.284)	(-6.384)	(-34.034)	(-13.282)	(-20.424)	(-21.16)
year_2	11,646	14,071	14,209	9,882	9,671	7,261	7,154	12,329	21,535	11,741
	(25.488)	(12.577)	(9.8)	(11.513)	(6.775)	(4.929)	(10.96)	(6.77)	(14.744)	(12.684)
year_3	25,937	31,659	29,143	22,434	21,551	16,105	17,663	30,059	45,967	25,754
	(57.362)	(28.265)	(20.415)	(26.473)	(15.293)	(11.024)	(27.367)	(16.835)	(31.683)	(27.816)
year_4	38,508	46,635	41,790	33,843	33,144	24,281	25,502	43,969	63,120	36,925
	(87.875)	(43.382)	(30.33)	(41.431)	(24.307)	(17.162)	(40.38)	(25.533)	(45.069)	(40.958)
year_5	62,656	78,377	66,946	55,718	53,814	39,501	42,396	76,166	102,447	62,036
	(139.925)	(70.396)	(47.732)	(66.71)	(38.466)	(27.433)	(66.731)	(43.208)	(70.966)	(67.303)
year_6	78,385	92,430	82,125	71,841	70,136	53,365	57,266	86,890	119,619	77,241
	(177.693)	(85.011)	(58.896)	(86.418)	(50.6)	(37.031)	(91.728)	(51.699)	(84.079)	(85.048)
year_7	103,603	117,207	108,696	97,199	95,131	75,774	79,716	115,574	149,752	104,476
	(239.438)	(108.61)	(79.245)	(119.867)	(69.915)	(53.644)	(130.72)	(69.092)	(107.491)	(118.264)
year_8	126,141	134,924	135,036	121,311	117,498	97,926	101,516	139,089	170,638	130,208
	(281.424)	(120.361)	(94.663)	(144.287)	(83.008)	(68.022)	(161.555)	(80.999)	(117.781)	(142.363)
year_9	144,818	147,300	154,837	137,817	133,488	123,034	123,616	155,666	191,102	147,792
	(324.949)	(133.351)	(108.626)	(165.322)	(94.851)	(85.538)	(196.569)	(91.53)	(133.193)	(162.432)
Constant	94,438	107,290	89,247	102,815	95,582	56,001	76,074	111,554	130,428	97,742
	(81.966)	(44.804)	(22.84)	(41.112)	(20.384)	(12.963)	(59.176)	(25.621)	(32.203)	(35.826)
N	1,599,856	162,949	141,321	177,912	326,185	31,910	304,300	92,962	206,309	156,008
r2_a	0.268	0.394	0.352	0.386	0.126	0.448	0.399	0.302	0.362	0.455

Table 6-7 Hedonic Regressions for Each County in S. East Adjusted for Sample Selection Bias

Variable	OLSALL1	OLScoun~23	OLScoun~24	OLScoun~25	OLScoun~26	OLScoun~27	OLScoun~28	OLScoun~29	OLScoun~30	OLScoun~31
type_1	80,649	89,435	102,172	32,532	58,605	40,022	75,002	64,429	116,893	46,902
	(68.215)	(38.027)	(26.586)	(13.274)	(12.582)	(9.431)	(60.220)	(15.294)	(24.713)	(14.409)
type_2	-74,768	-82,507	-75,959	-85,757	-79,509	-43,912	-58,289	-82,141	-117,400	-88,173
	(-62.763)	(-34.817)	(-19.445)	(-35.088)	(-16.974)	(-10.275)	(-45.669)	(-19.119)	(-24.701)	(-27.014)
type_3	-27,123	-32,244	-32,782	-39,447	-32,429	-14,099	-13,359	-43,956	-58,172	-34,224
	(-22.968)	(-13.757)	(-8.533)	(-16.092)	(-6.957)	(-3.317)	(-10.810)	(-10.472)	(-12.303)	(-10.507)
type_4	-54,335	-59,631	-59,589	-50,960	-56,542	-27,477	-41,322	-57,221	-84,926	-55,255
	(-46.028)	(-25.441)	(-15.463)	(-20.829)	(-12.166)	(-6.447)	(-33.529)	(-13.564)	(-17.855)	(-16.958)
LRpcAll9604W	-58,922	-75,758	41,730	-78,353	-56,578	-22,081	-56,875	29,948	-186,800	-69,362
	(-7.763)	(-4.287)	(2.055)	(-6.568)	(-2.407)	(-0.893)	(-4.953)	(0.993)	(-6.868)	(-3.690)
year_2	3,243	6,327	-26,878	19,417	-5,732	20,477	19,573	10,140	-40,858	-19,695
	(0.350)	(0.292)	(-1.069)	(1.353)	(-0.201)	(0.687)	(1.407)	(0.273)	(-1.230)	(-0.848)
year_3	-14,544	-10,439	-85,747	28,603	-9,598	16,203	-8,989	28,939	-65,143	-36,590
	(-1.616)	(-0.490)	(-3.487)	(2.031)	(-0.349)	(0.569)	(-0.673)	(0.821)	(-2.008)	(-1.590)
year_4	29,181	50,019	16,015	50,488	2,894	58,650	37,661	136,978	-12,887	-14,184
	(3.475)	(2.536)	(0.699)	(3.849)	(0.112)	(2.181)	(3.015)	(4.206)	(-0.425)	(-0.666)
year_5	976	19,483	-73,323	44,565	9,463	10,129	20,166	90,271	-55,804	-52,066
	(0.110)	(0.915)	(-3.052)	(3.246)	(0.343)	(0.354)	(1.526)	(2.607)	(-1.723)	(-2.319)
year_6	17,814	-16,183	-45,175	87,407	14,680	34,924	22,126	180,798	-3,573	-42,790
	(2.010)	(-0.769)	(-1.874)	(6.297)	(0.536)	(1.197)	(1.685)	(5.366)	(-0.112)	(-1.886)
year_7	35,082	14,240	1,049	98,230	15,283	40,464	41,367	187,866	16,676	-42,687
	(4.033)	(0.691)	(0.044)	(7.290)	(0.566)	(1.436)	(3.210)	(5.633)	(0.531)	(-1.929)
year_8	39,777	-7,088	-13,979	95,066	4,916	74,969	63,055	238,060	-1,046	-24,496
	(4.444)	(-0.330)	(-0.566)	(6.879)	(0.175)	(2.647)	(4.794)	(6.958)	(-0.033)	(-1.087)
year_9	-58,577	-115,100	-136,800	43,117	-66,975	-6,093	-16,570	67,422	-168,700	-123,400
	(-6.476)	(-5.387)	(-5.404)	(3.077)	(-2.373)	(-0.211)	(-1.237)	(1.972)	(-5.228)	(-5.365)
LRpcAll9604W_97	8,860	8,492	45,874	-10,890	16,738	-14,746	-13,762	2,555	68,817	32,698
	(0.868)	(0.354)	(1.643)	(-0.683)	(0.533)	(-0.449)	(-0.907)	(0.063)	(1.882)	(1.277)

Table 6-7 Hedonic Regressions for Each County in S. East Adjusted for Sample Selection Bias – *continued*

Variable	OLSALL1	OLScoun~23	OLScoun~24	OLScoun~25	OLScoun~26	OLScoun~27	OLScoun~28	OLScoun~29	OLScoun~30	OLScoun~31
LRpcAll9604W_98	44,118	46,156	128,206	-7,235	34,054	-479	28,986	1,138	123,303	66,660
	(4.445)	(1.957)	(4.692)	(-0.462)	(1.122)	(-0.015)	(1.985)	(0.029)	(3.449)	(2.627)
LRpcAll9604W_99	8,977	-4,698	29,521	-20,001	32,975	-38,824	-14,447	-102,900	83,446	51,527
	(0.967)	(-0.215)	(1.156)	(-1.368)	(1.156)	(-1.305)	(-1.056)	(-2.877)	(2.489)	(2.186)
LRpcAll9604W_00	67,010	64,937	156,689	11,569	48,448	32,156	24,084	-16,342	177,498	119,973
	(6.824)	(2.756)	(5.871)	(0.758)	(1.594)	(1.020)	(1.666)	(-0.431)	(4.977)	(4.845)
LRpcAll9604W_01	65,639	120,059	141,180	-18,135	60,787	20,258	38,096	-102,900	138,263	124,723
	(6.727)	(5.164)	(5.280)	(-1.177)	(2.017)	(0.630)	(2.655)	(-2.797)	(3.950)	(4.987)
LRpcAll9604W_02	74,039	113,468	119,088	-1,966	87,377	38,710	41,409	-79,628	149,619	154,898
	(7.736)	(4.983)	(4.520)	(-0.132)	(2.940)	(1.245)	(2.943)	(-2.188)	(4.337)	(6.358)
LRpcAll9604W_03	93,739	156,228	164,925	28,510	123,563	25,028	41,816	-108,600	191,932	164,838
	(9.520)	(6.595)	(6.030)	(1.860)	(4.001)	(0.801)	(2.912)	(-2.910)	(5.443)	(6.644)
LRpcAll9604W_04	221,380	288,333	319,261	104,814	219,232	141,540	152,563	94,323	397,534	289,899
	(22.333)	(12.265)	(11.442)	(6.764)	(7.091)	(4.453)	(10.467)	(2.539)	(11.247)	(11.495)
Constant	146,360	174,197	52,454	172,773	146,741	76,150	127,456	84,257	301,163	160,145
	(20.957)	(10.790)	(2.821)	(15.681)	(6.731)	(3.329)	(12.032)	(3.027)	(12.007)	(9.242)
N	1,599,856	162,949	141,321	177,912	326,185	31,910	304,300	92,962	206,309	156,008
r2_a	0.261	0.398	0.357	0.395	0.126	0.451	0.400	0.310	0.354	0.440

Table 6-8 Impact of Sample Selection Bias on House Price Inflation Estimates

(Based on a total sample of 1.5 million Housing Transactions in S.East over the Period 1996-2004)

(1) County	(2) Year	(3) Unadjusted CQ Price	(4) Annual % Change	(5) Cumulative % Change since 1996	(6) Adjusted CQ Price	(7) Annual % Change	(8) Cumulative % Change since 1996	(9) % point difference in Annual Inflation	(10) % Difference in Cummulative Inflation since 1996
Berkshire	1996	£ 73,430			£ 66,195				
	1997	£ 87,501	19.2%	19.2%	£ 81,014	22.4%	22.4%	-3.2%	
	1998	£105,089	20.1%	43.1%	£101,912	25.8%	54.0%	-5.7%	
	1999	£120,065	14.3%	63.5%	£111,517	9.4%	68.5%	4.8%	
	2000	£151,807	26.4%	106.7%	£150,615	35.1%	127.5%	-8.6%	
	2001	£165,860	9.3%	125.9%	£170,072	12.9%	156.9%	-3.7%	
	2002	£190,636	14.9%	159.6%	£193,903	14.0%	192.9%	0.9%	
	2003	£208,353	9.3%	183.7%	£215,335	11.1%	225.3%	-1.8%	
	2004	£220,730	5.9%	200.6%	£239,399	11.2%	261.7%	-5.2%	-30.4%
							Average:	-2.8%	
Bucks	1996	£ 58,235			£ 61,402				
	1997	£ 72,444	24.4%	24.4%	£ 80,398	30.9%	30.9%	-6.5%	
	1998	£ 87,377	20.6%	50.0%	£103,861	29.2%	69.1%	-8.6%	
	1999	£100,025	14.5%	71.8%	£106,938	3.0%	74.2%	11.5%	
	2000	£125,181	25.1%	115.0%	£144,768	35.4%	135.8%	-10.2%	
	2001	£140,360	12.1%	141.0%	£157,407	8.7%	156.4%	3.4%	
	2002	£166,930	18.9%	186.7%	£181,539	15.3%	195.7%	3.6%	
	2003	£193,271	15.8%	231.9%	£212,347	17.0%	245.8%	-1.2%	
	2004	£213,072	10.2%	265.9%	£243,907	14.9%	297.2%	-4.6%	-11.8%
						Average:	-1.6%		
East Sussex	1996	£ 61,798			£ 54,973				
	1997	£ 71,680	16.0%	16.0%	£ 63,500	15.5%	15.5%	0.5%	
	1998	£ 84,232	17.5%	36.3%	£ 76,341	20.2%	38.9%	-2.7%	
	1999	£ 95,640	13.5%	54.8%	£ 85,460	11.9%	55.5%	1.6%	
	2000	£117,516	22.9%	90.2%	£111,108	30.0%	102.1%	-7.1%	
	2001	£133,638	13.7%	116.3%	£124,245	11.8%	126.0%	1.9%	
	2002	£158,997	19.0%	157.3%	£151,236	21.7%	175.1%	-2.7%	
	2003	£183,108	15.2%	196.3%	£178,549	18.1%	224.8%	-2.9%	
	2004	£199,615	9.0%	223.0%	£202,904	13.6%	269.1%	-4.6%	-20.7%
							Average:	-2.0%	

Table 6-8 Impact of Sample Selection Bias on House Price Inflation Estimates – *continued*

(Based on a total sample of 1.5 million Housing Transactions in S.East over the Period 1996-2004)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
County	Year	Unadjusted CQ Price	Annual % Change	Cumulative % Change since 1996	Adjusted CQ Price	Annual % Change	Cumulative % Change since 1996	% point difference in Annual Inflation	% Difference in Cummulative Inflation since 1996
Hampshire	1996	£ 62,850			£ 57,734				
	1997	£ 72,521	15.4%	15.4%	£ 68,739	19.1%	19.1%	-3.7%	
	1998	£ 84,401	16.4%	34.3%	£ 82,189	19.6%	42.4%	-3.2%	
	1999	£ 95,994	13.7%	52.7%	£ 93,603	13.9%	62.1%	-0.2%	
	2000	£116,664	21.5%	85.6%	£115,645	23.5%	100.3%	-2.0%	
	2001	£132,986	14.0%	111.6%	£133,200	15.2%	130.7%	-1.2%	
	2002	£157,982	18.8%	151.4%	£160,393	20.4%	177.8%	-1.6%	
	2003	£180,348	14.2%	186.9%	£186,213	16.1%	222.5%	-1.9%	
	2004	£196,338	8.9%	212.4%	£209,991	12.8%	263.7%	-3.9%	-24.2%
							Average:	-2.2%	
Isle of Wight	1996	£ 41,918			£ 39,970				
	1997	£ 49,179	17.3%	17.3%	£ 45,701	14.3%	14.3%	3.0%	
	1998	£ 58,023	18.0%	38.4%	£ 55,695	21.9%	39.3%	-3.9%	
	1999	£ 66,199	14.1%	57.9%	£ 59,796	7.4%	49.6%	6.7%	
	2000	£ 81,419	23.0%	94.2%	£ 82,255	37.6%	105.8%	-14.6%	
	2001	£ 95,283	17.0%	127.3%	£ 95,152	15.7%	138.1%	1.3%	
	2002	£117,692	23.5%	180.8%	£119,145	25.2%	198.1%	-1.7%	
	2003	£139,844	18.8%	233.6%	£139,967	17.5%	250.2%	1.3%	
	2004	£164,952	18.0%	293.5%	£175,417	25.3%	338.9%	-7.4%	-15.5%
							Average:	-1.9%	
Kent	1996	£ 61,783			£ 57,222				
	1997	£ 68,938	11.6%	11.6%	£ 63,033	10.2%	10.2%	1.4%	
	1998	£ 79,446	15.2%	28.6%	£ 77,219	22.5%	34.9%	-7.3%	
	1999	£ 87,285	9.9%	41.3%	£ 80,437	4.2%	40.6%	5.7%	
	2000	£104,179	19.4%	68.6%	£101,472	26.2%	77.3%	-6.8%	
	2001	£119,050	14.3%	92.7%	£117,444	15.7%	105.2%	-1.5%	
	2002	£141,499	18.9%	129.0%	£139,998	19.2%	144.7%	-0.3%	
	2003	£163,300	15.4%	164.3%	£162,093	15.8%	183.3%	-0.4%	
	2004	£185,400	13.5%	200.1%	£193,215	19.2%	237.7%	-5.7%	-18.8%
						Average:	-1.8%		

Table 6-8 Impact of Sample Selection Bias on House Price Inflation Estimates – *continued*

(Based on a total sample of 1.5 million Housing Transactions in S.East over the Period 1996-2004)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
County	Year	Unadjusted CQ Price	Annual % Change	Cumulative % Change since 1996	Adjusted CQ Price	Annual % Change	Cumulative % Change since 1996	% point difference in Annual Inflation	% Difference in Cummulative Inflation since 1996
Oxfordshire	1996	£ 67,348			£ 70,250				
	1997	£ 79,677	18.3%	18.3%	£ 82,945	18.1%	18.1%	0.2%	
	1998	£ 97,407	22.3%	44.6%	£100,328	21.0%	42.8%	1.3%	
	1999	£111,317	14.3%	65.3%	£104,310	4.0%	48.5%	10.3%	
	2000	£143,514	28.9%	113.1%	£144,179	38.2%	105.2%	-9.3%	
	2001	£154,238	7.5%	129.0%	£148,160	2.8%	110.9%	4.7%	
	2002	£182,922	18.6%	171.6%	£178,489	20.5%	154.1%	-1.9%	
	2003	£206,436	12.9%	206.5%	£199,752	11.9%	184.3%	0.9%	
	2004	£223,013	8.0%	231.1%	£231,995	16.1%	230.2%	-8.1%	0.4%
							Average:	-0.2%	
Surrey	1996	£ 73,021			£ 56,191				
	1997	£ 94,555	29.5%	29.5%	£ 84,149	49.8%	49.8%	-20.3%	
	1998	£118,988	25.8%	63.0%	£114,350	35.9%	103.5%	-10.0%	
	1999	£136,141	14.4%	86.4%	£126,749	10.8%	125.6%	3.6%	
	2000	£175,467	28.9%	140.3%	£177,885	40.3%	216.6%	-11.5%	
	2001	£192,639	9.8%	163.8%	£190,880	7.3%	239.7%	2.5%	
	2002	£222,772	15.6%	205.1%	£222,485	16.6%	295.9%	-0.9%	
	2003	£243,659	9.4%	233.7%	£247,076	11.1%	339.7%	-1.7%	
	2004	£264,123	8.4%	261.7%	£284,988	15.3%	407.2%	-6.9%	-55.6%
							Average:	-5.7%	
West Sussex	1996	£ 64,602			£ 56,558				
	1997	£ 76,343	18.2%	18.2%	£ 69,562	23.0%	23.0%	-4.8%	
	1998	£ 90,356	18.4%	39.9%	£ 86,628	24.5%	53.2%	-6.2%	
	1999	£101,527	12.4%	57.2%	£ 93,901	8.4%	66.0%	4.0%	
	2000	£126,638	24.7%	96.0%	£124,465	32.5%	120.1%	-7.8%	
	2001	£141,844	12.0%	119.6%	£138,492	11.3%	144.9%	0.7%	
	2002	£169,078	19.2%	161.7%	£168,769	21.9%	198.4%	-2.7%	
	2003	£194,811	15.2%	201.6%	£196,900	16.7%	248.1%	-1.4%	
	2004	£212,394	9.0%	228.8%	£223,018	13.3%	294.3%	-4.2%	-28.7%
							Average:	-2.8%	

7 Conclusion

7.1 Summary

This report set out to address five questions about the reliability of house price measurement:

1. Does it matter whether house prices are measured accurately or not?

2. Where does the sample come from?

3. What is the mix adjustment?

4. What about properties that have not recently sold?

5. What can be done to correct for transactions bias?

We discussed in Chapter 2 the question of whether the reliability of house price measurement is of any great import. We concluded that, given the great array of factors and decisions affected by this variable, reliability of measurement was indeed of great importance. House prices are used in the analysis of demand and supply imbalances, 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. Via these channels, distortions in published indices could significantly affect the efficiency of personal financial decisions, investment choices, and planning policy.

In Chapter 3 we summarised the differences between existing measures of house price change in the UK. We looked, in particular, at the differences in the samples used in these measures, and differences in the index-calculation techniques. We found that, while other sources of bias are corrected for, perhaps the most important source of bias – transactions bias – is not currently addressed in any of the existing UK indices.

We then conducted a brief review of the literature on sample selection bias. The results of this survey, presented in Chapter 4, revealed a dearth of UK research on this topic. This was partly because existing approaches to the correction problem, developed in the US, require data that are not currently available in the UK. As a result, we set ourselves the challenge of developing new ways of correcting for transactions bias that could potentially be constructed from available UK data, and which could conceivably also improve on the existing approaches in terms of theoretical robustness.

Chapter 5 marked a change in the nature of our investigation from being one that was largely theoretical/hypothetical, to one that was applied and empirical. We looked in particular at the issue of duration dependence (the tendency for the duration of stay in a dwelling to affect the current probability of moving) and how it undermines existing approaches to sample selection correction but could, ironically, point to new avenues of research that would yield methods amenable to UK application. A preliminary analysis was carried out on Scottish data, but we noted that the approach would not be useful south of the Border for some time to come because it is only very recently that it has been possible to identify repeat sales in English and Welsh Land Registry data.

In Chapter 5 we focussed on the spatial nature of the probability of sale, an aspect that has been almost completely overlooked, not only in UK research, but in the US and European literatures also. We argued that spatial concentrations of particular property types, neighbourhood types and socio-economic factors, could conspire to cause non-randomness across space in the probability of sale. If this were found to be the case, then the possibility emerges of finding a method of correction that could be applied to all parts of the UK. This opportunity arises because detailed data now exist on the characteristics of areas down to the level of individual postcodes.

Given the potential practical importance of being able to identify and measure systematic spatial variation in the probability of sale, we devoted considerable energy and time to investigating whether evidence could be found for systematic variation across space in the probability of sale. An unprecedented integration of data sources on the South East housing market was analysed to produce very large sample results on how the probability of sale varied by neighbourhood density, average elevation, and typical neighbourhood dwelling size, age and type. The result was an overwhelming body of evidence demonstrating that the probability of sale does indeed vary systematically across space.

In the final stage of our investigation we examined the extent to which spatial variations in the probability of sale could cause bias in existing methods of house price measurement. Bias-correction would be somewhat redundant if the final effect of the bias turns out to be negligible, so this final step was crucial to our argument. Our results were unambiguous. Based on very large samples, our estimates indicated very large sample selection bias in unadjusted house price inflation series in certain counties. We found that sample selection bias varied greatly between counties which means that comparisons of unadjusted price inflation series between counties is potentially highly misleading.

To illustrate, we compared the unadjusted cumulative inflation results for Oxfordshire (231%) with those of Surrey (262%) over the period 1996-2004. The story told by the unadjusted indices was that the rate of inflation in the two areas was not that dissimilar. We then compared the inflation results derived from the transactions-bias corrected procedure and found that, for Oxfordshire, the *adjusted* rate of 230% was almost identical to the *un*adjusted estimate. In contrast, the adjusted figure for Surrey (407%) was massively greater than the unadjusted value. Taken together, the results suggested that the appreciation of the housing stock in the two counties was in fact very different, leading to potentially profound implications for planning decisions and housing policy. For example, based on the *unadjusted* estimates, we might have concluded that both counties needed a similar proportionate increase in new build to ameliorate house price inflation. Using the *adjusted* series we would arrive at the very opposite conclusion: Surrey is likely to need a far more radical boost to housing supply if price stability is to be restored.

Perhaps the most positive and important implication of our results is that they demonstrate that it is indeed feasible to construct an effective sample selection measure from existing data, not only for the South East, but potentially for all UK regions. This correction term could easily be incorporated into the main measures of house price inflation currently published.

7.2 Recommendations

RECOMMENDATION 1

We recommend that the investigation of transactions bias be extended to examine other house price series (such as those based on mortgage lender data), and that a variety of index computation methods be investigated to assess the extent to which sample selection bias persists under different sampling regimes and computation methods.

RECOMMENDATION 2

We recommend that sample selection correction variables for the South East be made freely available to other housing economists and providers of house price information so that they can conduct their own analysis of the impact of including this correction term.

RECOMMENDATION 3

We recommend that more research be done on alternative correction terms. For example, the probability of non-selection could be predicted from Fractional Logit regression methods, and combined with duration-based methods (applied to survey data) to provide a comprehensive measure of the probability of non-selection.

RECOMMENDATION 4

This report has provided a compelling case for sample selection correction in house price calculation. We recommend that analysis of sample selection bias be extended to all other UK regions. By developing corrected price indices for all regions, it would be possible to estimate the extent to which transactions bias distorts existing estimates of differences between regions.

RECOMMENDATION 5

We recommend further investigation into the nature of spatial variation in house price inflation using “inflation surfaces” rather than indices for administrative areas. Such approaches could help avoid some of the misleading effects of transactions bias. More work also needs to be done on the *causes* of diverging price trajectories, particularly between low- and high-density areas.

8 References

Aoki, Kosuke; Proudman, James; Vlieghe, Gertjan. House prices, consumption, and the monetary policy: a financial accelerator approach. *Bank of England Quarterly Bulletin*, Spring 2003, Vol. 43 Issue 1, p47

Attanasio, O., Blow, L., Hamilton, R. and Leicester, A. (2005) "Consumption, house prices and expectations." Working paper 271, quotes taken from the summary provided in the *Bank of England Quarterly Bulletin;* Autumn 2005, Vol. 45 Issue 3, p378.

Bailey, M. H., R. F. Muth, and H. O. Nourse. (1963). "A Regression Method for Real Estate Price Index Construction," *Journal of the American Statistical Association* 4, 933-942.

Barker, K. "The housing market and the wider economy." *Bank of England Quarterly Bulletin*, Spring 2005, Vol. 45 Issue 1, p108-116.

Barker, K. (2003) *Review of Housing Supply (Interim Report – Analysis)*, HM Treasury, London.

Barker, K. (2004) *Review of Housing Supply (Final Report – Recommendations)*, HM Treasury, London.

Case, B., H. O. Pollakowski, and S. M. Wachter. (1997). "Frequency of Transaction and House Price Modelling," *Journal of Real Estate Finance and Economics* 14, 173-187.

Case, B., and J. M. Quigley. (1991). "The Dynamics of Real Estate Prices," *Review of Economics and Statistics* 73(1), 50-58.

Case, K. E., J. M. Quigley, and R. J. Shiller. (2001). "Stock Market Wealth, Housing Market Wealth, Spending, and Consumption." Paper prepared for the NBER Summer Institute, Cambridge, MA, July 2001.

Case, K. E., and R. J. Shiller. (1987). "Prices of Single-Family Homes Since 1970: New Indexes for Four Cities," *New England Economic Review* 45-56.

Calhoun, C. A. (1996). *OFHEO House Price Indexes: HPl Technical Description,* Office of Federal Housing Enterprise Oversight, 1996.

Clapp, J. M., and C. Giacotto. (1992). "Estimating Price Trends for Residential Property: A Comparison of Repeat Sales and Assessed Value Methods," *Journal of Real Estate Finance and Economics* 5, 357-374.

DiPasquale, D., and C. T. Somerville. (1995). "Do House Prices Based on Transactions Represent the Entire Housing Stock? Evidence from the National and Metropolitan HS, 1974 1991," *Journal of Housing Economics* 4, 100-110.

Englund, P., J. M. Quigley and C. Redfearn. (1998). "Improved Price Indexes for Real Estate: Measuring the Course of Swedish Housing Prices," *Journal of Urban Economics* 44, 171-196

Englund, P., J. M. Quigley, and C. Redfearn. (1999). "The Choice of Methodology for Computing Housing Price Indices," *Journal of Real Estate Finance and Economics* 19(2), 91-112.

Fik, T.J., Ling, D.C., Mulligan, G.F. (2003) Modelling spatial variation in housing prices, *Real Estate Econ.*, 31 (4): 623-646

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 and Repeat Sales Index Estimates," *Journal of Real Estate Finance and Economics* l4(1-2). 33-50.

Gatzlaff, D. H., and D. R, Haurin. (1998). "Sample Selection Biases in Local House Value Indices." *Journal of Urban Economics* 43(2), 199-222.

Gelfand, A.E., Kim, H.J. Sirmans, C.F. and Banerjee, S. (2003) "Spatial Modelling with Spatially Varying Coefficient Processes," *Journal of the American Statistical Association*, 98, 462, 387-396.

Goetzmann. W. N. (1993). "Accounting for Taste: Art and the Financial Markets over Three Centuries," *American Economic Review* 83(5), 1370-1377.

Goetzmann, W. N., and M. Spiegel. (1995). "Non-Temporal Components of Residential Real Estate Appreciation," *The Review of Economics and Statistics* 77, 199-206.

Goodman, A.C. and Dubin, R.A. (1990) Sample stratification with non-nested alternatives. *Rev. Econ. Stat.* 72, pp. 168-173.

Griliches, Z. (1971). *Price Indexes and Quality Change.* Cambridge. MA: Harvard University Press.

Hausman, J. A. (1979). "Individual Discount Rates and the Purchase and Utilization of Energy Using Durables," *Bell Journal of Economics* 1(1). 33-54.

Heckman. J. (1979). "Sample Selectivity Bias as a Specification Error," *Econometrica* 47(1), 153-161.

Hwang. M., and J J. M. Quigley. (2002). "Price Discovery in Time and Space: The course of Condominium Prices in Singapore," Berkeley Program on Housing and Urban Policy, Working Paper W02-002.

Jud,D.G., and T.G. Seaks. (1994). "Sample Selection Bias in Estimating Housing Sales Prices," *Journal of Real Estate Research* 9(3). 289-298.

Kain, J. R, and J. M. Quigley. (1970). "Measuring the Value of Housing Quality," J*ournal of the American Statistical Association* 65(44), 532-548.

Kain, J. F., and J. M. Quigley. (1975). *Housing Markets and Racial Discrimination.* New York, NY: Columbia University Press.

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.

Meese, R. A., and N. E. Wallace. (1997). "The Construction of Residential Housing Price Indices: A Comparison of Repeat-Sales, Hedonic-Regression, and Hybrid Approaches," *Journal of Real Estate Finance and Economics* 14(1-2), 51-73.

Otha, M.,and Z. Giliches. (1975). "Automobile Prices Revisited: Extensions of the Hedonic Price Hypothesis," In N. E. Terleckyj (cd.). *Household Production and Consumption.* Chicago, IL: University of Chicago Press.

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

Rosen, S. S. (1974). "Hedonic Prices and Implicit Markets." *Journal of Political Economy* 82(1). 34-55.

Thomas, B. and Dorling, D. (2004) *Know your Place*, Shelter.

Zellner, A. (1962). "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests of Aggregation Bias." *Journal of the American Statistical Association* 57. 500-509.