

(Conference in Honour of Pat Hendershott, Ohio, July 2006)

Controlling for Transactions Bias in Regional House Price Indices

Gwilym Pryce & Philip Mason

Introduction

- Aim:
 - To establish a method for correcting transactions bias in house price indices that could be applied to countries and regions where info on individual dwellings is not available for the whole stock.
- Funded by Office of the Deputy Prime Minister (now called DCLG):
 - Pryce, G. and Mason, P. (2006) *Which House Price? Finding the Right Measure of House Price Inflation for Housing Policy* - Technical Report, Office of the Deputy Prime Minister, ISBN: 05 ASD 03771/a.
 - Available from the *Housing Resources* page of www.gpryce.com

(i) Does it matter whether HP indices are reliable & meaningful?

- macro policy
- estimating the impact of new supply
- landlords and investors
- lenders
- estimation of wealth inequality...
 - Emerging policy debate about long-term impacts of divergent house prices

“Misguided British Preoccupation with Housing”?

- month on month and place by place reporting of house prices disguises an increasingly inequitable housing market.
- **Danny Dorling:**
 - “We have been labouring under the misapprehension that the housing boom has been providing an easier way up the social ladder. However, our research reveals that children born into the poorest households in 2004 are now far less able than previous generations to escape poverty. In other words housing is taking us back towards the deep social divisions of Victorian society - a moment in history than no-one wants to see repeated.”
- Whatever your political perspective on this, house price measurement is set to be crucial to the debate.

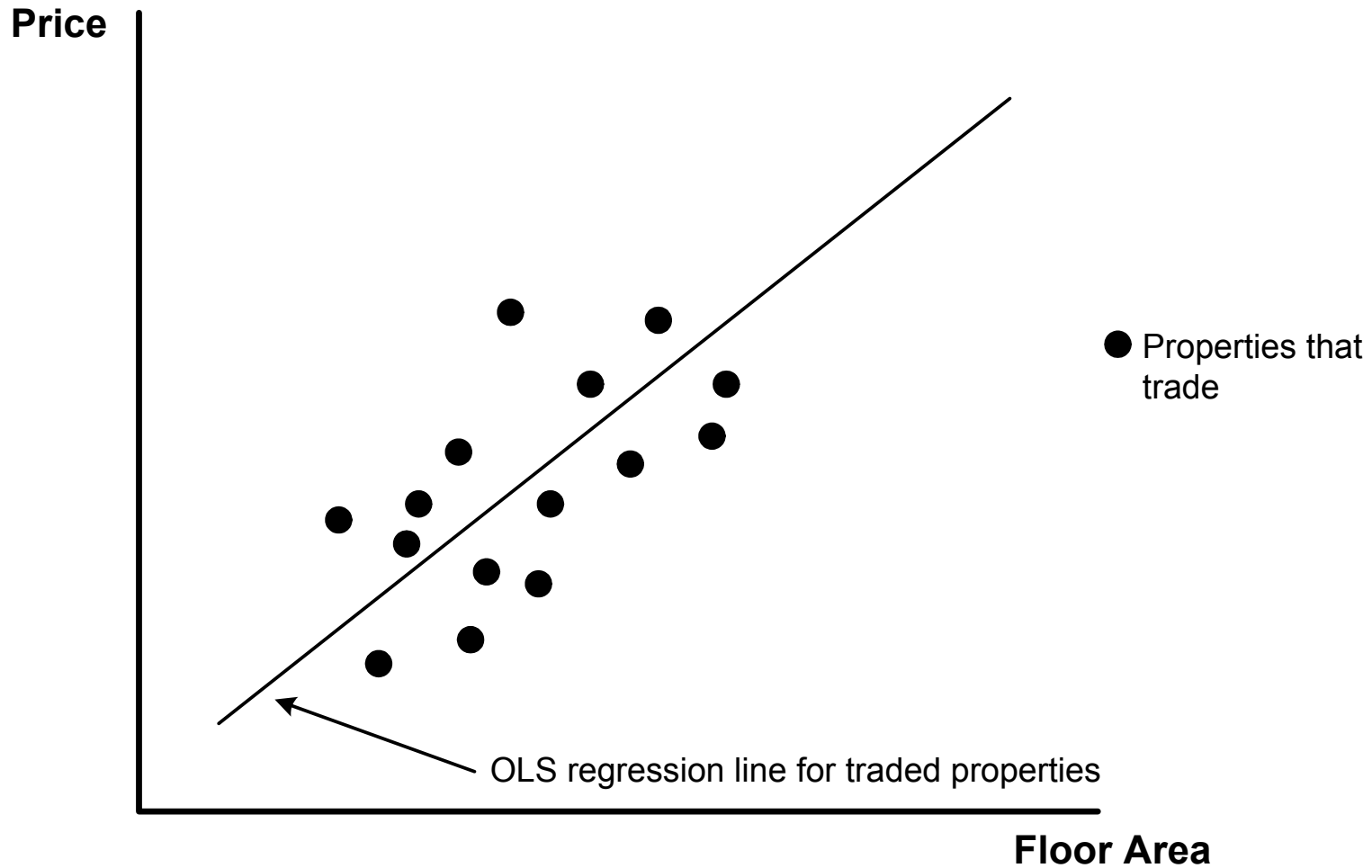
(ii) Existing Measures – in order of robustness:

- RICS
- Hometrack
- Rightmove
- Nationwide
- Halifax
- Land Registry
- ODPM/SML
- FT
 - uses Land Registry data as the benchmark, but what about properties that have not recently sold?

(iii) Impact of Untraded Properties on Hedonics:

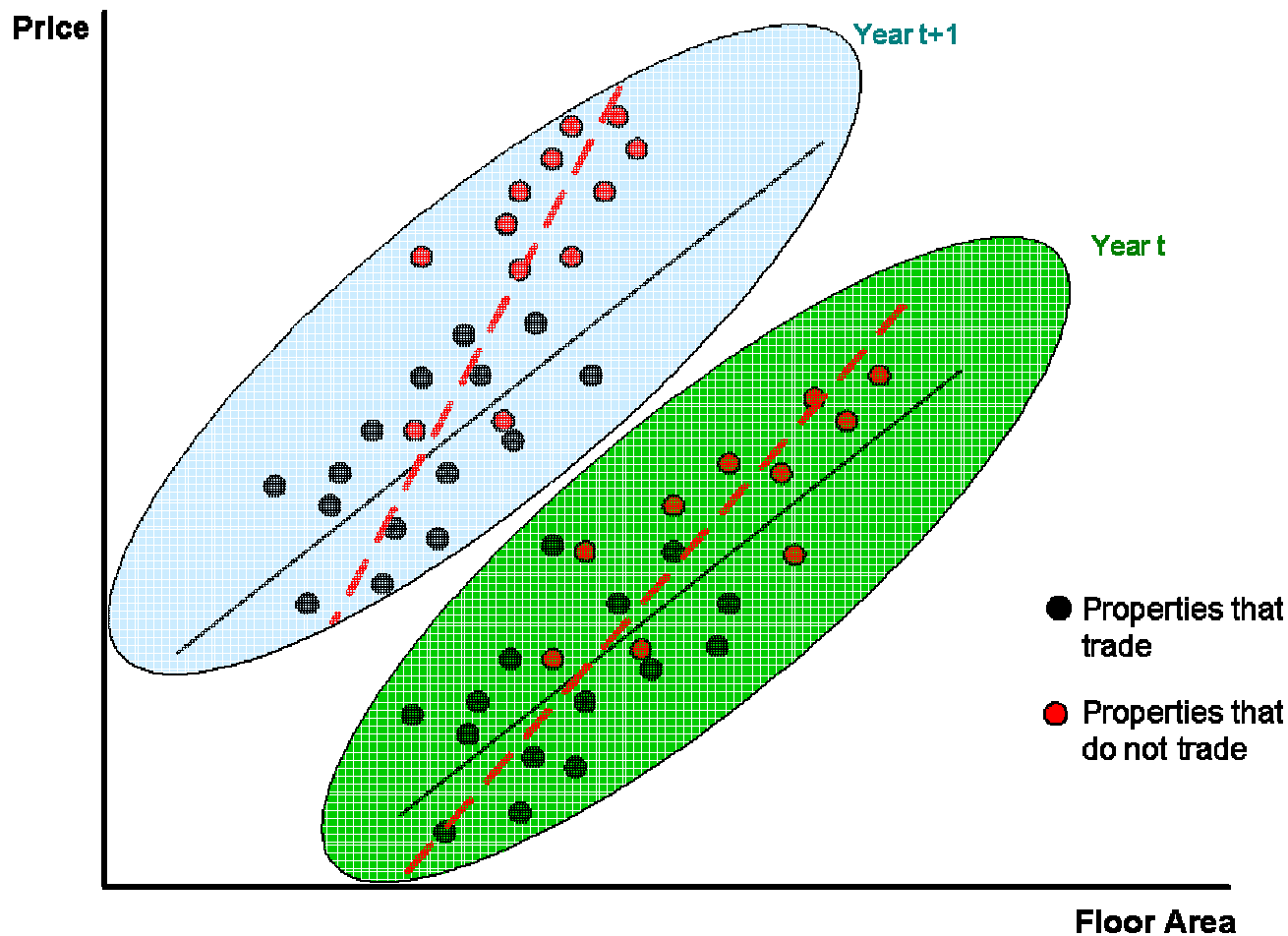
- If properties that do not sell, are on average similar to those that do,
 - then hedonic estimation will be unbiased
- If, however, properties that do not sell are different,
 - then hedonic estimation may be biased
 - Particularly if marginal price of attributes is different for untraded properties
 - E.g. high quality properties in desirable surroundings
 - And particularly if *price appreciation rates are different for traded and untraded properties.*

Regression Line: Traded properties only



Suppose Untraded Properties have different rates of inflation?

Price change intercept dummy not pick this up \Rightarrow underestimate HP inflation



(iii) Methods for Correcting Bias

(a) Gatzlaff, Haurin, Hwang, Quigley (GHHQ)

- Heckman: Probit selection equation \Rightarrow predicted hazard of non-selection.
- Requires info on entire housing stock:
 - Whether each dwelling has sold or not sold in each period
 - Dwelling attributes of both traded & untraded properties

\Rightarrow **not feasible to apply technique in UK**

(b) Fractional Logit Regression

(e.g. Hendershott and Pryce, 2006)

- Use FLR to create an instrument for probability of non-selection
- Requires only info on traded properties & size of stock:
 - Total number of sales in each postcode sector in each period
 - Total number of dwellings in each postcode sector (PAF)
=> % properties that sell in each postcode sector in each period
 - Dwelling attributes of traded properties only
 - Neighbourhood Information
- FLR yields the predicted probability of non-selection in each postcode sector for each year which can be entered on the RHS of the hedonic regression to reduce sample selection bias.

(iv) Structural Model & Estimation Strategy

$$p = a_0 + a_1 \textit{detached} + a_2 \textit{semi} + a_3 \textit{terraced} + a_4 \textit{pnonselect} \quad [1]$$

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

where:

- p = $\ln(\text{price})$,
- $\textit{pnonselect}$ = probability of non-selection (i.e. not trading),
- \mathbf{B} = barriers to sale, particularly public ownership,
- \mathbf{A} = attributes of dwellings,
- \mathbf{N} = neighbourhood quality (e.g. school performance, density, and crime),
- \mathbf{E} = employment factors,
- \mathbf{D} = life-cycle factors, such as age of household, and population change.

Estimation Strategy:

- Step 1: estimate FLR *pselect* regression
 - Expected signs?...
 - *pnonselect* = 1- predicted(*pselect*)
- Step 2: Include *pnonselect* on RHS of hedonic
 - regressions run on each month to create index I_t :

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

Table 1 Turnover Rate Scenarios:

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

(v) Data Description

Variables at Dwelling level:	n	Mean	Std.Dev.
price	1,418,153	£ 137,160	£ 141,427
detached	1,418,153	26%	0.437
flat	1,418,153	18%	0.382
semi	1,418,153	27%	0.447
terraced	1,418,153	28%	0.450
year_1996	1,418,153	10%	0.296
year_1997	1,418,153	11%	0.317
year_1998	1,418,153	12%	0.324
year_1999	1,418,153	14%	0.345
year_2000	1,418,153	13%	0.331
year_2001	1,418,153	13%	0.341
year_2002	1,418,153	15%	0.355
year_2003	1,418,153	12%	0.330

Variables at postcode sector level:

Social rented	1,241	12%	0.090
Economically active	1,241	65%	0.068
Average Education score	1,241	55.3	5.272
Violent Crime	1,241	0.9%	0.004
Burgulary	1,241	0.5%	0.002
Distance between dwellings	1,241	20.6	18.927
Dwellings pre 1920	1,241	24%	0.149
Semi detached	1,241	25%	0.102
Population change	1,241	6%	0.039
Population over 65	1,241	17%	0.060
Proportion of stock that trades in a given year	1,241	3%	0.005

(vi) Results: FLR Selection Regression

Table 4 Estimation of the Selection Equation: FLR

	1996	1997	1998	1999	2000	2001	2002	2003
Social rented	-1.518 (-13.576)	-1.324 (-14.578)	-1.269 (-10.970)	-1.274 (-11.924)	-1.428 (-9.091)	-1.289 (-15.963)	-1.336 (-10.221)	-1.216 (-8.663)
Economically active	0.336 (1.666)	0.143 (1.022)	0.345 (2.327)	0.321 (1.475)	-0.378 (-0.920)	0.212 (1.551)	-0.226 (-0.661)	-0.135 (-0.356)
Education score	0.006 (3.439)	0.002 (1.332)	0.002 (1.737)	0.005 (2.763)	0.006 (3.353)	0.001 (1.036)	0.004 (2.894)	0.002 (1.275)
Violent Crime	0.114 (0.051)	2.392 (1.164)	0.368 (0.203)	-2.891 (-1.437)	2.572 (0.984)	2.091 (1.020)	1.287 (0.593)	-0.116 (-0.050)
Burglary	2.840 (1.043)	-0.785 (-0.330)	-1.399 (-0.558)	0.577 (0.237)	1.141 (0.344)	-1.277 (-0.543)	-5.981 (-2.234)	-4.737 (-1.595)
Dist. between dwells	-0.005 (-9.514)	-0.005 (-8.444)	-0.003 (-6.561)	-0.003 (-5.165)	-0.004 (-6.521)	-0.001 (-1.403)	-0.001 (-1.585)	-0.002 (-5.293)

	1996	1997	1998	1999	2000	2001	2002	2003
Continued...								
Dwellings pre 1920	-0.048 (-0.815)	-0.018 (-0.315)	0.113 (2.149)	0.190 (3.288)	0.225 (3.289)	-0.041 (-0.607)	0.061 (1.003)	0.080 (1.361)
Semi-detached	0.035 (0.556)	-0.016 (-0.235)	-0.143 (-2.040)	-0.105 (-1.452)	-0.149 (-1.876)	-0.169 (-2.948)	-0.188 (-3.122)	-0.240 (-3.583)
Population change	0.346 (2.454)	0.309 (2.098)	0.254 (1.787)	0.761 (5.048)	1.253 (6.085)	0.719 (4.438)	0.788 (4.885)	0.821 (5.684)
Population over 65	0.883 (4.675)	0.637 (4.999)	0.369 (2.711)	0.921 (5.099)	0.191 (0.581)	0.204 (1.692)	-0.003 (-0.009)	-0.070 (-0.222)
Constant	-4.244 (-22.490)	-3.808 (-27.414)	-3.879 (-27.079)	-4.053 (-20.876)	-3.661 (-10.521)	-3.680 (-28.568)	-3.421 (-11.725)	-3.407 (-10.373)
n	1,198	1,198	1,205	1,241	1,263	1,267	1,280	1,280
ll	-100.4	-106.7	-111.4	-120.6	-117.0	-122.5	-130.0	-122.3

Dependent variable = proportion of the total housing stock that trades in a given year.

T-ratios, presented in parentheses, are based on Papke and Wooldridge (1996) robust standard errors.

(vi) Results: Hedonic Regression

- Is the selection term significant?
 - As a simple test we run the regression on all years with *pnonselect* on the RHS (& also attributes & intercept year dummies).
 - Then, to allow the coefficient on *pnonselect* to vary over time, we also include it in hedonic regressions run separately on each month.

Table 5 Hedonic Estimates on all years combined:

	Without Correction Term	With Correction Term
<i>detached</i>	0.989 (747.217)	0.981 (737.697)
<i>semi</i>	0.448 (374.005)	0.439 (361.958)
<i>terraced</i>	0.212 (176.953)	0.206 (170.707)
<i>pnonselect</i>	-	11.040 (45.456)
<i>_cons</i>	10.678 (6712.393)	-0.061 (-0.257)
+ year dummies		
N	1,418,153	1,418,153
r2_a	0.510	0.511

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

Figure 1: Results from Monthly Hedonic Regressions

Coefficient on Pr(non-selection) in Hedonic Ln(Price) Equation
(With 95% Confidence Interval)

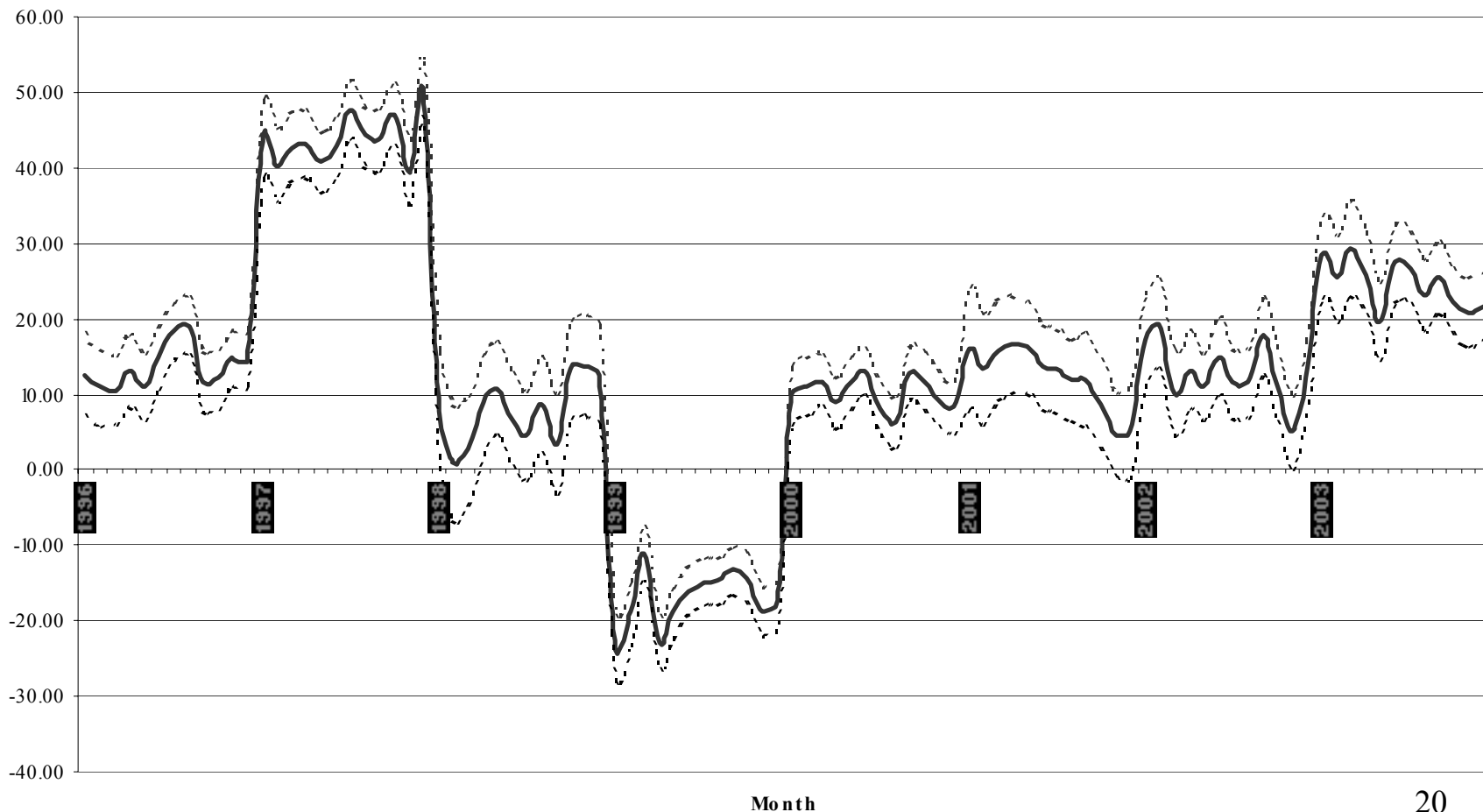
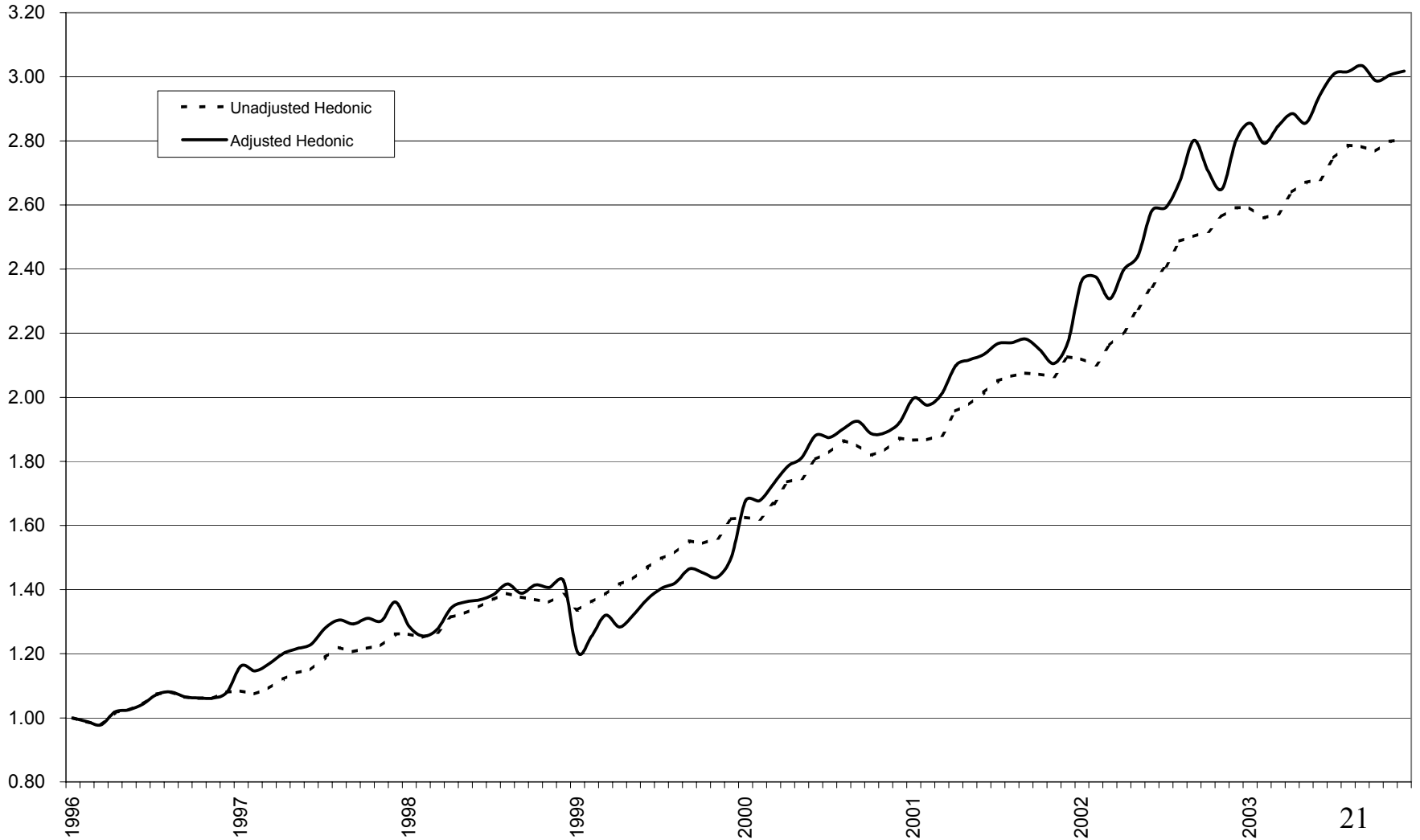


Figure 2:

Adjusted and Unadjusted Monthly Nominal Constant Quality House Price Indices
(S.East England 1996 to 2003)



Summary:

- Aim:
 - To establish a method for correcting transactions bias in house price indices that could be applied to countries and regions where info on individual dwellings is not available for the whole stock.
- Method:
 - FLR used to derive an instrument for the prob(non-selection)
- Results:
 - Estimated probability of non-selection was statistically significant in hedonic regression (both all years & monthly).
 - Effect tended to vary over time, even changing sign in 1999.
 - Overall, *unadjusted index* tended to *underestimate* the true rate of price appreciation of the stock of private housing.