International Real Estate Society Conference ’99

Co-sponsors: Pacific Rim Real Estate Society (PRRES)
Asian Real Estate Society (AsRES)

Kuala Lumpur, 26-30 January 1999

Price Segmentation, Size Effects and Information Diffusion in Housing Markets

Gregory J Costello
Department of Property Studies, Curtin University of Technology,
Perth, Western Australia
Phone: +61 8 9266 7778, Fax: +61 8 9266 3026, Email: costellog@cbs.curtin.edu.au

Abstract: This paper examines issues of information diffusion in housing markets. A large transaction data set for the city of Perth, Western Australia is used to construct market aggregate and price quartile repeat-sales and hedonic indexes for the period 1988-96. Significant size effects are observed. Cheaper properties exhibit higher individual rates of real price change in the short term but the lowest rates in the longer term. These results are consistent for tests with both index models and individual property price changes. These results are shown to cause bias in transaction based indexes. Further tests confirm that information from past price changes diffuses temporally within price segments at varying rates. Abnormal price changes are used to test for information diffusion between price segments. Weaker evidence is presented that lagged information of past price changes diffuses temporally between price segments. The diffusion of information from lower to higher price segments appears to be positive with diffusion from higher to lower segments appearing to be negative.

1.0 Introduction

Does information from past price changes diffuse on a consistent temporal scale for different price segments in housing markets? Do cheaper houses increase in price before more expensive properties during market upturns or is there a dominant mean price market segment from which information on price changes diffuses "outwards". In this paper these questions are applied to transaction data for a specific housing market segment in the city of Perth, Western Australia for the period 1988 – 1996.

Surprisingly, there appears to be little empirical work examining information diffusion and the influences of market segmentation in housing markets according to price criteria. The "size effect" where returns on the shares of small companies exceed the returns on the shares of larger companies is frequently discussed in studies of market efficiency in securities markets. Intuition would suggest that there could be similar effects in housing markets. Cheaper house price segments are characterised by greater affordability and therefore increased participation. Does this lead to more rapid diffusion of information? An alternative scenario is that increased affordability causes larger variations in demand and supply for cheaper housing than expensive housing. This in turn creates more volatility and speculative price "bubbles" in cheaper market segments. These are important questions because if inconsistencies between price segments can be observed and predicted, trading strategies might be used to exploit the information contained within past price changes for different price segments.

Real estate markets diverge significantly from the theoretical construct of perfect markets. Lagged price information should not affect prices in perfect markets. In real estate markets lagged price information is very important and will affect prices even in the absence of irrationality. The study of house price changes provides a productive source for learning about the informational role of prices. The objectives of this paper are to examine characteristics of house price dynamics that are consistent with rational learning in housing markets. How do buyers and sellers in housing markets obtain and use information? Are there differences in the role of information for past price changes between market segments? These questions are addressed with respect to temporal information diffusion and diffusion between price segments. The influence of lagged information is tested within the aggregate market and price quartile market segments.

This paper makes three specific contributions to the existing literature. First, it demonstrates that significant size effects exist in this housing market and therefore price segmentation is an important house price dynamic. Second, size effects can cause significant bias in house price indexes for aggregate housing markets. Third information on past price changes diffuses temporally between price segments and at varying rates in own price segments. The next section of this paper reviews the body of theory and related literature pertaining to the topic. Several hypotheses are suggested concerning relevant information sets applied to definitions of real estate market efficiency. The middle sections of the paper review the data and demonstrate the existence of size effects. The final sections establish the empirical method used to analyse information diffusion processes, review the empirical results and provide some conclusions and suggestions for related research.

2.0 Theory- Related Literature

This paper examines information diffusion in housing markets and hence the *informational* efficiency of those markets. Fama (1970) developed the efficient market hypothesis (EMH) and defined an efficient capital market as one in which the price of an asset fully reflects all available information. With this rigid definition there is an emphasis on the relationship between market prices and information. With the assumption of perfect markets this emphasis requires that the reaction of prices to new information is both "instantaneous" and unbiased. In an efficient market, competition between informed buyers and sellers should fully reflect the available information set, therefore there can be no biased price reactions. The EMH framework is closely aligned to the theory of rational expectations. Under perfect competition it is assumed that there are sufficient buyers and sellers in a market so that each participant cannot individually determine market prices. Prices reflect consensus opinions about the market value of assets although participants may not have homogenous expectations about the future benefits of the assets. In reality, asset markets do not completely satisfy all of the assumptions of perfect competition. Most asset markets have some imperfections that inhibit the flow of capital and information and reduce market efficiency below the perfect competition optimum. In this sense market efficiency represents a relative rather than an absolute concept.

Homeowner-investors lack perfect information in housing markets. Typical buying or selling behaviour requires participants to rationally expend resources to obtain information. These costs in the form of time and money have been discussed in the context of sunk costs as rigorously defined in studies of market structure (Clapp Dolde and Tirtiroglu: 1995:242). These costs lead to the development of thin markets where bilateral negotiations between buyers and sellers lead to prices reflecting information obtained by only a few of the most active participants within a housing market at any one time.

In discussing housing search behaviour, Wheaton (1990) suggests that over time households experience stochastic demographic changes which mismatch some households with their current house. This process generates a desired sale and new purchase for a mismatched household. These households undertake a costly search among houses for sale, deciding whether to negotiate for a given unit or continue the search process. These market participants learn about market fundamentals from the observation of housing units for sale, information from brokers and by negotiation with sellers. Sellers will learn in the same way, and by observing the behavior of potential buyers. Transaction prices of comparable units provide important information used by both sides of the market to set parameters for negotiation. This information may be available in a variety of forms. Typically, the most recent information on sale prices will be conveyed by word-of-mouth between the most active market participants, i.e., brokers, buyers and sellers. For this reason it is expected that active market participant's pricing decisions will be strongly influenced by lagged price changes. Is this trading rational? In theory, assuming perfect markets, market efficiency implies that feedback trading on lagged price information is irrational. If markets are efficient then feedback traders will provide incentives for rational investors to enter those markets and exploit the activities of irrational traders. One objective of this paper is to examine characteristics of housing price dynamics that are consistent with rational learning and not simply irrational feedback trading. How do buyers and sellers obtain and use information? This question is addressed with respect to temporal information diffusion and diffusion between price segments. The influence of lagged price changes is examined for effects within the aggregate market and market segments.

The question of the appropriate lag period that is consistent with real estate market efficiency is important and is also an empirical issue examined in this paper. Unlike security markets where stock prices are available daily, real estate price information is subject to considerable delay before becoming available as public information. Typically, real estate transactions are negotiated with a settlement period between date of transaction and final settlement. In Western Australia full public information is only available after settlement, typically involving a lag period of 60-90 days from the negotiation date of the transaction.

In this environment of information diffusion it is possible to hypothesise that there are in fact several relevant information sets and time lags appropriate to tests of the EMH in housing markets. One information set is what may be termed the *local* information set. The term local is used because it is likely that this information set is associated with spatial characteristics that may define a local market.[1] If brokers, buyers and sellers learn about prices from interacting with other active market participants then it is likely that market efficiency in housing markets should be consistent with shorter lag periods. This is because the local information set involves a word-of-mouth process of information diffusion that is very important to active buyers and sellers. This word-of-mouth process is independent of when transaction information becomes available as a public record. It is assumed that in an efficient housing market active market participants will transmit and receive this information and negotiations will proceed on the basis of this information. In this scenario, the most recent information will be the most scarce and also the most valuable. It is not necessary for all market participants to have access to this information. It is only valuable to those participants making price decisions likely to be influenced by this new information (i.e., the local market). It is assumed that this local information set is predominantly information of past prices and is therefore consistent with Fama's (1970) definition of weak-form market efficiency.

Another information set is the *full* information set consistent with a longer lag period and Fama's (1970) definition of semi-strong form market efficiency where the information set consists of all available public information. This longer lag period is determined by the period between negotiation and final settlement of a transaction. Subsequent to the transaction, full public details of the transaction (including hedonic characteristics) become available. This is the information set desirable for rational investors looking to exploit any systematic inefficiency in a housing market. Consistent with this definition of the full information set, the data used in this study covers the full city of Perth, Western Australia for the period 1988 – 1996 and includes important hedonic characteristics. The full information set is also influential in terms of institutional factors affecting real estate markets. Government authorities, lenders and the valuation profession all have access to the full information set. It is from this information set that important policy decisions likely to influence a housing market will originate.

Gatzlaff and Tirtiroglu (1995) summarise the results for a number of studies either directly or indirectly examining market efficiency in housing markets. Typically these studies have used aggregate market indexes with little analysis of the influences of market segmentation. One of the most influential studies completed by Case and Shiller: 1989) estimated indexes for single family housing in the cities of Atlanta, Chicago, Dallas and San Francisco for the period 1970-1988. They report significant serial correlations in price changes and interpret the results as an indication of irrational feedback trading. Meese and Wallace (1994) construct aggregate market indexes for several localities in San Francisco and report violation of market efficiency in the short run, but not the long run. They suggest that transaction costs could account for this pattern. More recently Clapp Dolde and Tirtiroglu (1995) (CDT) apply spatial market segmentation criteria to provide evidence of lagged effects of price changes both within own towns and non-neighbouring towns in Connecticut and the San Francisco area. They interpret their results as being consistent with rational learning rather than irrational feedback trading. Dolde and Tirtiroglu (1997) apply generalised autoregressive conditional heteroskedasticity (GARCH) and moving average (MA) estimators to the same CDT data and report strong evidence of negative serial correlations at short lags. This is contrary to previous results for the same data. The authors attribute this in part to the use of quarterly index differences as opposed to previous annual differences where moving average effects were omitted.

In many of these studies the requirements for index accuracy meant that large data sets were required and hence aggregate city wide indexes are used to report general results. These tests are analogous to tests of market efficiency on a composite stock market index such as the FTSE 100 or

[1] The term "local" may not be appropriate. The information set could also apply on a broad spatial basis to particular property market segments e.g., city-wide industrial land between one and two hectares. In housing markets however it is likely that the local spatial dimension is the most important cause of market segmentation in determining this information set.

Australian All Ordinaries. These studies are useful in examining aggregate market efficiency but it is also possible that the numerous variations in results are due to specification errors. Because of the heterogeneous characteristics of housing it is generally accepted that market segmentation is important. Acceptance of the hedonic pricing model implicitly assumes various segmentation criteria. This study examines influences associated with price segmentation. Hedonic analysis confirms that price segmentation is correlated with other segmentation criteria, most notably geographic regions and building ages. If homeowner-investors are constrained by price, then the relevant information set for individual homeowner-investors at the time of purchase is also likely to be constrained by price. In simple terms, a potential purchaser of a housing unit priced in a cheaper price segment of the supply of available housing units is not likely to consider price information for housing units priced in the highest price segments. It is likely that information asymmetry will exist between market segments. Because of these effects serial correlation tests using aggregate market indexes may provide spurious or ambiguous results due to the noise within the index from irrelevant price segments (information sets). It may be that serial correlation tests using more accurately measured price segment indexes are more appropriate in analysing patterns of information diffusion in housing markets.

Numerous studies of securities markets examine the influence of size effects (Banz: 1981), (Beedles, Dodd, Officer: 1988). These studies analyse the effect on returns of a company's "size" as measured by the total market value of a company's shares. The *size effect* is the observation that returns on the shares of small companies exceed the returns on the shares of larger companies, both before and after adjusting for beta-risk. One reason suggested for this is the influence of illiquidity. Because shares in small companies trade less frequently than shares in larger companies, investors in smaller companies require a higher expected return to compensate for the illiquidity of their investment. Surprisingly, there appears to be little empirical literature examining this issue in housing markets. There are good reasons to suspect that size effects may exist in housing markets. Liquidity effects are important and are likely to be contrary to the observations in securities markets. Cheaper homes are more affordable and the starter home hypothesis suggests that these homes trade more frequently. Increased demand for cheaper homes may lead to lower rates of price appreciation for these homes. Holding period effects are also likely to be important. Participants in housing markets maybe homeowners, consuming housing services or absentee landlord investors. Short holding periods are more likely to be associated with speculative investors. Typically consumers of housing services will occupy their homes for longer holding periods. The increased liquidity associated with cheaper properties suggests that mortgage finance may be more readily available for short-term speculation in these market segments. Given this scenario it is more likely that irrational feedback trading is prevalent in cheaper market segments.

3.0 Data

This study uses housing transaction data for the capital city of Perth, Western Australia. Perth is a city of approximately 1.3 million people. In Australia, important local services (education, health, police etc) are financed at the state government level thereby minimising the influence of Tiebout factors[2] associated with local government financing. Although spatial characteristics are still influential in house price dynamics, these influences tend to be correlated with structural or age characteristics of buildings and are not as influential as apparent in some US studies. Transaction details were taken from the Western Australian Valuer General's Office for the period 1988 – 96. The sample selected is strata title sales. This is a segment of the housing market covering a variety of housing styles and densities, with accurate hedonic variables for building age and area. Transaction details for the three most recent sales within the sample period were available for each single property. The full sample comprises 68,799 single transactions with approximately 72% of these being single sales, 17% once only repeat-sales and 11% being multiple repeat-sales.

[2] The "Tiebout hypothesis" refers to the observation by Charles Tiebout in 1956 that home-owners base their location preferences according to the standards of local government services.

From this transaction data two aggregate[3] market price indexes were constructed. Repeat-sales data were used to estimate a *quarterly* Weighted Repeat Sales (WRS) index according to methodology used by Case & Shiller (1989) and a *monthly* Hedonic index model was used for the full sample of transactions. More complete detail of index methodology is contained in the appendix. The aggregate data was then segmented according to price. The selling prices during the relevant sampling period (quarter or month) were ranked into quartile groupings and new price indexes were estimated for the price segment sub-samples. For the WRS indexes the initial selling price of a repeat-sale was the criterion for selection of quartiles.

Figure 1 provides descriptive statistics for the repeat-sales and full hedonic samples, together with the price quartile sub-samples. These results provide support for the "starter home hypothesis". Cheaper properties tend to sell more frequently and hence repeat-sales indexes tend to be prone to sample selection bias towards cheaper homes. The reported mean sale price is considerably lower in the WRS index (part A) than the hedonic index (part B). One advantage of using the full hedonic sample together with the repeat-sales sub-sample is that alternative price segmentation levels are apparent. Various theories of house price dynamics suggest that demographic characteristics such as average annual income will influence house price movements. Because of this, the use of price quartiles in specifying market segments is arbitrary. It is likely that the most influential price segments are not identified by simple quartiles. The alternative specification achieved by using two index samples broadens this simple approach to selection of price segments.

Figure 2 provides results for mean quarterly changes in the real log price indexes. The variable analysed; *z* is the deflated first (quarterly) difference for the real log price index. For *WRS* indexes, $z = W_{it} - W_{it-1}$ where; $W_{it} = WRS\ \hat{c}_{it} - LN\left(CPI_t\right)$. A similar methodology applies to H_{it}, for the hedonic index.[4] From the WRS index the growth in real house prices for the aggregate market during the sample period was approximately one third of 1 per cent per quarter or 1.2% per annum, whereas for the real hedonic index, growth in real house prices was approximately 1.6% per annum. The standard deviation in quarterly real price changes for both indexes was slightly higher than 3 percent per quarter. During the same period the standard deviation of quarterly real price changes in the Australian All Ordinaries stock market index was of the order of 6.5 per cent.

Seasonal influences in indexes must also be considered. In some US studies (Case & Shiller: 1989) (Clapp Dolde and Tirtiroglu: 1995), seasonal influences have been observed in quarterly real house price changes. Figure 2 provides results for tests of seasonality both on full sample indexes and individual price quartile sub-samples. It appears that there is a general trend of first quarter changes being highest and third quarter changes lowest. These differences appear very small and no statistically significant seasonal influences were observed for either the full samples or price quartile sub-samples. Individual quarterly differences were subjected to a two-sample *t* test, the null hypothesis being that the mean difference for any quarter was not different from the mean difference for all quarters. The null hypothesis could not be rejected at a level of significance of 10% for any quarterly period for either the full samples or price quartile sub-samples. A one way analysis of variance was also used, including all quarterly periods, the null hypothesis being that all quarters had the same mean. The null hypothesis could not be rejected at a level of significance of 10%.

---

[3] The term *aggregate market* refers to either the full sample for hedonic or repeat-sales including all price segment sub-samples. The term is used synonymously with *full sample*.

[4] The results in Figure 2 are for an hedonic index using *quarterly* sampling periods. This is primarily for direct comparison with WRS results and to facilitate tests of seasonal influences. All other results in this paper for the hedonic index use monthly sample periods.

Housing price changes were also tested for relationships with financial markets. A beta coefficient was estimated by regressing the quarterly change in the logarithmic aggregate WRS index on the corresponding change in the logarithmic Australian All Ordinaries composite stock market index. The resultant beta coefficient was virtually zero. These results confirm that housing markets do not appear to be significantly influenced by movements in securities markets and are consistent with results for US data reported by Gau (1987), Case & Shiller (1989) and (Clapp Dolde & Tirtiroglu: 1995).

4.0 Size Effects

The results in Figures 1 & 2 confirm varying rates of real house price changes according to price quartile sub-samples from two index models. An advantage of using repeat-sales data is that actual trading activity for individual properties is contained within the data. The variable R^* shown in Figures 1, 3 and 4 is the effective real annual rate of price change for *individual* repeat-sales. This variable is constructed by using the initial and subsequent selling prices together with the holding period of individual transactions. Formulae used in constructing this variable are contained in the appendix. In Figure 1, part A, the median R^* is reported for the full WRS sample and all price quartiles. The median effective annual real rate of price change for the full sample is approximately 1.5% and it appears that the most expensive homes in the fourth quartile are achieving the highest rates of appreciation, approximately 2.3% per annum. This median figure masks positive skewness in the distribution of rates of price change due to the influence of short-term trading. This is evident in the results reported in Figure 3. When the WRS sample is further segmented into short holding periods (one year or less) it can be seen that the full sample and all quartiles have short term mean rates of price change considerably higher than for the samples of all holding periods.

The *Mean* R^* results reported in Figure 3 part A provide strong evidence of the influence of size effects. Statistical testing indicates that quartile group means are significantly different. One sample *t* tests for the sub-samples of all holding periods indicate that means for all quartile groups except the fourth quartile are significantly different than the mean for the aggregate market sample. The null hypothesis for these tests is rejected at a level of significance of one per cent or greater. When similar tests are applied to short holding periods, all quartile groups are significantly different except the second quartile. For long holding periods all quartile sub-samples have mean rates of price change significantly different from the aggregate market sample. One way analysis of variance (ANOVA) tests reject the null hypothesis that all quartile groups have the same mean rates of price change. This result is consistent for the full sample of holding periods together with the reduced short and long holding period samples. The non-parametric Kruskal-Wallis test (results not reported) confirms these results. Multiple comparison tests (results not reported) indicate that for all holding periods, mean rates of price change for the first and fourth quartiles are significantly different than all other quartile groups. Means for the second and third quartiles are significantly different from the first and fourth quartiles but they are not significantly different from each other. For short holding periods, the mean rate of price change for the first quartile is significantly different from all other quartiles. All other quartiles are significantly different from the first quartile but not from each other. For long holding periods the mean rate of price change for each quartile sub-sample is significantly different from every other quartile.

An important conclusion from these results is that short-run speculative traders operate in all price quartiles and are able to exploit potential high positive rates of price change during certain market periods. Significantly, it appears that these short-run speculative traders achieve the highest rates of return in the first (cheapest) quartile. Furthermore, the lowest long-run rates of price appreciation, approximately one third of one per cent per annum, also apply to the first quartile. The highest long-run rates of price appreciation, approximately 2.3% per annum are achieved by the fourth quartile. These apparent return regularities may constitute anomalies in the efficient markets theory as applied to housing markets.

Rates of price change for individual properties cannot be ascertained for all properties within the hedonic data as a number of observations are once-only sales. However, size effects are also verified from the monthly hedonic sample as shown in Figure 3 part B. Real log index changes are used to test for statistical differences between price quartile sub-samples. Quarterly, half year and annual differences are tested. The variable z is the relevant difference for the real log price index for monthly sampling periods. For quarterly changes: z = $H_{it} - H_{it-3}$, where: $H_{it} =$ Hedonic $\hat{c}_{it} - LN(CPI_t)$. For half year and annual changes, z = $H_{it} - H_{it-6}$, and z = $H_{it} - H_{it-12}$ respectively.

The results reported in Figure 3 part B confirm that the rate of price change is lower for the cheapest price segments. This is most apparent in annual price changes. The first quartile sub-sample indicates a slight negative real return with all higher price quartile sub-samples showing results generally consistent with those reported in part A. The ANOVA results confirm that it is not until annual differences are taken that the null hypothesis (mean z is the same for all individual price quartile samples) can be rejected at a level of statistical significance.

5.0 Index Accuracy and Measurement Error

The data for this study was selected with the intention of maximising the accuracy of measurement of index period differences for the purpose of conducting serial correlation tests using lagged index differences. Two index methods were used; a quarterly WRS index and a monthly hedonic index. When using the WRS index for serial correlation tests spurious correlation may arise due to an errors-in-variables problem identified by Case & Shiller (1989). In summary, noise in the estimated index contaminates both dependent and independent variables used for serial correlation tests. Case & Shiller developed a simple expedient for this problem that has been used in this study. The methodology is to split the sample of individual house sales into two individual random samples and then estimate new WRS indexes, denoted indexes A & B. Serial correlation tests can completed by regressing index differences from index A on lagged index differences from index B or vice versa. Figure 1, Part A provides index accuracy ratios for the full sample, quartile samples and respective random sub-samples A & B for the WRS indexes. Index accuracy ratios are calculated from the weighted least squares (WLS) regression procedure used to construct the WRS index. Figures given are ratios of the standard deviation of a variable to the average WLS standard error for that variable. Index accuracy ratios are given for the WRS quarterly index in levels and for the first order (quarterly) difference. Higher ratios indicate more accurately measured index characteristics. Case & Shiller (1989) describe similar figures for the log index in levels as "accurate", and ratios in the vicinity of 2.7 – 4.0 for *annual* differences as "fairly accurate". Ratios in the vicinity of 1.0 – 2.0 for quarterly differences were discussed: "we thus cannot accurately describe the quarterly changes in the log prices, though the index will give a rough indication." Case & Shiller (1989: 127). Using this criterion it is evident that quarterly levels for the full WRS sample and respective price quartile sub-samples are well measured. For data used in this study, quarterly first differences for the full WRS sample are measured with ratios higher than for some annual differences in Case & Shiller's (1989) US study. For individual WRS quartiles, the levels of all indexes are well measured. For first differences, the first and second price quartiles are quite accurately measured, as are the respective random sub-samples. The third and fourth price quartiles are not as accurately measured, with the third quartile being marginal in terms of accuracy for the random sub-samples A & B. The fourth quartile is quite well measured in terms of the index level but with low ratios for the first (quarterly) difference.

Another useful diagnostic for assessing index accuracy is reported with tests for serial correlation using the WRS index shown in Figure 6. Individual WRS indexes are estimated for random samples as discussed above. Individual first differences for real log index A are regressed on the contemporaneous difference in index B. The β_1 coefficient for these regressions should be 1.00 if the indexes are measured perfectly but will vary from one due to the errors-in-variables problem. Figure 6 shows the estimated β_1 coefficient for quarterly index first differences for the full WRS sample is .953 with R^2 = .912. This confirms that quarterly first differences for the full WRS sample are well measured. Other quartiles appear to be quite well measured, with the fourth quartile being the least accurate, $\beta 1$ = 0.596, R^2 = 0.484.

An advantage of using hedonic indexes is the greater number of observations contained within the sample. The WRS indexes decline in accuracy as the sample size declines. This is most obvious when individual quartiles are split into random sub-samples A & B. Clapp, Dolde & Tirtiroglu (1995) demonstrate that hedonic indexes are not prone to the same problems of spurious correlation present with repeat-sales indexes. Index accuracy ratios for the monthly hedonic index are shown in Figure 1, part B. These ratios are estimated in a similar manner to ratios for the quarterly WRS index although an ordinary least squares (OLS) procedure is used in estimating the index and OLS standard errors are therefore used in the ratios. Ratios for index levels, first (monthly) differences and third (quarterly) differences are shown. These results are interesting in that the second and third quartiles appear to be more accurately measured than the full sample. Most of the noise in the index appears to emanate from the first and fourth quartiles and this appears to corrupt the accuracy of the full sample index. Figure 1 part B shows that for the second and third quartiles the hedonic index level and third differences are very accurately measured and that even the first differences are more accurately measured than third differences for the full hedonic sample.

Figure 4 provides additional information on the accuracy of the WRS index and the influence of holding periods. Figure 4 provides the *Mean R** results from Figure 3 together with regression results for, $R^* = \alpha + \beta WRS^* + \varepsilon$. The dependent variable R^* is the effective real annual rate of price change for individual repeat-sales. The independent variable WRS^* is the contemporaneous real annual rate of change in the WRS index for the holding period (formulae in appendix). The regressions denoted *Short holds* and *Long holds* are for reduced sub-samples. Short holds are defined as holding periods of one year or less. Long holds are all holding periods greater than one year. As discussed above, the results for *Mean R** confirm that short-term speculative traders operate in all market segments. The regression results show that for the full sample and all quartiles, WRS index accuracy is improved when only long holds are used for index estimation. If the WRS indexes accurately depict individual rates of price appreciation the α coefficients should be close to zero and the β coefficients should be close to 1.0. It can be seen that the full WRS sample estimates are biased by the short-term trading activity. When only long holding periods are used both coefficients are very close to these figures, α = 0.006, β = 1.005. The R^2 also improves significantly although a significant proportion of variation between actual individual property and WRS changes is unexplained. This is evident in the large standard deviations for the *Mean R** returns. The accuracy of WRS measurement for all quartiles improves when only long holds are included. These results are important in demonstrating the influence of a size effect in this housing market. This effect appears closely related to trading activity. It appears that most of the bias in the WRS index is created by the first price quartile. This is also the quartile where there is most volatility in the form of short-run trading. Another important consideration relates to housing market structure where the important distinction in housing markets between homeowners (occupiers - consumers) and absentee landlord (investors) must be recognised. If it is assumed that holding periods for owner-occupiers are predominantly long-term, then the majority of short-term trading is by investors.

6.0 Information Diffusion - Empirical Method

A general difficulty exists in applying models of total returns to housing in empirical tests for information diffusion in housing markets. Total returns on housing include implicit rental dividends in addition to price changes. This data does not exist at a level of accuracy that would warrant inclusion in these tests. In addition, differential tax treatments of owner status (owner-occupier or absentee landlord investor) make the task of measuring after tax effects prone to observation error. It is generally agreed that the volatility of total housing returns is dominated by the price change component. (Dolde & Tirtiroglu: 1997) (Capozza & Seguin: 1994). As a consequence of the difficulties in measuring total returns to housing, the efficient markets hypothesis (EMH) has been empirically tested in numerous real estate studies using the martingale model applied to price changes estimated from large aggregate transaction based indexes. Estimated prices denoted P_t , are the result of a stochastic process, termed a martingale with respect to the sequence of information sets, Φ_t , so that $E_t\left(P_{t+1} \mid \Phi_t\right) = P_t$ where E_t denotes the expectations operator conditional on the information set Φ_t available at time t. This indicates that the best forecast of P_{t+1} is P_t, given the relevant information set Φ_t (note that P_t is assumed to be in Φ_t). The martingale model of efficiency implies that the market is in equilibrium and that investors cannot consistently earn excessive or abnormal returns on investments based on any information set, Φ_t.

The maintained hypotheses that home-owner investors make inferences from lagged price changes from similar or different price segments can be tested by regressing price changes on lagged price changes from similar or different price segments:

$$\Delta P_{it} = \beta_0 + \sum_k \beta_k \Delta P_{it-k} + \varepsilon_{it} \qquad (1)$$

where ΔP_{it} is the observed price change determined from the order of the logarithmic index difference, ΔP_{it-k} are lagged index differences of a similar order, ε_{it} is an identically, independently distributed error term with mean zero and $k = 1,2,\ldots,K$, the maximum lag length. With this model serial correlation is modeled explicitly, therefore if positive diffusion of information occurs, $\beta_k > 0$ and the observed coefficients should be statistically significant.

This is the general model applied to different indexes in this study, however several econometric issues need to be considered. First, when using repeat-sales data, spurious correlation may occur due to an errors-in-variables problem identified by Case & Shiller (1989). Second, when using index differences of a higher order than the index sample periods an "overlapping" of sample periods occurs violating the assumption of independent error terms. Third, when regressing price changes from one price segment on lagged price changes from a different price segment spurious correlation will occur due to general trends in the data. All of these issues are specifically addressed below.

Figure 6 provides results for serial correlation tests for various WRS quarterly indexes. Case & Shiller (1989) used similar tests and discussed the issue of errors-in-variables causing spurious correlation due to noise in the index corrupting both dependent and independent variables. This problem would arise if a model as shown in equation 1 were applied to a single WRS index. Case & Shiller's solution was to split the repeat-sales sample into two individual random samples and then estimate new WRS indexes denoted indexes A & B. Serial correlation tests can be completed by regressing index differences from index A on lagged index differences from index B or vice versa. This procedure is followed for the full repeat-sales sample and the respective price quartile sub-samples. An OLS regression procedure is used to estimate the serial correlation coefficients. The estimating model is of the form:

$$\Delta W_{it} = \beta_0 + \beta_1 \Delta W_{j,t-1} + \beta_2 \Delta W_{j,t-2} + \varepsilon_{it} \qquad (2)$$

where the variable W_{it} is the quarterly WRS log index deflated by the contemporaneous CPI index $W_{it} = WRS\ \hat{c}_{it} - LN\left(CPI_t\right)$. The dependent variable ΔW_{it} is the first (quarterly) difference of the real log index, $\Delta W_{it} = W_{it} - W_{i\,t-1}$. The subscripts *i* and *j* denote the random sub-sample *A* or *B* used as either dependent or independent variable.

Figure 6 provides results for serial correlation tests using a procedure similar to Case & Shiller's (1989) study with several important differences[5]. First, where Case & Shiller report fourth order (annual) differences this study uses the *first* order (quarterly) difference. Recall from the index accuracy statistics in Figure 1 that the aggregate WRS index is more accurately measured than the WRS indexes used by Case & Shiller. Seasonality is not significant in these indexes whereas in other studies seasonality has necessitated a requirement that annual differences be used for serial correlation tests. Second, Case & Shiller used only a single fourth order (annual) lag, whereas this study uses two first order (quarterly) lags. Third, because Case & Shiller used quarterly data with annual differences, the overlap of error terms necessitated the Hansen & Hodrick (1980) method of moments correction be used to correct standard errors of the OLS estimates. For the results in Figure 6 all estimates and standard errors have been calculated with an OLS procedure.

While Case & Shiller's method is useful for aggregate data, the loss of observations associated with specifying random sub-samples means that accuracy declines when examining market segments. This is confirmed by the index accuracy diagnostic results in Figure 1 part A and Figure 6. The hedonic index includes many more observations and is estimated for monthly periods therefore providing more observations for the application of serial correlation tests.

The problem of overlapping sample periods can be explained with the following example. Assuming a quarterly index, let $\hat{c}_{it}$ represent the natural logarithm of the index *i* at quarter *t*. The first order difference for the index, ΔI_{it} is the continuously compounded quarterly rate of price change for house prices. If monthly series are used and third order differences are taken, the available number of observations (index differences) for serial correlation tests can be increased by overlapping monthly periods:

$$\Delta I_{it} = \left(\hat{c}_{it} - \hat{c}_{it-3}\right) = \left(\hat{c}_{it} - \hat{c}_{it-1}\right) + \left(\hat{c}_{it-1} - \hat{c}_{it-2}\right) + \left(\hat{c}_{it-2} - \hat{c}_{it-3}\right) \quad (3)$$

The problem with overlapping data arises because ΔI_{it} will overlap with its two predecessors, sharing two monthly changes with ΔI_{it-1} and one with ΔI_{it-2}. As a result, the ΔI_{it} are not independently distributed and standard errors calculated using ordinary least squares will be incorrect, invalidating tests of statistical significance for serial correlation coefficients. Whereas unbiased estimates of the coefficients may be obtained using OLS, the overlapping error terms necessitate estimation of variances by a method of moments (Hansen and Hodrick: 1980). This method is more completely discussed in the appendix.

Figure 7 provides results for serial correlation tests for various hedonic monthly indexes. The estimating model is of the form:

$$\Delta H_{it} = \beta_0 + \sum_k \beta_k \Delta H_{it-k} + \varepsilon_{it} \quad (4)$$

where the variable H_{it} is the monthly hedonic log index deflated by the contemporaneous CPI index. The dependent variable ΔH_{it} is either the third (quarterly) or fourth difference of the real hedonic log index, $H_{it} - H_{it-3,4}$ .as specified. These varying order differences have been used for two reasons. First, the index accuracy results in Figure 1 confirm that index accuracy for price quartiles does vary and higher order index differences are more accurately measured. Quartiles 2 and 3 are the most accurately measured price segment sub-samples. If the index is accurately

[5] Case & Shiller's (1989) procedure with annual differences was tested. In this study, whereas the index diagnostic for annual differences is very accurate, serial correlation coefficients for annual differences with quarterly lags were virtually zero. These results are not reported.

measured, lower order differences measure the influence of partial serial correlation coefficients (individual monthly lags) more accurately than higher order differences. For less accurately measured price segments the fourth difference provides an alternative estimation. Second, and more importantly an occurrence of terms subject to observation error, common to both sides of equation (4) may bias the coefficient estimates. From equation (3), if using the third order difference both ΔH_{it} and ΔH_{it-3} contain $\hat{c}_{it-3}$, with opposite sign and are therefore negatively correlated. Therefore if a serial correlation coefficient is calculated for a third monthly lag with a similar third order difference, the coefficient will be biased. To overcome this problem coefficients for lags of up to two months are estimated with third order differences and coefficients for lags of up to three months are estimated with fourth order differences. The results indicate few significant correlations at the third and longer lags, therefore differences or lags of more than four monthly periods are not used as explanatory variables in these tests.

The size effects discussed previously raise the possibility that lagged information of price changes from specific price segments may influence price changes in other price segments. Spurious correlation will occur if price changes are used for serial correlation tests as applied in equation (4). Overall trends in the data will cause price segments to increase or decrease at or about the same time and it will not be evident if price changes are occurring at different rates. This may be overcome by using a transformation to differentials. The procedure followed in this study is similar to the abnormal price change procedure used by Clapp Dolde & Tirtiroglu (1995) and Dolde & Tirtiroglu (1997) when analysing patterns of spatial information diffusion in housing markets. The abnormal price change for any price segment quartile ΔPAQ_{it} is the price change for an individual price quartile over and above the unweighted average price change for all price quartiles within the aggregate market ΔP_{Mt} :

$$\Delta PAQ_{it} = \Delta PQ_{it} - \Delta P_{Mt} = \Delta PQ_{it} - \frac{1}{4}\sum_{i=1}^{4} \Delta PQ_{it} \qquad (5)$$

Where for quarterly abnormal price changes $\Delta PQ_{it} = H_{it} - H_{it-3}$. The transformation in equation (5) has the effect of removing any trend common to all price segments within the aggregate market. The sum of all abnormal price changes for all price segments in the aggregate market will be zero for every period. Clapp Dolde & Tirtiroglu (1995) use the term *abnormal* price changes as opposed to excess price changes because the presence of a risk-free rate is eliminated in the transformation to differentials in equation (5). The aggregate market price changes serve as a benchmark from which the differential in an individual market segment might be interpreted as an abnormal return. The procedure is analogous to that used in empirical stock market event studies, where returns of security *X* are compared with returns on a market aggregate *M*, which includes *X*.

Figure 5 presents statistics for abnormal price changes calculated from the real log price monthly hedonic indexes. These results confirm the earlier results for size effects presented in Figure 3. It is evident that the cheapest properties (quartile 1) have negative abnormal price changes. For annual differences the price changes in the first quartile are approximately 2.3% below the average annual price change for the aggregate market. All other quartile groups have positive abnormal price changes and it appears the highest abnormal price changes were achieved within the third price quartile, approximately 1.3% above average as measured by annual differences. The highest volatility as indicated by standard deviation of abnormal price changes is observed within the first and fourth price quartiles. The results for one way ANOVA tests confirm that there are no statistically significant relationships between the mean abnormal price changes for price quartiles until sixth order (half year) differences are applied.

Figure 8 provides results for serial correlation tests for abnormal price changes between own and different price quartile sub-samples. The abnormal price changes are constructed from the hedonic monthly indexes as shown in equation (5). The serial correlation estimating model is of the form:

$$\Delta PAQ_{it} = \beta_0 + \sum_{i=1}^{5} \Delta\beta_k PAQ_{jt-k} + \varepsilon_{it} \quad (6)$$

where the subscripts *i* and *j* applying to the dependent and independent variables refer to the different price quartiles and *k* refers to the relevant monthly lag period. In constructing ΔPAQ_{it} the real log index change $\Delta PQ_{it} = H_{it} - H_{it-6}$, the sixth order (half-year) difference. The results in Figure 5 for ANOVA tests confirm that half-year differences are required before abnormal price changes between price quartiles are statistically different. Because sixth order differences were used, lag periods of up to five months have been used to prevent spurious correlation caused by an occurrence of terms common to both sides of equation (6). Because overlapping sample periods were used in this model the (Hansen & Hodrick: 1980) method of moments estimation procedure was used to estimate standard errors for β_k coefficients.

7.0 Information Diffusion – Empirical Results

Figure 6 provides results for serial correlation tests using the quarterly WRS indexes. Regressions 1,4,7,10 and 13 are index accuracy diagnostics previously discussed. They confirm that the overall aggregate (full sample) indexes are well measured, with the price quartile sub-samples being less accurately measured. All other regressions are serial correlation tests utilising the two quarterly lags for regressions of random sub-sample A on random sub-sample B and the corresponding reverse regression.

These results confirm positive information diffusion for β_1, the first lag (quarter). The aggregate (full sample) index and all price quartile segments have at least one regression showing positive diffusion of information at a level of statistical significance of 5% or greater. The β_1 coefficients for regressions 3 (full sample) and 5 (first quartile) are statistically significant at a level of 1% or greater. Regressions 2, 9, 11 and 14 are statistically significant at a level of 5%, and regressions 6 and 12 are statistically significant at a level of 10%. The only regressions with β_1 not displaying positive serial correlation at a level of statistical significance of 10% or less are regressions 8 (second quartile) and 15 (fourth quartile).

The results for β_2, the second (quarterly) lag indicate some negative serial correlation, although none of these coefficients are statistically significant at a level of 10%. These results are useful in confirming the existence of positive diffusion of information in the aggregate market and all price segment sub-samples. One problem arising from this methodology is that because quarterly differences are used we cannot test if positive information diffusion is more pronounced in the first second or third month of the quarter. Recall from the discussion of efficient markets theory in section 2 the question of relevant lag periods consistent with real estate market efficiency. If weak-form efficiency is consistent with short-term lag periods of one or two months then it becomes difficult to ascertain the significance of results using quarterly differences. The serial correlation tests for the monthly hedonic index results displayed in Figure 7 help overcome this problem.

An immediate difference is observed with the greater explanatory power of the results for monthly price changes shown in Figure 7. All regressions have *R* squared results significantly higher than for the corresponding regressions in Exhibit 6. One reason for this is the greater accuracy of the index differences measured within some price quartiles. Higher positive diffusion at the first (monthly) lag is evident for all regressions. The β_1 coefficients for all of these regressions are statistically significant at a level of 1% or greater. Regressions 1, 3, 5, 7 and 9 report regression results for the full sample and price quartile sub-samples using a third order difference with two monthly lags. It is evident that positive diffusion of information as measured at the first lag is most pronounced in the full sample, β_1 = 0.803 with lower coefficients applying to price quartile sub-samples. It is unlikely that this is due to greater index accuracy for the full sample. Recall from Figure 1 part B that for monthly hedonic indexes, the third order difference for the second and third price quartile sub-samples is measured more accurately than for the full sample. Yet the results in Figure 7 indicate β_1 coefficients of 0.662 and 0.574 for the second and third price quartile sub-samples respectively. One possible reason for the higher coefficient applying to the full sample is that information is diffusing between price quartile sub-samples. Information from adjoining price segments is only reflected in results for the full sample. Another significant result is that the only price quartile showing statistically significant positive diffusion of information at the second lag is the first quartile. Regression 3 indicates that when using the third order difference, the β_2 coefficient is 0.359, higher than for the first lag and statistically significant at 1%. Recall from the discussion of size effects that short-term speculative trading is evident within this first quartile. The results reported in Figure 4, confirm that short-term speculative returns are highest in this first quartile. Price change volatility as measured by standard deviation is also highest. These results indicate that for the first price quartile sub-sample, price changes are more persistent and therefore more predictable than for other price quartile segments. In terms of Fama's (1970) market efficiency classification, the cheaper housing segments do not appear as weak-form efficient in that there is greater potential to use the information contained within past price changes.

Regressions 2, 4, 6, 8 and 10 report regression results for the full sample and price quartile sub-samples using a fourth order difference with three monthly lags. These results confirm previous results that positive diffusion of information as measured at the first lag is most pronounced in the full sample. Regressions 4 and 8 confirm positive diffusion of information at the second monthly lag for the first and third quartile sub-samples at a significance level of 5%. Regression 4 indicates that the first price quartile is the only sample where positive diffusion of information is observed at the third monthly lag, although the *t* statistic is quite low ($p = 0.210$). All other samples have negative coefficients for the third monthly lag, confirming that the information value from past price changes is most pronounced within the first quartile.

The results reported in Figure 7 raise the question as to whether information of price changes from an individual price quartile, influences price changes in other quartile segments. Figure 8 provides regression results for abnormal price changes previously discussed. Each price quartile sub-sample is the subject of regressions using own abnormal price changes, and abnormal price changes from other price quartile sub-samples. A sixth order difference is used with five monthly lags.

Regressions 1,5,9 and 13 use own quartile lagged abnormal price changes as the independent variables. All price quartiles display positive diffusion of information for own abnormal price changes at the first lag although only the second quartile (regression 5, $p = 0.088$) and the third quartile (regression 9, $p = 0.042$) are at a level of statistical significance. This corresponds with the most accurately measured price quartiles for the monthly hedonic index. This could indicate that the greater measurement error created by a transformation to differentials using an average of all price quartiles might be responsible for the lower (and less statistically significant) positive coefficients for the first and fourth price quartiles.

The results for regressions using different price quartiles as dependent and independent variables are harder to interpret. In Figure 8, ten statistically significant serial correlation coefficients are shown in bold script. After elimination of the regressions for own lagged abnormal price changes (regressions 1,5,9,13) there are eight statistically significant serial correlations. In summary, five of these serial correlations are negative and three are positive. Of the three positive serial correlations all are between adjoining price segments and are at either the first or second lag. This supports the argument that in Figure 7, the β_1 serial correlation coefficients are higher for the full sample than for price quartile sub-samples because information diffuses positively between adjoining price quartiles. The highest positive serial correlation coefficient is for regression 16. The dependent variable is the fourth quartile and the independent variable is the third quartile. The β_2 coefficient is 0.303, indicating positive diffusion "upwards" from the lower price quartile at a lag period of two months. A similar result is evident for regression 11 where the β_1 coefficient is 0.260 confirming positive diffusion of information from the lower second quartile to the higher third quartile at a lag period of one month. Regression 12 confirms the results from regression 16. The β_2 coefficient of 0.123 indicates weaker positive diffusion of information between the third and fourth quartiles with this regression indicating that the diffusion "downwards" appears weaker. Interestingly, there are no statistically significant positive serial correlation coefficients for the first price quartile sub-sample.

The most enlightening information is contained in the statistically significant negative serial correlations in Figure 8. There are five significant negative serial correlations, all except one are between non-adjacent price quartiles, and all except one include the first price quartile as either the dependent or independent variable. The strongest negative serial correlation occurs in regression 2. The dependent variable is the first quartile and the independent variable is the second quartile. The β_1 coefficient is -0.403, indicating a negative relationship for abnormal price changes between the first and second price quartiles. This appears to be a relationship that operates from the second quartile "downwards". The corresponding reverse regression (regression 6) has a coefficient on the first lag that is very close to zero with some weak negative relationships at later lags but none that are statistically significant. A similar result is evident for regression 3 where the first price quartile is the dependent variable and the third price quartile is the independent variable. The β_2 coefficient is -0.301 and the β_3 coefficient is -0.295. Both of these coefficients are statistically significant at the 5% level and are the only consecutive lag periods that display significant positive or negative serial correlations for any of the regressions using abnormal price changes. Regression 3 supports the results from regression 2 in that there appears to be a negative relationship for information on abnormal price changes between the first and higher price quartiles. The corresponding reverse regression for regression 3, (regression 10) indicates that this relationship impacts more on lower priced properties. The β_2 coefficient for regression 10 is negative but lower, -0.188 and statistically significant at the 5% level. The other statistically significant serial correlation is in regression 8. The dependent variable is the second quartile and the independent variable is the fourth quartile. The β_5 coefficient is -0.113, supporting the negative relationship for abnormal price changes between lower and higher price quartiles. This result is statistically significant ($p = 0.069$), however the lack of any similar results for longer lag periods for other price quartiles indicate that this result may be due to sampling error.

8.0 Conclusions

Price segmentation is important in explaining house price dynamics. The results reported in this study confirm that size effects exist according to the price hierarchy of housing units sold. These results are similar to observations of securities markets in the short-run. Cheaper properties have highest real annual rates of price change for short holding periods of one year or less. In the long run the highest real annual rates of price change are observed for the highest priced properties. These results are demonstrated both for repeat-sales and hedonic data with statistics for individual property rates of price change together with accurately measured transaction based indexes.

Size effects impact upon the construction of transaction based indexes particularly repeat-sales indexes. Price segmentation confirms that the highest volatility in annual price changes occurs within the first and fourth price quartiles of properties sold. Most noise is contained within the first quartile where volatility induced by short term speculative trading causes bias in the index.

Information diffusion processes are examined for the influence of lagged price changes within own price segments and on other price segments. In terms of Fama's (1970) classification of market efficiency it appears that the housing market analysed in this study presents some anomalies. First, the size effects observed constitute return regularities. Successful short-term speculative trading is observed for all price quartiles. It appears that speculators are able to achieve consistent high positive returns. This influence is most pronounced within the first price quartile. It is difficult to interpret the level of information used by these traders. It is likely that they use more than past price information in developing short-term trading strategies. What constitutes full public information in a housing market? In this study the full information set is defined as all available public information and is the data used in construction of these indexes. The appropriate lag period for this information becoming available is approximately three months. It is highly unlikely that this data could be used to develop profitable trading strategies after allowances for transaction costs. The majority of tests reported using three monthly lags confirm serial correlation coefficients close to zero. This indicates little value for information of past price changes at the time frame in which this information becomes available to the public. Positive diffusion of information for past price changes occurs strongly at shorter lag periods, predominantly one or two months. Linneman (1986) proposed a trading strategy for housing markets whereby an investor could identify those properties that were under priced in comparison to similar properties by estimating hedonic regression models and pricing a property so that the regression residual was negative. Such a "below the line" trading strategy seems a plausible explanation for these high short term positive returns. Real estate buyers and sellers are not price takers. Bilateral negotiations mean that selective bargain hunting behaviour is very likely to occur from informed investors. In some cases this may conform with Fama's strong-form definition of market efficiency. Market participants may possess selective "inside" information likely to influence a seller's motivation.

Tests indicate that significant positive diffusion of information occurs for lag periods of one and two months. The structure of housing markets mean that lagged information for price changes is important and trading based on lagged information may be rational. The results in this study indicate that the majority of trading within price segments could be rational. The lack of significant positive or negative serial correlation coefficients for lag periods of longer than two months confirms that information on past price changes is rapidly capitalised into future house prices. The exception is within the first quartile where positive information diffusion is observed for longer lag periods indicating that prices for cheaper homes are more predictable. These results together with the observed size effects suggest that some irrational feedback trading does occur within the cheaper price segments.

These results appear to be quite consistent with those of Meese & Wallace (1994) who reported violation of market efficiency in the short run, but not in the long run. The results contrast significantly with those reported by Case & Shiller (1989) who reported positive diffusion of information for periods of one year. This is not altogether surprising. Case & Shiller's data was for single-family homes in four major US cities for the period 1971-1988. If data could be found for a similar study in Australia for the same period similar results may occur. Intuition suggests that it is likely that housing markets in Australia have become more efficient in the diffusion of information during the period of this study, 1988 – 1996. First, the rise of the secondary mortgage market has lead to the absence of credit "squeezes" that previously inhibited market activity and the flow of information. Second, the rise of information technology has facilitated more rapid diffusion of information on past prices within housing markets. In Western Australia, brokers have access to data services that provide information for reported sale prices in the previous week.

The influence of information on lagged past price changes between price segments is more difficult to interpret. Some results for tests of abnormal price changes confirm that positive diffusion of information does occur between adjoining price quartiles. Significantly, negative diffusion of information also occurs. Positive abnormal price changes in higher priced properties tends to be accompanied by negative changes in lower priced properties at lag periods of one two and three months. This may be the product of demand and supply relationships where there is a general trend of market participants trading up, thereby increasing demand in the higher priced market segments and at the same time increasing the available supply of cheaper homes for sale. There may also be significant relationships between structural and spatial characteristics of housing and these negative relationships.

It is very likely that the most influential levels of price segmentation have not been identified in this study. Segmentation by price quartiles is a simplistic statistical approach to a housing market with many determinants of relevant price sub-markets. This methodology could be improved with segmentation criteria identified on the basis of relevant housing market theory. One attractive option is specification of price segments according to demographic variables. Census data could be used to identify relevant demographic determinants of the housing stock. This data could then be matched with transaction data in specific spatial sub-markets with similar demographic profiles. Price segmentation according to specific spatial markets could then be used in similar tests to these. The explanatory power of these proposed tests could be improved by using a "stacked" regression procedure as used by Clapp Dolde and Tirtiroglu (1995) thereby increasing the number of observations available for use in serial correlation tests.

References

Bailey, M. J., Muth, R. F., Nourse, H. O. (1963). "A Regression Method for Real Estate Price Index Construction," *Journal of the American Statistical Association*. 58: 933-942.

Banz, R.W., (1981) "The Relationship between Return and Market Value of Common Stock", *Journal of Financial Economics,* March 1981, pp. 3-18.

Beedles, W., Dodd, P., Officer, R., (1988) "Regularities in Australian Share Returns", *Australian Journal of Management*, June 1988, pp. 1-29.

Capozza, D. R., Seguin, P.J., (1994) "Expectations, Efficiency and Euphoria in the Housing Market. *Regional Science and Urban Economics* 26: 369-386

Case, K. E., Shiller, R.J. (1989). "The Efficiency of the Market for Single-Family Homes," *The American Economic Review* 79, 125-137.

Clapp, J.M., Dolde, W., Tirtiroglu, D. (1995). "Imperfect Information and Investor Inferences from Housing Price Dynamics," *Real Estate Economics*, Fall 1995, Vol 23 No 3, 239-269.

Clapp, J.M., Giacotto, C., Tirtiroglu, D. (1991). "Housing Price Indexes Based On All Transactions Compared to Repeat Subsamples," *Journal of the American Real Estate and Urban Economics Association* 19 (3):270-285.

Dolde, W., Tirtiroglu, D. (1997). "Temporal and Spatial Information Diffusion in Real Estate Price Changes and Variances," *Real Estate Economics*, Winter 1997, Vol 25 No 4, 539-565.

Fama, E.E (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance* 25, 383-420.

Gatzlaff, D. H., Ling, D.C. (1994). "Measuring Changes in Local House Prices: An Empirical Investigation of Alternative Methodologies," *Journal of Urban Economics.* 35: 221-244.

Gatzlaff, D.H., Tirtiroglu, D. (1995). "Real Estate Market Efficiency: Issues and Evidence" *Journal of Real Estate Literature*, Vol 3, No 2, pp 157 189.

Gau, G.W. (1987). "Efficient Real Estate Markets: Paradox or Paradigm?" *Journal of the American Real Estate and Urban Economics Association* 15, 1-15.

Hansen, L.P., Hodrick, J.R. (1980). "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy*, October 1980, 88, 829-53.

Linneman, R (1986). "An Empirical Test of the Efficiency of the Housing Market," *Journal of Urban Economics* 20, 140-154.

Meese, R., Wallace, N., (1994) "Testing the Present Value Relation for Housing Prices: Should I Leave My House in San Francisco?" *Journal of Urban Economics* 35(3): 254-266.

Wheaton, W. C., (1990). "Vacancy, Search, and Prices in a Housing Market Matching Model," *Journal of Political Economy* 98(6):1270-1292.

Appendix

A1.0 Index Methodology

The index types used in this study are the Hedonic (explicit-time variable) method, and the Weighted Repeat Sales (WRS) method as used by Case & Shiller (1989). The Hedonic explicit-time variable method groups all data for adjacent time periods and includes discrete time periods as independent (dummy) variables. The following is a functional form acknowledged by Gatzlaff & Ling (1994) and Clapp Giacotto and Tirtiroglu (1991):

$$\ln P_{it} = \sum_{j=1}^{k} \beta_j \ln X_{jit} + \sum_{t=1}^{T} c_t D_{it} + e_{it}, \qquad \text{(A.1.1)}$$

where P_{it} is the transaction price of property i at time t, i= 1,...,n, and t = 1,...,T; β_j, j = 1,...,k, are a vector of coefficients on the structural and locational attributes, X_{jit}; c_t the time coefficients of D_{it}, time dummies with values of 1 if the ith house is sold in period t and 0 otherwise; and e_{it} is the random error with mean, 0, and variance σ_e^2. The sequence of coefficients $c_1,...,c_T$ represents the logarithm of the cumulative price index over the time period T.

For the sake of simplicity and consistency between price segments, the Hedonic indexes used in this study are *simple* hedonic indexes, that utilise only two structural variables, being the natural logarithms of building area and building age at the time of sale. These variables are accurately measured and were available for all transactions. The area of the building is obtained from a survey and is expressed as a discrete square metre number. This variable was also used to screen for structural alterations of repeat-sales. Location attributes were not used in the hedonic indexes. The reason for this is that price segmentation is correlated with specific localities. In effect different index models for different price segments would result from the inclusion of location variables. For time dummy variables discrete *monthly* periods were used for all hedonic indexes. The simple hedonic model provided robust estimates for the full sample and price segment sub-samples. For the full sample the adjusted R squared was .661 with coefficients on structural variables being *Ln Area* 1.011 t = 293.45 and *Ln Age* –0.015 t = -14.203.

The construction of repeat-sales indexes was pioneered by Bailey Muth and Nourse (1963) (BMN). The essential data required for a single property to be used in a repeat-sales index is an initial sale and date and a subsequent sale and date. Due to the same property transacting in different time periods it is assumed that property attributes remain unchanged and the resultant price difference is due to the intervening time period. The repeat-sales technique avoids many of the problems associated with hedonic explicit-time models. It can be shown that the repeat-sales model is a variant of the explicit-time variable approach by contrasting equation A1.2 below with equation A1.3.

$$\ln P_{it} - \ln P_{i\tau} = \left(\sum_{j=1}^{k} \beta_j \ln X_{jit} + \sum_{t=1}^{T} c_t D_{it} \right) - \left(\sum_{j=1}^{k} \beta_j \ln X_{ji\tau} + \sum_{\tau=1}^{T} c_\tau D_{i\tau} \right) + \varepsilon_{it\tau} \qquad \text{(A.1.2)}$$

where P_{it} and $P_{i\tau}$ are the prices of repeat-sales, with the initial sale at time τ and the second sale at time t; X_τ and X_t denote the structural and locational attributes at each respective sale, c_τ and c_t ,the time coefficients of D_{it} and $D_{i\tau}$. If it is assumed that the quality of the housing unit is constant between transactions then the difference between transaction prices at the two dates can be considered as a function solely of the time period, equation (A1.2) reduces to:

$$\ln P_{it} - \ln P_{i\tau} = \sum_{t=1}^{T} c_t D_{it} - \sum_{\tau=1}^{T} c_\tau D_{i\tau} + \varepsilon_{it\tau} \qquad \text{(A1.3)}$$

The dependent variable becomes the logarithm of the price ratio from the property having sold twice. The log price relatives are then regressed on a set of dummy variables corresponding with the time periods. The estimating equation becomes:

$$\ln\left(\frac{P_{it}}{P_{i\tau}} \right) = \sum c_t D_{it} + \varepsilon_{i\tau t} \qquad \text{(A1.4)}$$

where; P_{it} / $P_{i\tau}$ is the price relative for property *i*; D_{it} is a dummy variable which equals -1 at the time of initial sale and +1 at the time of the second sale, and 0 otherwise; c_t is the logarithm of the cumulative price index in period *t*; and $\varepsilon_{i\tau t}$ is a disturbance term. The logarithm of the initial value of the index is normalised by setting initial values in D_1 equal to zero, (i.e. omitting base period time category) and the *T* subsequent coefficients are estimated by Ordinary Least Squares (OLS) regression (Gatzlaff & Ling: 225).

Case & Shiller (1989) recognised the problem of heteroscedasticity induced by holding period in repeat-sales indexes. It can been shown that holding periods are not uniformly distributed through the sample period. Specifically, short holding periods are under represented in the beginning and end periods of the index. Case & Shiller (1989) use a three step Weighted Least Squares (WLS) correction. They model the Weighted Repeat Sales (WRS) method on the assumption that the log price P_{it} of the *i*th house at time *t* is given by:

$$P_{it} = C_t + H_{it} + N_{it} \qquad \text{(A1.5)}$$

where P_{it} is the log price of the *i*th house at time *t*, C_t is the log of the city wide level of housing prices at time *t*, H_{it} is a Gaussian random walk (where ΔH_{it} has zero mean and variance σ_h^2) that is uncorrelated with C_t and H_{jt} $i \neq j$ for all *T*, and N_{it} is an identically distributed normal noise term (which has zero mean and variance σ_N^2) and is uncorrelated with C_t and H_{jT} for all *j* and *T* with N_{jT} unless $i = j$ and $t = T$. Here, H_{it} represents the drift in individual housing value through time, and N_{it} represents the noise in price due to imperfection in the market for housing. The introduction of the two noise terms is in recognition of heterogeneous characteristics of real estate markets and market imperfections in the selling process such as the random arrival of interested purchasers. The WRS index utilises a three-step weighted (generalised) least squares procedure. In the first step the BMN procedure is followed exactly, and a vector of regression residuals is calculated. In the second step the squared residuals from the first step are regressed on a constant and the time interval between sales (holding period). The constant term is the estimate of σ_{N}^2, the noise in price due to imperfections in the market for housing. The slope term is the estimate of σ_H^2, the drift in individual housing unit value through time. In the third step a generalised least squares regression (weighted regression) is run by first dividing each observation in the step-one regression by the square root of the fitted value in the second-stage regression and running the regression again. The WRS indexes in this study use discrete *quarterly* time dummy periods.

<u>A2.1 Formulae for Calculation of *R** - Effective Real Annual Rate of Price Change for Individual Repeat-Sales</u>

In the text *R** is the notation for the effective real annual rate of price change for individual repeat-sales. Full notation is represented by:

$R*_{i|h}$ the effective real annual rate of price change for individual repeat-sale *i* given holding period *h*. The holding period for an individual repeat-sale is expressed as a discrete number of quarterly periods between the initial sale and subsequent sale dates.

$R*_{i|h}$ is calculated as $R_{i|h} - \Delta CPI\big|h$ where $R_{i|h}$ is the annual effective rate of price change for individual repeat-sales unadjusted for inflation. The deflator $\Delta CPI\big|h$ is the contemporaneous consumer price index (CPI) change for the holding period.

$R_{i|h}$ is calculated as [{1 + ((((*Sale_t* + *h* / *Sale_t*) ^ (1 / *h*) - 1) × 4) / 4)} ^ 4] – 1.

Sale_t is the initial selling price and *Sale_t* + *h* is the subsequent selling price.

$\Delta CPI\big|h$ is calculated as [{1 + (((((exp(*ln_cpi_t* + *h*) / exp(*ln_cpi_t*)) ^ (1 / *h*)) - 1) × 4) / 4)} ^ 4] - 1

ln_cpi_t is the logarithm of the CPI index number corresponding with the initial selling price period
ln_cpi_t + *h* is the logarithm of the CPI index number corresponding with the subsequent selling price period.

A2.2 Formulae for Calculation of *WRS** - Effective Real Annual Rate of Change for the Weighted Repeat-Sales (WRS) Index

In the text *WRS** is the notation for the effective real annual rate of price change for the weighted repeat-sales (WRS) index. Full notation is represented by; $WRS*|h$ and is calculated as $WRS - \Delta CPI|h$ where *h* represents the discrete number of quarterly index periods.

WRS is calculated as ((1 + (((((exp(*WRS_t* + *h*) / exp(*WRS_t*)) ^ (1 / *h*)) - 1) × 4) / 4)) ^ 4) - 1

WRS_t is the cumulative logarithmic index coefficient for the period corresponding with the initial sale index period. It is taken from the WLS repeat-sales price index procedure.

WRS_t + *h* is the cumulative logarithmic index coefficient taken from the WLS repeat-sales procedure for the period corresponding with the subsequent sale index period.

$\Delta CPI|h$ is calculated in the same manner as discussed above.

A3.0 Hansen and Hodrick's (1980) Correction for Overlapping Sample Periods

The data used in this study comprises estimated quarterly price index series for WRS index methods and monthly series for hedonic index methods. Indexes are estimated for full samples and sub-samples based on price quartiles.

Assuming a quarterly index, let $\hat{c}_{it}$ represent the natural logarithm of the index for price segment *i* at quarter *t*. The first difference for the index is the continuously compounded quarterly rate of price change for house prices. If monthly series are used and third order differences are taken, the available number of observations (index differences) for serial correlation tests can be increased.

$$\Delta I_{it} = (\hat{c}_{it} - \hat{c}_{it-3}) = (\hat{c}_{it} - \hat{c}_{it-1}) + (\hat{c}_{it-1} - \hat{c}_{it-2}) + (\hat{c}_{it-2} - \hat{c}_{it-3}) \qquad \text{(A3.1)}$$

The problem of "overlapping" data arises because ΔI_{it} will overlap with its two predecessors, sharing two monthly changes with ΔI_{it-1} and one with ΔI_{it-2}. As a result, the ΔI_{it} are not independently distributed and standard errors calculated using ordinary least squares (OLS) will be incorrect thereby invalidating tests of statistical significance for serial correlation coefficients. Whereas unbiased estimates of the coefficients may be obtained using OLS the overlapping error terms necessitates estimation of variances by a method of moments (Hansen and Hodrick:1980). Application of the method in this study is as follows.

$y_1, y_2, y_3, \quad y_q$ are dependent variables for index differences in sample size *Q*.

$y_i = \hat{c}_{it} - \hat{c}_{it-n}$ where c_{it} is the logarithmic index coefficient at time *t* and *n* represents the order of the index period difference.

$x_1, x_2, x_3, \quad x_q$ are corresponding lagged independent variables for similar index differences.

$x_i = \hat{c}_{it-L} - \hat{c}_{it-L-n}$ where *L* is the relevant lag period.

The estimating equation for serial correlation tests can be represented $y_q = \hat{\alpha} + \hat{\beta}x_q + u_q$

where it is assumed that $\hat{\alpha}$ and $\hat{\beta}$ are the unbiased Ordinary Least Squares (OLS) estimates of a constant term and the serial correlation coefficients respectively and u_q is the regression residual.

In matrix algebra notation, OLS coefficients, represented $\hat{\beta}$ can be calculated:

$$\hat{\beta} = (X'X)^{-1} X'Y$$

where *X* is a $Q \times K$ matrix, *K* representing the constant, plus the number of independent variables or lag periods used in estimating the model. *Y* is a $Q \times 1$ vector of *Q* rows and 1 column. The matrix of error terms u_q is of similar dimensions to *Y*. The matrices for an OLS regression assuming two independent variables (x_1 first lag, x_2 second lag) can be represented as:

$$Y = \begin{bmatrix} y_{1(t1)} \\ y_{2(t2)} \\ . \\ . \\ . \\ y_{q(tq)} \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{1(t1-1)} & x_{2(t1-2)} \\ 1 & x_{1(t2-1)} & x_{2(t2-2)} \\ . & & \\ . & & \\ . & & \\ 1 & x_{1(tq-1)} & x_{2(tq-2)} \end{bmatrix} \quad U = \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_q \end{bmatrix}$$

u_1, u_2, u_3, u_q are calculated $u_i = y_i - \hat{y} = y_i - \alpha + \beta_1 x_{i(ti-1)} + \beta_2 x_{i(ti-2)}$

Under OLS assumptions $E(U) = 0$

Under OLS the matrix for calculation of variances of the coefficients is calculated:

$Var(\hat{\beta}) = (X'X)^{-1} X'UU'X(X'X)^{-1}$

Where the error terms u_i are independently distributed, the UU' variance-covariance matrix is consistent and the calculated variances under OLS are applicable. However with overlapping data, serial correlation is created in the u_i terms.

Hansen and Hodrick's (1980) correction involves constructing $\hat{\Omega}$, an alternate variance-covariance matrix:

$\sigma^2_{\hat{\beta}} = (X'X)^{-1} X'\hat{\Omega}X(X'X)^{-1}$

$\hat{\Omega}$ is the $Q \times Q$ matrix

$$\begin{bmatrix}
R_u^Q(0) & R_u^Q(1) & .. & R_u^Q(k-1) & 0 & .. & .. & .. & 0 \\
R_u^Q(1) & R_u^Q(0) & & & R_u^Q(k-1) & & & & .. \\
.. & .. & .. & & & R_u^Q(k-1) & & & .. \\
R_u^Q(k-1) & .. & & .. & & & .. & & .. \\
0 & R_u^Q(k-1) & & & .. & & & R_u^Q(k-1) & 0 \\
.. & & .. & & & .. & & .. & R_u^Q(k-1) \\
.. & & & R_u^Q(k-1) & & & .. & .. & .. \\
.. & & & & R_u^Q(k-1) & & & R_u^Q(0) & R_u^Q(1) \\
0 & .. & .. & .. & 0 & R_u^Q(k-1) & .. & R_u^Q(1) & R_u^Q(0)
\end{bmatrix}$$

where k is the number of sampling periods included in the index difference

$R_u^Q(j) = \frac{1}{N} \sum_{i=j+1}^{N} u_i \cdot u_{i-j}$, $j = 0 \ldots k-1$, and $N = Q - j$

Calculation of $\sigma^2_{\hat{\beta}} = (X'X)^{-1} X'\hat{\Omega}X(X'X)^{-1}$ will result in a $K \times K$ matrix with the variances for coefficients on the main diagonal.

$$\begin{bmatrix} Var(\hat{\alpha}) & & \\ & Var(\hat{\beta}_1) & \\ & & Var(\hat{\beta}_2) \end{bmatrix}$$

Tests of statistical significance can be completed using the t distribution by deriving standard errors from these values.

Figure 1

Quarterly Weighted Repeat Sales Index and Monthly Hedonic Index Summary Statistics

Part A: Quarterly Weighted Repeat Sales (WRS) Indexes

Sample	N	Mean Sale Price ($'000) *s*	Median R^*	*Index Accuracy Ratios* Index Level	First Difference
Full Sample	18,583	89.8 *52.5*	0.015	15.9	5.8
A				*11.4*	*4.2*
B				*11.2*	*4.2*
First Quartile	4,634	47.6 *8.2*	0.007	6.5	3.8
A				*4.7*	*3.1*
B				*4.7*	*2.8*
Second Quartile	4,643	70.7 *5.3*	0.011	8.2	3.1
A				*5.9*	*2.3*
B				*5.7*	*2.4*
Third Quartile	4,675	89.7 *6.7*	0.017	8.0	2.8
A				*5.9*	*2.3*
B				*5.7*	*2.4*
Fourth Quartile	4,631	151.4 *70.6*	0.023	8.1	2.1
A				*5.7*	*1.5*
B				*5.9*	*1.7*

Notes:
1. The median *R** is the median effective real annual rate of price change for individual repeat-sales. The formula for calculation of this variable is given in the appendix.
2. Index accuracy ratios are calculated from the weighted least squares (WLS) regression procedure used to construct the WRS index. Figures given are ratios of the standard deviation of a variable to the average WLS standard error for that variable. Further detail of index procedures is available in the appendix.
3. Index accuracy ratios are given for the full sample or quartile sample together with the respective split random sub-samples *A* and *B*. The split random sub-sample methodology is described in the text.

Part B: Monthly Hedonic (H) Indexes

Sample	N	Mean Sale Price ($'000) *s*	Index Accuracy Ratios Index Level	First Difference	Third Difference
Full Sample	68,799	104.8 *63.4*	5.8	1.2	2.5
First Quartile	17,153	54.7 *12.1*	3.9	1.4	2.2
Second Quartile	17,210	81.6 *11.6*	13.5	3.1	5.0
Third Quartile	17,248	105.8 *18.3*	15.7	3.0	4.7
Fourth Quartile	17,188	177.0 *84.9*	4.7	1.5	2.0

Notes:
1. Index accuracy ratios are calculated from the ordinary least squares (OLS) regression procedure used to construct the Hedonic index. Figures given are ratios of the standard deviation of a variable to the average OLS standard error for that variable. Further detail of index procedures is available in the appendix.

Figure 2

Quarterly Changes in Real Log Price Indexes - Seasonal Influences						
Index Quarterly Sample	All Quarters *Mean* z *std. z*	*Seasonality* Two Sample *t* tests Mean *z* for Quarter *t*				One Way ANOVA
		t = 1 Mean *z* *(t)*	t = 2 Mean *z* *(t)*	t = 3 Mean *z* *(t)*	t = 4 Mean *z* *(t)*	F *Prob.*
WRS Full Sample	0.003 *0.031*	0.015 *(0.974)*	-0.006 *(-0.701)*	-0.010 *(-1.033)*	0.009 *(0.534)*	1.177 *0.338*
Hedonic Full Sample	0.004 *0.033*	0.010 *(0.409)*	0.003 *(0.042)*	-0.017 *(-1.502)*	0.017 *(0.918)*	1.346 *0.281*
WRS First Price Quartile	0.001 *0.039*	0.019 *(1.116)*	-0.006 *(-0.456)*	-0.020 *(-1.371)*	0.009 *(0.469)*	1.550 *0.225*
WRS Second Price Quartile	0.002 *0.030*	0.016 *(1.142)*	-0.010 *(-0.891)*	-0.004 *(-0.436)*	0.002 *(0.038)*	0.954 *0.429*
WRS Third Price Quartile	0.004 *0.033*	0.013 *(0.726)*	-0.004 *(-0.589)*	-0.009 *(-0.906)*	0.011 *(0.569)*	0.812 *0.499*
WRS Fourth Price Quartile	0.003 *0.028*	0.014 *(0.874)*	-0.005 *(-0.675)*	-0.005 *(-0.717)*	0.007 *(0.327)*	0.785 *0.513*

Notes:

1. The variable *z* is the first difference from the real log price index, for *WRS* indexes:
   $z = W_{it} - W_{it-1}$ where; $W_{it} = WRS\ \hat{c}_{it} - LN(CPI_t)$
   A similar methodology applies to the hedonic index.
2. The two sample *t* test, tests the null hypothesis that the mean for a quarterly period sample is the same as the mean for the full sample (All quarters Mean *z*).
3. The One Way ANOVA tests the null hypothesis that the Mean *z* is the same for all quarterly period samples.

Figure 3

Size Effects						
Part A: Individual Properties – Mean *R Statistics for Full WRS Sample and Price Quartiles**						
Sample	All Quartiles *Mean* R* *s*	Two Sample *t* tests Mean *R** for Quartiles 1 - 4				One Way ANOVA
		Quartile 1 Mean *R** *(t)*	Quartile 2 Mean *R** *(t)*	Quartile 3 Mean *R** *(t)*	Quartile 4 Mean *R** *(t)*	F *Prob.*
All Holding Periods	0.047 *0.259*	0.067 *(3.800)*	0.035 *(3.584)*	0.035 *(4.306)*	0.050 *(1.001)*	16.243 *0.000*
Short Holds	0.220 *0.592*	0.369 *(5.376)*	0.192 *(1.249)*	0.146 *(4.220)*	0.160 *(3.911)*	24.709 *0.000*
Long Holds	0.013 *0.070*	0.003 *(8.273)*	0.010 *(3.608)*	0.018 *(4.080)*	0.023 *(7.998)*	58.229 *0.000*

Notes:
1. The variable *R** is the effective real annual rate of price change for individual repeat-sales. The formula for calculation of this variable is given in the appendix.
2. The two sample *t* test, tests the null hypothesis that the mean for a quartile sample is the same as the mean for the full WRS sample (All quartiles Mean *R**).
3. The One Way ANOVA tests the null hypothesis that the Mean *R** is the same for all price quartile samples.

Part B: Mean Real Log Index Changes for Full Monthly Hedonic Sample and Price Quartiles						
Sample	All Quartiles *Mean* z *s*	Paired Sample *t* tests Mean *z* for Quartiles 1 - 4				One Way ANOVA
		Quartile 1 Mean *z* *(t)*	Quartile 2 Mean *z* *(t)*	Quartile 3 Mean *z* *(t)*	Quartile 4 Mean *z* *(t)*	F Prob.
Quarterly Change	0.004 *0.040*	0.002 *(0.657)*	0.006 *(0.829)*	0.008 *(1.486)*	0.007 *(0.724)*	0.306 *0.821*
Half Year Change	0.005 *0.062*	0.001 *(1.256)*	0.012 *(2.251)*	0.017 *(3.321)*	0.016 *(2.580)*	1.218 *0.303*
Annual Change	0.0003 *0.082*	-0.007 *(-1.899)*	0.017 *(4.225)*	0.029 *(8.165)*	0.025 *(6.517)*	3.140 *0.026*

Notes:
1. The variable z is the relevant index period difference for the monthly hedonic real log price index. For quarterly changes:
   $z = H_{it} - H_{it-3}$, where: H_{it} = Hedonic $\hat{c}_{it} - LN(CPI_t)$
   For half year and annual changes, $z = H_{it} - H_{it-6}$, and $z = H_{it} - H_{it-12}$ respectively.
2. The paired sample *t* test, tests the null hypothesis that the mean for an individual price quartile sample is the same as the mean for the full sample (All quartiles Mean *z*).
3. The One Way ANOVA tests the null hypothesis that the Mean *z* is the same for all individual price quartile samples.

Figure 4

Regression of Individual Property Price Changes on WRS Index Changes

$$R^* = \alpha + \beta WRS^* + \varepsilon$$

Sample	Mean R* *s*	No obs. S.E.E.	α *(t)*	β *(t)*	R^2
Full Sample	0.047	18,583	0.028	1.488	0.151
	0.259	0.238	*(15.764)*	*(57.551)*	
Short Holds	0.220	3,030	0.160	1.502	0.114
	0.592	0.557	*(15.158)*	*(19.786)*	
Long Holds	0.013	15,553	0.006	1.005	0.381
	0.070	0.055	*(13.029)*	*(97.870)*	
First Quartile	0.067	4,634	0.046	1.863	0.264
	0.365	0.313	*(9.952)*	*(40.740)*	
Short Holds	0.369	811	0.231	1.708	0.200
	0.790	0.707	8.687	14.220	
Long Holds	0.003	3,823	0.007	1.009	0.360
	0.074	0.059	*(6.833)*	*(46.381)*	
Second Quartile	0.035	4,643	0.025	1.156	0.098
	0.226	0.215	*(7.928)*	*(22.478)*	
Short Holds	0.192	646	0.154	1.007	0.049
	0.561	0.548	*(6.844)*	*(5.780)*	
Long Holds	0.010	3,997	0.006	0.975	0.410
	0.064	0.049	*(7.723)*	*(52.684)*	
Third Quartile	0.035	4,675	0.020	1.094	0.139
	0.184	0.171	*(7.665)*	*(27.455)*	
Short Holds	0.146	650	0.114	1.076	0.090
	0.449	0.429	*(6.615)*	*(7.985)*	
Long Holds	0.018	4,025	0.006	0.987	0.324
	0.067	0.050	*(6.957)*	*(55.532)*	
Fourth Quartile	0.050	4,631	0.028	1.114	0.081
	0.222	0.213	*(8.526)*	*(20.218)*	
Short Holds	0.161	923	0.135	1.103	0.051
	0.461	0.449	*(8.819)*	*(7.000)*	
Long Holds	0.023	3,708	0.003	1.055	0.394
	0.072	0.056	*(2.984)*	*(49.053)*	

Notes:

1. The dependent variable *R** is the effective real annual rate of price change for individual repeat-sales. The independent variable *WRS** is the contemporaneous real annual change in the WRS index for the holding period. The formulae for calculation of these variables are given in the appendix.
2. The regressions denoted *Short holds* and *Long holds* are for reduced sub-samples. Short holds are defined as holding periods of one year or less. Long holds are all holding periods greater than one year.

Figure 5

Abnormal Price Changes in Real Log Price Indexes						
Index Period Difference	Unweighted Average *Price Change* All Quartiles ΔP_{Mt} *s*	Mean *PA* for Quartile *Q*				One Way ANOVA
		Q = 1 Mean *PA* *s*	*Q* = 2 Mean *PA* *s*	*Q* = 3 Mean *PA* *s*	*Q* = 4 Mean *PA* *s*	F *Prob.*
Quarterly	0.006 *0.040*	-0.004 *0.028*	0.0004 *0.017*	0.002 *0.015*	0.001 *0.035*	1.083 *0.356*
Half Year	0.012 *0.056*	-0.010 *0.026*	0.001 *0.016*	0.005 *0.019)*	0.004 *0.029*	8.613 *0.000*
Annual	0.016 *0.079*	-0.023 *0.028*	0.001 *0.018*	0.013 *0.019*	0.009 *0.032*	34.567 *0.000*

Notes:

1. The variable *PA* is the abnormal price change taken from the real log price change, for monthly hedonic indexes. Full notation is represented by ΔPAQ_{it} , the price change for an individual price quartile over and above the unweighted average price change for all price quartiles within the aggregate market ΔP_{Mt} :

   $$\Delta PAQ_{it} = \Delta PQ_{it} - \Delta P_{Mt} = \Delta PQ_{it} - \frac{1}{4}\sum_{i=1}^{4} \Delta PQ_{it}$$

   where for quarterly abnormal price changes $\Delta PQ_{it} = H_{it} - H_{it-3}$ and $H_{it} = \text{Hedonic } \hat{c}_{it} - \ln(CPI_t)$. For half year and annual changes, sixth and twelfth order index differences are taken for construction of ΔPQ_{it}
2. The One Way ANOVA tests the null hypothesis that the Mean *PA* is the same for all price quartile samples.

Figure 6

Serial Correlation Tests – WRS Quarterly Indexes							
$\Delta W_{it} = \beta_0 + \beta_1 \Delta W_{j,t-1} + \beta_2 \Delta W_{j,t-2} + \varepsilon_{it}$							
Sample	Regression	Parameters: † = Index accuracy diagnostic	No. obs. S.E.E.	β_0 *(t)*	β_1 *(t)*	β_2 *(t)*	R^2 Adj R^2
WRS Full Sample	1	† *i = A, j = B, Lag = 0*	30 0.010	0.001 *(0.398)*	0.953 *(17.007)*		0.912 0.909
	2	*i = A, j = B*	28 0.021	-0.002 *(-0.631)*	0.412 *(2.454)*	-0.193 *(-1.293)*	0195 0.130
	3	*i = B, j = A*	28 0.020	-0.003 *(-0.735)*	0.537 *(3.215)*	-0.234 *(-1.613)*	0.296 0.239
WRS First Quartile Q_1	4	† *i = A, j = B, Lag = 0*	30 0.029	0.000 *(-0.046)*	0.810 *(5.941)*		0.558 0.542
	5	*i = A, j = B*	28 0.030	-0.005 *(-0.899)*	0.487 *(2.762)*	-0.237 *(-1.496)*	0.235 0.174
	6	*i = B, j = A*	28 0.026	-0.005 *(-1.075)*	0.251 *(1.902)*	-0.023 *(-0.193)*	0.134 0.065
WRS Second Quartile Q_2	7	† *i = A, j = B, Lag = 0*	30 0.021	0.000 *(0.076)*	0.750 *(6.336)*		0.589 0.574
	8	*i = A, j = B*	28 0.024	-0.004 *(-0.761)*	0.334 *(1.969)*	0.123 *(0.792)*	0.164 0.098
	9	*i = B, j = A*	28 0.024	-0.003 *(-0.674)*	0.420 *(2.502)*	-0.085 *(-0.545)*	0.206 0.143
WRS Third Quartile Q_3	10	† *i = A, j = B, Lag = 0*	30 0.018	0.000 *(-0.027)*	0.679 *(7.575)*		0.672 0.660
	11	*i = A, j = B*	28 0.021	-0.002 *(-0.602)*	0.337 *(2.361)*	-0.102 *(-0.872)*	0.183 0.117
	12	*i = B, j = A*	28 0.027	-0.002 *(-0.479)*	0.348 *(1.738)*	-0.156 *(-0.854)*	0.108 0.036
WRS Fourth Quartile Q_4	13	† *i = A, j = B, Lag = 0*	30 0.021	0.001 *(0.374)*	0.596 *(5.124)*		0.484 0.466
	14	*i = A, j = B*	28 0.023	-0.001 *(-0.297)*	0.310 *(2.357)*	0.011 *(0.082)*	0.183 0.117
	15	*i = B, j = A*	28 0.028	-0.001 *(-0.244)*	0.231 *(1.145)*	-0.141 *(-0.707)*	0.054 -0.022

Notes:

1. The variable W_{it} is the quartertly WRS log index deflated by the contemporaneous CPI index: $W_{it} = WRS\ \hat{c}_{it} - LN(CPI_t)$
2. The dependent variable ΔW_{it} is the first (quarterly) difference of the real log index, $\Delta W_{it} = W_{it} - W_{it-1}$
3. The subscripts *i* and *j* denote the random sub-sample *A* or *B* used as either dependent or independent variable. Equal random sub-samples are used to avoid spurious correlation. This methodology is more completely described in the text.
4. Regressions denoted † are index accuracy diagnostics where contemporaneous first differences for sub-sample A are regressed on sub-sample B. This methodology is more completely discussed in the text.

Figure 7

Serial Correlation Tests – Hedonic Monthly Indexes

$$\Delta H_{it} = \beta_0 + \sum_k \beta_k \Delta H_{it-k} + \varepsilon_{it}$$

Sample	Regression	Parameters	No. obs. S.E.E.	β_0 *(t)*	β_1 *(t)*	β_2 *(t)*	β_3 *(t)*	R^2 Adj R^2
Hedonic Full Sample	1	Third Difference	87 0.024	-0.001 *(-0.250)*	0.803 *(7.489)*	-0.067 *(-0.640)*		0.604 0.594
	2	Fourth Difference	85 0.023	-0.002 *(-0.642)*	0.767 *(7.242)*	0.054 *(0.390)*	-0.099 *(-0.989)*	0.666 0.654
Hedonic First Quartile	3	Third Difference	87 0.030	-0.002 *(-0.612)*	0.352 *(3.787)*	0.359 *(3.918)*		0.467 0.455
	4	Fourth Difference	85 0.030	-0.004 *(-1.147)*	0.372 *(3.359)*	0.250 *(2.245)*	0.132 *(1.265)*	0.564 0.547
Hedonic Second Quartile	5	Third Difference	87 0.031	0.001 *(0.196)*	0.662 *(6.232)*	-0.048 *(-0.451)*		0.414 0.400
	6	Fourth Difference	85 0.032	0.000 *(-0.095)*	0.531 *(4.971)*	0.200 *(1.622)*	-0.118 *(-1.169)*	0.431 0.410
Hedonic Third Quartile	7	Third Difference	87 0.034	0.001 *(0.252)*	0.574 *(5.175)*	0.039 *(0.345)*		0.338 0.322
	8	Fourth Difference	85 0.035	0.002 *(0.410)*	0.507 *(4.708)*	0.253 *(2.064)*	-0.183 *(-1.731)*	0.376 0.353
Hedonic Fourth Quartile	9	Third Difference	87 0.057	0.003 *(0.459)*	0.443 *(3.924)*	-0.076 *(-0.670)*		0.158 0.138
	10	Fourth Difference	85 0.052	0.004 *(0.747)*	0.471 *(4.553)*	0.025 *(0.204)*	-0.157 *(-1.486)*	0.239 0.211

Notes:

1. The variable H_{it} is the monthly Hedonic log index deflated by the contemporaneous CPI index: $H_{it} = \text{Hedonic}\,\hat{c}_{it} - \ln(CPI_t)$
2. The dependent variable ΔH_{it} is either the third or fourth order difference of the real log index, $H(t) - H(t)_{t-3,4}$ as specified in the parameters column.

Figure 8

Serial Correlation Tests – Abnormal Price Changes for Monthly Hedonic Indexes

$$\Delta PAQ_{it} = \beta_0 + \sum_{i=1}^{5} \Delta\beta_k PAQ_{jt-k} + \varepsilon_{it}$$

Sample	Regression	Parameters	No. obs. S.E.E.	β_0 (t)	β_1 (t)	β_2 (t)	β_3 (t)	β_4 (t)	β_5 (t)	R^2 Adj R^2
First Quartile	1	*i = PA_Q1*	81	-0.009	0.143	0.161	-0.021	0.003	0.038	0.063
		j = PA_Q1	0.024	*-3.103*	*1.274*	*1.482*	*-0.186*	*1.032*	*0.349*	0.001
	2	*i = PA_Q1*	81	-0.013	**-0.403***	0.214	0.141	0.189	0.046	0.093
		j = PA_Q2	0.024	*(-4.417)*	*(-2.362)*	*(1.274)*	*(0.861)*	*(1.126)*	*(0.280)*	0.033
	3	*i = PA_Q1*	81	-0.009	-0.032	**-0.301***	**-0.295***	0.020	0.083	0.129
		j = PA_Q3	0.023	*(-3.525)*	*(-0.218)*	*(-2.084)*	*(-2.093)*	*(1.058)*	*(0.588)*	0.071
	4	*i = PA_Q1*	81	-0.013	0.035	-0.071	0.080	-0.083	-0.068	0.029
		j = PA_Q4	0.025	*(-4.367)*	*(0.357)*	*(-0.786)*	*(0.879)*	*(1.111)*	*(-0.712)*	-0.035
Second Quartile	5	*i = PA_Q2*	81	0.000	**0.202***	0.140	0.029	-0.002	0.076	0.092
		j = PA_Q2	0.015	*(-0.073)*	*(1.729)*	*(1.186)*	*(0.241)*	*(1.033)*	*(0.685)*	0.032
	6	*i = PA_Q2*	81	0.000	0.018	-0.021	-0.038	-0.063	0.106	0.035
		j = PA_Q1	0.016	*(-0.076)*	*(0.259)*	*(-0.305)*	*(-0.549)*	*(1.056)*	*(1.476)*	-0.029
	7	*i = PA_Q2*	81	0.000	-0.054	-0.017	0.043	-0.024	0.034	0.008
		j = PA_Q3	0.016	*(-0.020)*	*(-0.529)*	*(-0.179)*	*(0.447)*	*(1.062)*	*(0.338)*	-0.058
	8	*i = PA_Q2*	81	0.001	-0.041	-0.015	-0.031	0.044	**-0.113***	0.058
		j = PA_Q4	0.016	*(0.197)*	*(-0.668)*	*(-0.244)*	*(-0.522)*	*(1.066)*	*(-1.845)*	-0.005
Third Quartile	9	*i = PA_Q3*	81	0.005	**0.233***	0.015	0.159	-0.001	-0.076	0.092
		j = PA_Q3	0.018	*(2.068)*	*(2.069)*	*(0.132)*	*(1.369)*	*(0.979)*	*(-0.693)*	0.032
	10	*i = PA_Q3*	81	0.005	-0.091	**-0.188***	0.045	0.161	-0.091	0.124
		j = PA_Q1	0.017	*(1.780)*	*(-1.168)*	*(-2.566)*	*(0.583)*	*(1.120)*	*(-1.193)*	0.066
	11	*i = PA_Q3*	81	0.007	**0.260***	-0.072	-0.095	-0.020	-0.049	0.052
		j = PA_Q2	0.018	*(2.769)*	*(2.012)*	*(-0.570)*	*(-0.757)*	*(1.130)*	*(-0.389)*	-0.011
	12	*i = PA_Q3*	81	0.007	-0.112	**0.123***	-0.060	-0.101	0.111	0.109
		j = PA_Q4	0.017	*(2.755)*	*(-1.616)*	*(1.896)*	*(-0.918)*	*(1.092)*	*(1.615)*	0.049
Fourth Quartile	13	*i = PA_Q4*	81	0.005	0.119	-0.037	0.012	0.140	0.071	0.048
		j = PA_Q4	0.028	*(1.434)*	*(1.066)*	*(-0.341)*	*(0.109)*	*(1.043)*	*(0.642)*	-0.015
	14	*i = PA_Q4*	81	0.004	-0.070	0.047	0.014	-0.101	-0.053	0.017
		j = PA_Q1	0.029	*(1.007)*	*(-0.538)*	*(0.375)*	*(0.106)*	*(1.045)*	*(-0.415)*	-0.049
	15	*i = PA_Q4*	81	0.006	-0.060	-0.282	-0.075	-0.167	-0.074	0.059
		j = PA_Q2	0.028	*(1.796)*	*(-0.283)*	*(-1.353)*	*(-0.347)*	*(1.055)*	*(-0.367)*	-0.004
	16	*i = PA_Q4*	81	0.005	-0.147	**0.303***	0.092	0.005	-0.041	0.047
		j = PA_Q3	0.028	*(1.112)*	*(-0.828)*	*(1.724)*	*(0.526)*	*(1.039)*	*(-0.232)*	-0.016

Notes:

1. The dependent variable ΔPAQ_{it} is the half year abnormal price change, defined as the price change for an individual price quartile over and above the unweighted average price change for all price quartiles within the aggregate market ΔP_{Mt} for the haly year:

   $$\Delta PAQ_{it} = \Delta PQ_{it} - \Delta P_{Mt} = \Delta PQ_{it} - \frac{1}{4}\sum_{i=1}^{4} \Delta PQ_{it}$$

   where for half year abnormal price changes $\Delta PQ_{it} = H(t)_{it} - H(t)_{it-6}$

   and $H(t) = \text{Hedonic } \hat{c}_{it} - \ln(CPI(t))$
2. The subscripts *i* and *j* denote the price quartile sub-sample (*PA_Q1…4*) used as either dependent or independent variable in the regression. These details are shown in the parameters column.
3. Coefficient statistics denoted * and in bold script are statistically significant at a level of 10%.