Heterogeneous Asset Pricing: An Examination of the Australian Residential Real Estate Market

Danika Wright

A dissertation submitted in partial fulfilment

of the requirements for the degree of

Doctorate of Philosophy

Discipline of Finance

Faculty of Business and Economics

University of Sydney

October 2010

Certificate of Originality

I hereby declare that this submission is my own work. To the best of my knowledge the content of this thesis contains no materials previously published or written by another person, without due acknowledgement. This thesis has not been submitted for any other degree or purpose.

I certify that the intellectual content of this thesis is the product of my own work and that all assistance received in preparing this thesis and sources have been acknowledged.

Signature of Candidate

…………………..……………………………………….

Danika Wright

Acknowledgements

I extend my most sincere thanks to my supervisors, Dr. Maurice Peat and Professor Alex Frino, for their expert guidance and ongoing encouragement. Many thanks also to Associate Professor Graham Partington, Professor David Johnstone and Professor Michael McKenzie for their advice and assistance. This thesis has benefited from the comments of Dr. Andrew Lepone, Professor David Gallagher, Dr. Vito Mollica and Dr. Matthew Hardman, to whom I extend my most heartfelt gratitude.

I gratefully acknowledge the financial support provided through the Australian Research Council Linkage Grant scheme, the Australian Postgraduate Award scheme and Rismark International. I additionally acknowledge the generosity of Rismark International, RP Data, the Australian Securities Exchange, and the Capital Markets Cooperative Research Centre. The excellent facilities, resources and opportunities provided by each of these institutions have added immeasurably to this thesis.

Sincere thanks go to my friends and colleagues. A price can not be put on the value of the comments and company of Kiril Alampieski, Abhishek Das, Elisa Di Marco, Andrew Grant, Jen Kruk, Steve Lecce, George Li, Teddy Oetomo, Talis Putnins, Jiri Svec, Brad Wong and Hui Zheng over the course of my candidature. Finally I thank my closest friends and parents for their unending support and belief in me.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ABS	Australian Bureau of Statistics
ACF	Autocorrelation Function
ACR	Annual Compound Return
ADF	Augmented Dickey-Fuller
AIC	Aikaike Information Criterion
ANOVA	Analysis of Variance
APM	Australian Property Monitors
ARCH	Autoregressive Conditional Heteroskedasticity
ARIMA	Autoregressive Integrated Moving Average
ASGC	Australian Standard Geographical Classification
ASX	Australian Securities Exchange
CDF	Cumulative Distribution Function
CME	Chicago Mercantile Exchange
CSI	Case-Shiller House Price Indices
FHOG	First Home Owner's Grant
GDP	Gross Domestic Product
HIA	Housing Industry Association
IMR	Inverse Mills Ratio
IPO	Initial Public Offering
MLE	Maximum Likelihood Estimation
NAR	National Association of Realtors
NSW	New South Wales
OLS	Ordinary Least Squares
OTC	Over-the-Counter
PACF	Partial Autocorrelation Function
PDF	Probability Distribution Function
RBA	Reserve Bank of Australia
RPX	RP Data
SBC	Schwarz-Bayes Criterion
SMH	Sydney Morning Herald
UK	United Kingdom
USA	United States of America
VG	Valuer General
WRS	Weighted Repeat-Sales

Synopsis

This thesis assesses the methods for pricing heterogeneous assets, with specific examination of the Australian residential real estate market. Each of the three primary methods for controlling for heterogeneity in residential real estate pricing research – median-price analysis, hedonic regression and repeat-sales regression – is susceptible to bias. This thesis represents the first broad study of the impact of these alternative methodologies on research outcomes.

Methodological comparisons are made in the study of residential real estate market efficiency in Chapter 4. Using an extensive database of property sale and attribute data, indices are independently estimated following the three alternative methods, from which market efficiency, in its weakest form, is tested. Specifically, the autoregressive predictability and seasonality of returns is measured using Box-Jenkins methodology and regression analysis.

Access to more comprehensive data than has previously been available to researchers enables the estimation of indices following the three alternative methodologies. This allows a controlled cross-methodology comparison of results and represents a major extension to the previous work into residential real estate market efficiency (Case and Shiller, 1989; Rosenthal, 2006). Furthermore, this study represents the first major study of weak-form efficiency and seasonality in returns to the major Australian property markets of Sydney and Melbourne.

The results indicate that a predictable autoregressive lag structure exists in the returns, thus rejecting the hypothesis of market efficiency in the Australian residential real estate market in its weakest form. There is, however, no deterministic monthly seasonality in returns although the results from the alternative indices are not consistent. Median-price index returns, for example, exhibit significant heterogeneity-induced seasonality and negative first-order autocorrelation. This contrasts with the returns to the repeat-sales index, which exhibit significant *positive* first-order autocorrelation.

Chapter 5 examines the effect of sample selectivity in the presence of asset heterogeneity on hedonic pricing models. A substantive literature has demonstrated the existence of a price impact to properties sold by auction relative to those that sell by private treaty (Dotzour et al., 1998; Lusht, 1996; Mayer, 1998). The work presented in this thesis extends the research in several ways. Firstly, sample selectivity between sale methods is tested and controlled for using the Heckman two-stage procedure and a matched sampling technique. Secondly, a larger sample of sales than has previously been considered is fitted to a more completely specified hedonic function. This is made possible by access to a database containing the virtual population of sales from the Sydney house market. Finally, this represents the most thorough study of the impact of auctions on prices in the Australian market.

Using selectivity-corrected regression analysis, it is found that the auction sale mechanism has no effect on prices in the Sydney house market. This runs contrary to the results of past research in the Australian and New Zealand residential real estate markets which document a price premium (Dotzour et al., 1998; Lusht, 1996). Unadjusted hedonic-regression analysis is also performed, the results of which support an auction premium. This difference in results demonstrates the selectivity bias which has influenced past research, and a further methodological issue in pricing heterogeneous assets.

The pricing and performance of new properties is considered in Chapter 6. This is a further examination of the ability of hedonic pricing models to counter sample selectivity. Sample selectivity-controlled methods are applied to a comprehensive sample of sales from the Perth residential real estate market. Differences in attributes between the sample of new and existing properties are observed. After controlling for this sample selectivity, the results indicate a significant price premium to new houses and units of around 10% and 7%, respectively. An examination of the returns to new properties at their first subsequent sale, however, demonstrates significant underperformance of new residential real estate assets relative to the market. This is the first study to empirically and theoretically assess the pricing of new residential real estate assets.

Finally, Chapter 7 applies alternative index estimation methodologies to a study of the value of public and private information in the residential real estate market. Heterogeneity in the market is a major source of information asymmetry. Indices for the Sydney residential property market are estimated from lead sources of information, both public (newspaper reporting and advertisements) and private (agents' sales records). Correlation and regression analysis demonstrates that private information is a strong predictor of movements in an index estimated from the population of sales, in line with *a priori* expectations. This is the first study to assess the value of information in this market and represents a new direction for residential real estate research, particularly as housing derivatives markets continue to evolve.

1. Introduction

1.1 Overview

Valued at over A$3.4 trillion, residential real estate is the largest investable asset class in Australia. Despite the importance of this market to the macroeconomy, investors, regulators and academics alike, it is also one of the least researched asset classes.

The complexity of measuring residential real estate prices is the fundamental limitation to past research. This complexity is driven by several features inherent to the market: heterogeneity, illiquidity and data accessibility. No two houses are identical, less than 10% of the Australian housing stock turns over annually, and data are difficult to obtain, and rarely complete or current, as a result of privacy regulations and the highly decentralised nature of the market. While several methodologies have developed in response to these endemic issues, none are perfect.

Median-price analysis relies on the median observation of all prices in a given market and period. This approach, while straightforward to apply and parsimonious in data requirements, is affected by variations in the types of properties observed to sell across time periods or markets. For example, if there is a higher proportion of more expensive properties selling in one period than another, the observed change in the time-series of median prices will be influenced by this change in traded composition.

Two regression-based alternatives designed to overcome the effects of composition bias in residential real estate price analysis are the hedonic regression and repeat-sales methodologies.

Hedonic-regression analysis involves modelling residential property prices as a function of the set of attributes that differentiate properties, such as location and property characteristics. In this way, heterogeneity between properties is statistically accounted for. Repeat-sales techniques are a specialised form of hedonic regression. This approach models prices from the sample of trade pairs for all properties that sell more than once and, assuming properties remain unchanged between sales, explicitly controls for the heterogeneity in the housing stock.

Both of these approaches rely on the various specification and distributional assumptions of regression analysis. Omitted variable bias, in particular, is likely to affect these methods. The ability to fully specify hedonic models is often limited by data availability and the complex interrelationships between assorted explanatory variables. The standard repeat-sales model, on the other hand, by assuming that properties remain constant between sales, omits any age effects. A further issue specific to the repeat-sales technique is the potential selectivity among the sample of properties that sell more frequently, and are thus represented in this model.

The different methodological biases associated with these various approaches are likely to influence the results of research into behaviour and performance of prices in the residential real estate market. However, outside of the application of these methods to price index estimation, a complete void exists in the comparative research concerning how use of one method over another may influence the outcomes of empirical housing market price research. This thesis provides the first broad study of the impact of methodology on residential real estate market price research.

To demonstrate the impact of the potential biases associated with various methodologies, this dissertation applies median pricing, hedonic regression, and repeat-sales techniques to assess the weak-form efficiency of the residential real estate market, the impact of sale method on prices, the pricing of residential properties at their first sale, and derivative market structure. To each of these studies, the choice of alternative methodologies for controlling for heterogeneity across assets is of vital importance to the results. This question has not previously been considered outside of the effect various measures have on the bias and precision of house price index estimates.

This thesis benefits from access to an extensive and previously unexplored database, allowing the application of new econometric modelling techniques to residential real estate market data. This database covers a broader cross-section of properties and longer time series than has previously been used in Australian residential real estate market research. Supplied by RP Data (RPX), a public Australian property information and reporting company, the database is built upon the sales reported to the Valuer-General (VG) for each state and territory; in effect, capturing the total population of sales in the Australian market. RPX augments this with attribute data collected from print and internet property listings and its own real estate agent clients. The richness of this data enables the use of data-intensive regression-based techniques for analysing house prices and, in so doing, overcomes a major limitation to the past research in this market.

1.2 Research Topics

Residential Real Estate Market Weak-Form Efficiency

The first empirical comparison of the alternative methodologies is made through an analysis of the weak-form efficiency of the Australian residential real estate market. Disagreement exists in the conclusions of past research which has tested for weak-form efficiency in residential real estate markets. The seminal work undertaken in the United States (USA) by Case and Shiller (1989) indicates that a significant degree of predictability exists in the returns to residential real estate. This conclusion, made using repeat-sales methods, violates the efficient markets hypothesis. More recent research however, such as Rosenthal (2006), has used hedonic-regression approaches and demonstrated that weak-form efficiency holds in housing returns. No study has previously considered the impact of methodology on the results of such research.

The relevant literature which has considered residential real estate market efficiency and seasonality is reviewed in Section 2.2 and is empirically examined in the Australian market in Chapter 4. The review identifies that research into the weak-form efficiency of returns to residential real estate is largely restricted to the North American housing markets. In Australia there has not previously been any broad research into the housing market's efficiency. The implications of the results of this research, however, are of practical significance to investors and regulators alike, particularly given the recent innovation of housing derivatives products, such as the Chicago Mercantile Exchange (CME)-listed Case-Shiller House Price Indices (CSI).

The Effect of Sale Method on Prices

The Australian residential real estate market presents a unique opportunity to study the impact of market structure on prices, given the coexistence of multiple sale methods: auctions and private treaty sales. Taking a rational expectations perspective, property price should be unrelated to the sale method. That is, if a price premium to auctions persisted, relative to private treaty sales, then private treaty sales would disappear, as all sellers would choose to auction their properties, and vice versa.

Yet the consensus finding of previous research into the residential real estate market design is that a systematic price impact results from the use of auctions. Some researchers have found a price discount for auction sales (Allen and Swisher, 2000; and Mayer, 1998, both in the USA) based on the argument that the auction mechanism is typically used for disposal sales following defaults on financing. Other research has observed a price premium for auction sales (Dotzour et al., 1998, in New Zealand; Lusht, 1996; and Newell et al., 1993, in Australia). An issue in the comparability of the results of these studies is that they have all relied on different models. Mayer (1998), for example, uses a repeat-sales approach while Dotzour et al. (1998) estimate a hedonic-regression. The past research to have assessed whether auctions have an impact on prices of residential real estate is critiqued in Section 2.3.1. Empirical testing of the effect of sale method on prices in the Australian residential real estate market is reported in Chapter 5.

In line with this exploration of rational expectations pricing, this thesis examines the pricing of residential property assets at their first sale. No previous research has considered this area of residential real estate pricing, although similarities with the information asymmetry explanations for equity initial public offering (IPO) underpricing are identified. The relevant equity IPO literature is reviewed, including the limited research to have touched on information asymmetry in the residential property pre-sale market in Section 2.3.2. Chapter 6 reports the empirical results from assessing the pricing and performance of new residential real estate assets in the Australian market.

The results are of importance from both practical and academic perspectives. The pricing of new properties and their subsequent performance should affect investment decisions. Taxation and regulatory policies designed to encourage new home purchases and developments need to consider any pricing anomalies at the first sale. Such monetary incentives include the recent grants, and concessions on stamp duty applicable to new property purchases from federal and state governments.

Academic research concerned with residential real estate valuation is often complicated by the management of property age effects. In hedonic-regression analysis, for example, property age is often omitted from the pricing function due to limited data availability. If age and prices are shown to be related, this can lead to specification error in the hedonic form, and specifically, the case of omitted variable bias.

Residential Real Estate Derivatives Markets

The recent innovation of residential real estate derivative products motivates many new research questions – one of which is how informationally efficient the housing market is, and what value can be put on public and private information. Section 2.3.3 reviews the structure of housing derivatives markets. The value of information is then examined empirically in Chapter 7.

Traditionally, residential real estate information is difficult and costly to obtain, and rarely timely. This is a result of the highly disaggregated market, and of privacy laws which restrict public access to residential real estate sales information. Whereas in the direct investment market, real estate transactions typically require weeks to conclude, derivatives markets for housing may offer higher-frequency trading. The publication of house price indices which underlie such products, however, will still be lagged due to delays in sales reporting and data processing. Consequently, current transactions information – such as the sample of sales results reported in newspapers – may be used to predict the underlying index movement.

1.3 Structure of the Dissertation

The remainder of the dissertation is organised as follows.

Chapter 2 reviews the literature relevant to residential real estate pricing and market design. This chapter first defines residential real estate, with particular reference to the Australian market, and demonstrates its position as a financial asset. The literature to have examined the role of residential real estate in the macroeconomy is reviewed, as is the past research to have demonstrated the portfolio diversification gains from including residential real estate assets. Through these sections, the importance of accurate and unbiased research into the measurement of residential real estate prices is first identified. This is followed by a review of the statistical methodologies developed to measure residential real estate price dynamics.

The literature review proceeds by examining several areas of research which have applied these methods. Section 2.2 considers the weak-form market efficiency of the residential real estate market. Section 2.3 reviews literature related to the trading and pricing of residential real estate assets; specifically, the effect on prices of alternative sale methods in the secondary market, the pricing of new properties, and the design of housing derivatives markets.

From the literature review, Chapter 3 develops a set of testable hypotheses. These hypotheses are designed to further explore topics which previous research has disagreed upon and to address gaps in the academic research identified through the literature review. Testing of the hypotheses developed in this Chapter then forms the basis for the research conducted in Chapter 4 through to Chapter 7.

Chapter 4 empirically assesses the weak-form efficiency of the Australian residential real estate market. Focusing on the Sydney and Melbourne property markets, the returns from three alternative price index estimates are modelled to test for forecastability and deterministic seasonality.

Chapter 5 examines whether the method by which a property is sold – that is, auction or private treaty – has an effect on the price achieved. Several statistical approaches are applied to a comprehensive sample of sales from the Sydney housing market to test for a sale method price effect.

Chapter 6 analyses prices and performance of residential properties at their first sale. The chapter focuses on transactions of residential real estate sales in Perth, as this is the largest market for which construction year data are available.

Chapter 7 is concerned with the market design of residential real estate derivatives markets. Drawing from the existing structure of the listed property derivatives on the CME, this chapter assesses whether the underlying house price index figures may be predicted from leading sources of property sale information in the Sydney market.

Chapter 8 concludes the thesis. It summarises the findings of the research undertaken in each of Chapters 4 through 7, and ties this together to form a larger conclusion as to the effect of alternative methods on residential real estate price research. In this final chapter, the implications of the main conclusions made in this dissertation are discussed, and directions for future research are proposed.

2. Literature Review

2.1 Background

The objective of this dissertation is to examine the significance of methodology choice in influencing the conclusions of empirical studies of residential real estate prices. In this section, residential real estate is defined. From an outline of its broad characteristics, its relation to the macroeconomy and investment methods are explored in detail. Lastly, this section provides an overview of the methodologies for measuring residential real estate price dynamics.

2.1.1 What is Residential Real Estate?

In its most fundamental definition, real estate encompasses all buildings, the land upon which they are built and all vacant land. This is obviously a very broad definition. The buildings or land may be used on the one hand by any number of commercial, agricultural, industrial, community or government organisations or, on the other, serve as a place of residence. A broad range of uses and market participants exist across the different real estate types. Consequently, the pricing and market design in these alternative market sectors often differ. For this reason, the residential real estate market is often segregated for analytical purposes from the commercial and other non-residential sectors.

Though similarities exist between commercial and residential real estate – they both, for example, are large, tangible and durable assets – their markets are not necessarily influenced by the same economic conditions. While major macroeconomic factors such as interest rates have an effect on the prices and activity in both markets, this influence is unlikely to be equal or contemporaneous (DiPasquale and Wheaton, 1996). Furthermore, demand for non-residential property is driven by more volatile economic variables, such as employment, gross domestic product (GDP) and money supply, but household income and financing are more significant determinants of residential real estate prices and demand. Montezuma and Gibb (2006) argue that,

> "Housing demand (owner-occupied and indirectly rented sector) depends [as much] on availability, cost and flexibility of mortgage financing as on demographic factors, which are less important in explaining non-housing property demand" (Montezuma and Gibb, 2006: 341).

The main reason for these contrasting driving factors is the differences in economic motives between participants in the residential market compared to those in the commercial, agricultural, or industrial property markets: since the different assets provide different 'physical' services, the agents acting in these markets are different. A growing literature demonstrates, in fact, that the correlation between returns to commercial and residential real estate is close to zero or even negative (Montezuma and Gibb, 2006). The key empirical research to have considered the correlation between returns to the alternative real estate sectors and other financial assets is discussed in the more general framework of portfolio diversification in Section 2.1.3.

This dissertation considers pricing efficiency in the residential real estate market. Residential real estate is defined as the total of land, buildings and other improvements that exist for the purpose of providing people residence. As such, the terms residential real estate, residential property and housing assets are used interchangeably in this dissertation.

Residential real estate itself comprises submarkets that may be defined by geographical region or property type. In Australia, the Sydney housing market is often talked of separately from the Perth housing market, for example. Similarly, the dynamics of freehold housing and strata-units, while related, are not perfectly co-integrated. Although often aggregated together to form some estimate of the national housing market performance, the factors affecting residential property prices differ in each capital city (Brown, Li and Lusht, 2000).

Research into the dynamics and prices of the residential real estate market is of importance to retail and institutional investors, economists, and policy makers, not least as a result of its size. Residential real estate globally represents one of the largest asset classes (Himmelberg, Mayer, and Sinai, 2005). Residential real estate in Australia is estimated to be worth A$3.4 trillion.[1] By comparison, the Australian domestic equity market has a capitalisation of approximately A$1.3 trillion.[2]

Residential real estate accounts for a substantial proportion of wealth at both the individual and economy-wide levels. The Reserve Bank of Australia (RBA) estimates that approximately 65% of the average Australian household's wealth is invested in their home, and that residential real estate investment accounts for around 85% of household borrowing.[3] The implication of these figures to the broader macroeconomy is discussed in further detail in Section 2.1.2.

---

[1] Estimate current as at 31 January 2010. Source: RPX.

[2] Estimate current as at 31 January 2010. Source: *Historical Market Statistics*, ASX.

[3] Source: *Reserve Bank Bulletin*, July 2004.

Finally, residential real estate is the most pervasively held investable asset class in Australia; it is estimated that approximately 65% of Australians have some level of exposure to residential property assets,[4] compared with an estimated share ownership rate of 46%.[5] While not unique among developed countries in this respect – similar homeownership rates are reported for the USA (67.5%)[6] and United Kingdom (UK) (68%),[7] – homeownership rates are significantly lower in many other developed countries, including Germany and France, with homeownership rates of 43%[8] and 57%,[9] respectively. This is largely due to cultural factors, such as the often discussed 'Great Australian Dream' (Moran, 2006), and government-backed schemes to increase homeownership, such as the First Home Owners Grant (FHOG).

Often, in Australia and other countries with high levels of residential property investment, the property that is owned is the investor's primary place of residence. That is, the investors are 'owner-occupiers'. Australia has one of the highest homeownership rates in the world, with less than a third of households renting.[10] When the residential property investment is not the investor's primary place of residence it is commonly described as an 'investment' property. This is a somewhat

---

[4] The ABS estimates that 32.6% of the Australian adult population own residential real estate outright (that is, without any financing), while 32.3% of the population have debt-financed residential real estate investment exposure (typically through mortgage-debt instruments). Source: *Year Book Australia*, ABS, Catalogue 1301, 2008.

[5] Source: *Share Ownership Study*, ASX, 2007.

[6] Source: USA Census Housing Bureau, 2009.

[7] Source: *Housing and Planning Statistics*, UK Statistics Authority, 2009.

[8] Source: Statistisches Bundesamt Deutschland, 2009.

[9] Source: European Commission Eurostat, 2001.

[10] Source: *Australian Census Data*, ABS, 2006.

misleading term as all wealth directed to residential property should be considered a financial investment; the distinction is only in the component of that investment which is used for consumption purposes.

That the purchase and sale of residential real estate represent the largest financial transaction that most individuals will undertake in their lifetimes (Can, 1992) further warrants the need for research into its pricing and market dynamics. Despite this, data relating to the fundamental investment characteristics of this market is limited. There is, for example, no simple answer to the question ‘How many houses are there in Australia?’ There is also little consensus among house price index providers as to the level of growth over a given historical period. This is demonstrated by the returns to the Australian housing market estimated by alternative data providers from the February 2010 *Statement on Monetary Policy* produced by the RBA presented in Table 2-1; the estimates of growth to the September 2009 quarter range from 2.8% to 4.4%.

The two major reasons why more financial research has not to date been undertaken in this market are: (1) the difficulties that develop in accounting for the features that set it apart from other financial asset markets; and (2) issues with data accessibility, typically arising from privacy concerns. These affect the ability to develop robust and reliable pricing models for residential real estate assets, a tool at the very basis of financial research. The methods for measuring prices and pricing residential real estate are reviewed in Section 2.1.4.

Table 2-1:

Australian House Price Growth

This table presents the recent returns to Australian housing as calculated by three different index providers: the Australian Bureau of Statistics (ABS), Australian Property Monitors (APM), and RPX.

Index Provider	September 2009 Quarter Return (%)	December 2009 Quarter Return (%)	Year to September 2009 Quarter Return (%)	Aggregate return: Peak-to-trough (%)	Aggregate return: Trough-to-latest (%)
Panel A: Capital Cities					
ABS	4.4	5.2	13.6	-5.5	14.5
APM	3.7	4.3	11.7	-3.6	11.8
RPX	2.8	2.8	10.2	-3.0	10.2
Panel B: Regional Areas					
APM	2.6	5.9	10.6	-4.8	10.6
RPX	1.0	1.9	5.1	-3.8	5.1

Source: *Statement on Monetary Policy*, RBA, February 2010

Several features of residential real estate assets set them apart from other financial investments. These include: (1) high transaction costs; (2) spatial fixity and immobility; (3) parcel size, durability and indivisibility; and (4) heterogeneity. Physical transactions of residential real estate incur high costs as a result of stamp duty, mortgage duty, legal conveyance fees and building inspection fees. The Housing Industry Association (HIA) estimates that the round-trip costs on the median-priced Australian residential property are over $200,000.[11] The average cost of residential real estate investment in Australia is illustrated in Figure 2-1.

Residential real estate is an immobile asset: it is fixed to its location. This feature is also referred to as spatial fixity. Any physical investment in the asset requires the potential buyer to go to the asset since the asset cannot realistically be delivered. This results in a highly decentralised market for trading, which in turn creates the potential for very large information asymmetries to exist and persist.

Real estate is a large, durable, tangible good: it is indivisible and single units require large initial investments of wealth. As a result, direct investments in the physical housing stock are typically highly leveraged and form a major component of the investor's wealth portfolio. Financing of housing purchases are typically secured by using the property being purchased as collateral – this arrangement is so common, in fact, that these specific loans are distinguished from other borrowing as *mortgages*. To the loan provider, the immobility and durability that real estate assets provide are preferred qualities in collateral.

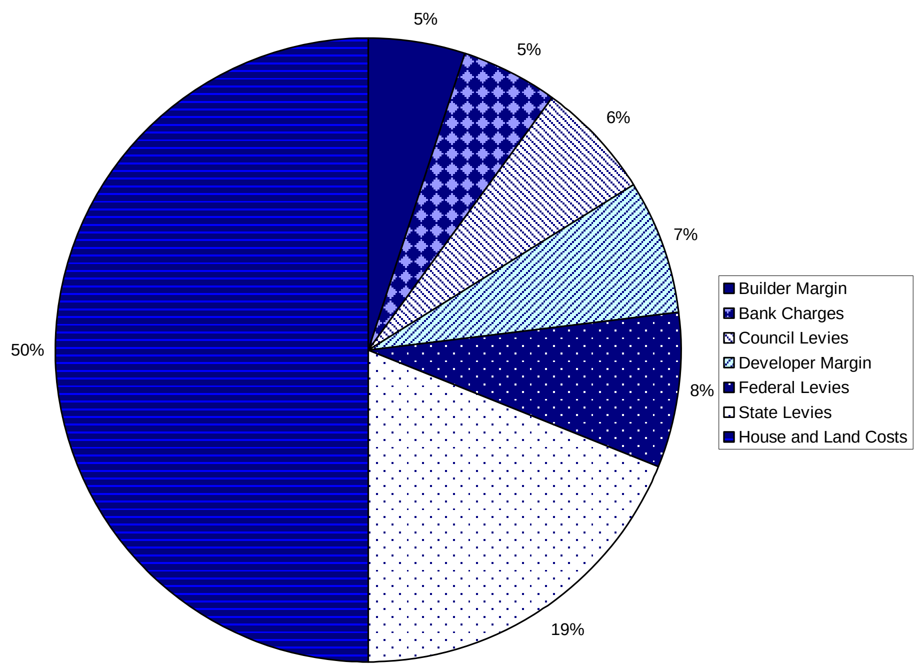
---

[11] Source: *HIA National Outlook*, Housing Industry Association, 2007.

Figure 2-1:

The Cost of Residential Real Estate Investment in Australia

Figure 2-1 demonstrates the breakdown of costs associated with investment in Australian residential real estate.



Source: HIA

The decision to take such a long, undiversified position in an asset and finalise financing arrangements, however, requires time; The average lag time in Australia between price agreement at the exchange of contracts and the settlement date when ownership is transferred is six weeks. As a result of this time cost, in addition to the high monetary costs, in Australia the average holding period of residential real estate is relatively long and the annual turnover rate of the housing stock is very low, estimated at 7.5 years and 6%,[12] respectively.

[12] Source: RPX.

Arguably the most significant feature of residential real estate that sets it apart from traditional financial assets is asset heterogeneity. No two physical housing investments are identical since every unit of residential real estate has unique attributes and spatial characteristics, and is indivisible (DiPasquale and Wheaton, 1996). ‘Property attributes’ refers to quantifiable features such as the number of bedrooms, age or size of a property, as well as qualitative features such as its overall condition. The spatial characteristics of a property, often summarised as its *location*, include its given aspect, elevation, and proximity to neighbourhood features. As a durable good, units of housing are exposed to age, a feature affected in part by prior ownership (Goetzmann and Spiegel, 1995). Thus, even two units built to be identical will have differences, and these differences will increase with time.

Given the absolute heterogeneity of the housing stock and the infrequent turnover of assets, the price formation process of residential real estate assets is obscured. The resulting information asymmetry is exacerbated by the unknown physical changes which may occur to the housing stock; “Many durable goods are either new to the market or are subject to a high level of technical change. Hence, information about them may take time to diffuse through the population” (Deaton and Muellbauer, 1983: 346). This further compounds the information asymmetries existent in the market.

Heterogeneity affects the ability to price housing assets and to assess the price performance of the residential real estate market as a whole. While econometric techniques have developed to overcome pricing issues associated with illiquidity in the equity and other financial markets (Lo and MacKinlay, 1990), specific real estate pricing methods have evolved to control for heterogeneity, such as hedonic regression and repeat-sales methods. However, their implementation in empirical research into the housing market has been limited by their intensive data requirements. These methods, their advantages and disadvantages are discussed in more detail in Section 2.1.4.

Accessibility to residential real estate sales information is very limited. This is driven partly by the size of the market and complexity in collating trades from a non-centralised market. While some information relating to prices and dates can be sourced from state and territory VG offices, to which all sales are eventually reported, accessibility is limited and expensive, and does not include detailed attribute or spatial data. Yet, due to heterogeneity across assets, it is of vital importance to know more than sales information; to adequately conduct unbiased analysis in the residential real estate market the researcher requires data on the property's location and characteristics.

Evaluating the price performance of residential real estate is further complicated by the need to account for its income stream – typically the rental yield – and the costs to ownership. In addition to its capital gains, the income residential real estate assets generate, *rent*, determines the value of housing. This holds for investment properties as well as those considered owner-occupier where the rent stream is an implied consumption cost. Thus, even if all *observable* data are available, the income to owner-occupied properties would remain unknown.

Academics and practitioners have attempted to account for rental yields despite this data limitation, typically by predicting an imputed rental yield for all properties in the market using regression analysis and data from the sample of rental properties. This has enabled the estimation of accumulation indices for research purposes and housing derivatives markets. However, an additional limitation is in differentiating between gross and net rental yield, as the costs to investment in residential real estate are non-trivial and detailed data for these costs is scarce.

The recent development of derivative contracts written over portfolios of residential real estate, allowing *synthetic investment*, and the increasing use of securitisation as a means of gaining exposure, known as *indirect investment*, may reduce the impact of these issues and increase the liquidity of the market. These developments improve the ability of residential real estate to be considered a financial asset, furthering the need for research into the dynamics of this market. The next section, Section 2.1.2, considers the role of residential real estate in the macroeconomy and is followed in Section 2.1.3 by a discussion of the alternative methods of residential real estate investment and its treatment as a financial asset.

2.1.2 Residential Real Estate and the Macroeconomy

This section discusses the findings of research that has sought to identify the relationship between the economy and the residential real estate market. The key channel identified in the literature is the effect house prices have on consumption patterns. This in turn is used as a premise to explain the relationship between housing market activity, the business cycle and interest rates.

House prices and the economy are related through many channels. In many cases this relationship is endogenous; changes to the state of the housing market, for example, have an effect on the economy, and vice versa. The channels by which housing and the economy are related may be direct or indirect, and the consequences may be positively or negatively correlated. However, these channels and their specific working are far from understood (Iacoviello, 2005).

In response to this gap in academic understanding a large literature, dedicated to empirically testing hypotheses that attempt to explain the close relationship between house prices and the economic business cycle, has developed. A non-trivial component of the motivation behind this research has been the anecdotal 'smoking gun' position real estate has had in a number of major economic and financial crises, particularly when aggregate credit levels are affected by falling house prices, including the Asian financial crisis of 1997 (Quigley, 2001) and the 2009 global financial crisis (Shiller, 2009).

Leamer (2007) claims in his NBER working paper title that 'Housing *is* the Business Cycle'. This bold statement is supported – albeit in varying degrees of significance, lags, and robustness – by a growing body of recent theoretical and empirical economic research. Iacoviello (2005) in particular has made significant contributions in this area. Building on the work of Kiyotaki and Moore (1997), Iacoviello (2005) incorporates the role of house prices in a model of the business cycle. Specifically, his research demonstrates the significant effect movements in residential real estate prices have on aggregate credit levels. He shows that this arises largely as a result of the use of real estate assets as collateral against a large proportion of borrowing. As discussed earlier, real estate is a desirable asset for financing collateral due to its durable and immovable characteristics.

The interconnectivity of housing markets and the economy is further amplified through its effect on government revenues, production and investment activity, and wealth effects.

As a result of the taxes and transfer duties imposed by governments on residential real estate, approximately 40% of the New South Wales (NSW) state government's taxation revenues, accounting for over 10% of total state treasury income, are derived from housing markets. In housing downturns, however, reliance on this income stream can contribute to significant budget imbalances. The NSW state government, for example, reported a 2008-09 budget shortfall in expected revenue from stamp duty of $1.068 billion resulting from lower property prices and turnover.[13]

A strong housing market, and the correlated housing demand, boosts construction activity and wider economic output. In Australia, for example, it is estimated that A$64 billion (approximately 5.3% of GDP) is directly contributed to the economy from the building and renovation industries. Factoring the indirect impact of housing investment, such as to the primary materials and transport industries, the ABS estimates a total multiplier effect of 2.866 to the residential real estate construction and investment industry.[14]

---

[13] Source: *Report on State Finances 2008-09*, NSW Government Treasury, 2009.

[14] Source: *Year Book Australia*, ABS, Catalogue 1301, 2002.

The wealth effect refers to the changes in spending levels that are associated with a corresponding change in either real or perceived wealth. Across a range of markets, changes in house prices are shown to affect consumption levels via this wealth effect. To measure the wealth effect of residential real estate markets, the elasticity of consumption to house prices – that is, the expected percentage change in consumption in response to a 1% change in house prices – is estimated. The estimates of long-run elasticity of consumption to house price vary from around 0.06 in Case, Quigley, and Shiller (2005) and 0.08 in Davis and Palumbo (2001), to 0.2 in Iacoviello (2005). For further empirical research in this area see Campbell and Cocco (2007) and Greenwood and Hercoqitz (1991). The common finding across this research is a significant and positive wealth effect attributable to house prices.

Resulting from this positive relationship between consumption and house prices, rising house prices are consequently associated with rising consumer prices, inflation, and interest rates. As such, investment in residential real estate can serve as a useful hedge against inflation and interest rates (Bond and Seiler, 1998). From such an observation, a number of researchers have examined the more general investment and diversification possibilities of residential real estate. The next section, Section 2.1.3, reviews the literature that has considered the role residential real estate can play in the investment portfolio.

The large volume of research that has been applied to residential real estate markets across a range of disciplines and across the boundaries of academic, corporate and regulatory research is matched by the interest in investing in its assets. The next section describes the ways in which investors may gain exposure to this market.

2.1.3 Residential Real Estate Investment

This section gives an overview of the alternative methods of residential real estate investment – direct, securitised, and synthetic. This is followed by a review of the results of past research that has empirically examined the potential diversification benefits of residential real estate assets to a multi-asset portfolio.

Direct investment is the most common method of transacting real estate assets. Direct investment involves taking a capital position in the physical housing asset. In practice, only long positions – that is, investments that are a positive function of the asset price – can be acquired in direct real estate investment. Taking a short position over direct residential real estate investment is unheard of as a result of the impracticalities such an investment would involve – the borrowing of a third party's real estate asset, a round-trip sale, including transaction costs, and return to the original owner within a given time-frame. Furthermore, as discussed in Section 2.1.1, direct investment requires taking a relatively substantial and undiversified exposure in residential real estate assets.

Securitised investment, also referred to as indirect investment, is enabled by institutions that create funds from a pool of real estate assets. Examples include real estate investment trusts and mortgage-backed funds. A number of these investment trusts list on exchanges, including the Australian Securities Exchange (ASX), with the theory that investors are able to take equity positions in the securitised real estate. The degree to which such vehicles provide investors with the true dynamics of real estate returns is further complicated by their level of co-movement with the overall equity market (Oikarinen, Hoesli and Serrano, 2009).

The demand for liquidity in residential real estate market investments has most recently seen the emergence of housing derivatives. Futures, options and other derivatives written over residential real estate are all methods of synthetic investment. The major centre for the development of such products has been the USA, where exchange-traded options and futures have been listed through the CME since May 2006. Various over-the-counter (OTC) derivative products, including forwards, options, swaps and spread bets, have been launched in the USA, UK, Hong Kong and Singapore (Ong and Ng, 2009; Shiller, 2008), with similar OTC and exchange-traded products expected to launch in Australia in the near future.[15]

The emergence of housing derivative products creates a relatively low-cost, more liquid, synthetic alternative to direct investment residential real estate investors to gain exposure to price movements in the market. Furthermore, these developments enable opportunities for diversification, hedging and negative net exposure in the residential property area, which currently are impractical. The academic research that has considered – often from a theoretical viewpoint – the potential benefits of residential real estate in a multi-asset portfolio is now reviewed.

---

[15] 'New Way to Buy A Piece of Property', Maurice Dunlevy, *The Australian*, September 11, 2008.

The early research into the potential diversification gains of investments in residential real estate drew from what had then recently been discovered about commercial property's investment potential (Friedman, 1971; Fisher and Sirmans, 1994). As data accessibility improved, researchers increasingly focused on the role of residential real estate in the investment portfolio. Table 2-2 summarises the key papers that have considered the role of residential real estate in the institutional investor's portfolio.

Ibbotson and Siegel's (1984) work on multi-asset portfolio optimisation is one of the earliest papers to include residential real estate in a comparison with other investment classes. For the period 1947 to 1982, an annual total returns[16] index for real estate is constructed as the market-value weighted composite of residential, agricultural, and (from 1960) commercial real estate. The returns to this real estate composite index and equity returns, based on the S&P500 index, are found to be negatively correlated.

Ibbotson and Siegel (1984) also document negative correlation between the returns to real estate and the returns to both long-term corporate and government bonds. While not including any further portfolio optimisation analysis in their research, Ibbotson and Siegel (1984) conclude that real estate can add value to a multi-asset portfolio based on the negative correlation it has with other major asset classes. This broad finding is supported in almost all subsequent research in the area.

---

[16] Capital gain and income estimates are sourced from a variety of sources for each type of real estate, including the home purchase component of the consumer price index, the USA Department of Agriculture, the Engineering News Record's Building Cost Index, and various surveys. Various assumptions are made relating to the rate of depreciation, yield to commercial property, and real estate taxes and leverage.

Table 2-2:

Portfolio Diversification and Residential Real Estate

This table summarises the key papers to have examined the diversification potential of residential real estate in a multi-asset portfolio. The authors and publication date of these papers are reported, along with a description of the real estate index methodology, market and period used. A summary of the empirical findings of each paper is also provided.

Study Authors	Market, Period and Data	Findings
Ibbotson and Siegel (1984)	USA, 1947-1982: Appraisal-based market-value weighted national composite index of agricultural, residential and commercial real estate	Negative correlation between real estate returns and equities, long-term corporate and long-term government bonds
Webb, Curcio, Rubens (1988)	USA, 1947-1983: The disaggregated Ibbotson and Siegel (1984) index	Low/negative correlation between housing and equity, bonds and other real estate
Goetzmann and Ibbotson (1990)	USA, 1969-1987: Repeat-sales based equally weighted composite index of residential real estate returns to four major cities	Negative correlation between real estate returns and equities and both long- and short-term government bonds
Hutchison (1994)	UK, 1984-1993: Appraisal-based equally weighted national composite	Low correlation between housing and equities, and negative correlation between housing and short-term government debt
Hoesli and Hamelink (1997)	Switzerland, 1981-1992: Total return hedonic indices	Estimate an optimal allocation to housing of between 20% and 30% of wealth in multi-asset institutional portfolio
Lee (2008)	Australia, 1996-2007: ABS Stratified-median national composite price index	Low/negative correlation between house price index and equity, bonds and commercial property indices

Webb, Curcio, and Rubens (1988) undertake portfolio optimisation in the presence of real estate assets using a stylised single index model. This research considers the various real estate sectors – agricultural, residential, and commercial – as separate asset markets. The findings of Webb et al. (1988) show that for any combination of stock, bonds and real estate assets, residential real estate is given a significant positive weight in the mean-variance optimal portfolio. Interestingly, the results presented in Webb et al. (1988) support the theory of low – or even negative – correlation between returns to residential and non-residential property as discussed in Section 2.1.1.

The limitations to the conclusions of their research arising from the available house price index data are acknowledged in both Ibbotson and Siegel (1984) and Webb et al. (1988). The house price index data used in both studies is based on valuations, or 'appraisal data'. While overcoming the liquidity and compositional issues associated with the use of transactions data, appraisal-based data are prone to bias in valuer accuracy. Geltner (1993) shows that the primary bias in appraisal-based indices is smoothing, resulting in a spuriously large first-order autocorrelation in returns.

Goetzmann and Ibbotson (1990) attempt to avoid the potential issues associated with appraisal-based indices by applying the repeat-sales house price index estimation technique (Bailey, Muth and Nourse, 1963; Case and Shiller, 1987). A national composite house price index representing four cities in the USA for the period 1970 to 1986 is estimated and compared with commercial real estate, equities, and government securities. Annual returns to residential real estate are observed to be negatively correlated with equities and both long- and short-term government bonds, adding to the argument for real estate to be included in a well-diversified portfolio.

Regional diversification in residential real estate is also touched upon by Goetzmann and Ibbotson (1990). Separating the cities underlying their national composite index – Atlanta, Chicago, San Francisco and Dallas – into four distinct price indices, the authors demonstrate that potentially large risk-return gains can be obtained from geographic diversification. This is analogous to the work first undertaken by Solnik (1974) and many since that has shown international diversification of equities in the stock portfolio adds value. For a further discussion of regional diversification in equities markets see Driessen and Laeven (2007) and Campa and Fernandes (2006).

The notion of diversification within the housing market is extended by Brown et al. (2000). In line with Markowitz's (1952) prediction that combining assets affected by less common factors increases diversification gains, Brown et al. (2000) consider the grouping of housing submarkets by common economic characteristics, such as industry, employment or demography, rather than purely geographical proximity. The authors use quarterly transaction-based indices for the Hong Kong housing market covering the period January 1984 to December 1995, finding that efficient portfolios from active intra-city diversification are superior to portfolios based on equally weighted strategies. This finding is broadly supported by the other research to have considered alternative strategies for intra-market residential real estate diversification, including Eichholtz, Hoesli, MacGregor and Nanthekumaran (1995), Chua (1999) and Wilson and Zurbruegg (2003).

The reported benefits of residential real estate's inclusion in multi-asset portfolio strategies are not confined to research in the USA residential property market. The main research articles focusing on multi-asset portfolio optimisation using markets outside of the USA to be discussed in this dissertation are Hutchison (1994) in the UK; Hoesli and Hamelink (1996) in Switzerland; and Lee (2008) in Australia.[17]

Hutchison (1994) looks at the performance of residential real estate in the UK over the period 1984 to 1992 relative to equities and government bonds (gilts). While providing no specific examination as to the effect on a portfolio of combining these assets, diversification potential of residential real estate can be inferred from the statistics reported. In particular, Hutchison's (1994) observation of low correlation between housing and equities (0.0772) and negative correlation between housing and gilts (-0.2627) suggests that in the UK there may be potential portfolio diversification gains to allocating a portion of wealth to residential real estate. This result supports the findings of earlier work in the UK real estate market undertaken by MacGregor and Nanthakumaran (1992).

Hoesli and Hamelink (1996) use efficient frontier analysis to assess the level of diversification benefit to an institutional investor from investment in Zurich and Geneva residential property. Their findings suggest an optimal allocation to housing assets of between 20% and 30% of portfolio value. The authors also provide some evidence to support the regional diversification gains found by Brown et al. (2000).

---

[17] Other notable research into portfolio optimisation in the presence of residential real estate inside the USA includes Grauer and Hakannson (1995) and Kallberg, Liu and Greig (1996). For further reading into multi-asset portfolio diversification incorporating residential real estate in markets outside the USA, the articles by Ben-Shahar (2003) in Israel; Montezuma and Gibb (2006) in Switzerland and The Netherlands are worth considering.

Lee (2008) reports similar findings on the potential diversification benefits of residential real estate in the Australian market in his study of inter-asset correlation. Using quarterly returns over the period 1996 to 2007 from the ABS House Price Index, and comparing the inter-asset correlation between it and various equity, bonds, and commercial property indices, Lee (2008) finds that Australian housing has diversification potential.[18]

The major impediment to the uptake of – and subsequent poor market depth in – housing derivative products has been the quality and availability of adequate measures of market-wide price movement. This in turn has limited the academic empirical research into the profitability and optimality of residential real estate investment. The seeming lack of monetary interest in these emerging markets is widely attributed to investors' misunderstanding and suspicion of the processes involved in the estimation of the underlying index. In the next section, Section 2.1.4, the main index constructions methodologies are reviewed.

---

[18] The discussion to this point has considered residential real estate as part of the institutional investor's portfolio. Portfolio optimisation at the individual or household level is a more challenging research question as it requires consideration of an existing undiversified residential property asset. The main research articles that have examined portfolio diversification from the perspective of *a priori* homeownership are Goetmann (1993), Flavin and Yamashita (2002), Englund, Hwang and Quigley (2002) and Cocco (2005).

2.1.4 Measuring Residential Real Estate Prices

This dissertation, while examining the existence of pricing biases and anomalies in the residential real estate market, has a broader objective of assessing the implications of the biased and inefficient methodologies that have been applied in the past to such research. This section provides an overview of the key methodologies that have evolved to examine the dynamics of residential real estate prices. Specifically, a description and critique of the main house price index construction methodologies are given. These include mean- and median-price based measures, as well as both non-parametric and regression-based constant-quality approaches.

The growth in academic interest concerning the influence residential real estate market activity and prices may have on the macroeconomy, and vice versa, has largely been enabled by recent developments in and availability of house price indices. The three main methods by which indices for house prices are measured are: (1) the median-price series; (2) the hedonic regression; and (3) the repeat-sales regression. These are outlined in Table 2-3.

The median-price series forms an index of price levels in the residential real estate market by tracking the median price of transacted properties in a given time period. Mark and Goldberg (1984) strongly support the use of the median-price series on the grounds that it is relatively straightforward to implement, is inclusive of all sales, and has extremely parsimonious data requirements (see also Crone and Voith, 1992).

Table 2-3:

Price Index Estimation Methodologies

Table 2-3 provides a summary of the three key methods for estimating price indices for residential real estate. These methods are: (1) the median-price index; (2) the hedonic regression; and (3) the repeat-sales regression. For each, the leading academic papers are cited, a description of the data required and methodology are provided, and the main advantages and disadvantages of the method are listed.

	Median-Price Index	Hedonic Regression	Repeat-Sales Regression
Leading Academic Papers	Mark and Goldberg (1984), Crone and Voith (1992), Prasad and Richards (2006)	Griliches (1971), Rosen (1974), Wallace (1996), Triplett (2004)	Bailey, Muth, and Nourse (1963), Case and Shiller (1987), Webb (1988), Goetzmann and Spiegel (1995)
Data Required	Sale price and date	Sale price, date, and detailed hedonic attribute data concerning structural characteristics and features of the property	Sale price and date for those properties that have sold more than once over the sample period
Method	Non-parametric; time-series of median prices	Parametric; multivariate regression to estimate quality-controlled market price changes	Parametric; regression of the price appreciation on repeat-sales (trade pairs of the same property) to estimate quality-controlled market price changes
Advantages	Parsimonious data requirements	Quality-controlled	Parsimonious data requirements
	Relatively uncomplicated		Quality-controlled
	Overcomes illiquidity		
Disadvantages	Not quality-controlled	Extensive data requirements	Sample selectivity concerns
	Affected by heterogeneity	Concerns regarding specification of hedonic function	Concerns regarding heteroskedasticity and multicollinearity

Mark and Goldberg (1984) acknowledge also that median prices are more robust to extreme observations than mean prices.[19] The ability of the median-price series to produce a periodic estimate of price levels or growth is not affected by the illiquidity of the residential real estate market, since it can be calculated for any period and region in which at least one sale has occurred.

Despite its seeming advantages in overcoming residential real estate market data issues, a major shortcoming in the application of the median-price series is its susceptibility to severe 'compositional bias' (Prasad and Richards, 2007).

Compositional bias refers to movements in an index that are the result of changes in the average quality of homes sold in a given period rather than movements in the true underlying index. That is, movements in the median-price series may simply reflect the quality (and hence, prices) of houses sold at a point in time, rather than actual price movements in the total housing stock. The effects of compositional bias are commonly assessed in the literature by the relative volatility of the alternative indices. Hansen (2006), for example, shows the median-price series to exhibit significantly higher volatility than regression-based constant-quality index measures such as the hedonic and repeat-sales methodologies.

The hedonic regression is a method that attempts to overcome the issue of compositional bias associated with median-price measures. The premise for this lies

---

[19] Mean price-based indices have similar properties to the median but, given the large variability of house prices and the mean's exposure to extreme observations, is a less preferred measure in residential real estate based studies.

in hedonic theory which suggests that the value of a composite good – such as a house – is the sum of its components (Griliches, 1971; Rosen, 1974). Thus, by decomposing the sample of houses into their various structural and location attributes, the differences in these attributes across houses can be controlled. The main criticisms of this method are the issues associated with its intensive data requirements and the most appropriate specification of the hedonic pricing model itself.

The repeat-sales regression method is a special form of the hedonic-regression approach in which the change in price between sales of property is modelled. First proposed by Bailey et al. (1963), this model was later refined by Case and Shiller (1987), among others, and now is the index used to underlie the CME listed housing derivatives market. By considering the change in price between periods of trade pairs of the same property, the issue of heterogeneity between assets is mitigated. That is, by comparing prices of the same property, homogeneity is created. As such there is no requirement for the detailed attributes data that exists in the estimation of hedonic indices.

The major impediment to repeat-sales based index accuracy is the likely sample selectivity and inefficiency in data treatment induced by restricting the sample to properties that have sold at least twice. Clapp and Giacotto (1992), for example, claim that these high-turnover properties are typically starter-homes, and consequently smaller and less expensive. Additionally, they suggest that repeat-sales samples may be subject to a 'lemons' phenomenon (Akerlof, 1970). An alternative view, proposed by Steele and Goy (1997), is that repeat-sale properties are more likely to have experienced above average growth. A further concern to the repeat-sales methodology is the potential for it to induce smoothing; that is, spuriously large positive serial correlation in returns and low volatility. Wright (2006) demonstrates this point, arguing that the requirement of a second sale to occur for an observation to be included in the index estimation results in an index that lags the true underlying market index.

It can be seen that none of the existing methods for estimating residential real estate price indices is completely without bias. This has severe implications for research derived from house price indices and for broader residential real estate market research that has adopted similar methodologies. An example of this is in research that has considered the effect of alternative sale methods on prices. While the use of these various methodologies is driven by the need to control for quality in a market of heterogeneous assets, the choice of one over another is typically determined by the availability of data. A fundamental objective of this dissertation is to investigate the impact of methodology choice on the findings of residential real estate market research. The next section reviews the literature to have examined efficiency and seasonality in residential real estate returns with emphasis given to the authors' choice of methodology. A later section presents a review of the previous research to have applied similar research in studies of residential real estate market trading and pricing.

2.2 Residential Real Estate Market Efficiency and Seasonality

As discussed in Section 2.1, the residential real estate is characterised by illiquidity, a lack of transparency, and high transaction costs. Consequently, this market is often categorised as being highly inefficient relative to more liquid financial markets (Roulac, 1976). The main argument against real estate market efficiency is that significant positive autocorrelation exists in returns; that is, an observed price movement tends to be followed in the subsequent period by a price movement in the same direction.

This argument against residential real estate market efficiency is further supported by the perceived existence of a distinct seasonality in returns. As synthetic real estate investment grows as a viable and liquid alternative to direct investment in the housing market, so too does the significance of predictable price patterns in the market, driven by autocorrelation and seasonality. This section reviews the literature that has explored the concepts of autocorrelation and seasonality in equity returns, and compares the results to similar empirical work that has been applied to the residential real estate market.

2.2.1 Do Returns to Residential Real Estate Follow a Random Walk?

An efficient market may be described as one in which “prices reflect information to the point where the marginal benefits of acting on information (the profits to be made) do not exceed the marginal costs” (Fama, 1991: 1575). In his seminal work on this topic, Fama (1965) outlines the three forms of market efficiency and the conditions required for each to hold. The three states of market efficiency, in ascending order of necessary requirements, are: weak, semi-strong, and strong form.

Weak-form efficiency describes a market in which prices fully reflect all public historical price information. The requirements of weak-form market efficiency must be met in order for semi-strong and strong-form market efficiency to hold. Consequently, testing of the efficient market hypothesis in its weakest form is a common starting point in market efficiency studies: if it is rejected then the market is not efficient in any state.

Semi-strong form market efficiency requires prices to fully reflect all public historical and current information. If semi-strong form market efficiency holds then weak-form efficiency is implied.

Strong-form market efficiency requires that *all* historical and current information, including privately held information, is reflected in prices. The presence of either deterministic autocorrelation or seasonality in an asset's prices or returns violates the efficient markets hypothesis developed by Fama (1965) in its weakest form.

The question of weak-form efficiency in equity and other financial markets has received a significant amount of academic research. In the Australian equities market, for example, conjecture as to its efficiency has continued from the earliest work by Groenewald and Kang (1993) to the contemporary study undertaken by Worthington and Higgs (2009). Only recently has similar research been conducted in the residential real estate market.

As discussed in Section 2.1.1, the residential real estate market has a number of characteristics which may hinder attempts to take advantage of existing informational inefficiencies. To summarise this situation,

> "[M]ost real estate markets are not subject to the same forces of arbitrage and to the rapid exchange and processing of information that characterises the stock market (and to a lesser extent the bond market), we might not be surprised to find that real estate portfolio returns are somewhat predictable from one period to the next" (Goetzmann and Ibbotson, 1990: 71).

In line with this reasoning, a body of literature has evolved to test for the presence of a random walk or predictability in residential real estate returns. The majority of research has rejected the hypothesis of housing market efficiency in its weakest form, instead finding support for persistence, or inertia, in returns. Table 2-4 summarises the key papers that have examined residential real estate market efficiency.

The earliest literature in the area of residential real estate market efficiency found support for the random walk hypothesis. Claiming his study as the "first rigorous testing of real estate market efficiency" (1984: 301), Gau (1984) constructs three price series from transactions of apartments in the City of Vancouver, Canada, for the period 1971 through 1980. Attempting to control for asset heterogeneity, Gau (1984) standardises the observed transactions by square footage, gross income, and number of rooms. A single observation from the geographical submarket with the highest transaction frequency is then selected at random per month to create the price series.

Table 2-4:

Real Estate Market Weak-Form Efficiency

This table presents the key academic papers that have tested the principle of weak-form efficiency in the various residential real estate markets. For each paper, the author and publication date are reported as well as a description of the market and period considered in the empirical analysis. An overview of the methodology used in the paper is provided and the main findings are summarised.

Author/Year	Data/Market	Methodology	Findings
Gau (1984)	Monthly observation of a randomly selected sale for Vancouver, Canada, 1971-1980	Autoregressive models to determine serial correlation. Box-Jenkins ARIMA model	Support for weak-form efficiency in returns
Rayburn, Devaney, Evans (1987)	Monthly standardised mean return to 10 submarkets within Memphis, USA, 1970-1984	Box-Jenkins ARIMA modelling	Support for long-horizon weak-form efficiency in returns
Case and Shiller (1989)	Quarterly repeat-sales indices for four cities in the USA, 1970-1986	Forecastability measured from autoregression between split-sample repeat-sales indices	Reject weak-form efficiency
Hosios and Pesando (1991)	Quarterly repeat-sales indices for Toronto, Canada, 1974-1989	Replicate the methodology of Case and Shiller (1989)	Reject weak-form efficiency
Clayton (1998)	Quarterly appraisal-based index for Vancouver, Canada apartment sales, 1982-1994	Excess return autocorrelation at quarterly and annual intervals	Support for weak-form efficiency in returns
Hill, Sirmans and Knight (1999)	Repeat-sales data underlying Case and Shiller (1989)	Test for homoskedasticity in the error term of the repeat-sales index regression model	Reject weak-form efficiency
Ito and Hirono (1993)	Annual total-return hedonic-imputation index for the Tokyo housing market 1981-1992	Autoregression of imputation index excess return, calculated over three alternative financial assets	Reject weak-form efficiency
Rosenthal (2006)	Quarterly hedonic index for 81 cities in the UK 1991-2001	Box-Jenkins ARIMA modelling	Support for weak-form efficiency

Gau (1984) estimates autocorrelation functions on first differences of his estimated price series, finding little evidence of serial correlation in two out of the three series; only the series standardised by price per number of rooms showed significant serial correlation, arising at the first, fourth, fifth, and ninth lags. Utilising any of the three standardised price series, however, in an autoregressive integrated moving average (ARIMA) model, Gau (1984) finds support for weak-form efficiency of the Vancouver apartment market.[20]

Rayburn, Devaney and Evans (1987) find further support for the existence of weak-form efficiency in residential real estate. Rather than use a single observation to form a price series as Gau (1984) undertook, the authors in this study estimated a monthly index for each of ten identified submarkets of Memphis, USA, covering the period 1970 to 1984, using a truncated and standardised mean transactions price. The authors employ similar tests to those used by Gau (1984). While finding evidence of weak serial correlation in seven submarkets, Rayburn et al. (1987) conclude that after taking account of transaction costs in the market, weak-form market efficiency holds.

While both Gau (1984) and Rayburn et al. (1987) acknowledged the effect of heterogeneity across residential real estate assets in their attempts to control for it using standardised non-parametric price measures, the subsequent research in this area has been dominated by regression-based constant-quality index estimation procedures.

[20] Gau (1984) also finds support for weak-form efficiency in the Vancouver commercial real estate market and, extending upon this work, Gau (1985) finds evidence to support the semi-strong form efficiency hypothesis in the Vancouver apartment market.

In particular, researchers have estimated housing market returns from various repeat-sales indices and, more recently, hedonic-regression techniques.

Case and Shiller (1989) rely on indices estimated from the weighted repeat-sales (WRS) house price index estimation methods, described in Section 2.1.4, that they pioneered. An issue with the use of this particular index estimation technique in autoregressive studies is the potential for spurious correlation to arise as a result of its endemic regressing of trade pairs. That is, the regression error for a property at the time period of its first sale will likely be correlated with the error value for the same property at the time of its second sale. In an autoregression on this data, the underlying error terms will be correlated among trade pairs, violating an assumption of the ordinary least squares estimation technique.

To overcome this issue, Case and Shiller (1989) use a 'split-sample' method. This involves partitioning the trade-pair data into two distinct subsamples, such that both sales of the same property are observed in only one subsample. Repeat-sales indices are then estimated from these subsamples using the weighted repeat-sales methodology, yielding two independent series. The serial regressions are fitted using returns to one subsample index as the dependent variable, and returns to the other as the independent variable. In this way, the same house will not be regressed upon itself, and there is no spurious correlation induced by the index method.

Case and Shiller (1989) reject the hypothesis of weak-form efficiency in their time-series analysis of returns. Interestingly, however, this is at odds with the very assumption regarding housing prices and returns that the same authors use to justify the weighted repeat-sales index for which they are best known (Case and Shiller, 1987; Hill, Sirmans and Knight, 1999). That is, the purpose of the weighting method used in WRS index methodology is to counter the supposed heteroskedasticity induced by the presence of a random walk, yet if returns do not follow this process, correcting for it will not improve the model (Greene, 2003).

Hosios and Pesando (1991) replicate the split-sample repeat-sales index forecastability model of Case and Shiller (1989) in the housing market of Toronto, Canada. Using data for the period 1974 through 1989, their results broadly support those of Case and Shiller (1989). Despite low explanatory power to the fitted models of annual real percentage change in the index, at a quarterly level the authors find significant persistence in returns.

Hill et al. (1999) develop an experiment for detecting the presence of a random walk component in house prices which they apply to the same data set used by Case and Shiller (1987). As a result of the high illiquidity in residential real estate markets, a useable time-series of transactions data does not exist for individual properties. The authors argue that this limits the suitability of traditional autoregressive tests for weak-form market efficiency in analyses of these markets.

Instead of using a market-wide price index, Hill et al. (1999) use raw transaction level data and consider the properties of the repeat-sales model's error term. Specifically, they propose testing for homoskedasticity in the normalised error term. By making the assumption of a first-order autoregressive component in the model, the authors determine that if evidence for homoskedasticity of the error term is found then the hypothesis of a random walk in house prices is supported.

Hill et al. (1999) show, through a series of Monte Carlo based experiments, this test to be consistent and robust. When applied to data from repeat-sale properties in the four cities first presented in Case and Shiller (1989), as well as data from the Baton Rouge area of Louisiana, Hill et al. (1999) conclude there is "strong evidence against the null hypothesis of a random walk in these four cities" (Hill et al., 1999: 101). The authors apply their findings to the index methodology literature, suggesting WRS measures may be improved by accounting not for a random walk, but for a stationary autoregressive component. This supports the earlier proposal of an autoregressive conditional heteroskedastic (ARCH) component to house price returns put forward in Mahieu and van Bussel (1996).

Clayton (1998) is one of the few more recent studies to find support for weak-form efficiency in residential real estate returns. Using a quarterly appraisal-based index for apartment sales in Vancouver, Canada, for the period 1982 to 1994, Clayton (1998) performs autocorrelation tests of excess returns at quarterly and annual intervals. His results detect no significant autocorrelation at the quarterly level and, interestingly, significant *negative* autocorrelation when testing at the annual level.

The limited research that has been conducted into the weak-form efficiency of returns to the residential real estate market outside of North America also yields conflicting results. For example, Ito and Hirono (1993) reject weak-form efficiency in their case study of the Tokyo housing market. Using an annual total return index for the period

1981 to 1992 constructed by a hedonic imputation method, significant coefficients are estimated in an autoregression of excess returns over their mean.

By contrast, Rosenthal (2006) states that there is little support for the argument "that the housing market shows gross and clearly exploitable informational (weak) efficiency" (2006: 309). This conclusion comes from the modelling of the autoregressive lag structure of returns to residential real estate in the UK over the period 1991 to 2001. Specifically, Rosenthal (2006) considers the returns to monthly hedonic indices for 81 cities across the UK, finding that beyond the most recent two to three months there is little predictive power in past returns. Once transaction costs and time delays are considered, these findings support weak-form efficiency since no exploitable trading opportunities can exist.

Weak-form market efficiency in the sense of autocorrelated returns represents just one component of the efficient market paradigm. For a market to be considered weak-form efficient there can be no value from the analysis of past returns in the prediction of future returns. Systematic abnormal returns to given events or times, commonly referred to as 'seasonality', represents a violation of weak-form efficiency. The next section discusses this phenomenon in detail.

2.2.2 Seasonality in Residential Real Estate Prices

In addition to serial correlation and persistence in residential real estate market returns, a further inefficiency regularly espoused by market agents and commentators is that there are better times to sell one's house through the year as a result of housing market price seasonality.[21] The empirical research that has been conducted in the area of housing market price seasonality, however, actually suggests that on average there is no month in which returns to residential real estate are consistently above expectations.

Rozeff and Kinney's (1976) paper is one of the earliest studies to identify deterministic monthly seasonality in equities. Using monthly returns on the New York Stock Exchange for the period January 1904 through December 1974 the authors show, using regression analysis, that the mean stock index return over the month of January is significantly higher than any other month, all else being equal.

This finding supported the earlier results made in the Australian market by Officer (1975) and Praetz (1973). Further evidence for this turn-of-the-year effect or, more commonly, the 'January effect,' has also been found in European and Asian stock markets. Gultekin and Gultekin (1983) examine stock market returns in 16 industrial countries, finding a January effect in all 16 and the statistical significance of the result being strong in 15. Aggarwal and Rivoli (1989) look at the four East Asian stock markets of Hong Kong, Singapore, Malaysia and the Philippines and find a significant January effect in all but the last. The effect is also apparent in other financial markets, including bonds (Schneeweis and Woolridge 1979) and futures (Gay and Kim, 1987).

Given the relative structural inefficiencies of the residential real estate market, it may be expected that it too would demonstrate some form of seasonality. A brief review of

[21] 'Don't Let Housing's Seasons Scare You,' *Business Week*, November 9, 2005.

the opinions from real estate agents and market commentators – often relying on average and median-price patterns – would support this expectation, with offerings of advice as to the best times to buy and sell your home. Little academic support exists, however, for this expectation; the research to have applied analogous research from other financial markets to residential real estate markets has been unable to confidently detect seasonality in returns.

The main academic articles to have examined residential real estate seasonality are summarised in Table 2-5. The earliest such work was conducted by Case and Shiller (1989) in their examination of the wider topic of real estate market efficiency described in the previous section.

Case and Shiller (1989) test for seasonality in the housing markets of the four cities of Atlanta, Boston, Chicago, and Dallas in the USA. Using indices of house prices estimated from their WRS methodology, the authors don't detect any definitive seasonal pattern. Using an F test for equality in average index returns per quarter, Case and Shiller (1989) reject the hypothesis of equality in returns to a given quarter only in Chicago at the 2% level of significance: the results suggest a positive return to Chicago housing in the first quarter of the year, and a negative average return in the last quarter. This finding is supported by the results of a regression of change in index on lagged index values and seasonal dummy variables.

Table 2-5:

Seasonality in Residential Real Estate Prices

This table presents the key academic papers that have tested for seasonality in residential real estate prices. For each, the author/s and publication date are reported with a brief description of the market, period and methodology underlying the house price index used in the empirical analysis. The methodology used to test for residential real estate return seasonality is provided and the key results are provided.

Paper/Year	Data./Market	Methodology	Results
Case and Shiller (1989)	Quarterly repeat-sales indices for four cities in the USA, 1970-1986	*F* test for equality; Seasonal dummy-variable regression	One city, Chicago, demonstrates positive (negative) average return in the first (fourth) quarter
Kuo (1996)	Case and Shiller (1989) indices	Replicate and extend partitioned sample methodology of Case and Shiller (1989)	Weak evidence to support Case and Shiller (1989) findings of seasonality in Chicago
Costello (2001)	Quarterly hedonic index for Perth, 1988-1995	Analysis of Variance (ANOVA) and *t* test analysis	No significant evidence supporting seasonality in returns
Rossini (2002, 2000)	Quarterly stratified index for Adelaide, 1984-2001	Seasonal dummy-variable regression	No evidence in support of quarterly seasonality in Adelaide housing market
Rosenthal (2006)	Quarterly hedonic index for 81 cities in the UK, 1991-2001	Seasonal dummy-variable regression	No robust evidence in support of seasonality in various UK housing markets

Hosios and Pesando (1991), in a brief examination of seasonality in the Toronto housing market, detect a low level of seasonality in a similar pattern to Case and Shiller's (1989) Chicago result. They attribute these findings to the colder climates of these areas.

Kuo (1996), using the same quarterly index data as Case and Shiller (1989), replicates their partitioned sample regression method, and compares the results to those obtained from a Bayesian model that treats the partitioned indices as random variables, since, depending on how the data are partitioned, the results are random variables in themselves. His results show weak evidence of a superior month in Chicago using both nominal and inflation-adjusted returns in the Bayesian model, and inconclusive support for seasonality in the other cities considered.

Rosenthal (2006) in the UK, and Rossini (2000, 2002) and Costello (2001) in Adelaide and Perth, respectively, use alternatives to the repeat-sales method of Case and Shiller (1987) to estimate house price indices from which they test for market seasonality. Costello (2001) and Rosenthal (2006) use hedonic index estimation techniques. Hedonic methods use regression analysis to statistically control for differences in the quality and attributes of the traded housing stock, extracting constant-quality estimates of market-wide price changes. By contrast, Rossini (2000, 2002) uses a stratified index estimation methodology which measures median-price movements from partitions of the sales sample in an attempt to overcome heterogeneity bias.

Rosenthal (2006) examines the issue of house price seasonality in the UK market over the 11-year period 1991 to 2001. He estimates monthly hedonic price indices for 81 cities across England, Wales and Scotland. The returns to these indices are then regressed on quarterly dummies to test for differences in the average return across the year. Rosenthal's (2006) results do not consistently or robustly find evidence to reject the hypothesis of quarterly equality in constant-quality prices.

Costello (2001) performs seasonality tests on transaction volume and returns for the Perth housing market over the period 1988 to 1995. Given the relatively short sample period of his study, Costello (2001) 'stacks' the indices of the 13 regional subdivisions he identifies for the Perth market in order to increase the robustness of his tests. Using ANOVA and *t* test analysis, he finds that while there is a definite peak and trough in market activity in the first and fourth quarters of the year, respectively, there is no statistically significant evidence of such seasonality in prices.

Using stratification methods, Rossini (2002) finds little evidence of seasonality in the Adelaide housing market between January 1984 and September 2001. The process of stratification is a non-parametric alternative to constructing quality-adjusted price indices, by which similar properties are grouped into strata and market-wide movements are estimated as some average of the price growth for the individual strata. Rossini (2002) groups properties into strata based on the number of rooms, building age, and property type, and compares the results to those derived from a cluster-analysis that grouped properties by their suburb.

Rossini's (2002) results show no significant difference between quarters in index growth for estimates derived from either stratification method. This builds on the work by Rossini (2000). In this earlier paper, the author similarly failed to find support for seasonality in residential real estate prices for a smaller sample of suburbs in Adelaide using time-series analysis and a cross-sectional regression model.

Rossini (2002, 2000) and Costello (2001) represent the only research to consider Australian residential real estate market efficiency. The external validity of these results is limited by the small segment of the market that Adelaide and Perth represent; approximately 8% and 12% of the Australian market, respectively.[22] Further research using the larger and more liquid Sydney and Melbourne markets is required to determine the efficiency of Australian residential real estate.

Of more fundamental concern to the conclusions of past research into residential real estate market efficiency is the effect that the choice of index methodology may have on the results. As discussed in Section 2.1.4, mean- and median-price based indices typically display excess volatility in returns. Repeat-sales indices, on the other hand, are shown in the literature to exhibit a smoothing bias. Do such biases in the alternative index methodologies explain the divergent results of past research? A comparison of the results of market efficiency tests from the different indices has not previously been undertaken. Such research would provide insight into the validity of the conclusions made in previous research and add to the overall understanding of efficiency in the residential real estate market.

---

[22] Source: RPX.

2.3 Pricing Residential Real Estate

This section examines pricing bias and trading opportunities in transactions of residential real estate. Whereas the previous section dealt with efficiency at a market-wide level, this section considers three different approaches to gaining investment exposure to the housing market. The first two subsections deal with direct investment strategies, concerning the resale, or 'secondary', market design and the market for 'new' properties, respectively. The third section is concerned with information transfer in a hypothetical synthetic market.

2.3.1 Secondary Market Design

This section discusses the design of the secondary market for trading residential real estate in Australia and reviews the international theoretical and empirical literature that has examined the effect of market design on trading and prices.

As discussed in Section 2.1.3, there are three main ways in which an investor can gain exposure to residential real estate assets: (i) direct investment; (ii) indirect, or securitised, investment; or (iii) via a synthetic position using derivative instruments.

In the context of direct investment, the two ways by which the vast majority of real estate is transacted in Australia is by private treaty and by open auction. The English auction proceeds as "an ascending, sequential-bid auction in which bidders observe the bids of others, and decide whether or not to increase the bid" (Baye, 2009: 457). Optimal behaviour in the English auction for the bidder is to determine the maximum they would be prepared to pay for the given good, and bid that amount.

The other well-known auction formats are the Dutch, first-sealed and second-sealed auctions. The Dutch auction commences with a high bid-price, which is sequentially lowered by the auctioneer until a buyer accepts. Under a first-sealed auction, potential buyers place their bid privately and/or simultaneously. The successful bidder is the highest bidder who pays their submitted price. A second-sealed auction works similarly to the first-sealed format, differing in that the successful bidder pays the second-highest bid price. These auction systems, however, are rarely applied to transactions of residential real estate in Australia.

Auctions currently account for close to 10% of all sales in the Australian housing market.[23] However, in Melbourne it has been estimated that as many as 75% of property sales occur via auctions (Reed, Robinson and Williams, 2002). Maher (1989) posits that the reason there are so many more in Melbourne than anywhere else - both in Australia and in other countries – is a factor of the real estate industry there itself, which often puts forward claims that auctions are the best measure for establishing 'true' value and obtaining a decisive result: "That the [Melbourne] real estate industry is actively promoting auctioning is beyond doubt. That they are succeeding is also evident" (Maher, 1989: 504).

Recently, the auction sales mechanism has become more commonplace in other Australian cities as well as several other countries. Growth in the use of auctions has occurred during periods of high capital appreciation and market activity (see, for

---

[23] Source: RPX.

example, Susilawati and Lin, 2006). Given this increase in the use of auctions as a viable residential real estate transaction method, research into the price effects of the alternative sale mechanisms – namely, private treaty sales and auctions – is of fundamental importance to investors, academics and regulators.

The majority of the literature into this area has observed a significant impact on residential real estate sales price arising from the use of the auction market mechanism, violating the economic principle of the law of one price. This finding potentially has broader implications for the market design of other financial assets, including equities and derivatives. Some researchers have found a price discount for auction sales (Allen and Swisher, 2000; and Mayer, 1998, in the USA) based on the argument that they are typically disposal sales following defaults on financing. Other research, however, such as Dotzour et al. (1998) in New Zealand, Lusht (1996) and Newell et al. (1993) in Australia and Quan (2002) for vacant land in the USA, has identified a significant price premium for auction sales.

Three distinct strands of theory have evolved in the academic literature to predict the effect of the auction sale method on asset prices: (1) there is a systematic difference in prices and use of the auction mechanism achieves a premium; (2) there is a systematic difference in prices and use of the auction mechanism achieves a discount; and (3) there is no systematic difference between the prices of properties sold by auction and private treaty. These theories are summarised in Table 2-6.

Table 2-6:

Theoretical Auction Pricing Literature

This table presents the key theoretical arguments for the three auction pricing possibilities. The alternative economic concepts are listed, as are the key academic papers, with a summarised explanation of the theory as applied to the auction sales mechanism. Panels A and B present the concepts justifying the existence of an auction premium and discount, respectively. Panel C summarises the auction pricing indifference theories.

Concept	Theory	Key Papers
Panel A: Auction premium		
Monopolistic seller	Drawn from economic theory, if a vendor is selling a unique property, there is asymmetric control on pricing, in the vendor's favour	Milgrom (1987); Riley and Samuelson (1981)
Vendor bargaining power	The right of refusal lies with the vendor; if a price does not meet their (potentially inflated) expectation the sale does not proceed	Milgrom (1987)
Winner's curse / Loss aversion	In order to win the competitive bidding, the buyer must bid more than the other bidders, which can exceed rational value; to avoid the 'pain' of missing out, buyers overbid to secure their purchase	Milgrom (1989), French and McCormick (1984)
Panel B: Auction discount		
Liquidity value	If the vendor seeks a timely sale they are likely to sell for less and use the relatively quicker auction mechanism; the value of a timely sale increases with the presence of holding costs	Mayer (1998); Mayer (1995)
Market mismatch	If buyers spend a shorter amount of time searching by buying at auction, the house they buy is unlikely to be the best match to their preferences, and so pay less as a compromise.	Mayer (1998)
Panel C: Auction indifference		
Auction equivalence	Mathematically, the prices achieved under different auction settings are shown to be equivalent	Vickrey (1961)
Market efficiency	If the anomaly of an auction premium/discount is known, then this information should fully be incorporated into price expectations; buyers will not attend auctions where there is a perceived premium, and the mechanism collapses, and vice versa	Fama (1970)

Theories justifying the presence of an auction price premium are related to the imbalance of rights and information, in favour of the vendor (Milgrom, 1987; Milgrom, 1989; French and McCormick, 1984). On the other hand, it is the imbalance in available time separating potentials buyers from vendors that underlies the main theories supporting an auction discount (Mayer, 1995; Mayer 1998). The work by Vickrey (1961) that demonstrates the equivalence of prices under alternative auction formats, taken with the viewpoint that private treaty sales are a slow-moving Dutch auction in themselves (Adams et al., 1992), and the requirements of Fama's (1970) market efficiency, suggests that prices should not differ under various sale methods, all else being equal.

Several approaches have been taken in the academic literature to empirically test the effect of the auction sale mechanism on price in property markets. The key studies, their methodology, data, and results are summarised in Table 2-7. The two main empirical approaches taken to assess the price impact of the auction mechanism on sales of residential real estate have been: (1) comparisons of prices achieved at auction with private treaty sales results or imputed property values; and (2) hedonic-regression analysis.

In one of the earliest papers to compare prices for residential properties at auction with those arrived at by private treaty, Newell et al. (1993) observe a 3.6% price premium to auctions in Sydney. In their paper, Newell et al. (1993) rely on the median price of sales achieved by the two methods. While conceptually uncomplicated and undemanding in its data requirements, such an approach is susceptible to heterogeneity bias driven by sample selectivity (discussed later in this section).

Table 2-7:

Auctions and Prices

This table presents the key academic papers to have empirically tested for a pricing bias in sales of residential real estate attributable to the auction sales mechanism. The authors and publication of year are given, with a summarised description of the methodology and data used in the study. The results are presented indicating whether the auction sales mechanism resulted in a price premium or discount in the given study

Authors/Year	Methodology	Data	Results
Ashenfelter and Genesove (1992)	Post-auction price behaviour	New Jersey, USA, 1990, condominium sales	13% premium
Gau and Quan (1992)	Hedonic regression	Austin, USA, vacant residential land, 1991	40% discount
Newell et al. (1993)	Median values	Sydney, Australia, 1992	3.6% premium
Lusht (1996)	Hedonic regression	Melbourne, Australia, 1988-1989, detached house sales	8% premium
Mayer (1998)	Repeat-sales; trade-pair comparison	Dallas and Los Angeles, USA, 1970-1991	Discount
Dotzour et al. (1998)	Hedonic regression	Christchurch, New Zealand, 1991-1992, detached house sales	No significant premium in lower priced areas; however, a 5.9-9.5% premium in high-priced areas
Allen and Swisher (2000)	Assessed value comparison	Florida, USA, repossessed properties, 1998	13-21% discount
Stevenson and Young (2004); Gurdgiev et al. (2010)	Assessed value comparison	Dublin, 1997-2001	24% premium
Ong (2006)	Post-auction price behaviour	Singapore, 1995-2000	Premium to auctions over failed auctions subsequently sold by private treaty

To overcome such potential bias, later papers that have examined whether an auction premium (or discount) exists in housing markets have compared auction results with the assessed values of those same properties. The use of assessed values, also widely referred to as appraisals or valuations, of the same properties that sell using the auction mechanism minimises the impact of any sample selectivity.[24]

Allen and Swisher (2000) utilise this more robust methodology. Using auction results and independent valuations for 170 properties that had been acquired by the mortgage provider following loan foreclosure in Florida in 1998, the authors estimate a statistically significant discount of between 13% and 21.5% attributable to the auction mechanism. This result, however, must be accepted with some caution: the vendors in this sample were highly motivated and seeking liquidity.

Stevenson and Young (2004), in their analysis of auction results in the Greater Dublin housing market over the period 1997 to 2001, present a more typical empirical study of the auction mechanism. The authors compare both auction and private treaty sales results with their respective valuations. While, on average, both auction and private treaty sales are found to sell at a premium, this result is only statistically significant for auction properties, for which Stevenson and Young (2004) estimate a premium of approximately 24% over their valuations. These findings are reaffirmed for the Dublin market in Gurdgiev, Stevenson and Young (2006).

[24] This approach is similar to the methodologies undertaken in studies of residential real estate valuation accuracy. See, for example, Matysiak and Wang (1995).

An alternative to the use of valuations in auction price comparisons is the pricing of failed auctions (auctions in which the bid does not reach the reservation price) in the post-auction market. Using a unique sample of auctions which 'fell through' and were subsequently sold at a privately negotiated price, Ashenfelter and Genesove (1992) observe a 13% auction premium for condominiums in the USA state of New Jersey for the month of April in 1990. Ong (2006) also observes an auction-driven price premium, making use of a detailed sale, bid and property attributes dataset for auctions in the Singapore housing market over the period 1995 to 2000.

Hedonic regression is the main alternative methodology used to directly compare prices achieved by auction with those from private treaty sales. Hedonic regression attempts to control for the various attributes and locational characteristics of a property. This method thereby statistically accounts for heterogeneity between properties in order to derive an unbiased estimate of the effect of sale method on price, thus avoiding the biasing effect of endogeneity in the choice of sale method and subsequent sample selectivity. Several studies have empirically tested for a systematic price bias induced by the sale method this way with divergent results.

Dotzour et al. (1998) consider the dynamics of the Christchurch, New Zealand, housing market. The authors estimate a price premium to relatively high-priced areas of unique properties of between 5.9% and 9.5% using a hedonic regression. In areas of relatively homogenous properties; however, prices are statistically insignificantly different between those properties that sell at auction and those sold using the private treaty method. This finding is broadly supported by Lusht (1996) who estimates an 8% price premium in the Melbourne housing market during the period 1988 to 1989.

Evidence for an auction discount, however, has also been supported in hedonic studies. For example, Gau and Quan (1992) apply hedonic-regression methodology to vacant land sales in Austin, USA, during 1991, estimating a 40% price discount attributable to use of the auction sales mechanism.

These breakdowns in the law of one price are not unique to the use of the sequential English ascending open-bid auction method in transactions of residential property. A price discount to the auction mechanism is reported in several papers that have considered the Taiwan market, where first-price sealed bids have been used in residential real estate transactions (Lin, Tsai and Chang, 1997; Lin and Huang, 2005). On the other hand, an auction premium is observed for first-price sealed bids in the Lagos, Nigeria, residential land market (Amidu and Agboola, 2009).

A common issue across the alternative methodologies in this line of research is the potential for sample selectivity, or endogeneity, bias in the results. That is, that certain types of properties with certain characteristics, related, positively or negatively, to price, are more likely to sell by a given means.

In fact, most researchers acknowledge the uneven quality of properties sold by the auction mechanism. It is generally argued, based on anecdotal evidence, that auctions "in the USA involve foreclosures of bankruptcies, especially at the lower end of the market" (Stevenson and Young, 2004: 47), while in markets such as Australia and New Zealand and much of Europe, auctions "are more likely to be used in booming markets and for desirable properties" (Mayer, 1998:42).

Dotzour et al. (1998) demonstrate empirically that certain hedonic qualities determine whether a property is more or less likely to be taken to auction. Using probit analysis, their results suggest that vendors of properties which are more unique in their features or located in more desirable areas are more likely to attempt to sell by the auction mechanism than by private treaty.

Consequently, any research attempting to isolate the price effect of the auction sale mechanism needs to completely account for the different attributes of properties. When certain attributes that are correlated with the auction decision are not included in a hedonic regression, the results are prone to omitted variable bias (Greene, 2003).

A useful econometric technique to account for sample selectivity is the Heckman two-stage procedure. Developed by Heckman (1978, 1978) to counter the expected omitted variable bias arising from sample selectivity, the Heckman procedure involves the estimation of a sample selectivity variable from a discrete choice model, such as a probit regression. This sample selectivity variable, designed to reflect unobserved variables that are influencing the choice, is then included in the hedonic price regression. Dotzour et al. (1998) employ a two-stage Heckman procedure. Despite the strong indications of sample selectivity achieved by their probit model, the authors find no sample selectivity in their adjusted regression model.

The effect of omitted variable bias is examined empirically in Mayer (1998), where the auction-price effect is tested using a hedonic regression and a repeat-sales based method. Mayer (1998) argues that by considering the same property through time, this method avoids the issue of unobserved quality and demanding data requirements of previous research that has relied on hedonic regressions.

Mayer's (1998) data consists of condominium sales in Los Angeles for the period 1970 to 1991 and both condominium and house sales in Dallas covering 1979 to 1991. It is found that the auction mechanism has no statistically significant price effect in the Los Angeles market, where the attributes between the two subsamples are relatively constant. There is a much higher variation in the quality of properties between the two subsamples in Dallas, however, which presents some interesting results. From the repeat-sales regression, Mayer (1998) estimates a discount of 24% to auctions in Dallas, justifying this result as a reflection of the prevailing 'bust' market conditions. This discount is estimated at 37% using the hedonic method, significantly larger than the repeat-sales based estimate, and biased in line with the less desirable attributes of auctioned properties in this market.

As discussed in Section 2.1.4, a major limitation to traditional repeat-sales measures is that depreciation is not accounted for. Mayer (1998) acknowledges this, specifying a dummy variable for 'newness' in his model. While not presenting the fitted coefficient for this variable, property age, and specifically whether it is new, appear to be important factors in hedonic modelling of residential real estate prices. The price of properties at their first sale is the focus of the next section.

2.3.2 The Primary Market for Real Estate

Having explored the secondary market sales mechanisms available for trading residential real estate assets in the previous section, the primary market for housing is now considered. This section reviews the theoretical and empirical literature related to the pricing and performance of equity initial public offerings in order to develop hypotheses for the price behaviour of residential real estate assets at their first sale.

The issue of fair asset prices is essential to the neoclassical financial theory. While it is accepted that over- and underpricing can occur, the core of the theory states that they cannot persist; market forces will eliminate mispricing through arbitrage. However, a significant and persistent positive return to stocks when they are first publicly traded is well-documented in the equity IPO research.

Underpricing – the pervasive pattern of IPOs closing at a significantly higher price than their offer price – presents a puzzle[25] to the efficient markets hypothesis prediction that prices reflect past information, including observable patterns (Fama, 1970).

It is a well-documented phenomenon in the equities market, and legitimately so, with $27 billion 'left on the table' in IPOs[26] over the period 1990-1998 in the USA alone (Loughran and Ritter, 2002). At the extreme, VA Linux posted a 698% first-day return in December 1999 (Bodie, Kane and Marcus, 2002).

The dominant theory that has evolved to explain the presence of IPO underpricing is based on asymmetric information between different pairings of market participants: the underwriter and the issuer, the underwriter and investors, and uninformed and informed investors. Other theories that have been proposed as explanations for equity IPO underpricing include post-market liquidity risk (Booth and Chua 1996),

---

[25] As part of a two-part puzzle involving positive initial returns, and underperforming long-run returns (Copeland, Weston and Shastri, 2005).

[26] Money left on the table is defined as "the number of shares offered multiplied by the first day capital gain, measured from the offer price to the closing price" (Ritter, 2003: 427).

decreased monitoring (Brennan and Franks, 1997) and behavioural biases. Most empirical research finds these to be significant factors in determining IPO underpricing, but struggle to explain the existence of the underpricing phenomenon in isolation.

Baron and Holmström (1980) note that an investment bank acting as the IPO underwriter on behalf of the issuer has the opportunity to obtain private information from potential investors (through preselling activities, for example). This creates an asymmetry in information between the issuer and underwriter. Unlike earlier IPO underpricing models, which assumed symmetric information to exist between the two parties (Mandelker and Raviv, 1977; and Baron, 1979), this model requires a level of asymmetric information when investment banks act as underwriters.

Under this assumption that asymmetric information between the issuer and underwriter exists prior to contract, Baron (1982) develops a model that explains underpricing as a result of this asymmetry: issuers rationally delegate the offer price setting to the better-informed underwriter. This underwriter, facing a moral hazard issue, sets prices sub-optimally to those that would prevail in the absence of asymmetries. The less uncertainty on the part of the issuer regarding the issue's demand reduces the level of this delegation, and hence the underpricing. This is consistent with the empirical findings of Ibbotson (1975) and his contemporaries.

A further implication of Baron's (1982) proposition is that underpricing reduces marketing costs of an underwriter, and will lead investors to engage in rent-seeking behaviour (Loughran and Ritter, 2002).

Muscarella and Vetsuypens (1989) empirically test the 'Baron model' in analysing the performance of 'self-marketed' IPOs – the situation where investment banks issue their own securities. This approach, of course, totally eliminates any information asymmetry between issuer and underwriter. They find that significant underpricing persists, and conclude that Baron's asymmetry alone cannot explain underpricing in the equity IPO market.

Alternatively, Beatty and Ritter (1986) explain underpricing as the consequence of ex-ante price uncertainty and information asymmetry between potential investors and the underwriter. That is, when investors submit purchase offers, there is a level of uncertainty about what the actual post-market price will be. Critical to this model is the fact that some issues are overpriced. Because of the institutional setting of most IPOs – that once the offer price is set, excess demand is managed by rationing – investors are more likely to receive shares in an undersubscribed, and therefore probably overpriced issue; a phenomenon often described as the winner's curse.

In the housing market, first offer underpricing is implicit since the offer price is typically not made public. Consequently, the anomaly observed is of a price premium relative to the property's 'fair value,' estimated as a function of its characteristics and location. The information asymmetry arguments of Baron offer an explanation for why a price premium would be observed at first sale in the residential real estate market. In the real estate market the developer of the property takes on the role of the underwriter and the potential buyers are analogous to the post-listing equity investors.

Chang and Ward (1993) provide one of the few academic insights into the primary market pricing of physical residential real estate assets. In their paper, Chang and

Ward (1993) consider the pre-sales market[27] – where properties are sold by developers before their construction has commenced – of Taipei, Taiwan, over the period 1988 to 1990. They demonstrate theoretically that, in response to uncertainty regarding the building quality (and even its actual completion) investors in the pre-sales market should expect to earn a premium for the risk they bear. This is supported in the empirical statistics reported by Chang and Ward (1993), which show that the average price of pre-sale properties is higher than the average price of existing residential properties, by a factor of between 25% and 35%.

A complete gap exists in the literature as to the more general pricing behaviour of residential real estate assets at their first sale. This presents an interesting line of research for three reasons. Firstly, investment in residential real estate outside of the primary place of residence (that is, investment properties) has experienced sustained growth in the Australian market, and developers have responded with speciality investment property offerings and pre-sale offerings similar to the market design in Chang and Ward's (1993) study.[28] The behaviour of property prices at their first sale would need to be taken into consideration by these investors and developers alike. Secondly, if a component of price and subsequent capital growth is directly related to the first offering of the property, then mortgage providers should consider such factors in their loan criteria. Finally, the presence of a first-sale price effect may influence the decision to include such data in the estimation of residential real estate indices for research objectives and the purpose of underlying housing derivatives markets, such

[27] Pre-sales are a method for developers to finance their constructions while hedging out a level of associated risk, and are commonly used in Taiwan, Korea, and China. The analogy in the Australian market is 'off the plan' sales.

[28] *Housing Finance Australia*, ABS, Catalogue 5609, January 2010.

as have been listed on the CME. An abnormal proportion of first-sale observations for a given month being included in the estimation of a house price index, for example, could bias the observed index return for that month if there is a bias in first prices. The inclusion of the full set of hedonic attributes – using the hedonic-regression technique, for example – would not prevent this bias from occurring.

Residential real estate derivative markets and the indices that underlie their products are the topic of the next section. Specifically, the market structure and the flow of information in the emerging housing derivatives markets are considered.

2.3.3 Residential Real Estate Derivatives Markets

Residential real estate is one of the largest investable asset classes. As reported in Section 2.1.1, its capitalisation in Australia alone is estimated at over \$3.4 trillion compared to a total capitalisation of \$1.3 trillion for listed equity. Yet despite these figures, timely dissemination of information and understanding of this particular asset market is lacking.

In Australia, for example, a lag of up to eight weeks can exist between the agreement of sales price at the exchange of contracts and the addition of this information to a centralised database such as the VG, which is itself organised separately in each state.

As such, the estimation and reporting of any house price indices based on observed sales are delayed. This includes indices underlying residential real estate derivatives, where the delay in reporting ranges from a minimum of four weeks to around three months. Consider the S&P/CSI which underlie the CME's housing derivative products. Values of the CSI are published and settled against on the last Tuesday of the month with a two-month lag; the September 2009 index values, for example, will be released on October 27, 2009.

The RBA, itself acknowledging the importance of housing markets to the economy and the potential of housing derivatives markets, has criticised the 'real-time problem' of lags in residential property sales reporting. Outlining the trade-off between prompt index estimation and insufficient or biased data samples (leading to revisions as the population of sales data becomes available), the RBA concludes that "the existing measures [of house price index estimation] would benefit from more timely availability of source data on housing transactions." [29]

This is not to say, however, that no current information reflecting the trend of house prices exists. Firstly, the weekend newspapers often report a sample of the past week's sales and auction results. Another source of up-to-date information is the set of listings of properties for sale. List prices can be thought of as a non-binding ask price; they are a potential seller's first offer to the market and represent the first bid in a slow-moving Dutch auction. As such, the population of sales observations and final prices will reasonably be some function of the sample of properties advertised for sale. Finally, the real estate agents themselves know some sample of the successful sales which could be sourced. Figure 2-2 illustrates the flow of information in a residential real estate market with related housing derivative products.

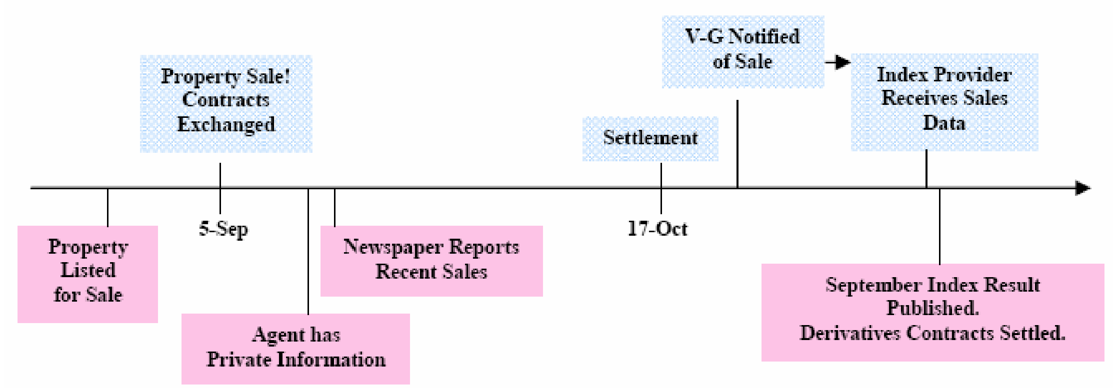
---

[29] 'Measuring House Prices: An Update', *RBA Bulletin*, June 2006.

Figure 2-2:

Information Flow in Sales of Residential Property

Figure 2-2 presents a stylised example of the chronological flow of residential real estate sales information relative to the timing of a market-wide house price index publication and derivatives contract settlement. Market events, such as the property sale and settlement date, are indicated above the timeline, while information publication dates are presented beneath the timeline.



Previous research in the area of residential real estate price prediction has attempted to identify the relationship between housing prices and macroeconomic fundamentals, such as interest rates, GDP, and employment (Scott, 1990; Himmelberg et al., 2005). The relationships between such variables, however, are typically complex, long-term and interdependent (see Section 2.1.2).

Given the illiquidity of the market, little has been done to examine whether the high level of informational inefficiency in the underlying market can be exploited in the presence of relatively liquid derivatives markets.

A second, more subtle, question embedded in such research explores the value of public and private information in this emerging market: whereas the newspaper sales and online listings are publicly available data, a sample of agents' sales information represents private information. While the value of information and impact of information asymmetry is well documented in markets for equities and their derivatives, a void exists in the literature that considers the value of such information in the residential real estate market. Such a question is well worth asking as the interest in housing as an investment class continues to grow.

2.4 Summary

This chapter has presented a thorough review of several important areas of academic research in the residential real estate market which can broadly be considered as: (1) market efficiency and seasonality; and (2) market design and information asymmetry. The development of a set of hypotheses from this literature review and the empirical testing of these hypotheses is the subject of the following chapters in this dissertation.

To establish this research within the broader finance literature, this chapter first presented a review of the literature that has considered: (1) the role of residential real estate prices and investment in the wider macroeconomy; (2) optimal institutional portfolio decisions in the presence of feasible investments in residential real estate assets; and (3) the methods that exist for assessing the price dynamics of this market, and the inherent limitations to traditional measures.

In the context of residential real estate market efficiency, detailed in Section 2.2, the literature has specifically considered the market's weak-form efficiency. That is, the ability of past patterns in prices and returns to predict future movements, an area with little clear-cut consensus.

The structures of the secondary, primary, and derivative residential real estate markets are considered in Sections 2.3.1, 2.3.2, and 2.3.3, respectively. With respect to the secondary market's trading apparatus, this literature review has primarily considered the effect of the auction mechanism on prices.

In considering the primary market for housing assets – that is, properties at their first sale – and the potential design of a housing derivatives market, this dissertation is concerned with informational asymmetries and dissemination efficiency.

The common feature of these strands of residential real estate market research is the significance that the methodology and data used can have on the empirical results. The accurate and unbiased research of such areas is of significant and growing importance.

3. Hypotheses Development

This chapter develops a set of testable hypotheses based on the literature review presented in Chapter 2. These hypotheses are designed to assess the impact that the choice of empirical methodology may have on the conclusions of research conducted into residential real estate prices. Specifically, the hypotheses seek to determine the significance the choice of methods – typically restricted by data limitations – used in previous studies of residential real estate market price dynamics may have in influencing the results

These hypotheses relate to four aspects of residential real estate market pricing bias and efficiency: (1) efficiency and seasonality in returns; (2) the auction mechanism; (3) information asymmetry at the first sale of a residential properties; and (4) information efficiency in housing markets.

3.1 Residential Real Estate Market Efficiency

The review of the empirical literature which tests for efficiency in residential real estate returns presented in Section 2.2 broadly indicates a rejection of the weak-form efficiency hypothesis. Studies such as Case and Shiller (1989) and Hill et al. (1999) demonstrate the existence of predictable patterns of persistence, or inertia, in returns to housing. Several papers in the research, however, report results which conflict with this finding. The findings of a more recent study by Rosenthal (2006) suggest that the housing market may be considered weak-form efficient, supporting the conclusions of earlier research by Gau (1984) and Rayburn et al. (1987).

This thesis begins by testing the weak-form efficiency hypothesis in the Australian residential real estate market. The issue of residential real estate market efficiency is of significant practical importance to investors and regulators. A highly inefficient market may discourage trade, further reducing liquidity. Given the close relationship between housing markets and the macroeconomy discussed in Section 2.1.2, particularly through lending activities and government revenue, the implications of housing market efficiency are broad and likely to impact upon other financial markets. The emergence of housing derivative products, such as those listed on the CME as outlined in 2.1.3, adds to the need to revisit the issue of residential real estate market efficiency. The uptake of such products and success of a derivatives market for housing assets will be determined by the level of market efficiency.

An examination of weak-form efficiency in the Australian residential real estate market also presents an interesting research opportunity. No previous research into real estate market efficiency has considered the Australian market. In fact, outside of independent case studies in Tokyo (Ito and Hirono, 1993) and the UK (Rosenthal, 2006), the research has been limited to the housing markets of the USA and Canada. An extension to the Australian market provides the opportunity to examine market efficiency in a previously unconsidered market.

The Australian market presents an interesting case study; given the relatively high levels of housing investment, information availability and mortgage market maturity (Catte, Girouard, Price and André, 2004), it may be expected that the Australian market is relatively more efficient than those previously considered.

Weak-form market efficiency requires that an asset's historical price movements cannot be used to profitably predict the future prices and price movements of that asset. This implies the absence of an autoregressive relationship in asset returns. Therefore, to assess whether the residential real estate market satisfies a necessary requirement for weak-form market efficiency, the following null hypothesis is tested:

> ***Hypothesis 4_1: The returns to Australian residential real estate are unrelated to past returns, ceteris paribus.***

The results from testing this hypothesis are presented in Chapter 4. Tests of this null hypothesis represent the first step in examining the efficient market hypothesis in its weakest form. If rejected, it implies that autoregressive forecasting models may provide valuable forecasting information. That is, the observed prices don't fully reflect available information relating to past market prices (Fama, 1970), which may consequently be profitably exploited. If, on the other hand, there is insufficient evidence to reject this hypothesis, a necessary condition of weak-form market efficiency is met. This has implications for research into asset pricing models for residential real estate (Gau, 1987) and the application of a rational expectations framework in assessing policies, such as the impact of first home owner's grants, and the market design.

The high degree of heterogeneity in the residential real estate market, as outlined in Section 2.1.1, places additional emphasis on the *ceteris paribus* condition in this hypothesis. To accurately test *Hypothesis 4_1* the differences in attribute quality in the housing stock must be controlled for through the use of a constant quality measure.

The next two hypotheses to be developed in this chapter are concerned with the influence of asset heterogeneity on the results of weak-form market efficiency tests.

The review of the literature to have examined residential real estate market efficiency presented in Section 2.2 highlights the critical impact the choice of methodology may have on results. As a result of the low liquidity in the market, market efficiency tests are undertaken using the returns from an estimated house price index. The main house price index estimation methodologies, however, are prone to a set of biases, as discussed in Section 2.1.4. The second hypothesis to be tested in this dissertation considers the effect of composition bias in the median-price index on the results of weak-form efficiency tests.

Composition bias occurs in the median (and mean) price index as a result of the high level of heterogeneity in quality and attributes of the underlying housing stock, which is not accounted for in the index methodology (Crone and Voith, 1992; Gatzlaff and Ling, 1994). That is, movements in such indices may reflect changes in the composition of properties that sell from one period to the next rather than a true underlying price movement in the housing market. For example, an observed increase in the index level from March to April may indicate an increase in overall market prices, or it may indicate a higher proportion of larger or more desirable properties selling in April than in May. This bias has been demonstrated in the previous literature as a spuriously high volatility and spurious negative first-order serial correlation in returns to median-price indices (Case, Pollakowski and Wachter, 1991; Conniffe and Duffy, 1999).

The impact of composition bias in the median-price index estimation methodology on the outcomes of weak-form efficiency studies, particularly the expectation of spurious negative correlation, is tested through the following hypothesis, the results of which are presented in Chapter 4:

> ***Hypothesis 4_2: The returns from a median-price based index of house prices exhibit negative first-order autocorrelation.***

Failure to reject this hypothesis has two possible implications. There is predictability in returns to the housing market based on past price information. That is, weak-form market efficiency is violated. This is tested in the earlier hypothesis, *Hypothesis 4_1*. The other possibility is that, due to composition bias, there is predictability in median-price indices. These possibilities are not mutually exclusive. By jointly considering a rejection of this hypothesis with the results from the earlier hypothesis, *Hypothesis 4_1*, the degree to which the median-price index methodology impacts weak-form market efficiency may be observed.

The implications of a rejection of *Hypothesis 4_2*, while particularly acute in studies of weak-form market efficiency, are of broader concern to all studies of residential real estate prices. Testing this hypothesis provides another opportunity to demonstrate the effect of heterogeneity-induced composition bias, adding to the work of Case et al. (1991) and Conniffe and Duffy (1999).

As identified in Section 2.2 of the literature review, little consideration has been given in studies of residential real estate market efficiency to how house price index choice affects empirical results. No previous study has jointly tested market efficiency using alterative index methodologies. To the validity of conclusions from previous research and the research design of future research, such a question is well worth considering. The potential for bias inherent in an index methodology to create the impression of market inefficiency also has severe consequences for the development of housing derivates markets. The housing futures and derivatives listed on the CME, for example, are written over a repeat-sales index. A concern in relation to repeat-sales indices, however, is their susceptibility to smoothing. A third hypothesis is developed to test the effect of smoothing bias in the repeat-sales index on the outcomes of weak-form market efficiency.

Geltner (1997) argues that smoothing bias – the result of aggregating data over time, which biases second-order moments towards zero – is particularly acute in repeat-sales based indices. This is because the incorporation of current transactions data are not included in the index estimation unless and until it is a trade pair. That is, the index can never be entirely contemporaneous and temporal aggregation is likely to be exaggerated in this index. As Campbell, Lo, and MacKinlay (1997) econometrically demonstrate:

> "[A]lthough expected returns time-aggregate linearly …variances do not. As a result of the negative serial correlation in returns the variance of a sum is less than the sum of the variances" (Campbell et al., 1997: 95).

The following hypothesis is designed to test for the presence of smoothing bias and its effect in tests of weak-form market efficiency that use the returns from repeat-sales indices:

Hypothesis 4_3: The returns from a repeat-sales index of house prices exhibit a high degree of positive autocorrelation at short intervals.

The results from testing this hypothesis are presented in Chapter 4. The repeat-sales index methodology has been used to estimate housing market returns that underlie the majority of past research into the weak-form efficiency properties of the residential real estate market. Failure to reject the above hypothesis would cast doubt over the findings of a number of previous studies which rejected weak-form efficiency in the housing market, such as Case and Shiller (1989). Such a result would also provide implicit support for the empirical findings of a smoothing bias in repeat-sales indices by Geltner (1997) and Capozza, Hendershott, Mack and Mayer (2002).

Seasonality in returns is an empirical anomaly shared by many financial assets, including equities, futures and bonds. Section 2.2.2 of the literature review outlines several examples of financial asset seasonality. The presence of seasonality represents a specific violation of weak-form market efficiency as it offers another opportunity to predict future price movements based on historical returns.

For several reasons it may be expected that the residential real estate market exhibits seasonality in returns. Firstly, the relative structural inefficiencies of the residential real estate market limit the ability of arbitrageurs to profitably take advantage of such an anomaly. Consequently, seasonality patterns will be allowed to persist in the market. As discussed in Section 2.1.1., inherent structural features – such as heterogeneity, high transaction costs and limited information availability – increase information asymmetries among market participants and reduce liquidity.

Secondly, returns to the housing market are shown to be highly cointegrated with seasonal macroeconomic time series. Section 2.1.2 details, for example, the strong relationship between factors such as building activity and productivity, and returns to residential real estate. Given the multi-directional nature of this correlation, the strong seasonality in these macroeconomic variables may either drive or be driven by seasonality in the residential real estate market. Finally, the observation of transaction volumes seasonality in residential real estate markets (see, for example, Costello, 2001; Ngai and Tenreyro, 2009) may explain the presence of seasonality in residential real estate returns if changing volumes are associated with supply-demand imbalances.[30]

In a weak-form efficient market, however, seasonality in returns cannot persist. Taking a rational expectations perspective, market participants will profitably take advantage of such inefficiency. That is, if market participants are aware of a systematic higher (lower) return in the 'hot' ('cold') seasons identified by Ngai and Tenreyro (2009), they will avoid buying (selling) at such times.

The competing possibilities motivate tests of seasonality in residential real estate markets. The following null hypothesis is designed to test whether deterministic seasonality in returns to certain months of the year exists in the Australian market:

[30] For further discussion of the relationship between turnover and demand in the residential real estate market see Berkovec and Goodman (1996), in which this relationship is thoroughly documented.

Hypothesis 4_4: Australian residential real estate returns are not related to the month of sale, ceteris paribus.

A rejection of this hypothesis would indicate a deterministic monthly pattern in returns to residential real estate, thus adding support to the literature that argues against weak-form efficiency in the housing market. Failure to reject this hypothesis, on the other hand, is a necessary (although not sufficient) condition for weak-form market efficiency.

The majority of empirical research has been unable to confidently detect residential real estate returns seasonality (Costello, 2001; Rosenthal, 2006). A thorough review of this research, as provided in Section 2.2.2 of the literature review, however, shows a conflict exists in the research. Some seasonality studies undertaken, notably in markets with relatively extreme climates, such as Chicago (Case and Shiller, 1989; Kuo, 1996) and Canada (Hosios and Pesando, 1991), indicate a degree of seasonality in residential real estate returns linked to the colder months of the year.

While behavioural arguments have been developed to attempt to explain this result, exploration of the Australian market provides the opportunity to examine seasonality in residential real estate returns in a market with a relatively constant climate. Examining seasonality in the Australian housing market, using larger samples than have been previously considered in the studies by Costello (2001) and Rossini (2000, 2002), will provide further evidence on the issue of residential real estate returns seasonality.

The findings from studies of residential real estate seasonality, like the issue of weak-form efficiency, are also important for determining the way in which research should be undertaken and predicting how market participants will behave. From a research perspective, if the month of sale is a determinant of price then seasonal factors should be included in valuation studies. In practical terms, the presence of seasonality in returns may influence the timing of sales and also affect the development and uptake of housing derivatives markets.

The next two hypotheses consider seasonality in the composition of properties that sell from one period to the next, and how this may impact upon the results of testing for seasonality in returns from alternative index estimation methodologies.

The previous literature has identified that seasonality in residential real estate sales volumes is related to seasonality in mobility (Goodman, 1993). Mobility refers to the propensity for owner-occupier housing investors to relocate their place of residence. This typically involves selling their existing housing investment to finance their subsequent property purchase. Furthermore, seasonal mobility may differ across demographic groups. An example may be the expectation that more families with dependents will move in the summer months if they are encouraged to coincide their moves with the commencement of the school year. Given different demographic groups will demand different qualities in the housing stock – for example – it is expected that seasonality exists in the composition of housing that transacts from one period to the next.

The next hypothesis is concerned with seasonality in the composition of housing transacted. Given the significant variation in price explained by location (see, for example, DiPasquale and Wheaton, 1996), Prasad and Richards (2006) demonstrate that traded composition may be proxied by the proportion of sales taking place in more expensive suburbs. A relationship between this proportion and the month of the year is an indication of composition seasonality. The presence of composition seasonality in the Australian residential real estate market is tested through the following null hypothesis:

> ***Hypothesis 4_5: The volume of residential real estate sold in more expensive suburbs as a proportion of the total sales volume in the Australian market is not related to the month of sale, ceteris paribus.***

The results of testing *Hypothesis 4_5* are presented in Chapter 4. From the preceding argument, it is expected that this null hypothesis will be rejected. That is, it is expected that the Australian market will demonstrate a level of seasonality in the composition of housing types that sell through the year. Such a finding has serious implications for the treatment of housing market data. Specifically, rejection of this hypothesis will further emphasise the importance of accounting for variations in the quality and characteristics of the traded housing stock in studies of house prices.

As a corollary to *Hypothesis 4_5*, the effect of composition seasonality in measures of house price growth is considered. If differences in the quality of the housing stock that transacts between periods is deterministic, a measure of changes in aggregate house prices which does not account for the composition of the housing stock that is traded will be biased. Composition bias and its effects on the properties of median-price based indices are defined in the introduction to *Hypothesis 4_2*. The following hypothesis explicitly tests whether seasonality in the composition of properties that transact transmits a composition bias induced seasonality into the median-price index:

> ***Hypothesis 4_6: The returns to residential real estate estimated from a median-price index are related to the month of the year.***

A failure to reject this hypothesis would support the preceding argument that has linked seasonality in transacted composition to seasonality in the median-price index. Such a result would be in line with the findings of previous empirical work. Case and Shiller (1987, 1989), for example, argue against the use of median-price indices in research, given the pronounced seasonality of the median-price based National Association of Realtors (NAR) index that they argue may be driven by the composition of houses sold over the year. Prasad and Richards (2006) perform a similar analysis in the Australian market, demonstrating seasonality in median-price indices for the housing market in several major cities. Testing of *Hypothesis 4_6* provides the opportunity to corroborate these results with an alternative sample. Importantly, this set of hypotheses considers the bias of alternative index methodologies – including the hedonic and repeat-sales – in a wider test of weak-form market efficiency.

3.2 Sale Method and Prices

The literature review presented in Section 2.3.1 demonstrated that there is no consensus as to the size or even the direction of any price effect of the alternative sale mechanisms – private treaty sales and auctions – in transactions of residential real estate. This is despite the large volume of empirical research that has been undertaken to assess the impact of the auction mechanism on prices. This thesis proceeds by developing a hypothesis that tests the effect of auctions on prices in the Australian residential real estate market. The results of testing this hypothesis are reported in Chapter 5.

If multiple methods for selling a property – such as auctions and private treaty sales – are to coexist in the market, property prices cannot be related to the sale method used. Under a rational expectations framework, a relative price premium to one method would lead to the disappearance of the other.

Previous research to have tested for auction price effects in sales of residential real estate, however, has not found this rational *a priori* expectation to hold. Instead, price discount to auction sales has been observed in several studies conducted in the USA (Allen and Swisher, 2000; Mayer, 1998), while studies based in Australia and New Zealand have observed an auction price premium (Dotzour et al., 1998; Lusht, 1996; Newell et al., 1993).

The following hypothesis tests whether a relationship exists between the method by which a property is sold and the price achieved:

Hypothesis 5_1*: *The price at which a property sells is not related to the method by which it is sold, ceteris paribus.

A failure to reject this null hypothesis is consistent with a rational expectations framework. The large academic interest in the effect of auctions on prices in residential real estate markets motivates its further consideration. The majority of the literature has found evidence inconsistent with rational expectations.

The competing theories that have been used to justify the auction premium and auction discount findings are outlined in Section 2.3.1. Those which have attempted to support the finding of an auction premium have drawn on economic theory underlying the imbalance of power between the vendor and potential buyer; the vendor has the option to set pricing and refuse bidder offers (Milgrom, 1987, 1989). From the prospective buyer's perspective, theories from behavioural finance such as the winner's curse and loss aversion have been appropriated to explain the phenomenon of an auction price premium (French and McCormick, 1984). The observation of an auction price discount, on the other hand, has been explained as the cost of liquidity in the residential real estate market (Mayer, 1998). This rationalisation argues that a vendor will be prepared to sell at a discount in order to avoid the continued holding costs of the holding asset.

While these theories are adequate as explanations of a price impact to the auction mechanism in individual cases, they are limited in explaining the presence of a persistent auction price premium or discount in a market where auctions and private treaty sales compete with each other as alternative sale mechanisms. If auctions, on average, continually achieve a higher price than sales by private treaty then vendors will only sell by auction, and vice versa.

A limitation to past research in this area undertaken in the USA residential real estate market is the segmented market space in which auctions and private treaties are used. As discussed in the literature review, auctions in the USA are typically used by distressed vendors for sales of cheaper properties (Stevenson and Young, 2004). The Australian residential real estate market setting presents an opportunity to study this issue of housing market structure, given the coexistence of the multiple sale methods: auctions and private treaty sales.

Understanding of the residential real estate market is of vital importance to investors, economists and regulators alike. As discussed in Section 2.1.1, residential real estate accounts for the largest and most pervasive allocation of personal wealth in Australia. Persistent pricing biases driven by sale method may impact upon the decisions of investors and loan-providers, and in turn have a significance impact on the regulatory setting. This research is also necessary for ensuring integrity in the results of future housing market valuation research, and in the design of housing derivatives markets.

3.3 Information Asymmetry and New Properties

Section 2.3.2 of the literature review discusses the equity IPO literature and how a similar underpricing phenomenon may occur at the first offer of residential real estate assets. Drawing from the information asymmetry explanations used to rationalise the behaviour of equity IPOs, the next hypotheses to be tested, the results of which are reported in Chapter 6, consider whether an initial price premium to sales of new residential properties exists and the subsequent investment performance of these properties.

Significant information asymmetries exist in the residential real estate market as a result of the high illiquidity and limited data accessibility. These characteristics of the market are discussed in further detail in Section 2.1.1. When a property is first offered to the public the existing information asymmetry is exacerbated. That is, the information held by potential buyers of the new property is incomplete, since they do not yet know: (1) the quality of construction and how it will hold over time; (2) the quality of making an investment in that property (that is, its growth and yield potential); and, (3) the quality of the location. This final point is particularly acute for new areas where most new developments occur and the public infrastructure is still being developed. The vendor, on the other hand, is expected to be better informed than the potential buyer on all of these details.

As with the equity IPO literature, the original vendors of new residential real estate assets – typically developers – must encourage market participation in the presence of these information asymmetries (Beatty and Ritter, 1986). This is done by setting initial offer prices, on average,[31] below the fair market value. Under the design of the residential real estate market, initial offers from the developer to the public are not observed, and in many cases are operated through private expressions of interest. Instead, only the successful bid price is observed. As with the model developed by Beatty and Ritter (1986) for initial public equity offerings, the winner's curse may be attached to the successful bid. That is, the successful bid may be overpriced, given the true valuation and latent demand of the given property.

The following hypothesis to be tested predicts the existence of a price premium in the first public sale of residential real estate assets:

> ***Hypothesis 6_1: The price of a residential property at its first sale is greater than the price of other properties, ceteris paribus.***

Research considering the specific price behaviour of new residential properties has not previously been undertaken in any real estate market. The only paper, in fact, to consider the primary residential real estate market is provided by Chang and Ward (1993). In their examination of the Taiwanese pre-sales housing market – a setting in which developers may finance their incomplete projects – the authors demonstrate empirically a significant overpricing in sales of new properties. A failure to reject *Hypothesis 6_1* would provide support for the findings of Chang and Ward's (1993) developer financing study while also creating a literature concerned with the more specific question of residential real estate demand and pricing at first sale.

---

[31] Note that, crucially, not all offers can be priced below their fair market value; some must be overpriced (Beatty and Ritter, 1986).

Given the practical importance of such research – for example, the decision of a mortgage provider to provide a loan to a new property – it is surprising that so little academic research has been undertaken in this area. The author posits that this may be a result of data limitations to potential researchers.

As a corollary to *Hypothesis* 6_1, it is of interest to consider the subsequent performance of new residential properties relative to the wider market. In the equity IPO literature, the initial underpricing and significant listing-day return are typically observed to be followed by long-run underperformance (Loughran and Ritter, 1995). To assess whether new residential real estate assets similarly underperform, the following hypothesis is tested:

> ***Hypothesis*** $\boldsymbol{6_2}$***: The return to new residential properties after their first sale is lower, on average, than the contemporaneous market return, ceteris paribus.***

The practical implications of a failure to reject this hypothesis are similar to those for *Hypothesis* 6_1. The decisions of loan providers and investors, for example, may reflect an expectation of the long-run performance of new properties. The issues for policy-makers from this question are significant. How, for example, is the significant component of individual and national wealth in residential real estate managed if there is a disadvantage to investing in new properties? This should be of particular concern given the recent government-based initiatives to encourage investment in new developments such as the 2008-09 stimulus boost to the FHOG.

The subsequent performance of new housing, as with the issue of its pricing, is an area of research also previously untouched in the academic literature. The peculiar nature of the Australian residential real estate market provides an interesting setting for such a study, given the policy initiatives to boost investment in new housing (through tax-incentives, grants, and relaxations to foreign investment laws) and the subsequent recent growth in new housing developments.

3.4 Residential Real Estate Information Efficiency

Section 2.3.3 described the lack of informational efficiency in the Australian residential real estate market. It is currently impossible to produce timely price indices for this market based on the population of sales. This can largely be attributed to the lag between contract and settlement dates in sales of residential property and the resulting delay in distribution of sales information from the central government agencies, such as the VG in each state.

However, several sources of contemporaneous residential real estate sales data exist. These include the sample of sales reported in newspapers, the sample of properties listed for sale (advertisements), and the sample of sales known to real estate agencies. Each of these reflects a sample of function of the population of sales for a given period. Thus, it is expected that an index based on these samples will be correlated with an index estimated from the population. The final hypothesis to be tested, the results of which are reported in Chapter 7, considers whether there is value to such advance information.

In line with market efficiency theory, the newspaper-published sales results and advertisements can be thought of as public information, while the sample of sales known to real estate agents is more private information. As such, it is expected that there will be higher predictive power in the agents' sample of known sales than either public source of sales information. This conjecture is tested in the following hypothesis:

Hypothesis 7_1: The correlation between returns to an index estimated from the population of sales and to an index estimated from an advance sample of sales is higher when the advance sample is based on private information than public information.

It is expected that this correlation is positive for all returns estimated from the samples of advanced sales information when measured against the returns estimated from the population of sales. This is because the advance sales samples – sales results published in newspapers, advertisements, and real estate agent sales information – each represent non-mutually exclusive subsets of the total set of sales.

The sales knowledge held by real estate agents can be considered private information relative to the easily accessible sales results publicised through newspaper sale results and listings (expected results in the case of listings). As such, it is expected that the agent's private information will represent a larger and more comprehensive sample of sales than the information available publicly. This expectation will be supported if there is insufficient evidence to reject *Hypothesis 7_1*.

The value of information in the residential real estate market has never before been investigated in the academic literature. To some extent it has been a null research question in the past, given the relatively long transaction times and costs involved with real estate investments. The recent development of housing derivatives markets, however, such as the futures and options listed over a residential property repeat-sales price index on the CME, stimulates the need for such research. Its outcomes can affect the design and ultimately the success of these developing markets.

4. Efficiency and Seasonality in Residential Real Estate Prices

4.1 Introduction

Residential real estate market efficiency is of importance to the range of market participants, particularly given the recent development of and interest in housing derivative products. While the majority of studies have rejected the hypothesis of weak-form efficiency in returns to residential real estate (Case and Shiller, 1989; Hill et al., 1999), this has not been accepted unanimously (Gau, 1984; Rosenthal, 2006). In Chapter 3, a set of hypotheses designed to test the weak-form efficiency of the Australian housing market are presented. This chapter outlines the methodology, data and results of testing these hypotheses as set out in Section 3.1.

Weak-form efficiency tests of residential real estate returns are undertaken by applying the Box-Jenkins methodology to differenced price indices for the Sydney and Melbourne housing markets. In an extension to previous work in this area, the methodology is applied to indices estimated by the three main residential real estate index estimation methodologies: (1) median-price; (2) hedonic regression; and, (3) repeat-sales.

The significant positive autocorrelation at the 12-month lag observed to returns from the median-price index is further examined. This issue of residential real estate price seasonality has received a relatively large amount of academic interest, in response to anecdotal belief of 'better' buying and selling months commonly held by agents and some market commentators. The empirical results of a seasonal regression support the observations of ARIMA modelling following the Box-Jenkins methodology.

Specifically, seasonality analysis based on a median-price index does demonstrate significant monthly trends, which is not observed for constant-quality hedonic and repeat-sales indices measures.

The difference in results between the median and the constant-quality indices is shown to be driven by changes in the type of housing traded in any given month. The results demonstrate that a lower (higher) proportion of expensive properties sell in January (February), which correlates with the lower (higher) average returns to those months in the median-price index. While the composition of properties trading is not constant, changes in it are unrelated to movements in constant-quality indices which produce 'truer' measures of price performance.

The remainder of this chapter is set out as follows. The next section outlines the methodology to be undertaken in testing the efficiency and seasonality hypotheses. This is followed in Section 4.3 by a description of the data to be used in this study and the index construction methodologies. Section 4.4 presents the results of the hypothesis testing and Section 4.5 concludes this chapter.

4.2 Research Design

This section outlines the methodology to be undertaken to test the six hypotheses related to weak-form efficiency and seasonality in the residential real estate market. *Hypotheses* 4_1, 4_2 and 4_3 are tested using Box-Jenkins methodology and *Hypotheses* 4_4, 4_5 and 4_6 are tested using regression analysis.

4.2.1 Box-Jenkins Methodology

Weak-form efficiency requires that all public historical information is reflected in current prices. This implies that the returns to residential real estate are unrelated to past returns in the market. This conjecture is captured by *Hypothesis* 4_1 which is tested empirically using the Box-Jenkins methodology.

The Box-Jenkins methodology is a commonly applied technique for modelling and forecasting from time-series data. The first stage is called the *Identification* stage, and involves testing for stationarity and estimation of the autocorrelation function (ACF) and partial autocorrelation function (PACF) in order to identify the autoregressive structure of the series. The second stage, *Estimation*, is an iterative process to fit the optimal ARIMA model to the time series – in this case, residential real estate price indices. The final stage of the Box-Jenkins methodology, *Forecasting*, is not undertaken in the present analysis. In testing *Hypothesis* 4_1 it is sufficient to determine from the *Estimation* stage whether an autoregressive relationship exists in returns to Australian residential real estate. Diagnostic-checking of the residuals is included to demonstrate the appropriateness of the model specifications in capturing the autoregressive structure of the indices.

The application of ARIMA models requires that the time series is stationary. A stationary series is one which displays statistical properties that are constant over time. Autoregressive modelling of non-stationary series often results in the estimation of spurious regressions (Granger and Newbold, 1974).

It is expected that residential real estate prices are non-stationary; and that a trend term is contained in prices and price levels as given by a market-wide price index. To ascertain if this is the case, a unit root in the log price index is tested for using the augmented Dickey-Fuller (ADF) test. This test requires fitting the model given by Equation 4.1 independently for each index time series (Greene, 2003):

$$\Delta I_{x,t} = \alpha_x + \beta_x t + \gamma_x I_{x,t-1} + \sum_{i=1}^{p} \delta_{x,i} \Delta I_{x,t-i} + \varepsilon_{x,t} \qquad 4.1$$

where,

$I_{x,t}$ is the value of index, x, at time t, for x = {*Median*, *Hedonic*, *Repeat-Sales*}

α_x is the regression intercept term

β_x is the coefficient on the time, t, representing the trend in the given index, Ix

$\gamma_{x,}$ is the coefficient on the lagged index value, $I_{x,t\text{-}1}$, representing the degree to which the preceding index value affects the current index value

$\delta_{x,i}$ is the coefficient to the i^{th} autoregressive term

p represents the order of autoregressive parameters

Δ denotes the one period change in the given variable

$\varepsilon_{x,t}$ is the random error term to the dependent observations in the model assumed to be distributed independently and identically through a Normal distribution with mean, E[ε_x], equal to 0 and variance, $\sigma^2(\varepsilon_x)$, equal to 1. Such distributions are henceforth notated as i.i.d. N[0,1].

The unit root test is performed under the null hypothesis that the change in the index level is unrelated to the value of the index in the preceding time period; that is, the null hypothesis tests whether γ equals zero. It is expected that, as with most macroeconomic series, house price indices are non-stationary; their levels do not exhibit a unit root. Instead, it is likely that the price indices are integrated to the order 1, denoted as I(1), indicating that first differences (that is, returns) of the indices should be used in ARIMA modelling. The remaining discussion of the methodology, unless otherwise specified, deals with index returns.

To determine the autoregressive structure of returns to the residential real estate market, the ACF and PACF of the log differences of the price index are calculated. These provide an indication of the significance of historical price movements in predicting future movements at various lags. Specifically, it is important to this study to identify whether there is information contained at significant lags, such as 1 and 2, indicating persistence, and 3, 6 and 12, indicating seasonality.

Using the ACF and PACF results, the appropriate order of moving average and autoregressive lags, respectively, for the ARIMA model may be identified. The generalised specification of the ARIMA(p,d,q) model to be fitted to each index is given by Equation 4.2:

$$(1-B)^d I_{x,t} = \mu_x + \frac{\theta_x(B)}{\phi_x(B)} e_{x,t} \qquad 4.2$$

where,

μ_x is the constant mean term representing the average change in the index I_x

B is the backshift operator, such that $B^i X_t = X_{t-i}$, for all i = 1, 2, …T

$\phi_x(B)$ is the autoregressive operator, represented as a polynomial in the back shift operator with p autoregressive orders, such that $\phi_x(B) = 1 - \left[\sum_{i=1}^{p} \phi_i B^i I_x \right]$

$\theta_x(B)$ is the moving-average operator, represented as a polynomial in the back shift operator with q moving average orders, such that $\theta_x(B) = 1 - \left[\sum_{i=1}^{q} \theta_i B^i I_x \right]$

q represents the order of moving average parameters

d represents the order of differencing

$e_{x,t}$ is the random error term to the model, assumed to be i.i.d. $N[0,1]$

all other variables are defined as for Equation 4.1.

If no predictable model may be fitted following this methodology, we will find that there is insufficient evidence to reject *Hypothesis* 4_1 and conclude that the residential real estate market meets a requirement of weak-form efficiency. This is formalised as:

$$H4_1^0 : \theta_x(B) = 0$$

$$H4_1^A : \theta_x(B) \neq 0$$

Hypothesis 4_2 posits that composition bias in median-price indices will introduce a spurious negative autocorrelation in returns from a median-price index at short lags. A finding of no statistical significance to the coefficient of the first autoregressive term from an ARIMA model would indicate a failure to reject the null of this hypothesis. This is formalised as:

$$H4_2^0 : \phi_x(B \mid p = 1) = 0$$

$$H4_2^A : \phi_x(B \mid p = 1) \neq 0$$

By contrast, *Hypothesis 4_3* is developed from the observation that repeat-sales indices are susceptible to smoothing bias. The results from the estimated ARIMA model are used to test this hypothesis. Positive estimates of moving average parameters at short lags in returns from repeat-sales indices will indicate a rejection of the null of this hypothesis. This is formalised as:

$$H4_3^0: \ \phi_x(B \mid p < 5) \ = 0$$

$$H4_3^A: \ \phi_x(B \mid p < 5) \neq 0$$

4.2.2 Regression Analysis

This section describes the process by which the hypotheses concerning residential real estate market seasonality – *Hypotheses 4_4*, *4_5* and *4_6* – will be further tested.

Regression analysis is used to test for the presence of seasonality in residential real estate returns. Specifically, returns to residential real estate are regressed on a set of dummy variables representing the months of the year. Use of this technique is similar to that used in studies of equity seasonality (Rozeff and Kinney, 1976; Aggarwal and Rivoli, 1989) and real estate market seasonality (Rosenthal, 2006). The regression model to be estimated by Ordinary Least Squares (OLS) is given by Equation 4.3:

$$R_{x,t} - \overline{R}_x = \sum_{i=1}^{12} d_{x,i} Month_i + \varepsilon_{x,t} \qquad 4.3$$

where,

$R_{x,t}$ is the return to index I_x (that is, $R_{x,t}$ is the standardised first difference of index I_x) for the interval [t-1,t], calculated as $\dfrac{I_{x,t} - I_{x,t-1}}{I_{x,t-1}}$

$\overline{R}_x$ is the mean return to index I_x, calculated as $\sum_{i=1}^{T} \frac{1}{T} r_{x,t}$

$d_{x,1}$ to $d_{x,12}$ are regression parameters to be estimated representing the expected excess return to each month where i = 1 represents January, i = 2 represents February, and so on, calculated from index I_x

$$Month_i = \begin{cases} 1 \text{ when month of sale equals } i \\ 0 \text{ otherwise} \end{cases}$$

$I_{x,t}$ represents the value of index I_x at time t

$\varepsilon_{x,t}$ is the random error component of the model, assumed i.i.d. $N(0,1)$

all other variables are defined as for Equation 4.2.

The model is defined without an intercept term in order to ensure full-rank of the regression matrix. Statistical significance to the fitted parameters $d_{x,1}$ to $d_{x,12}$ indicates the presence of a deterministic monthly factor in the returns to index I_x. This observation will lead to a rejection of *Hypothesis 4₄*, which is formalised as:

$H4_4^0: d_{x,1} = d_{x,2} = \ldots = d_{x,12} = 0$

$H4_4^A$: At least one $d_{x,i} \neq 0$, for $i = 1, 2, \ldots 12$

This chapter examines seasonality, not only in returns to residential real estate, but also in the composition of properties that are traded. To this end, a second monthly dummy variable regression is estimated using a measure of the change in composition as the dependent variable. The specific regression model to be estimated by OLS is given by Equation 4.4:

$$k_t = \sum_{i=1}^{12} \delta_i' Month_i + \varepsilon_t \qquad 4.4$$

where,

k_t represents the proportion of more expensive properties at time t

δ'_1 to δ'_{12} are regression parameters to be estimated representing the average proportion of expensive properties selling in month i

$Month_i$ is a set of monthly dummy variables, defined as for Equation 4.3

ε_t is the random error component of the model, distributed i.i.d. $N(0,1)$.

Expensive properties are identified as those in suburbs with median sales price in the top five deciles of all suburbs when ranked by median prices. This measure is adapted from Prasad and Richards (2006). Location as a proxy for quality is considered reasonable, as typical valuation models demonstrate that a significant component of variation in property prices can be explained by location (Di Pasquale and Wheaton, 1996).

Month is a determinant of the relative quality of properties selling in a given month if the estimates to $\delta'_{x,1}$ to $\delta'_{x,12}$ are statistically significant. That is, the relative composition of properties selling is not constant across the year, and may in fact be predicted for any given month. This observation indicates a rejection of the null hypothesis that the proportion of properties selling from more expensive suburbs is not related to the month of sale, as given by *Hypothesis* 4_5. This is formalised as:

$H4_5^0 : \delta'_1 = \delta'_2 = \ldots = \delta'_{12} = 0$

$H4_5^A$: At least one $\delta'_i \neq 0$, for $i = 1, 2, \ldots 12$

While the presence of seasonality in median prices is supported if any of the parameters $\delta_{MED,1}$ to $\delta_{MED,12}$ in Equation 4.3 are statistically significantly nonzero, the cause of such seasonality is the final consideration of this chapter. *Hypothesis* 4_6 posits that this is driven by the underlying change in transacted composition, which, in a heterogeneous market, is not controlled for in median-price based analysis.

To determine the extent to which any seasonality in returns is driven by changes in the composition of properties transacting from one period to the next, the model, adapted from Prasad and Richards (2006) and given by Equation 4.5 is estimated by OLS:

$$r_{x,t} = a_x + b_x \, \Delta k_t + \varepsilon_{x,t} \qquad 4.5$$

where,

$r_{x,t}$ is the change in the natural log of index I_x over the interval $[t\text{-}1,t]$ calculated as, $\log(I_{x,t}) - \log(I_{x,t-1})$

Δk_t represents the linear change in the proportion of properties selling in the most expensive suburbs, k, over the interval $[t\text{-}1,t]$

a_x is the regression intercept term

b_x is coefficient on the change in composition

$\varepsilon_{x,t,t-1}$ is the random error component of this model, assumed i.i.d. $N(0,1)$

all other variables are defined as for Equation 4.4.

A statistically significant positive estimate on the coefficient for the compositional change variable, b_x, indicates correlation between changes in the composition of houses sold, k, and changes in the index, I_x. This represents the test of *Hypothesis* 4_6 which is formalised as:

$H4_6^0 : b_x = 0$

$H4_6^A : b_x \neq 0$

It is expected that b_{MED} will be statistically significant and positive, representing a rejection of the null hypothesis. It is expected that estimates of b using constant-quality indices – which account for heterogeneity in the transacted housing stock – will not be statistically significantly different from zero.

4.3 Data and Index Estimation

This section describes firstly the data to be used in this chapter's empirical section, its source and the filters that have been applied to it. This is followed by a description of the residential real estate price indices estimated from this data and which will underlie the subsequent analysis.

The data to be used in this chapter's index estimation is sourced from RPX, a publicly listed Australian-based company that specialises in collecting and aggregating housing market data from various sources including every state VG, real estate agents, council records, and internet listing tools. The database consequently contains details relating to virtually every transaction of residential real estate in the country.

All normal sales[32] of residential houses and units in the cities of Sydney and Melbourne between 1 January 1999 and 31 December 2008 are taken. These geographical bounds are defined as set out in the ABS Australian Standard Geographical Classification.[33]

Sydney and Melbourne are chosen as they account for 25% and 22% of Australia's residential dwellings, respectively, collectively over 40% of residential real estate asset wealth, and approximately 44% of national annual turnover value.[34] Given their significance to the Australian housing market and the importance of utilising data for

[32] Non-normal sales, such as non-arm's length transactions and sales identified as distressed, are removed.

[33] Source: *Standard Geography Volume I – Australian Standard Geographical Classification*, ABS, Catalogue 1216, July 2006.

[34] Source: RPX.

large, liquid markets in house price index estimation, the empirical section of this chapter is restricted to the Sydney and Melbourne housing markets.

Several filters are applied to the data to remove extreme and erroneous observations. Specifically, the following data are removed: sales with incomplete transaction, address, or attribute data; transactions of non-residential property;[35] sales classified as 'non-normal';[36] prices in the top and bottom 5% of observations; and houses with land sizes in the top and bottom 1% of observations.

Table 4-1 presents a set of descriptive statistics for the final data used to estimate the alternative house price indices required for this chapter's empirical section.[37] This includes data on the number of total sales observations, sales observations with complete hedonic information, and repeat-sale observations.

The final sample of sales comprises 512,441 and 357,901 observations for the Sydney houses and units markets, respectively, and 504,920 and 246,690 observations for the Melbourne houses and units markets, respectively. Of these, hedonic information was available for 367,179 (71.65%) house sales and 212,259 (59.31%) unit sales in Sydney, and 449,923 (89.10%) house sales and 205,927 (83.47%) unit sales in Melbourne.

---

[35] Property type definitions originate from the VG and are supplied to the author by RPX.

[36] In certain cases, such as transfers between relatives, this data are captured by the VG. To proxy for any additional abnormal sales, such as distressed sales and failed loans, properties that turn over repeatedly within six months are also removed.

[37] Data analysis throughout this thesis, unless otherwise specified, is undertaken using SAS software.

Table 4-1:

Descriptive Statistics

This table reports summary statistics of the sample of data underlying the estimated indices. This data covers the period January 1999 to December 2008 and statistics are reported separately for the four markets considered: Sydney houses, Melbourne houses, Sydney units, and Melbourne units. Panel A presents, respectively, the number of: (1) unique properties, (2) sales observations, (3) sales with hedonic information, and (4) repeat-sale observations. Summary attribute information is presented in Panel B.

	Sydney Houses	Melbourne Houses	Sydney Units	Melbourne Units
Panel A: Observations				
Properties	1,068,010	930,791	524,171	385,091
Sales	512,441	504,920	357,901	246,690
Hedonic Sales	367,179	449,923	212,259	205,927
Repeat Sales	117,625	108,709	93,329	52,452
Panel B: Attributes Details				
Land Size (m^2)	576.32	358.23	-	-
Bedrooms	3.25	3.09	2.11	2.18
Bathrooms	1.78	1.73	1.44	1.39
Parking (%)	74.20	57.02	57.65	37.80
Pool (%)	7.82	3.19	4.00	1.72
Waterfront (%)	1.44	0.19	1.96	0.30
Scenic View (%)	6.05	2.58	7.84	3.74
Air-Conditioning (%)	7.06	8.86	4.24	6.50

These figures give an indication of the relative liquidity of the selected markets. Repeat-sales account for around 21% to 26% of all sales over this period. This demonstrates that over a decade less than a quarter of properties in these relatively liquid property markets sold more than once. These figures also highlight the inefficient data use inherent to the repeat-sales index estimation method.

The average attribute statistics presented in Panel B of Table 4-1 provide some noteworthy results. The average land size to houses in Sydney (576 m^2) is much larger than that of the average Melbourne house (358 m^2). The average number of bedrooms, however, for houses is approximately three and approximately two for units in both cities. Lastly, 1.44 % of houses and 1.96% of units in Sydney are identified as waterfront, whereas only 0.19% of houses and 0.30% of units are waterfront in Melbourne. This difference is likely to be driven by the significant difference in geography between the two cities.

This chapter considers the monthly dynamics of property prices in examining the weak-form efficiency of the Australian residential real estate market. It is of interest, then, to note the relative monthly sales volumes. Table 4-2 reports the total sales volumes for each month over the sample period as well as the average proportion of transactions to that month during the sample period.

The monthly volumes demonstrate that residential real estate sales activity is not constant across the year. For all markets considered in this study, January accounts for the lowest average proportion of annual sales.

Table 4-2:

Aggregate Monthly Sales Volumes

This table reports the total sales observations for each sample market by month as well as the average proportion of all sales that occur in each month.

	Sydney Houses	Melbourne Houses	Sydney Units	Melbourne Units
January	25,893	29,293	16,076	13,430
	5.1%	5.8%	4.5%	5.4%
February	42,645	41,534	27,775	19,502
	8.3%	8.2%	7.8%	7.9%
March	50,306	46,878	33,769	22,624
	9.8%	9.3%	9.4%	9.2%
April	42,674	41,324	30,112	20,219
	8.3%	8.2%	8.4%	8.2%
May	49,646	46,496	35,538	22,805
	9.7%	9.2%	9.9%	9.2%
June	42,271	40,245	30,615	20,448
	8.2%	8.0%	8.6%	8.3%
July	43,313	42,784	32,232	22,080
	8.5%	8.5%	9.0%	9.0%
August	43,526	43,413	31,858	21,631
	8.5%	8.6%	8.9%	8.8%
September	43,970	41,688	30,398	20,261
	8.6%	8.3%	8.5%	8.2%
October	44,605	47,228	30,769	22,602
	8.7%	9.4%	8.6%	9.2%
November	46,543	45,505	31,645	21,858
	9.1%	9.0%	8.8%	8.9%
December	37,049	38,532	27,114	19,230
	7.2%	7.6%	7.6%	7.8%

Indices for the Sydney and Melbourne residential real estate market will be used in this chapter's hypothesis testing. The methodologies used to estimate these indices, introduced in Section 2.1.4, are now outlined. In total 12 indices are estimated. Price indices for houses and units in Sydney and Melbourne are estimated by three methodologies: (1) median-prices, (2) hedonic regression, and (3) repeat-sales.

The median-price index is formed by tracking the median price of all transacted properties in a given market and time period. For every time period, t, the median transactions price of all properties sold in that period is obtained as:

$$m_t = \left[P_{i,t} \mid i = \frac{N_t + 1}{2} \right] \qquad 4.6$$

where,

m_t is the median sales price in month t

$P_{i,t}$ is the price of the median sales observation, when sales are ranked by price

N_t is the number of sales observations in month t

t is the year and month of sale for all $t = 1, 2, \ldots, T$.

The specific hedonic index form to be estimated in this chapter is a log-linear model of sales price on observed property features and dummy variables for each period as given by Equation 4.7:

$$p_{i,t} = \alpha_t + \sum_{j=1}^{K} \beta_{j,t} Attribute_{i,j} + \sum_{m=1}^{N} \gamma_{j,t} .Suburb_{i,j} + \sum_{s=1}^{3} \delta_t Time_{i,s} + \varepsilon_{i,t} \qquad 4.7$$

where,

$p_{i,t}$ is the natural log of the sales price of house i at time t

α_t is the intercept term

β and γ are explanatory variable coefficients to be estimated reflecting the implicit value of each property attribute and suburb, respectively

$Suburb_{i,j}$ is a dummy variable equal to 1 if property i is located in suburb j, and zero otherwise

δ_s estimates the cumulative growth rate to time s

$Time_{i,s}$ is a set of dummy variables equal to 1 if the property sold in period s and zero otherwise

$\varepsilon_{i,t}$ is the random variation in price of house i at time t not captured by the model, distributed i.i.d. $N(0,1)$.

The adjacent-period model is an enhancement of the traditional hedonic models in that it pools data only from consecutive time periods, thereby allowing for non-constant attribute values. For each 'adjacent-period' subset of data, the hedonic function is estimated and the estimated growth rates to each period are chain-linked to form an index (Wright, 2006).

The alternative quality-controlled index methodology is the repeat-sales index. Estimation of the repeat-sales index follows the WRS two-stage least-squares method of Case and Shiller (1987) that regresses the trade-pair returns on a negative dummy variable for the month of the first sale and a positive dummy for the month of the second sale. The model to be estimated is given by Equation 4.8:

$$r_{t,s}^{i} = \sum_{t=1}^{T} \delta_t x_t + u_t^i \qquad 4.8$$

where,

$$x_j = \begin{cases} -1\, if \;\; j\, is\, the\, period\, of\, the\, first\, sale, s \\ 1\, if \;\; j\, is\, the\, period\, of\, the\, second\, sale, t \\ 0\, otherwise \end{cases} \qquad \text{for all } j = 1, 2, \ldots T$$

$r_{t,s}^{i}$ is the difference in log prices for house i between t and s for all $t > s$

δ_t is the logarithmic growth rate of the underlying house price index

T is the total number of monthly periods the index covers

$u_{i,t}^{2}$ is the regression error term, assumed i.i.d. $N(0,1)$.

Under the two-stage least-squares model of Case and Shiller (1987), the variables, $r_{t,s}^{i}$, x_j^i, and u_t^i, are weighted by the square root of the predicted values of, $u_{i,t}^{2}$, obtained from Equation 4.9:

$$\mathrm{E}[u_{i,t}^{2}] = A + B(t - s) \qquad 4.9$$

where,

$u_{i,t}^{2}$ is the square of the regression error obtained for each observation from the unweighted estimation of Equation 4.8

A is the fitted regression intercept term

B is the fitted coefficient on the time between sales

all other variables are as defined for Equation 4.8.

Indices, based at 100, are estimated by chain-linking the continuous growth rate between each quarter. This process is modelled by Equation 4.10:

$$I_{1,\mathrm{x}} = 100$$

$$I_{t,x} = I_{t-1,x} e^{r_{t,x}} \qquad \text{for } t = 2, 3, \ldots T \qquad 4.10$$

where,

$r_{x,t}$ is the continuous growth rate of the index-type x given by: $\ln(m_t) - \ln(m_{t-1})$ for the median-price index; δ_s for the hedonic index, and; δ_t for the repeat-sales index

e is the base of the natural logarithm

there are T unique monthly periods in the index series.

The estimated indices and returns for the house and unit markets in Sydney and Melbourne are charted in Figures 4-1 through 4-4. The indices are all based at 100 in January 1999.

The Sydney house price index value doubled in 10 years, reaching a value of around 200 by December 2008. Prices did not grow at a constant rate, however, instead exhibiting two distinct periods of house price growth.

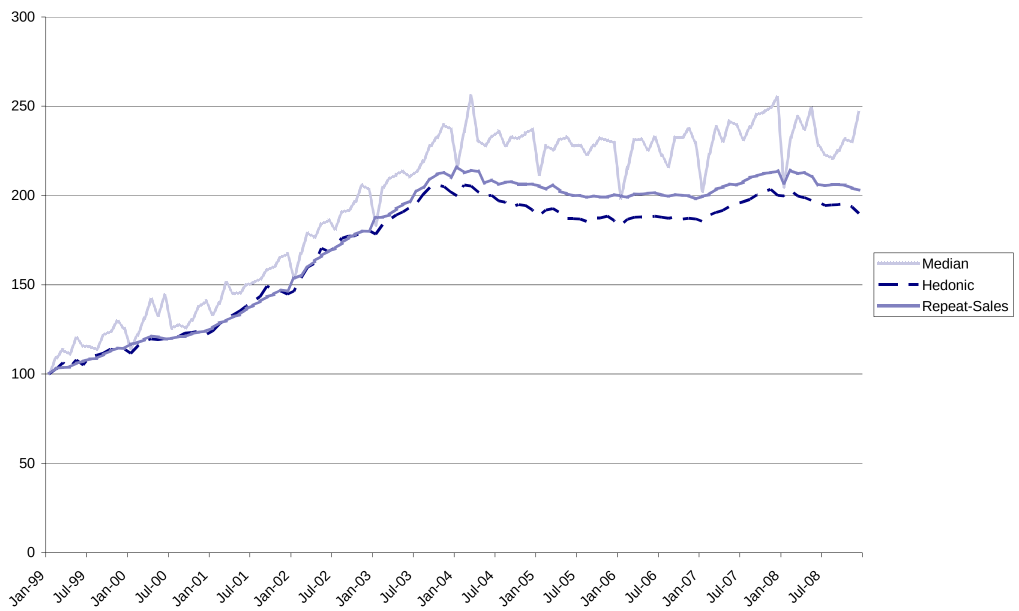
The first of these subperiods, from January 1999 to around July 2003, is characterised by relatively strong, monotonic growth. The index peaks at around 230 at the end of this subperiod in the second half of 2003. In the second period, July 2003 to December 2008 Sydney house prices have experienced a prolonged price deflation and stabilisation. The index values, in fact, are approximately the same at the end of this second period. The index indicates a short-lived price recovery towards the end of the sample period.

Figure 4-1:

Index – Sydney Houses

Panel A and Panel B present the monthly indices, based at 100 in January 1999 and their returns, respectively, as estimated by the median, hedonic, and repeat-sales methodologies for the Sydney house market.

Panel A: Sydney House Price Indices



Panel B: Sydney House Price Returns

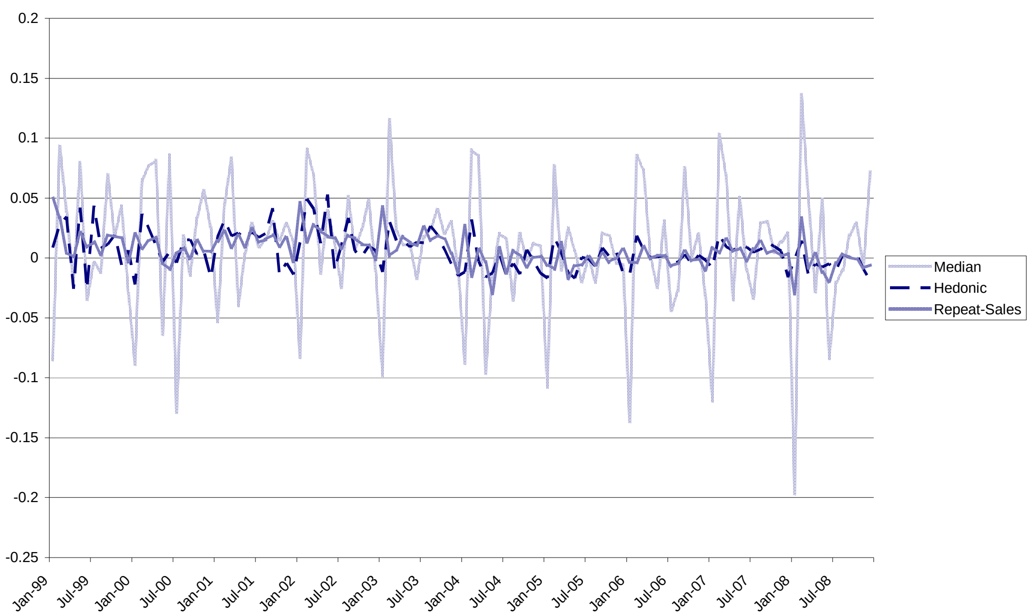
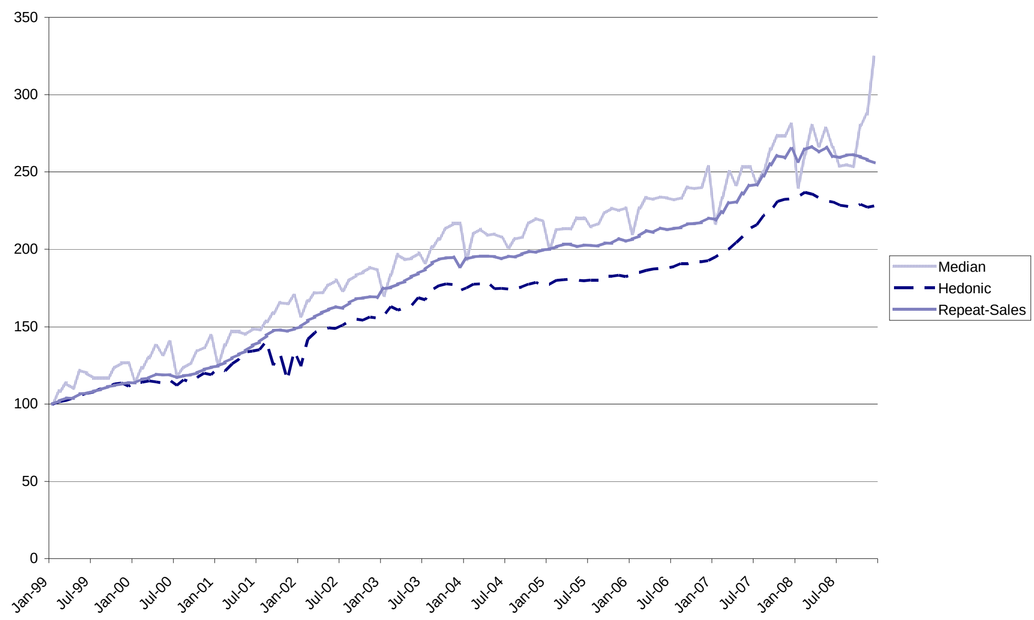


Figure 4-2:

Index – Melbourne Houses

Panel A and Panel B present the monthly indices, based at 100 in January 1999 and their returns, respectively, as estimated by the median, hedonic, and repeat-sales methodologies for the Melbourne house market.

Panel A: Melbourne House Price Indices



Panel B: Melbourne House Price Returns

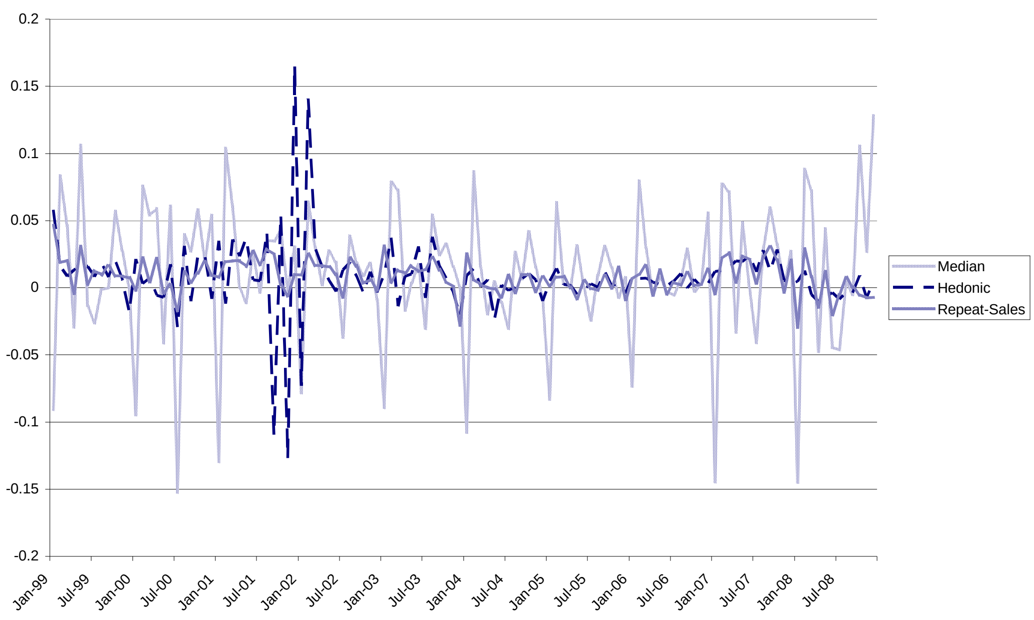
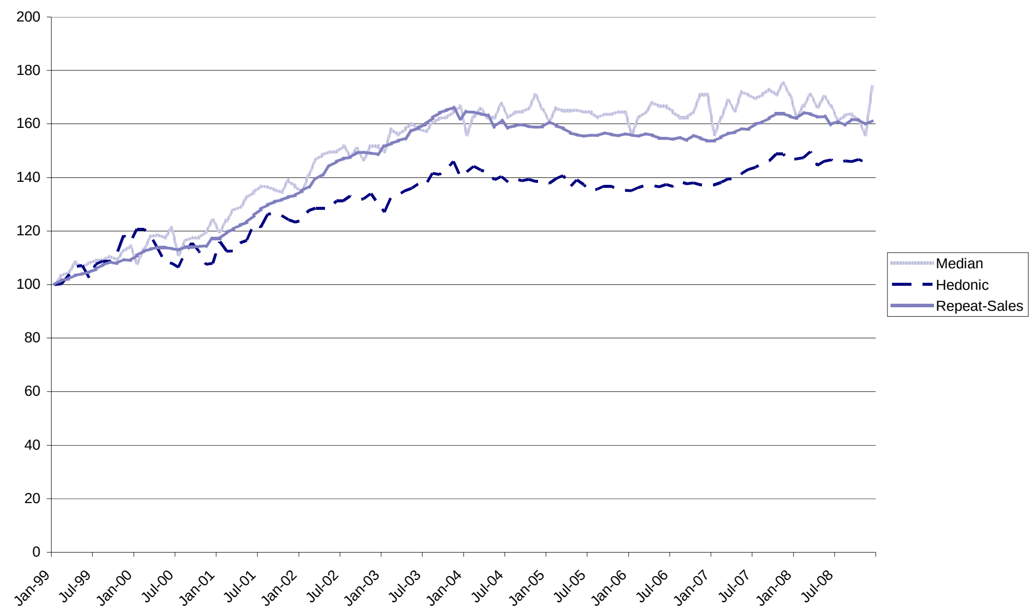


Figure 4-3:
Index – Sydney Units

Panel A and Panel B present the monthly indices, based at 100 in January 1999 and their returns, respectively, as estimated by the median, hedonic, and repeat-sales methodologies for the Sydney unit market.

Panel A: Sydney Unit Price Indices



Panel B: Sydney Unit Price Returns

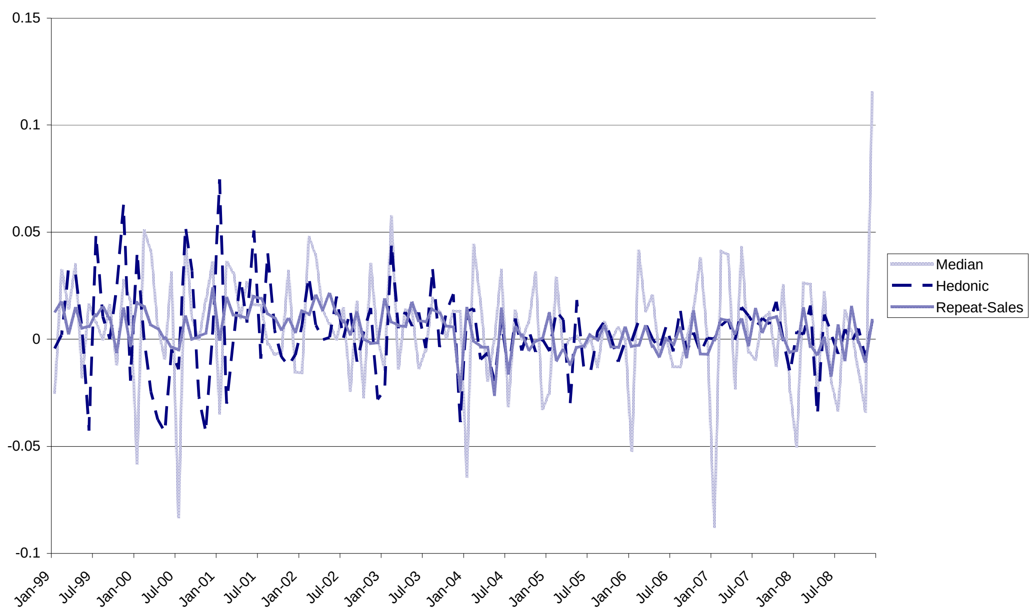
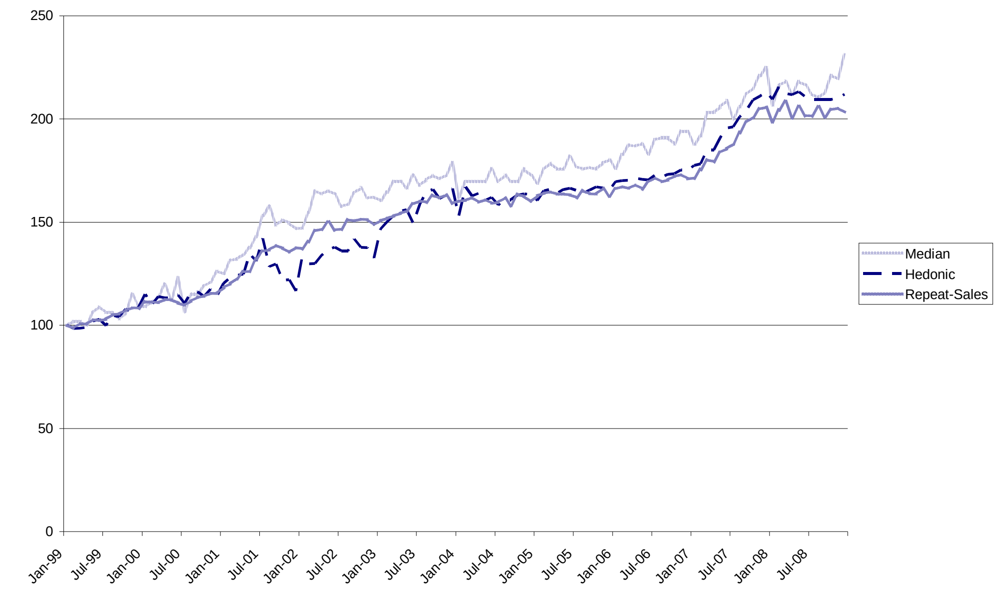


Figure 4-4:

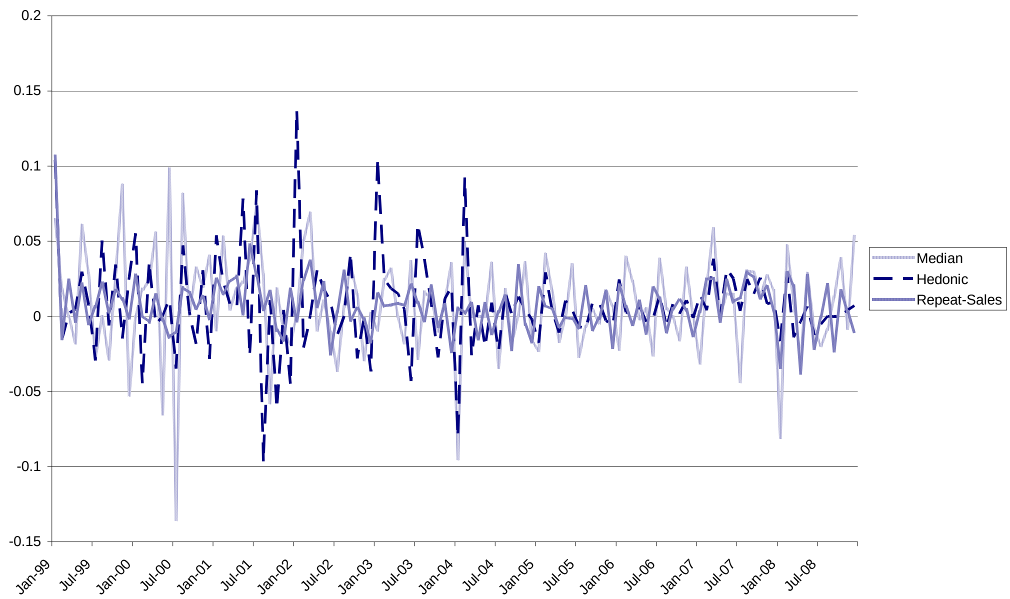
Index – Melbourne Units

Panel A and Panel B present the monthly indices, based at 100 in January 1999 and their returns, respectively, as estimated by the median, hedonic, and repeat-sales methodologies for the Melbourne unit market.

Panel A: Melbourne Unit Price Indices



Panel B: Melbourne Unit Price Returns



Compared to the hedonic and repeat-sales indices, the median produces a much more volatile index, with negative spikes occurring at every January. This observation is reflected in the time series of returns to the Sydney house market, presented in Panel B of Figure 4-1. The monthly returns to Sydney house prices support the observations drawn from the index values. For approximately the first half of the sample period considered the returns to each index trend around 0.025, and then around zero for the second half of the period. The median index returns are much more volatile than other indices, especially around the turn of the year.

Index results for Sydney units are relatively consistent with those for Sydney houses. The two subperiods of time are apparent, though units have underperformed houses, the index level rising to around 160 (compared with 200) at December 2008. It is important to note that these indices only account for capital gains.[38]

The Melbourne residential real estate market has grown more consistently than Sydney's: the Melbourne house and unit indices peak at around 250 and 200, respectively, at December 2008. Index methodology observations from the Sydney market, such as the higher volatility in the median index, may also be made about the Melbourne market.

The next section presents the descriptive statistics of the index dynamics and the results of applying the methodology described in Section 4.2 to each of these indices.

---

[38] Total return indices are not considered in this thesis due to data availability.

4.4 Results

This section presents the results of this paper. Firstly, the descriptive statistics of the index returns are presented. This includes the results from testing for stationarity in the indices. Secondly, the autocorrelation and partial autocorrelation functions are plotted and used to fit ARIMA models for the returns to each of the three indices.

4.4.1 Index Dynamics

Descriptive statistics for the returns to the alternative house price index methodologies are presented for Sydney and Melbourne in Tables 4-3. For the period considered, January 1999 to December 2008, the average annualised return to Sydney houses is estimated at 10.84% by the median-price index, 5.79% by the hedonic index, and 7.81% by the repeat-sales index. This is higher than the estimated average return to Sydney units, which ranges from 3.34% using the hedonic index to 6.32% using the median-price index.

Over this 10-year period, the Melbourne property market is observed to outperform the Sydney market. The average annualised returns to Melbourne houses (units) are estimated at 13.07% (10.42%) by the median-price index, 9.35% (9.44%) by the hedonic index, and 9.92% (8.34%) by the repeat-sales index.

The average returns to the median-price index are higher and more volatile for all markets than the returns to either constant-quality index: the standard deviation of monthly returns in the Sydney house market is estimated at 0.189 by the median-price index, 0.061 by the hedonic index and 0.045 by the repeat-sales index.

Table 4-3:

Index Dynamics

This table presents returns statistics for the house and unit markets in Sydney (Panel A) and Melbourne (Panel B) based on the three index estimation methodologies. The first four moments of the return's distributions are summarised and presented as the average annualised return, the annualised standard deviation, the kurtosis and skewness. Test statistics from the ADF τ test for unit root stationarity are also reported. Rejection of the τ test statistic at the 1% level of significance is represented by three asterisks, *** (Dickey, Hasza and Fuller, 1984).

	Houses			Units		
	Median-Price	Hedonic	Repeat-Sales	Median-Price	Hedonic	Repeat-Sales
Panel A: Sydney						
Average Annualised Return (%)	10.84	5.79	7.81	6.32	3.34	5.01
Standard Deviation	0.19	0.06	0.05	0.10	0.09	0.03
Kurtosis	1.37	0.55	1.35	2.41	5.73	0.26
Skewness	-0.72	0.19	0.35	-0.21	-0.52	-0.47
ADF τ statistic (Index)	-2.28	-1.97	0.64	-1.48	-1.97	-0.14
ADF τ statistic (First Difference)	-11.98 ***	-6.88 ***	-6.53 ***	-12.29 ***	-8.97 ***	-5.70 ***
Panel B: Melbourne						
Average Annualised Return (%)	13.07	9.35	9.92	10.42	9.44	8.34
Standard Deviation	0.18	0.10	0.04	0.12	0.11	0.06
Kurtosis	1.38	13.91	0.83	2.10	3.41	7.04
Skewness	-0.83	0.59	-1.90	-0.58	0.69	1.15
ADF τ statistic (Index)	-4.74 ***	-1.79	-1.52	-2.51	-2.39	-1.63
ADF τ statistic (First Difference)	-10.41 ***	-6.41 ***	-6.30 ***	-10.31 ***	-9.06 ***	-8.57 ***

Interestingly, the volatility to the returns of the repeat-sales index is lower in all markets than the volatility in returns to either the median of the hedonic index. These results are consistent with the findings of the past academic research which has compared the established alternative index estimation methodologies as discussed in Section 2.1.4. From the kurtosis and skewness estimates it is observed that no index indicates the returns to housing are normally distributed. These results demonstrate the extent to which the choice of index methodology may affect research which assumes normality.

Failure to reject the ADF τ statistic for the index levels indicates that they are non-stationary series. These tests use the critical values for the ADF test as presented in Dickey et al., 1984). Rejection of the τ statistic for the first-differenced series supports the *a priori* expectation that the house price indices are first-order integrated, I(1), processes. Consequently, the remainder of this chapter will be concerned with modelling first differences.

4.4.2 Identification Stage

The ACF and PACF are estimated for first differences for the three index methodologies considered: median, hedonic, and repeat-sales. These functions are important in analysing the autoregressive structure of a time series. Consideration of these is the *Identification* stage of the Box-Jenkins methodology for fitting ARIMA models.

The ACF and PACF for first differences of the Sydney house market median index, hedonic index, and repeat-sales index are presented in Tables 4-4, 4-5, and 4-6 respectively.[39]

The slow decay in the ACF using the index levels supports the finding of non-stationarity in the index series reported by the ADF statistic in Table 4-3. For all markets and index estimation methodologies this result occurs. The ACF for returns (calculated as first differences of the log index series) does not exhibit this pattern, in line with the stationarity test findings reported in Table 4-3. For this reason, ARIMA modelling is undertaken on the index returns series.

The Box-Jenkins methodology for fitting ARIMA models follows an iterative process. The autocorrelation patterns identified by the ACF and PACF to the returns series are used in various combinations to find a parsimonious model which minimises the Aikaike Information Criterion (AIC) and Schwartz-Bayes Criteria (SBC) statistics. A well-specified model for the given data requires there to be no pattern to the residuals. That is, there are no significant autocorrelations in the residuals of the fitted model, as determined by the Ljung-Box χ^2 statistic (2003). If the χ^2 statistic rejects the null hypothesis of zero autocorrelation in residuals then a better model is required to explain the autoregressive pattern. Diagnostic checking of the models using this test-statistic is reported in Section 4.4.3.

---

[39] Due to the space required to present these functions, the results for the other property markets considered in this chapter are reported in Appendix A. Their results are consistent with those reported in this section for the Sydney house market.

Table 4-4:

Correlogram – Median-Price Index

This table presents the ACF and PACF for returns to the median-price index for Sydney houses, estimated following the methodology set out in Section 4.3. Correlations are calculated over first differences in the log median-price index, with significance at the 5% level denoted by *.

	Autocorrelation Function			Partial Autocorrelation Function		
Lag	Correlation	Correlogram		Correlation	Correlogram	
1	-0.2724*	*****		-0.2724*	*****	
2	-0.1775*	****		-0.2719*	*****	
3	0.0694		*	-0.0772	**	
4	-0.1704	***		-0.2569*	*****	
5	0.0536		*	-0.1090	**	
6	0.2409*		*****	0.1627*		***
7	0.0454		*	0.2385*		*****
8	-0.1678	***		0.0198		
9	0.1559		***	0.2429*		*****
10	-0.2647*	*****		-0.1396*	***	
11	-0.0470	*		-0.1662*	***	
12	0.4885*		**********	0.3163*		******
13	-0.0500	*		0.2646*		*****
14	-0.2081	****		-0.0533	*	
15	0.0660		*	0.0040		
16	-0.0926	**		0.0498		*
17	0.0141			0.0654		*
18	0.1877		****	-0.0645		*

Table 4-5:

Correlogram – Hedonic Index

This table presents the ACF and PACF for returns to the hedonic index for Sydney houses estimated following the methodology set out in Section 4.3. Correlations are calculated over first differences of the log hedonic index levels, with significance at the 5% level denoted by *.

	Autocorrelation Function		Partial Autocorrelation Function	
Lag	Correlation	Correlogram	Correlation	Correlogram
1	-0.0156		-0.0156	
2	0.0842	**	0.0840	**
3	0.1983*	****	0.2024*	****
4	0.0082		0.0107	
5	0.2758*	******	0.2543*	*****
6	0.1609	****	0.1540*	***
7	0.0166		-0.0040	
8	0.1202	**	0.0108	
9	0.1350	***	0.09578	**
10	-0.0575	*	-0.1421*	***
11	0.1885*	****	0.0769	**
12	0.1303	***	0.1100	**
13	0.1211	**	0.1370*	***
14	-0.0205		-0.1576*	***
15	0.0061		-0.0263	*
16	0.0229		-0.0646	*
17	0.0173		-0.0828	**
18	0.2138*	****	0.1540*	***

Table 4-6:

Correlogram – Repeat-Sales Index

This table presents the ACF and PACF for returns to the repeat-sales index for Sydney houses estimated following the methodology set out in Section 4.3. Correlations are calculated over first differences of the log repeat-sales index levels, with significance at the 5% level denoted by *.

	Autocorrelation Function		Partial Autocorrelation Function	
Lag	Correlation	Correlogram	Correlation	Correlogram
1	-0.4946*	*********	-0.4946*	**********
2	-0.0044		-0.3296*	*******
3	0.0032		-0.2417*	*****
4	0.0024		-0.1845*	****
5	-0.0040		-0.1523	***
6	0.0059		-0.1188	**
7	0.0001		-0.0934	**
8	-0.0055		-0.0830	**
9	0.0081		-0.0603	*
10	-0.0003		-0.0434	*
11	-0.0040		-0.0369	*
12	0.0126		-0.0097	
13	-0.0091		-0.0056	
14	-0.0037		-0.0088	
15	0.0098		0.0053	
16	-0.0150		-0.0108	
17	0.0156		0.0048	
18	-0.0094		-0.0009	

Autocorrelations in the returns to the median index for Sydney houses, presented in Table 4-4, exhibit significant negative spikes at lags 1 and 2, and significant positive spikes at lags 6 and 12. The short-order negative autocorrelation is an indication of spurious volatility bounce, while the highly positive 6- and 12-month lags suggest seasonality in returns.

Table 4-5 presents the autocorrelation in returns to the hedonic index for Sydney houses. The results indicate significant positive spikes in returns at lags 3, 5 and 11.

The repeat-sales index behaves unlike either of the other indices. The autocorrelation in returns to the repeat-sales index for Sydney houses, reported in Table 4-6, is significant and negative at the first lag only, while the partial autocorrelations are negative and decay slowly with significant values to lags 1 through 4.

4.4.3 Estimation Stage

From the autocorrelation patterns exhibited by the ACF and PACF for each index, the models that are found to best fit the time series of returns for the Sydney houses market estimated by each of the three index methodologies are now presented. Table 4-7 reports the estimated parameters from these models.

Table 4-7:

ARIMA Models – Sydney House Price Indices

This table reports the estimated parameters and their *t* statistics for the ARIMA models that best fit the returns from the median, hedonic and repeat-sales indices. Statistical significance of the parameters at the 1% and 10% level of significance is represented by *** and *, respectively. The number of observations, AIC and SBC statistics are also reported.

	Median		Hedonic		Repeat-Sales	
Parameter	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic
μ	0.0063	1.12	0.0018	0.49	0.0065 *	2.5
Autoregressive Terms						
$\Phi 1$	-0.9553 ***	-8.89				
$\Phi 3$			0.8049 ***	7.92		
$\Phi 6$	-0.0086	-0.13				
$\Phi 12$	0.9914 ***	13.85				
Moving-Average Terms						
$\theta 1$	-0.5462 ***	-4.46			0.8346 ***	17.1
$\theta 2$	0.4357 ***	5.16				
$\theta 3$			0.6465 ***	5.6		
$\theta 5$			-0.2429 ***	-3.32		
$\theta 6$	-0.1582 ***	-1.27				
$\theta 12$	0.5706 ***	4.47				
Observations	131		131		131	
AIC	-497.378		-716.157		-82.322	
SBC	-474.376		-704.656		-76.572	

A factored ARIMA model with seasonal factors at 6 and 12 months is found to best fit the returns to the median-price index. This model is given by Equation 4.11:

$$Y_t^{MED} - Y_{t-1}^{MED} = \mu + \frac{\left(1-\theta_1 B - \theta_2 B^2\right)\left(1-\theta_6 B^6 - \theta_{12} B^{12}\right)}{\left(1-\phi_1 B\right)\left(1-\phi_6 B^6 - \phi_{12} B^{12}\right)} \alpha_t \qquad 4.11$$

with,

Autoregressive factor 1	1 + 0.9553 B(1)
Autoregressive factor 2	1 + 0.0086 B(6) – 0.9914 B(12)
Moving-Average factor 1	1 + 0.5462 B(1) – 0.4357 B(2)
Moving-Average factor 2	1 + 0.1582 B(6) – 0.5706 B(12)

A statistically significant negative coefficient to the first-order autoregressive term in Equation 4.11 indicates a predictable ‘bounce’ in median-price index returns. That is, from one month to the next the sign of the return to this index will change, but be of almost equal magnitude, *ceteris paribus*.

The first-order moving-average coefficient is also found to be statistically significant and negative, while the second-order moving-average coefficient is positive. This demonstrates the persistence of the sign reversal in monthly returns. From this result, there is insufficient evidence to reject *Hypothesis* 4_2. That is, the results suggest a forecastable pattern in median-price index returns as a result of predictable negative first-order autocorrelation.

Applying Box-Jenkins methodology to the returns from the hedonic index, an ARIMA model is fitted with an autoregressive term at the third lag and moving-average terms at the third and fifth lags. This model is given by Equation 4.12:

$$Y_t^{HED} - Y_{t-1}^{HED} = \mu + \frac{\left(1 - \theta_3 B^3 - \theta_5 B^5\right)}{\left(1 - \phi_3 B\right)} \alpha_t \qquad 4.12$$

with,

Autoregressive factor	1 – 0.8049 B(3)
Moving-Average factor	1 – 0.6465 B(3) + 0.2429 B(5)

A predictable pattern in the returns to the hedonic index for Sydney houses is found, with statistically significant positive coefficients to the third-order autoregressive term, and the third- and fifth-order moving-average terms. This result may suggest quarterly seasonality in house price returns, although it is difficult to rationally explain why this would occur.

Using the returns to the repeat-sales index, the following ARIMA model given by Equation 4.13 is fitted following Box-Jenkins methodology:

$$Y_t^{REP} - Y_{t-1}^{REP} = \mu + \left(1 - \theta_1 B\right)\alpha_t \qquad 4.13$$

with,

Moving-Average factor	1 – 0.834634 B(1)

Only a single factor is required to optimally fit an ARIMA model to the returns of the Sydney houses' repeat-sales index. This first-order moving-average term is highly statistically significant, although negative. This findings is in conflict with the results of earlier research to have claimed the repeat-sales index is smoothed relative to other index measures; that is, its second-order statistical moments are biased towards zero. This finding also fails to reject *Hypothesis* $H4_3$, since returns to the repeat-sales index are found to be highly positively autocorrelated at short intervals.

Diagnostic checking of these estimated models is undertaken through analysis of the Ljung-Box χ^2 test for autocorrelation of the residuals at 12 monthly lags (Greene, 2003). The statistics for the estimated models using each index method are presented in Table 4-8.

The χ^2 test-statistic tests the null hypothesis that the autocorrelation in residuals is equal to zero, based on the degrees of freedom in the model. From the results reported it can be seen that the hedonic and repeat-sales ARIMA models are well specified as there is no residual autocorrelation left in their return data. For the Sydney residential real estate market (both houses and units), the Ljung-Box χ^2 statistic weakly rejects the null hypothesis. This indicates that there is some residual autocorrelation in the specified ARIMA model for the returns to the median index. This is despite multiple iterative specifications of the model, further demonstrating the significant level of predictability to this data.

Table 4-8:

Residual Autocorrelation

This table reports the Ljung-Box χ^2 test-statistic for each index ARIMA model for each market considered. Three asterisks, ***, represent rejection of the null hypothesis of zero autocorrelation in the residuals to lag-12 at the 1% level of significance and one asterisk, *, represents rejection at the 10% level of significance.

	Sydney Houses	Sydney Units	Melbourne Houses	Melbourne Units
Median	14.33 *	18.64 *	7.77	7.01
Hedonic	7.28	4.39	11.53	10.49
Repeat- Sales	0.43	0.34	9.34	13.83

The objective of this chapter is not to prescribe the appropriateness of alternative index methodologies. It is apparent, however, that the hedonic index is the least biased measure of house price movements of the alternative methodologies considered. An ARIMA model, with significant positive terms at lags 3 and 5, is fitted for the hedonic index returns. This result indicates a rejection of *Hypothesis* $H4_1$. That is, the evidence presented in this chapter shows that the returns to Australian residential real estate can be forecast from past returns, *ceteris paribus*. This result violates the efficient markets hypothesis in its weakest form.

Predictable patterns in the time series of returns to the median and repeat-sales indices are also observed. Returns to the repeat-sales index for Sydney houses exhibit a significant positive first-order moving-average autocorrelation. Median index returns, by contrast, are negatively autocorrelated at lags 1 and 6, while positively correlated at lag 12. This 12(6)-month positive (negative) autocorrelation indicates a strong pattern of seasonality, not detected in the other index measures. Seasonality in returns is explored more thoroughly in the following section.

4.4.4 Seasonality

The average return to each month for the three alternative indices is depicted in Figures 4-5 for the Sydney property market and 4-6 for the Melbourne property market. On visual inspection, while for most months the alternative indices have average returns in the same direction, the median-price index return appears amplified compared to the others.

For all four markets, the average return to January and July based on the median-price index is negative; these are followed by pronounced positive average February (and March) and August returns, respectively. The patterns in the average monthly returns derived from the constant-quality indices are much less noticeable; in general the average for each month sits between 0% and 2%, and while a negative average January return can be seen, it is less than half the size of the return obtained from the median index.

Differences in the average return to each month are compared more formally using *t* test analysis. Table 4-9 presents the average return to Sydney houses and units of each index by month.[40] Statistical significance using the *t* test for difference in mean return for each month against the rest of the year is identified in this table.[41].

---

[40] The results for the Melbourne market, reported in Appendix A, are consistent with those presented in this chapter for Sydney.

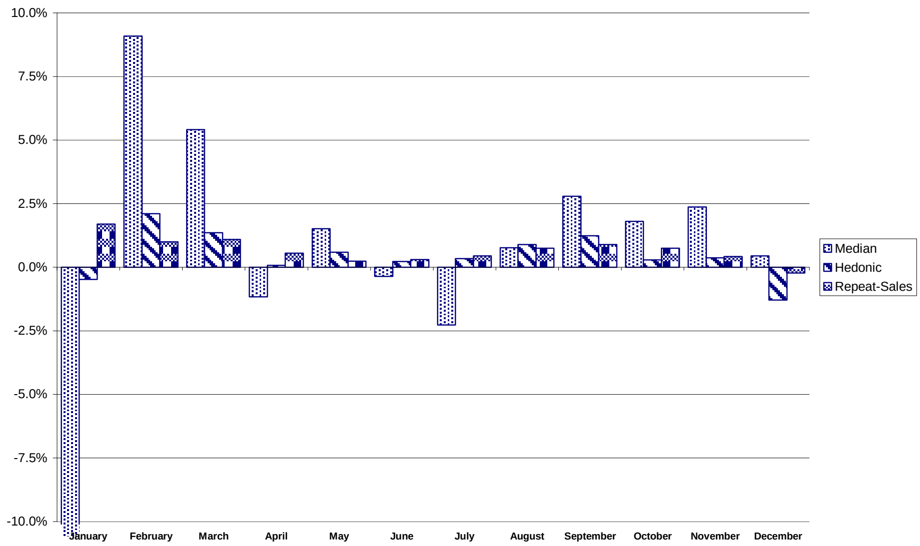
[41] All months are included in estimation of the annual mean against which the monthly average is tested and reported in Table 4-9. These results are robust to estimation of the annual mean excluding the month being tested.

Figure 4-5:

Average Monthly Returns – Sydney

This figure presents the average return to each month for the Sydney property market as estimated by each of the three index methods considered – median, hedonic, and repeat-sales.

Panel A: Houses



Panel B: Units

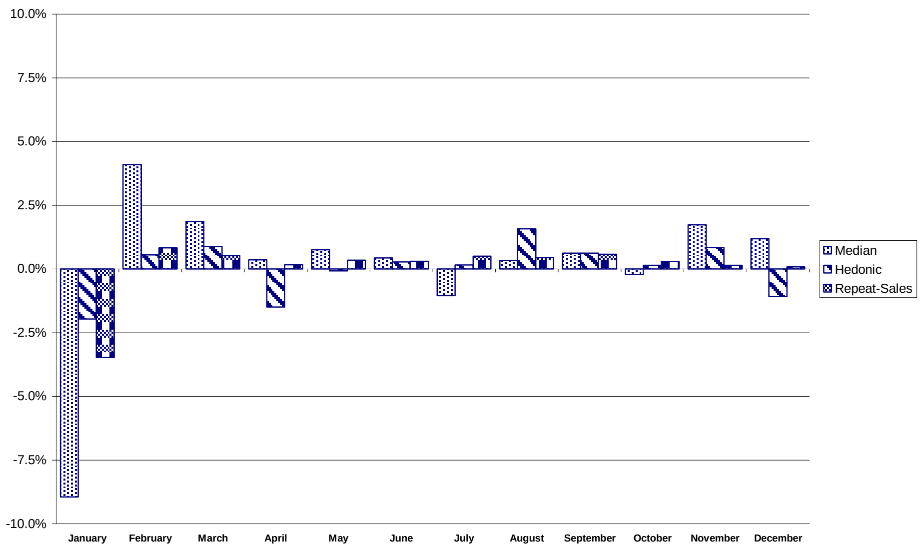
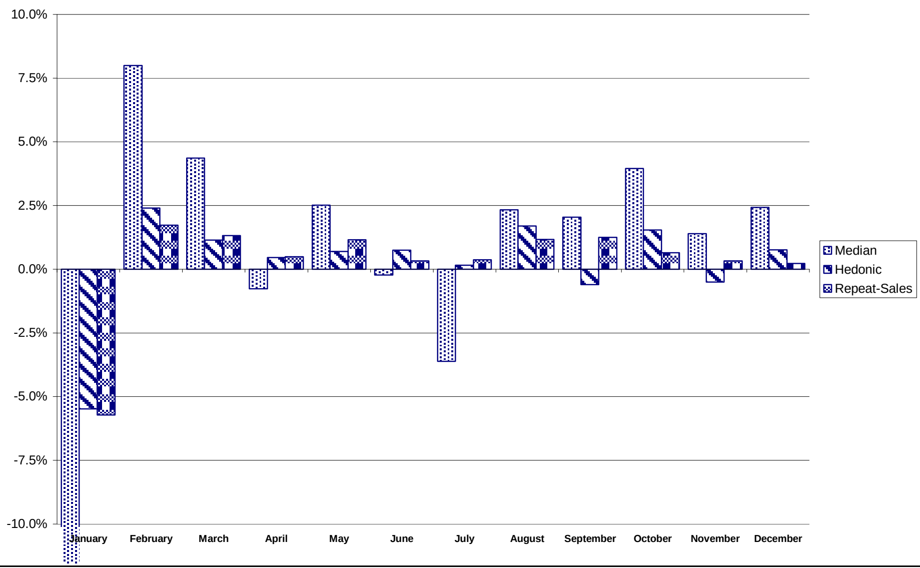


Figure 4-6:

Average Monthly Returns – Melbourne

This figure presents the average return to each month for the Melbourne property market as estimated by each of the three index methods considered -– median, hedonic, and repeat-sales.

Panel A: Houses



Panel B: Units

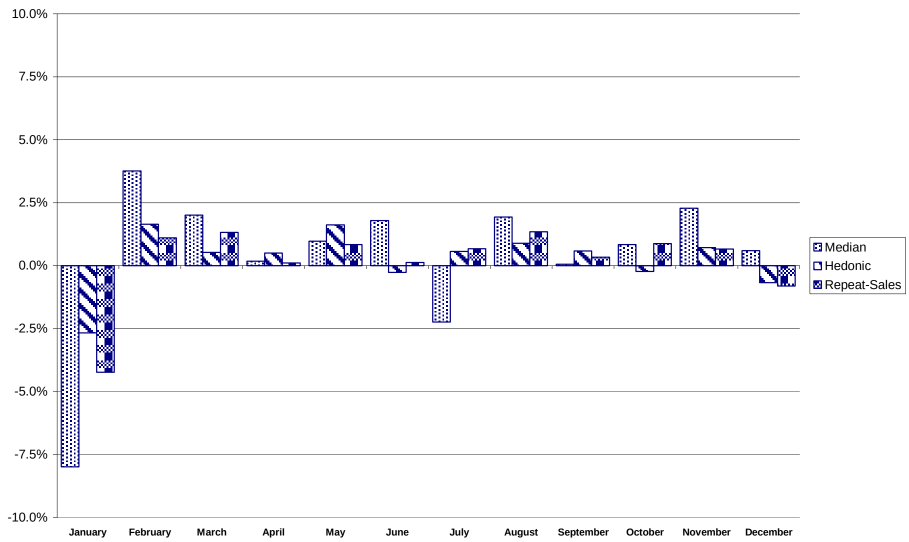


Table 4-9:

Average Monthly Returns – Sydney

This table reports the average return to each month from the alternative indices considered for each market and the statistical significance of equality between that month and all months based on the Student's *t* test. Three asterisks, ***, represent rejection of equality at the 1% level of significance and one asterisk, *, represents rejection at the 10% level of significance.

Month	Sydney Houses			Sydney Units		
	Median (%)	Hedonic (%)	Repeat-Sales (%)	Median (%)	Hedonic (%)	Repeat-Sales (%)
January	-10.58 ***	-0.48	1.70	-8.94	-1.96	-3.47
February	9.08 ***	2.11 *	0.99	4.09 ***	0.55	0.83
March	5.41 ***	1.36	1.10	1.86 *	0.89	0.53
April	-1.17	0.07	0.56	0.36	-1.49	0.17
May	1.52	0.58	0.24	0.76	-0.07	0.35
June	-0.36	0.22	0.30	0.43	0.27	0.30
July	-2.27 *	0.33	0.44	-1.05	0.15	0.51
August	0.77	0.88	0.74	0.34	1.58 *	0.44
September	2.79 *	1.24	0.89	0.62	0.62	0.58
October	1.80	0.28	0.74	-0.21	0.15	0.29
November	2.36	0.38	0.42	1.73 *	0.84	0.15
December	0.45	-1.28	-0.22 ***	1.19	-1.08 *	0.08

These results support the visual observations. Based on the median-price index, all markets demonstrate negative returns in January and positive returns in February which are significantly different from the annual average at the 1% level of significance. Average July returns are negative in all markets considered (significantly different from the average for houses in Sydney and Melbourne), and positive in August.

The hedonic and repeat-sales indices, by comparison, show much less variation in the average return to each month than the median. Houses in Sydney are the only market to show a consistently significant pattern in both constant-quality indices: a negative average return to December, not January!

The results of this parametric testing demonstrate considerable support for *Hypothesis 4_6* and insufficient evidence to reject *Hypothesis 4_4*. These hypotheses are now further tested using regression analysis.

The estimated parameters of fitting the regression model given by Equation 4.3 to the returns from the median, hedonic and repeat-sales indices for Sydney houses are reported in Table 4-10.[42] The results for other markets are consistent to those reported here and are presented in full in Appendix A.

---

[42] The results of regression analyses undertaken in this chapter are consistent with those achieved using White's (1980) heteroskedasticity correction following the procedure set out in SAS.

Table 4-10:

Return Seasonality

This table reports the estimated coefficients of Equation 4.3 and their significance. Significant non-zero estimates at the 1% and 10% level of significance are denoted by *** and *, respectively. The *F* statistic and Adjusted R^2 are also reported.

Month	Median		Hedonic		Repeat-Sales	
	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic
January	-0.1142 ***	-10.54	-0.0103	-2.32	0.0105 *	2.55
February	0.0818 ***	7.55	0.0221 ***	4.97	0.0023	0.56
March	0.0473 ***	4.37	0.0079	1.78	0.0033	0.80
April	-0.0199	-1.83	-0.0051	-1.14	0.0001	0.03
May	0.0052	0.48	0.0035	-0.80	-0.0046	-1.12
June	-0.0071	-0.66	-0.0043	-0.96	-0.0030	-0.74
July	-0.0297 ***	-2.74	0.0000	0.01	-0.0015	-0.37
August	-0.0051	-0.47	0.0041	0.93	0.0008	0.20
September	0.0214	1.97	0.0049	1.10	0.0024	0.59
October	0.0076	0.70	-0.0024	-0.54	-0.0003	-0.07
November	0.0157	1.45	-0.0054	-1.21	-0.0017	-0.41
December	-0.0030	-0.28	-0.0151 ***	-3.40	-0.0084 *	-2.12
F Statistic	17.12 ***		4.29 ***		1.17	

When housing attributes are left unaccounted for in estimates of the market's returns – that is, returns from the median-price index – the joint null hypothesis that, on average, excess monthly returns are equal to each other and equal to zero, is rejected by the *F* statistic in all four markets. Excess returns, based on this measure, are found to be significantly negative in January and July, and positive in February.

Such results do not persist when measured by the constant-quality indices. The results based on repeat-sales returns fail to reject the hypothesis equality between monthly returns for all markets. In all markets but Sydney houses the *F* statistic does not reject the hypothesis of equality using the hedonic index.

These findings confirm the visual observations and the results of *t* tests in the previous section. That is, while there is insufficient evidence to reject *Hypothesis* 4_4, that monthly seasonality is not present in residential real estate returns, there is strong evidence to support the *Hypothesis* 4_6 which posits that such seasonality may be observed in median-price indices. The anecdotal evidence of residential real estate market seasonality may be an artefact of biased market-return estimation methods.

Comparing the fitted regression results where returns to the median-price index are regressed, Panel A, for Sydney houses and units, with the Melbourne markets, their larger monthly coefficient estimates and *F* statistics are expected as Sydney property generally has a higher level of compositional variation. That is, as a result of its wider socio-economic, demographic and underlying topographical variation, there is a wider range of quality in Sydney's property than in Melbourne's.

This result may also explain why the hedonic index's results, Panel B, for Sydney houses suggest some level of monthly seasonality but the repeat-sales index's results do not; an incomplete or mis-specified hedonic function may not capture all this variation, potentially leading to some level of compositional bias as well.

The results of fitting Equation 4.4 are presented in Table 4-11. Two very interesting results emerge from this analysis. Firstly, the results indicate that different types of properties trade through the year. This finding supports *Hypothesis* 4_5. On average, the proportion of higher-quality homes that sell falls significantly in January and July, but increases in February and March.

The F statistics for all markets are highly significant, as are the Adjusted R^2 values at 0.781 (Sydney houses), 0.836 (Melbourne houses), 0.602 (Sydney houses), and 0.497 (Melbourne houses). This shows that the quality of real estate transacting is not constant across months. Instead, it is very closely related to the month in which it is observed to sell; in other words, there is compositional seasonality.

The second observation is that the direction and significance of the changes in the transacted composition of housing from month to month closely resemble the pattern of average monthly returns to the median-price index reported in Table 4-9.

Table 4-11:

Compositional Change

This table reports the estimated coefficients of the monthly dummy variables and their significance from fitting Equation 4.4 for each of the four markets. Significant non-zero estimates at the 1% and 10% level of significance are denoted by *** and *, respectively. The F statistic is also reported. Note that the R^2 statistic in such a model is meaningless given the intercept term has been set equal to zero.

	Sydney Houses	Melbourne Houses	Sydney Units	Melbourne Units
January	-0.2446 ***	-0.3333 ***	-0.0951 ***	-0.0908 ***
February	0.1623 ***	0.2284 ***	0.0676 ***	0.0628 ***
March	0.0849 ***	0.0777 ***	0.0362 ***	0.0234 *
April	-0.0051	-0.0233	-0.0108	0.0009
May	0.0145	0.0328	0.0098	0.0111
June	-0.0162	-0.0252	-0.0123	-0.0078
July	-0.0489 ***	-0.0750 ***	-0.0216 *	-0.0227 *
August	0.0064	0.0313	-0.0151	0.0145
September	0.0345 *	0.0155	0.0018	-0.0167
October	-0.0020	0.0404 *	0.0104	0.0159
November	0.0341 *	0.0100	0.0159	0.0096
December	-0.0022	0.0318	0.0170	0.0210
F Statistic	36.46 ***	52.73 ***	15.66 ***	10.21 ***

Table 4-12 reports the results from fitting the regression model given by Equation 4.5 using returns to each index. This model provides a test for compositional bias in the median house price index. Correlation between changes in the median-price index and the composition of houses trading, measured by the coefficient *b*, is large, positive, and significant for all markets. That is, as the proportion of houses in more expensive suburbs increases, so does the observed median sales price and, consequently, the median-price index. The R^2 values are also quite high for such a simple model, reaching 0.87 for Sydney houses and 0.81 for Melbourne houses, meaning that a high level of variation in median-price index movements is explained by composition change.

These results are not present for movements in the hedonic index. The R^2 values are reasonably low, and the estimates of *b* are insignificantly different from zero for all markets but Sydney houses which are slightly positive at the 10% level of significance. Repeat-sales index movements tell a similar story, though it is now the Melbourne markets that are displaying a weak relationship between index movements and the quality of the houses that are trading.

The conflict in results for the hedonic and repeat-sales indices show that there is no clear, robust relationship between constant-quality index movements and compositional change. The results presented here, however, do support argument for heterogeneity-induced seasonality in the median-price index.

Table 4-12:

Compositional Bias

This table presents the fitted parameter estimates and goodness of fit statistics of Equation 4.5 and their significance for each of the four markets and the alternative indices considered. Statistical significance at the 1% and 10% is denoted by *** and *, respectively.

	Houses			Units		
Parameter	Median	Hedonic	Repeat-Sales	Median	Hedonic	Repeat-Sales
Panel A: Sydney						
a	0.0069 ***	0.0060 ***	0.0061	0.0047 ***	0.0046 ***	0.0041
b	0.5165 ***	0.0181	-0.0646	0.4532 ***	0.0659 *	0.1982
F Statistic	810.18 ***	2.00	0.09	176.82 ***	5.83 *	0.33
Adjusted R^2	0.8729	0.0167	0.0007	0.5998	0.0471	0.0028
Panel B: Melbourne						
A	0.0098 ***	0.0078 ***	0.0083 ***	0.0080 ***	0.0066 ***	0.0068 ***
B	0.3788 ***	-0.0113	0.0254 ***	0.4237 ***	-0.0211	0.0668 *
F Statistic	507.78 ***	1.35	8.71 ***	67.75 ***	0.33	4.28 *
Adjusted R^2	0.8114	0.0113	0.0687	0.3647	0.00278	0.0350

4.5 Summary

This chapter empirically examines the weak-form efficiency of returns to the Sydney and Melbourne residential property markets. Firstly, this chapter tests whether past returns to each market may be used to predict future returns. If founded, this represents a core violation of the efficient markets hypothesis in its weakest form. No prior research of this direct nature has been conducted in the Australian market. Secondly, the returns to the Sydney and Melbourne market are analysed specifically for the presence of seasonality.

The major advantage to the research presented in this chapter over past research is the ability to make cross-methodological comparisons. The extensive database underpinning this research enables estimation of median-price, adjacent-period hedonic and repeat-sales indices. The same methodology in testing for weak-form efficiency is applied to the returns from each index, allowing for a controlled test of the effect of methodology on the results of such research.

ARIMA models are fitted to each index following Box-Jenkins methodology. The returns to all indices display a level of autoregressive predictability, indicating that the residential real estate market is not weak-form efficient. However, the optimal ARIMA specification varies significantly depending on the series it is modelling.

The median-price returns series is statistically significant and negative (positive) at the first (second) lag. This supports the findings of earlier research to have demonstrated excess spurious volatility to median-price indices (Prasad and Richards, 2006). The series also exhibits significant seasonal components at lags 6 and 12.

Returns to the hedonic index, by contrast, are best predicted, albeit weakly by comparison to the median and repeat-sales models, by the observation at lag 3. Repeat-sales index returns are statistically significantly and positively autocorrelated at short lags, indicating a large degree of persistence.

This disparity in results across methods sets the context in which to interpret the conflicting results of past research. Case and Shiller (1989), among others, rely on repeat-sales indices in rejecting residential real estate market efficiency. Rosenthal (2006), on the other hand, finds support for weak-form efficiency in housing market returns using hedonic indices.

Using regression analysis to further test for seasonality in the returns to residential real estate finds that the month of sale has no significant effect on returns in the Sydney and Melbourne house and unit markets, *ceteris paribus*. This is in line with the findings for other residential real estate markets by Case and Shiller (1989), Kuo (1996), Rossini (2000, 2002), and Rosenthal (2006).

However, the results show that seasonality analysis based on a median-price index does demonstrate significant monthly trends. The difference in results between the median and the constant-quality indices is shown to be driven by changes in the type of housing traded in any given month. The results demonstrate that a lower (higher) proportion of expensive properties sell in January (February), which correlates with the lower (higher) average returns to those months in the median-price index.

5. The Impact of Auctions on Prices

5.1 Introduction

This chapter addresses the issue of comparability in housing attributes when comparing prices of residential real estate achieved by alternative sale methods. Following the hypotheses developed in Chapter 3, it is argued that limitations in data and statistical methodology have prevented previous studies from fully controlling for the heterogeneity of housing and bias in the distribution of auctions. Specifically, the endogeneity of sale method choice to the characteristics of the property being sold may have a biasing effect in standard regression analysis. This bias has presented itself as an observed price premium in the housing markets of Australian and New Zealand where auctions are the more likely sale method for higher-priced properties.

This chapter is set out as follows. Section 5.2 details the residential real estate auction market in Australia, with particular reference to Sydney. The methodology to test for the presence of an auction price premium and the presence of endogeneity in the marketing choice decision are set out in Section 5.3. This is followed in Section 5.4 by a description of the data to be used. Section 5.5 presents the results of the hypothesis testing and Section 5.6 summarises the findings and concludes this chapter.

5.2 Institutional Setting

The empirical component of this chapter uses data for house sales in Sydney over the period May 2003 to August 2007. The Sydney housing market is chosen as it represents the largest and most liquid real estate market in Australia: it accounts for approximately 25% of the total Australian housing stock, over 30% of the Australian housing market capitalisation and 40% of gross dollar-value turnover.[43]

The majority of sales of residential real estate in Australia occur through either private treaty negotiations, commonly referred to as private treaty sales, or auctions. Private treaty sales typically occur in the following manner. Following the decision to sell their property, the vendor advertises the property for sale. This advertisement is referred to as property 'listing' and usually conducted through a real estate agent. It includes property details, such as its size and features, and an asking price.

Potential buyers will make bids determined by their personal estimation of the property's value. These bids, however, are not necessarily equal to the vendor's ask price. The vendor will then choose whether to accept the highest bid or wait for another bidder with a higher valuation to arrive.

By contrast, auctions take place in a limited period of time. As with private treaty sales, a property listing is created following the vendor's decision to sell. This listing now also includes the time and date at which the auction of the property will occur.

[43] Source: RPX.

Private bids by potential buyers typically may be made prior to the auction date; however, less than one-fifth of auction sales occur this way. The standard auction method used in sales of residential property in Australia on the auction date is the English auction. This follows a public, ascending bid process, after which the vendor chooses whether to accept the highest bid, measured against their privately held property valuation, referred to as the 'reservation price,' or pass in the property and relist as a sale by private treaty.

5.3 Research Design

This section outlines the methodology used to test *Hypothesis* 5_1, as developed in Chapter 3. This hypothesis posits that the impact of sales mechanism on price is insignificant.

Two approaches are used in the empirical testing of the first of this hypothesis. Firstly, differences in price attributable to the method of sale are examined using hedonic-regression analysis that replicates and extends the work of Dotzour et al. (1998). This approach may be affected by sample selectivity, however, if the choice of sale method is made endogenously with respect to the given property characteristics. Consequently, sample selectivity-corrected regressions following the Heckman two-stage procedure and a matched sampling technique are also applied.

5.3.1 Auctions and Prices

In this section a regression model designed to quantify the impact of economic conditions (external factors) and the characteristics of a property (internal factors) on the price achieved by a given property is developed.

The first empirical model presented, Equation 5.1, replicates the Dotzour et al. (1998) semi-logarithmic hedonic model with respect to the data available in this study:

$$\begin{aligned} p_{i,t} = \alpha + \beta Size_i + \rho Rate_t + \sum_{j=2}^{J} \gamma_j Suburb_{j,i} + \sum_{s=2}^{T} \delta_s Time_{s,t} \\ + \lambda Auction_{i,t} + \varepsilon_{i,t} \end{aligned} \qquad 5.1$$

where,

$p_{i,t}$ is the natural log of price of property i in period t

$Size_i$ is the natural log of the land size (measured in hectares) of the i^{th} property

$Rate_t$ is the average mortgage interest rate in the period of sale, t

$Suburb_{j,i}$ is a dummy variable equal to 1 if the i^{th} property is located in $Suburb_j$, and 0 otherwise

$Time_{s,t}$ is a dummy variable representing the year and month of sale; equal to 1 if the sale of the i^{th} property occurred in the s^{th} period, and 0 otherwise

$Auction_{i,t}$ is a dummy variable equal to 1 if the sales observation of property i at period t used the auction mechanism, and 0 otherwise

α is the fitted regression intercept term

β is the estimated coefficient to $Size_i$, reflecting the expected percentage price change of a one unit change in the land size variable, *ceteris paribus*

ρ is the estimated coefficient to $Rate_t$, reflecting the expected percentage price change of a one unit change in the interest rate variable, *ceteris paribus*

γ_j is the estimated coefficient to the j^{th} suburb, reflecting the expected relative percentage price difference attributable to a location in $Suburb_j$ over $Suburb_1$, *ceteris paribus*

δ_s is the estimated coefficient to s^{th} time period, reflecting the expected percentage price difference attributable to a sale in $Time_s$ over $Time_1$, *ceteris paribus*

λ is the estimated coefficient to $Auction_{i,t}$, reflecting the expected percentage change in price attributable to use of the auction mechanism over a private treaty sale, *ceteris paribus*

$\varepsilon_{i,t}$ is the vector of regression error terms, assumed i.i.d. $N[0,1]$.

The regression model presented in Equation 5.1 is to be estimated using the OLS estimation procedure. To ensure full-rank of the estimated model, note that the coefficients on the first suburb and the first time period are fixed to zero, as is the coefficient of the implicit private treaty variable.

The outcome of interest to this chapter is the significance, sign and size of the auction variable coefficient, λ. A statistically significant estimate of λ represents a rejection of *Hypothesis* 5_1. This hypothesis is formally tested as:

$$H5_1^0 : \lambda = 0$$

$$H5_1^A : \lambda \neq 0$$

A significant positive (negative) estimate of λ indicates the existence of an auction premium (discount), while a statistically insignificant estimate is indication of no price impact deriving from the sale mechanism.

It is expected that the signs of the coefficients of the *Size*, *Rate*, and *Auction* variables are positive. These *a priori* expectations are driven by: (1) the well-documented positive relationship between property size and value; (2) the contemporaneous positive correlation between house prices and interest rates (as discussed in Section 2.1.2); and (3) the use of a particularly parsimonious hedonic model which is likely subject to omitted variable bias and the presence of sample selectivity.

The sale method variable is further subdivided to account for properties which listed for auction but sold either: (1) before the posted auction date; or, (2) were passed in at the auction and subsequently sold in the private treaty submarket. These additional sale 'types' are captured in this chapter's second regression model, Equation 5.2:

$$p_{i,t} = \alpha + \beta Size_i + \rho Rate_t + \sum_{j=2}^{J} \gamma_j Suburb_{j,i} + \sum_{s=2}^{T} \delta_s Time_{s,t} + \lambda_1 Auction_{i,t} + \lambda_2 Before_{i,t} + \lambda_3 After_{i,t} + \varepsilon_{i,t} \tag{5.2}$$

where,

$Before_{i,t}$ is a dummy variable equal to 1 if the sales observation of property i at period t listed for auction and sold prior, and 0 otherwise

$After_{i,t}$ is a dummy variable equal to 1 if the sales observation of property i at period t listed for auction and sold after, and 0 otherwise

λ_1, λ_2 and λ_3 are the estimated coefficients to the three auction types: $Auction_{i,t}$, $Before_{i,t}$ and $After_{i,t}$, respectively, reflecting the expected percentage change in price attributable to a sale by the given auction mechanism over a private treaty sale, *ceteris paribus*

all other variables and coefficients are defined as for Equation 5.1.

It is expected that the results for the coefficients to the *Before* and *After* variables mimic the sequential-auction price-decline anomaly. That is, $\lambda_2 > \lambda_1 > \lambda_3 > 0$. This has been the finding of past research in auctions of multiple units, including real estate in the USA and Singapore, and artwork (Ong, 2006; Lusht, 1994; Beggs and Graddy, 1997).

Dotzour et al. (1998) include variables for property age and condition, design quality, floor-size and primary outer-wall construction material. These are omitted from this analysis as data for these specific variables is not available. Acknowledging this limitation, the relatively parsimonious models, Equations 5.1 and 5.2, are extended to incorporate a greater range of property attributes, as reflected in Equations 5.3 and 5.4:

$$p_{i,t} = \alpha + \sum_{k=1}^{K} \beta_k Attribute_{k,i} + \sum_{j=2}^{J} \gamma_j Suburb_{j,i} + \sum_{s=2}^{T} \delta_s Time_{s,t} + \rho Rate_t + \lambda Auction_{i,t} + \varepsilon_{i,t} \qquad 5.3$$

$$p_{i,t} = \alpha + \sum_{k=1}^{K} \beta_k Attribute_{k,i} + \sum_{j=2}^{J} \gamma_j Suburb_{j,i} + \sum_{s=2}^{T} \delta_s Time_{s,t} + \rho Rate_t + \lambda_1 Auction_{i,t} + \lambda_2 Before_{i,t} + \lambda_3 After_{i,t} + \varepsilon_{i,t} \qquad 5.4$$

where,

$Attribute_{k,i}$ represents a vector of k hedonic variables relating to the i^{th} property

β_k is the vector of estimated coefficient corresponding to each of the k hedonic variables, reflecting for each the percentage price change of a one unit change in that variable, *ceteris paribus*

all other variables and coefficients are defined as for Equation 5.2.

Specifically, the hedonic property variables included are: the property land size, the number of bedrooms, the number of bathrooms, the ratio of bedrooms to bathrooms, the number of additional rooms, the number of car spaces supported by the property, swimming pool, air-conditioning, scenic view, and waterfrontage. These final four variables are binary observations, entered into the model as dummy variables equal to 1 if the attribute is present in the i^{th} property, and 0 otherwise.

Implicit in the use of these models alone to estimate the value of the auction premium is the assumption that the decision to use the auction mechanism is made exogenous to property characteristics and other factors affecting the property price. If the properties that sell by auction are, on average, more expensive than those that sell by private treaty, least squares estimates of λ in Equations 5.1 through 5.4 will be positively biased. That is, the fitted coefficient will overestimate the value of the auction mechanism (Greene, 2003).

The next section details the methodology to be undertaken to determine if the same method choice is made endogenously using a discrete choice modelling technique.

5.3.2 Sample Selectivity

This section presents the methodology for the Heckman two-stage sample selectivity correction procedure. The two stages involved are: (1) discrete-choice modelling, and (2) selectivity-corrected regression.

Detecting and correcting sample selectivity in samples of properties that sell by different methods is motivated in part by the findings in the literature of sample selectivity among those properties that sell using the auction mechanism. That is, it is expected that certain types of properties are more likely to sell by auction than by private treaty. In residential real estate markets such as Australia, the tendency is for more expensive or unique properties to use the auction mechanism. This auction sample selectivity induces an omitted variable bias into least squares regression estimates, that presents as a positive bias to the auction dummy variable coefficient (Greene, 2003).

The discrete choice modelling to be undertaken examines the extent to which property quality and characteristics influence the choice of sale method. To test for this endogeneity in the choice of sale method with respect to attributes of the property that are jointly positively related to price, a binary probit model is estimated by maximum likelihood (MLE). The explicit model to be estimated in this dissertation is given by Equation 5.5:

$$\begin{aligned} Auction_i = & \sum_{k=1}^{K} \beta_k^* Attribute_{k,i} + \sum_{j=2}^{J} \gamma_j^* Suburb_{j,i} + \sum_{s=2}^{T} \delta_s^* Time_{s,t} \\ & + \rho^* Rate_t + \zeta_i \end{aligned} \qquad 5.5$$

where,

$Auction_i$ is the binary dependent variable, equal to 1 if the i^{th} property listed for sale by auction, and 0 otherwise

ξ is the error term of the probit model, assumed i.i.d. $N[0,1]$

the estimated coefficients – β^*, γ^*, δ^*, ρ^* – represent a measure of the predicted change in the probability of an auction of the i^{th} property in response to a given unit change in the respective independent variable, *ceteris paribus*

all other variables are defined as for Equation 5.4.

Specifically, under the structure of the probit model, the probability of the auction sale method being chosen can be estimated from Equation 5.5 given,

$Auction_i = 1$ if $z_i^* > 0$

$Auction_i = 0$ if $z_i^* \leq 0$

$$P(Auction_i = 1) = \Phi\left[\sum_{k=1}^{K} \beta_k^* Attribute_{k,i} + \sum_{j=2}^{J} \gamma_j^* Suburb_{j,i} + \sum_{s=2}^{T} \delta_s^* Time_{s,t} + \rho^* Rate_t\right]$$

$$P(Auction_i = 0) = 1 - \Phi\left[\sum_{k=1}^{K} \beta_k^* Attribute_{k,i} + \sum_{j=2}^{J} \gamma_j^* Suburb_{j,i} + \sum_{s=2}^{T} \delta_s^* Time_{s,t} + \rho^* Rate_t\right]$$

where z_i^* is an unobserved selection variable determining the method by which a property is sold and Φ denotes the cumulative distribution function (CDF) of the Gaussian Normal distribution.

It is expected that the estimated coefficients of attributes positively correlated with property sales price, such as the number of bedrooms, waterfrontage and certain locations, will be positive. This finding would support the results of similar modelling presented in Dotzour et al. (1998) and indicates a degree of sample selectivity in the choice of sale method. That is, certain types of properties (those with attributes that have positive probit model coefficients) are more likely to sell by auction. As a result, extra care needs to be taken in the interpretation of the results from least squares regression modelling described in Section 5.3.1. If the OLS model fails to fully and correctly account for all variables explaining price and sale method choice, the estimated coefficients will exhibit omitted variable bias (Greene, 2003). The next section discusses the method to be used to overcome the omitted variable induced specification error arising as a result of sample selectivity.

The Heckman two-stage procedure can be used to correct for endogeneity in treatment variables (Winship and Mare, 1992). To see how, consider the model of a continuous variable, Y_i, expressed as a linear combination of a set of independent continuous variables, X_i, and a binary [0,1] treatment variable, Z_i:

$$Y_i = X_i^{'}\beta + Z_i\lambda \qquad 5.6$$

This may be rewritten for the alternative outcomes of Z_i as:

$$\left[Y_i \middle| Z_i = 0\right] = Y_i^0 = X_i^{'}\beta \qquad 5.7$$

$$[Y_i | Z_i = 1] = Y_i^1 = X_i^{'}\beta + \lambda \qquad 5.8$$

The case where data are only available for either Z_i = 0 or Z_i = 1 frequently occurs in natural studies in sociology and economics. Under such conditions, only either Equation 5.7 or 5.8 may be estimated, but not both. This demonstrates the existence of an overt sample selection issue, as described by Heckman (1978, 1979). Greene (2003: 788) shows that the sample selectivity correction method applied in such circumstances can be extended to studies affected by selectivity in the regressor treatment effects.

To apply the Heckman method, two steps are undertaken: probit modelling of the treatment variable to estimate a sample selectivity variable, and re-estimation of a least squares regression including this variable as a regressor. From the probit coefficient estimates, obtained using the methodology outlined in Section 5.3.2, the Inverse Mills Ratio (IMR), denoted by κ, is estimated for every observation as,

$$\hat{\kappa}_i = \frac{\varphi\left[\sum_{k=1}^{K}\beta_k^* Attribute_{k,i} + \sum_{j=2}^{J}\gamma_j^* Suburb_{j,i} + \sum_{s=2}^{T}\delta_s^* Time_{s,t} + \rho^* Rate_t\right]}{\Phi\left[\sum_{k=1}^{K}\beta_k^* Attribute_{k,i} + \sum_{j=2}^{J}\gamma_j^* Suburb_{j,i} + \sum_{s=2}^{T}\delta_s^* Time_{s,t} + \rho^* Rate_t\right]} \qquad 5.9$$

where φ and Φ denote the probability density function (PDF) and the CDF of the Gaussian Normal distribution, respectively.

The estimates of κ_i are entered into the regression model given by Equation 5.10 as the sample selectivity variable:

$$p_{i,t} = \alpha + \sum_{k=1}^{K} \beta_k Attribute_{k,i} + \sum_{j=2}^{J} \gamma_j Suburb_{j,i} + \sum_{s=2}^{T} \delta_s Time_{s,t} + \rho Rate_t + \lambda Auction_{i,t} + \beta_\kappa \kappa_i + \varepsilon_{i,t} \quad 5.10$$

where,

κ_i denotes the sample selectivity variable, as calculated from Equation 5.9

β_κ the coefficient of the sample selectivity variable

all other variables are defined as for Equation 5.1.

Inclusion of the sample selectivity variable, κ_i, provides the sample selectivity correction. The coefficient of the sample selectivity variable, β_κ, in Equation 5.10 indicates the presence of sample selectivity. Statistical significance to this coefficient indicates the presence of sample selection bias in the sample with respect to the use of auctions as a sale method. As such, its inclusion in the price regression is justified and appropriate.

5.3.3 Matched Sampling

As a robustness test of *Hypothesis 5_1*, a matched sampling technique is used. The matched sample experimental design is commonly employed as an alternative to two-stage selectivity correction regressions to reduce bias in observational studies, where random assignment of the treatment variable of interest is not possible.

Matched sampling refers to the selection of treatment units and control units that have similar values for one or more matching variables. In studying residential real estate prices, properties are not assigned a sale method by random selection; it is observed in each case. A Monte Carlo study by Rubin (1979) demonstrates that combining matched sampling and regression adjustment is highly effective in controlling for bias induced by non-random sampling.

The matched sampling process involves several steps. Firstly, all auction sales in the database described below must be identified and separated. For each of the auction observations a single private treaty sale is then selected as a match. This requires the two observations to be taken from the same location and time period, proxied by statistical subdivision and the month of sale. For auction sales achieving multiple private treaty matches, the private treaty sale observation with the nearest land size to the auctioned property is selected.

Using the subset of observations in this matched sample, the prices achieved by alternative sale methods are analysed. Prices are compared using dependent *t* test analysis on the matched-pairs sample, and the regression models given by Equations 5.1 and 5.3 described in the previous section are re-estimated.

5.4 Data and Descriptive Statistics

The population of house sales in NSW, Australia, from 1 May 2003 to 31 August 2007 is obtained from RPX, having originated from the NSW VG and RPX's real estate agent client base. Sales are matched using a unique property identifier to a listings history and attributes database that is also obtained from RPX. This enables the hedonic attributes of a property and the method by which a property was sold to be identified to each sales record.

RPX has constructed the attributes database from information received from real estate agents and cross-referenced to data contained in newspaper and internet sales-listings. This is necessary due to the fact that in Australia land tax is calculated based on the unimproved value of land and, consequently, very few hedonic attributes are recorded by the VG. The sample of data used, that is to which hedonic information is able to be matched, represented approximately 80% of all sales.

The study focuses on the Sydney market. Sydney is taken as the Sydney statistical division, excluding Gosford-Wyong, as defined within the ASGC.[44] Several further filters are applied to the data to remove data errors and extreme outliers. For prices, erroneous observations (prices less than $1,000 or greater than $100 million) are removed. Remaining outliers in the upper and lower 10% of prices are removed. Observations with missing values for the necessary explanatory and matching variables are removed, except for the binary variables – pool, waterfront, view and

---

[44] *Standard Geography Volume I – Australian Standard Geographical Classification*, ABS, Catalogue 1216, July 2006.

air-conditioning – where null entries are treated as not present. Extreme land size entries are filtered from the sample by removing the upper and lower 5% of values. Bedroom and bathroom variables are filtered by removing entries less than 1 (that is, zero-bedroom houses are removed), and values in the top 5% of observations. Finally, sale observations identified as failed auctions are removed from the sample.[45]

A final variable required in this study is the monthly average standard mortgage variable rate. This is sourced from the RBA's *Bulletin of Statistical Tables*.[46]

Table 5-1 provides descriptive statistics for sales volume and prices. In the sample period, May 2003 to August 2007, there were 89,542 house sale observations with full sales and hedonic information. Of these, auctions represented 26% of the market.

---

[45] Failed auction sales, which sold after the auction date in the private treaty market, are defined as those sales that occurred by private treaty any time in the 180 days after an advertised auction date. It is not possible from the available data to distinguish those auctions which failed and were transferred to the private treaty market from those that were voluntarily moved to the private treaty market prior to the auction date.

[46] Interest Rates (Section F), Indicator Lending Rates, *Bulletin Statistical Tables*, RBA, December 2007.

Table 5-1:

Descriptive Statistics

This table presents the descriptive statistics for the sample of Sydney house sales for the period 1 May 2003 to 31 August 2007, including observation frequency, price distribution, and the average composition of hedonic attributes. Panel A contains statistics for the entire sample of sales. Panel B presents separate statistics for auction sale properties that sell at, before and after the nominated auction date.

Panel A: All Sales

	All Sales	Private Treaty	Auctions
Observations	89,542	65,926	23,616
Sale Method (%)		73.63	26.37
Average Price ($)	641,870	558,734	873,950
Median Price ($)	536,375	481,000	750,000
Standard Deviation ($)	397,152	311,340	504,285
Land Size (m^2)	593.33	618.33	523.43
Bedrooms	3.30	3.33	3.23
Bathrooms	1.66	1.65	1.70
Other Rooms	1.01	1.01	1.01
Car Spaces	0.98	1.00	0.91
Pool (%)	11.13	10.74	12.23
Air-Conditioning (%)	10.72	10.78	10.55
Scenic View (%)	6.22	5.06	9.44
Waterfront (%)	0.62	0.44	1.12

Panel B: Auction Sales

	Sold at Auction	Sold Before	Sold After
Observations	9,679	4,744	9,193
Average Price ($)	868,672	993,036	818,055
Median Price ($)	748,000	855,000	700,000
Standard Deviation ($)	506,777	535,309	473,910

The average price of Sydney houses during the sample period is $641,870. The median price is $536,375, indicating negative skew in the distribution of prices. Prices at auction are $873,950, on average, whereas private treaty sales achieve an average price of $558,734. Median prices show a similar difference, and on these measures alone auction sales appear to achieve a price premium. However, these raw figures are of little value, as it is expected that the decision to auction is not random, but rather a decision made endogenously with respect to the characteristics of the property.

Summary statistics of the hedonic characteristics by sale method for the sample of house sales are given in Table 5-1. The average property land size was almost 600 m^2, and the average house had 3 bedrooms, 1.6 bathrooms, 1 other room and 1 car space. It is observed that 11.1% of properties have a pool, 10.7% are air-conditioned, 6.2% have a scenic view and 0.6% have water frontage. Interestingly, one-fifth more auction properties have a pool than properties sold by private treaty, almost twice as many have a scenic view, and approximately three times as many have waterfrontage.

Panel B of Table 5-1 presents the disaggregated prices by auction sale timing. Separating auction sales into those which sold at auction, those which sold prior to auction,[47] and those which were passed in at auction and subsequently sold in the private treaty submarket reveals that while auctions on average sell at prices more than $300,000 (36%) over the average private treaty price, there is also a negative relationship between the timing of the sale and the price realised.

---

[47] A sale before auction is defined as a sale any time in the 30 days prior to an advertised auction date.

Sales that occur prior to auction sell, on average, at \$993,036, which is 14% higher than sales at auction (\$868,672), which in turn is 6% higher than failed auctions which subsequently sell in the private treaty submarket (\$818,055). This result supports the findings of sequential auctions in commercial real estate and art, although it runs contrary to the empirical results of price discovery in sequential auctions of residential real estate from the USA and Singapore markets (Ong, 2006).

Table 5-2 presents statistics describing the distribution of auction sales across the geographical regions of Sydney as defined by ASGC statistical subdivision codes. These statistics reveal some interesting relationships between prices and auction prevalence.

The majority of auctions in Sydney occur in the areas with the higher average prices. Calculated as the proportion of auction sales to a given area (Auction Distribution), collectively over 80% of auctions in Sydney occur in the East, Inner, Inner West, Lower North Shore, Central North, Sutherland-St George region and Northern Beaches. These areas, however, account for only 54% of all sales during the sample period.

The average private treaty and auction prices in these areas range from \$614,806 to \$1,520,764 and \$687,111 to \$1,597,994, respectively. Furthermore, the majority of sales in four of these seven most expensive areas in Sydney take place by auction: Sydney's East (70%), Inner (58%), Inner West (61%), and Lower North Shore (51%).

Table 5-2:

Regional Distribution of Auctions

This table presents the distribution of sales and auctions across the various subdivisions of the Sydney housing market. Sale by auction indicates the proportion of sale in that region undertaken by auction, and auction distribution represents the proportion of all auction sales in Sydney to occur in the given region. The table also includes average private treaty and auction price for each regional subdivision.

Subdivision	Total Sales	Auctions (%)	Auction Distribution (%)	Average Private Treaty Price ($)	Average Auction Price ($)
Eastern Suburbs	3,136	70.22	9.32	1,530,764	1,597,994
Inner West	3,367	60.71	8.66	797,758	832,652
Inner City	9,226	58.40	22.82	692,111	754,675
Lower North Shore	6,832	50.57	14.63	1,024,263	1,175,725
St George-Sutherland	9,593	27.60	11.21	614,806	687,111
Canterbury-Bankstown	4,592	26.02	5.06	459,884	488,347
Northern Beaches	6,634	24.83	6.97	875,463	1,127,640
Central Northern	9,793	21.73	9.01	694,429	898,116
Fairfield-Liverpool	6,443	14.90	4.07	387,519	375,343
Central West	6,040	12.65	3.24	467,687	472,986
Outer South West	5,514	6.18	1.44	338,065	314,453
Blacktown	8,732	5.12	1.89	383,762	345,999
Outer West	9,640	4.12	1.68	345,485	340,395

It is worth noting that the higher incidence of sales by auctions occurs in locations that are closer to the Sydney Central Business District, the harbour and the ocean coastline. For example, the four areas listed with over half of all sales taking place by auction border the CBD. Location, particularly proximity to the CBD in highly urbanised areas, is known to be a significance driver of property value. Harbour and coastline proximity are also considered attractive property features in Sydney. Properties in these areas are also expected to exhibit more unique features, such as waterfrontage, than areas away from the CBD.

By contract, less auction activity is recorded in the less expensive areas of Sydney. Auctions account for less that 15% of sales in five of the six cheapest regions: Fairfield-Liverpool (15%), Central Western (13%), Outer South Western (6%), Blacktown (5%), Outer Western (4%). Sales in these cheaper areas, with prices ranging from \$338,065 to \$467,687 for sales by private treaty and \$314,453 to \$488,347 for sales by auction, account for approximately 56% of all sales but only 17% of auctions for this sample.

The descriptive statistics presented in this section suggest that the average property to sell by auction is more expensive than the average property that sells by private treaty. The average auction property, however, is also more likely to have unique features (such as waterfrontage) and sell in more expensive locations. The next section reports the results of applying the statistical methodology outlined in Section 5.3 to determine whether the auction mechanism itself has an effect on price once such variations across properties are accounted for.

5.5 Results

This section presents the results of applying the data described in the previous section to the methodology and hypothesis testing outlined in Section 5.3.

5.5.1 Hedonic Regression

To assess the auction price impact, several additive hedonic-regression models are described in methodology section. The first of these, Equation 5.1, replicates the empirical model of Dotzour et al. (1998) given the data available to this study.

Treating the relatively parsimonious Dotzour et al. (1998) replication as a restricted model, a more fully specified hedonic model is developed, as given by Equation 5.3. This hedonic model incorporate a wider range of property attributes, including the number of bedrooms, bathrooms, car spaces and other rooms and features such as pool, air-conditioning, scenic views and waterfrontage.

Table 5-3 presents the empirical results from estimating these models using the data for the Sydney housing market over the period May 2003 to August 2007. These results indicate a significant price premium to auctions in both the parsimonious model and the more completely specified model.

Table 5-3:

Hedonic Regression

This table reports the fitted results of the model given by Equations 5.1 (restricted OLS model) and 5.3 (hedonic OLS model). The estimated coefficients and several diagnostic statistics are presented. Due to the number of suburb and monthly time dummy variables, a summarised description of the statistical performance of these variables is provided. Full results are available from the author upon request. Statistical significance at the 5% large sample size adjusted critical t value is given by *.

	Restricted OLS Model		Hedonic OLS Model	
Variable	Coefficient	t Statistic	Coefficient Estimate	t Statistic
Constant	13.1906 *	81.89	12.7200 *	94.68
Auction	0.0247 *	10.93	0.0227 *	12.01
Land Size	0.2905 *	115.01	0.2232 *	103.41
Interest Rate	0.1029 *	4.27	0.0787 *	3.91
Bedrooms			0.0829 *	55.67
Bathrooms			0.0609 *	26.70
Bed/Bath			-0.0064 *	-3.47
Other Rooms			0.0302 *	13.51
Car Spaces			0.0269 *	37.02
Pool			0.0375 *	19.43
Air-Conditioning			0.0274 *	14.47
Scenic View			0.0194 *	7.47
Waterfront			0.2036 *	24.69
	Significance at 1% (%)	Significance at 10% (%)	Significance at 1% (%)	Significance at 10% (%)
Suburbs (J = 630)	87.62	92.38	88.57	92.38
Months (T = 50)	62.00	80.00	58.00	84.00
Observations	75,605		75,605	
Critical t	3.337		3.336	
Adjusted R^2	0.8482		0.8956	
F Statistic	619.54 *		928.33 *	

The estimated coefficient of the auction dummy variable, λ, shows a statistically significant premium attributable to the auction sales mechanism of 2.3% (2.5% in the restricted model) measured against the sample-size adjusted critical t value.[48]

Sample-size adjusted t statistics are used to control for potential bias in the standard t statistics arising from the large sample size used in this paper, and ensure consistency in the interpretation of results from the smaller matched sample (Connolly, 1989; Davidson and Faff, 1999). This is consistent with the findings of research into the auction-price effect undertaken in Australia and New Zealand by Dotzour et al. (1998), Lusht (1996) and Newell et al. (1993).

The estimated coefficients on the hedonic variables have the expected signs and relative magnitudes, and are all statistically significant based on the sample size adjusted critical t value. Land size and waterfrontage, in particular, are highly positively correlated with price. The Adjusted R^2 statistics of the restricted and expanded models are 0.8482 and 0.8956, respectively, demonstrating the large proportion of price variation explained primarily by land size and location.

The results from this regression analysis are consistent with the findings in Dotzour et al. (1998) and other hedonic approaches in Australia, such as Lusht (1996). That is, that all else considered, there is an auction premium across house sales. Although

---

[48] These critical values are estimated as,

$$t^* = [(T-k)(T^{1/T}-1)]^{1/2}$$

where T and k represent the sample size and number of parameters to the model, respectively (Connolly, 1989).

smaller than the findings of earlier research, this result demonstrates that the data used in the study behaves similarly to the data used in previous studies.

In estimating the models given by Equation 5.1 and 5.3, only observations pertaining to authentic private treaty sales and auctions which sold on the date of the auction are included. That is, sales which listed for auction but sold either before or after the auction date are removed.

The results of estimating the regression models that account for the differences in prices to auction sales which occur before, on, and after the auction date are considered. Two models are again tested: a parsimonious restricted model reflecting the Dotzour et al. (1998) variables and a more completely specified hedonic model, given by Equations 5.2 and 5.4, respectively. The results of these fitted models are presented in Table 5-4.

The results from fitting Equations 5.2 and 5.4 demonstrate the persistence of a price premium to auctions that sell on the auction date of 2% to 2.3%. The expectation that the coefficient to *Before*, λ_2, would be larger than that to the dummy variable *Auction*, λ_2, is founded; the results to both the restricted and full hedonic models indicate a significant price premium in the range of 5.1% to 6.8% in auction sales that sell before the auction date.

Table 5-4:

Auction Timing Regression Results

This table reports the fitted results of the model given by Equations 5.2 and 5.4. The estimated coefficients and several diagnostic statistics are presented. Due to the number of suburb and monthly time dummy variables, a summarised description of the statistical performance of these variables is provided. Full results are available from the author upon request. Statistical significance at the 5% sample size adjusted critical *t* value is given by *.

	Restricted OLS Model		Hedonic OLS Model	
Variable	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic
Intercept	13.4173 *	88.24	12.846 *	101.58
Auction	0.0227 *	9.84	0.0205 *	10.68
Sold Before	0.0682 *	22.43	0.0514 *	20.31
Sold After	0.0067	2.95	-0.0142 *	-7.51
Land Size	0.3057 *	129.32	0.2316 *	20.31
Interest Rate	0.0784 *	3.44	0.0634	3.35
Bedrooms			0.0819 *	58.34
Bathrooms			0.0669 *	31.22
Bed/Bath			-0.0039	-2.19
Other Rooms			0.0312 *	15.58
Car Spaces			0.0278 *	39.94
Pool			0.0407 *	22.51
Air-Conditioning			0.0309 *	17.27
Scenic View			0.0271 *	11.58
Waterfront			0.2027 *	28.56
	Significance at 1% (%)	Significance at 10% (%)	Significance at 1% (%)	Significance at 10% (%)
Suburbs (J = 635)	87.72	91.97	88.50	91.97
Months (T = 50)	46.00	60.00	46.00	66.00
Observations	89,542		89,542	
Critical t	3.364		3.364	
Adjusted R^2	0.8507		0.8969	
F Statistic	740.25 *		1115.19 *	

The results for the prices of auction sales that sell after the auction date are less clear. The restricted model indicates the price premium to sales that are negotiated after the auction date is neither economically nor statistically significant. The results from the hedonic model, however, show a statistically significant estimate of -0.0142 to the coefficient of *After*, λ_3, indicating that sales after auction actually attract a discount relative to private treaty sales.

The sign, size, and significance of the other explanatory variables included in these models are consistent with the results from fitting Equations 5.1 and 5.3. The overall fit of the restricted and hedonic models presented in Table 5-4 is comparable to that reported for the models that exclude sales that occur before and after the auction date.

The results of the hedonic-regression modelling presented through this section broadly support the findings presented in earlier research undertaken in Australia and New Zealand of a price premium to auctions over private treaty sales. As identified in Section 2.1.4, a limitation of the hedonic-regression approach is the need to fully and correctly specify the model. If certain property attributes which are correlated with the decision to sell a property by auction are not included in the model, typically when such data are unavailable, the regression results may be affected by omitted variable bias (Greene, 2003). The statistics presented in Section 5.4 indicate that, on average, properties with pools, views, and waterfrontage are more likely to sell by auction, for example. Auctions are also more likely to take place in more expensive areas. The next section considers the impact of property attributes on the decision to auction and represents the first step in identifying and correcting for sample selectivity.

5.5.2 Probit Model

This section presents the results of probit modelling to determine if certain factors influence the sale mechanism decision.

Table 5-5 presents the results of the probit sale mechanism decision model, given by Equation 5.5. The results from a restricted model that includes only land size, location, and month of sale are also reported. Given these models take as their dependent variable the sale method, where $Auction_i$ equals 1 for properties sold by auction and $Auction_i$ equals 0 for those sold by private treaty, positive coefficients represent characteristics that increase the likelihood of a given property selling by auction, *ceteris paribus*.

The results indicate that larger properties, with more rooms and unique features, such as air-conditioning, swimming pool, a scenic view, or waterfrontage, are more likely to sell by auction than private treaty. The observation of a positive coefficient to land size in the probit model conflicts with the univariate average statistic reported in Table 5-1, emphasising the need for controls on the range of property features when conducting research of this kind using residential real estate data.

Table 5-5:

Probit Model

This table reports the results of MLE fitting of the probit model given by Equation 5.5. The coefficient estimates and related diagnostic statistics are presented. Due to the number of dummy variables used to account for the effect of suburb and month of sale in this model, these estimated coefficients are not presented. Instead, a summarised description of the statistical performance of these variables is provided. Full results are available from the author upon request. Coefficient estimate significance at 1% level of significance is denoted by *.

	Restricted MLE Model		Hedonic MLE Model	
Variable	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic
Intercept	-0.7641 *	-3.99	-4.2066 *	-3.57
Land Size	0.3528 *	19.15	0.3132 *	16.35
Interest Rate			0.6725 *	3.81
Bedrooms			0.0478 *	3.49
Bathrooms			-0.0300	-1.43
Bed/Bath			-0.0129	-0.74
Other Rooms			0.259 *	14.34
Car Spaces			-0.0055	-0.78
Pool			0.1416 *	8.09
Air-Conditioning			0.2360 *	13.32
Scenic View			0.2634 *	12.46
Waterfront			0.2096 *	3.43
	Significance at 1% (%)	Significance at 10% (%)	Significance at 1% (%)	Significance at 10% (%)
Suburbs (J = 634)	48.98	63.62	47.32	61.83
Months (T = 50)	100.00	100.00	16.00	34.00
Observations	89,542		89,542	
Log-Likelihood	-37,827		-37,409	
Maximum Absolute Gradient	2.479		4.081	

From the descriptive statistics presented in Section 5.4, it can be seen that only a small proportion of properties have a pool (11.13%), air-conditioning (10.72%), scenic view (6.22%), or waterfrontage (0.62%). Consequently, properties with any of these features are considered more unique. The results from probit modelling indicate that more unique properties – that is, properties with any of these less common features – are more likely on average to sell by auction, *ceteris paribus*.

This finding broadly supports the conclusions of previous research such as Dotzour et al. (1998) and indicates a degree of sample selectivity among those properties that sell by auction. If this selectivity is not completely accounted for through a completely and correctly specified regression model, a sample selectivity bias may be induced. As outlined in the methodology section of this paper, the Heckman two-stage procedure provides a method by which sample selectivity may be corrected in OLS regressions using the IMR from the probit model to form a sample selectivity variable. The results from this sample selection correction procedure are now presented and discussed.

5.5.3 Selectivity-Corrected Regression

This section presents and discusses the results of a two-stage Heckman model. Table 5-6 reports the Heckman second-stage regression results. The most important results from this are: (1) the highly significant estimated coefficient to sample selectivity variable; and (2) the absence of statistical significance to the auction dummy variable coefficient. These results contradict the findings of Dotzour et al. (1998) which may have been affected by small sample and hedonic data limitations.

Table 5-6:

Selectivity-Corrected Regression

This table reports the results of OLS fitting of the second-stage Heckman regression model given by Equation 5.10 on the full sales sample. The estimated coefficients of the fitted model and their significance are reported with several regression diagnostic statistics. Statistical significance at the 5% sample size adjusted critical *t* value is given by *. Due to the number of dummy variables used to account for the effect of suburb and month of sale in the regression, a summarised description of the statistical performance of these variables is provided. Full results are available from the author upon request.

Variable	Coefficient Estimate	*t* Statistic
Intercept	10.6614 *	77.72
Auction	0.0088	3.23
Sample Selectivity	0.5955 *	39.58
Land Size	0.3642 *	93.83
Interest Rate	0.3386 *	16.88
Bedrooms	0.1024 *	68.65
Bathrooms	0.0514 *	23.72
Bed/Bath	-0.0107 *	-6.12
Other Rooms	0.1401 *	41.08
Car Spaces	0.0245 *	35.14
Pool	0.1045 *	43.15
Air-Conditioning	0.1405 *	42.69
Scenic View	0.1394 *	37.82
Waterfront	0.2783 *	38.03
	Significance at 1% (%)	Significance at 10% (%)
Suburbs ($J = 630$)	93.49	96.35
Months ($T = 50$)	86.00	90.00
Observations	75,605	
Critical *t*	3.336	
Adjusted R^2	0.8980	
F Statistic	1130.77 *	

The rejection of the null hypothesis that β_κ is zero strongly indicates the existence of a sample selection bias between properties that sell by auction and private treaty in the data, and consequently that a sample selection bias correction in the hedonic regression is appropriate. With this correction, there is insufficient evidence to reject the null hypothesis that the auction coefficient, λ, is zero. All other hedonic variables maintain their relative significance and magnitude when the Heckman sample selectivity variable is included in this regression.

The implication of this result is that no price premium to auctions exists. As such the rational expectations requirement of pricing equality between auctions and private treaties is supported. Furthermore, the results suggest that uncorrected bias induced by sample selectivity may explain the price premium observed in previous studies of the Australian and New Zealand housing markets (Lusht, 1994; Dotzour et al., 1998).

The following section uses the matched sampling technique outlined in the methodology section to test *Hypothesis* 5_1.

5.5.4 Matched Sampling

This section uses the matched sampling technique to explore the effect of sale method on price. Table 5-7 provides the descriptive statistics of the prices for the matched samples obtained with replacement. The procedure matched 7,933 private treaty sales to auction sales, without replacement. Some observations are lost when there is low liquidity (no sales recorded) in certain areas and time periods. The private treaty average and median prices are now closer to the average and median prices of auctioned properties.

Table 5-7:

Matched Sample Descriptive Statistics

This table presents the descriptive statistics of the matched sample (obtained following the with-replacement rule). Panel A reports the number of matched observations as well as several price statistics. These include the average, median, and standard deviation of price in the matched sample, as well as between the post-matched auction and private treaty subsamples. Panel B provides information on the average composition of hedonic attributes in the matched sample and the post-matching auction and private treaty subsamples.

Panel A: Matched Sample Observations and Prices

		Price		
Sample	Observations	Average	Median	Standard Deviation
All Sales	15,866	785,754	680,000	443,346
Private Treaty	7,933	747,472	650,000	412,666
Sold at Auction	7,933	824,035	716,000	468,947

Panel B: Matched Sample Hedonic Attributes

Variable	All Sales	Private Treaty	Auctions
Land Size (m^2)	505.65	495.10	516.21
Bedrooms	3.14	3.14	3.15
Bathrooms	1.60	1.62	1.59
Other Rooms	1.01	1.01	1.01
Car Spaces	0.86	0.85	0.87
Pool (%)	9.16	8.89	9.43
Air-Conditioning (%)	8.34	8.09	8.60
Scenic View (%)	6.37	5.04	7.69
Waterfront (%)	0.81	0.61	1.01

The pattern of hedonic attributes in the matched sample, presented in Panel B of Table 5-7, shows that the differences in proportions of waterfront properties and those with views is now smaller between the auction and private treaty subsamples, when compared to the full sample breakdowns presented in Table 5-1. The representation of all other variables is approximately the same as observed in the full sample.

Given the strong relationship documented between location and auction incidence in the full sample of properties reporter in Table 5-2, it is worth considering the regional distribution of auctions in the matched sample. These statistics are presented in Table 5-8.

As expected, the areas with the highest proportion of sales which are included in the matched sample are the areas with the highest overall auction activity. As a result the areas identified as having the highest average prices in Section 5.4 are heavily represented in the matched sample. That is, sales from Sydney's Eastern (6.27%), Inner (26.35%), Inner West (9.47%), Lower North Shore (12.81%), and Central Northern (6.95%) suburbs, as well the St George-Sutherland Shire (12.76%) and Northern Beaches (4.80%) areas collectively account for approximately 80% of sales in the matched sample. By contrast, these areas account for only about 54% of all sales in the total sample under consideration in this chapter.

Across the different regions in the matched sample the average prices by both private treaty and auction are approximately the same as the average prices reported from the total sample for each area in Section 5.4, with no consistent pattern in the size or direction of any average price changes.

Table 5-8:

Matched Sample Auction Distribution

This table presents the statistics for the matched sample. The number of matched sales, the proportion of sales to be matched, the proportion of matched auctions to occur in each region and the average private treaty and auction price for each region of the matched sample are reported.

Subdivision	Matched Sales	Proportion (%)	Auction Distribution (%)	Average Private Treaty Price ($)	Average Auction Price ($)
Eastern Suburbs	996	31.76	6.28	1,496,923	1,648,333
Inner West	1,502	44.61	9.47	842,918	853,364
Inner City	4,180	45.31	26.35	728,052	767,797
Lower North Shore	2,032	29.74	12.81	956,280	1,151,364
St George-Sutherland	2,024	21.10	12.76	615,345	697,163
Canterbury-Bankstown	1,066	23.21	6.72	479,542	491,591
Northern Beaches	760	11.46	4.79	945,896	1,120,460
Central North	1,102	11.25	6.95	702,513	883,787
Fairfield-Liverpool	744	11.55	4.69	377,032	363,645
Central West	666	11.03	4.20	448,557	466,408
Outer South West	206	3.74	1.30	345,513	311,057
Outer West	234	2.43	1.47	345,268	324,427
Blacktown	354	4.05	2.23	381,495	338,779

Given there are a total of 9,679 auction sales which occurred on the date of the auction, the maximum size of the matched sample is 19,358. Only 7,933 auction observations are successfully matched using with-replacement sampling (this figure is lower when the sampling is conducted without replacement), yielding a matched sample of 15,866 observations.

Regression analysis is applied to the sample of matched observations. The results from fitting the models given by Equations 5.1 and 5.2 to the matched sample are presented in Table 5-9.

The results from this combination of matched sampling and regression analysis support those obtained from the sample selectivity-corrected regression modelling. That is, given the estimated auction dummy variable coefficient is insignificant, both economically and statistically, there is insufficient evidence to reject *Hypothesis* 5_1. This result occurs for both the restricted model, which replicates the specification of Dotzour et al. (1998) and the more completely specified hedonic model. The other hedonic coefficients have the expected signs and magnitude, although the interest rate variable is now also statistically insignificant. Both the restricted and hedonic models explain a large degree of variation in prices achieving Adjusted R^2 statistics of 79.25% and 85.8%, respectively.

Table 5-9:

Matched Sample Regression Results

This table reports the results of the re-estimated regression models given by Equations 5.1 and 5.3 using the sample obtained by the matching procedure. Coefficient estimates of the fitted model and several regression diagnostic statistics are presented. Coefficient significance at the sample size adjusted 5% level of significance is denoted by *. Due to the number of dummy variables used to account for the effect of suburb and month of sale in the regression, these individual estimated coefficients are not reported. Instead, a summarised description of the statistical performance of the suburb and month of sale explanatory variables is provided. Full results are available from the author upon request.

	Restricted OLS Model		Hedonic OLS Model	
Variable	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic
Intercept	13.4535 *	39.06	13.0424 *	45.70
Auction	0.0059	1.62	0.0026	0.85
Land Size	0.3969 *	65.32	0.3027 *	58.03
Interest Rate	0.1051	2.09	0.0433	1.04
Bedrooms			0.0732 *	18.71
Bathrooms			0.0964 *	16.21
Bed/Bath			0.0132	2.72
Other Rooms			0.0425 *	8.42
Car Spaces			0.0283 *	14.24
Pool			0.0426 *	7.97
Air-Conditioning			0.0505 *	9.39
Scenic View			0.0412 *	6.64
Waterfront			0.1814 *	10.63
	Significance at 1% (%)	Significance at 10% (%)	Significance at 1% (%)	Significance at 10% (%)
Suburbs ($J = 545$)	64.95	76.88	8.44	14.86
Months ($T = 47$)	6.38	29.79	21.28	27.66
Observations	15,866		15,866	
Critical t	3.051		3.051	
Adjusted R^2	0.7925		0.8580	
F Statistic	102.86 *		159.72 *	

The results from the combination of matched sampling and regression support the conclusions from applying sample selectivity-corrected hedonic regressions. That is, that sale mechanism does not influence price. These findings indicate that *Hypothesis* 5_1 should not be rejected.

This result leads to a very different conclusion from the one derived from the results of regression analysis used in isolation, as undertaken in Section 5.5.1 of this chapter and in much of the academic research. The belief, driven by this earlier research, of a significant price premium attributable to the auction mechanism in the Australian and New Zealand housing markets, appears to be the result of shortcomings in the methods of past research used to control for sample heterogeneity and selectivity in sales method.

5.5.5 Additional Robustness Tests

Several robustness tests are undertaken on the data, particularly with respect to the matched sample procedure. This section summarises the results of such testing.[49]

Firstly, the results of the regression analysis on the matched sample are robust to the choice of with- or without-replacement matching. The coefficient to the auction dummy variable under both procedures is found to be insignificant, and the results for all other variables are statistically similar.

[49] Full results of this additional analysis are presented in Appendix B.

Secondly, the hedonic model is re-estimated on several partitions of the matched sample, including high- versus low-valued properties and modal versus more unique properties. These partitions are motivated by the observation in the prior literature, and the findings of the probit modelling in this chapter, of an increased likelihood of auction being the sale mechanism for larger, more expensive, more unique properties. Under both partitioning methods the coefficient of the auction dummy variable remains insignificant.

Several alternative matching procedures are considered. These include a match solely on nearest land size, and one which uses the nearest number of bedrooms as the secondary filter after location and month of sale, instead of land size. Failure to reject *Hypothesis* 5_1, as reported in this chapter, is robust to these different methods.

Finally, a small sample of sales with attribute and sale method details is obtained for the housing market of Christchurch, New Zealand – the same market considered by Dotzour et al. (1998) – for the period January 2005 to December 2008. The results of applying our hedonic regression and matched sampling methodologies are consistent with those reported in the present study for the Sydney market. That is, once selectivity bias is mitigated, there is no significant price premium to auctions. This provides some early external validity to the results of this chapter, and suggests that the results of this paper are not limited to the Sydney housing market. A thorough re-examination of the price effect of auctions in markets previously considered is a well-motivated direction for in future research as more comprehensive data becomes available.

5.6 Summary

The results of this chapter support the null *Hypothesis* 5_1, that there is no difference between the prices of properties that sell at auction and those that sell by private treaty, *ceteris paribus*. This contradicts the observation of an auction price premium made in earlier studies of the Australian and New Zealand housing markets (Newell et al., 1993; Lusht, 1996; Dotzour et al., 1998).

Using a detailed sample of property sales and attributes from the Sydney housing market for the period May 2003 to August 2007, it is shown that the decision to sell by auction is made endogenously. That is, certain types of houses – those which are larger, or exhibit more unique features – are more likely to sell by auction. As a result, this study controls for sample selectivity through the Heckman two-stage correction technique and use of a matched sampling procedure.

In contrast to the result of an auction premium of between 2% and 2.5% obtained from unadjusted hedonic-regression analysis, the coefficient to the auction variable is found to be insignificantly different from zero in regressions incorporating a sample selectivity variable (as prescribed by the Heckman two-stage procedure) and in regressions on a matched sample (Rubin, 1979). This result is robust to alternative matching procedures and sample partitions.

The conclusion made in this chapter is required under a rational expectations perspective if sales by both auctions and private treaties are to coexist in the housing market.

6. Pricing New Properties

6.1 Introduction

This chapter examines the relationship pricing of new properties and their subsequent performance. Regression analysis is performed to determine whether a bias exists in the pricing of new residential real estate assets relative to the prices of existing real estate. This tests *Hypothesis* 6_1 developed in Chapter 3. Following from this, the longer-term return performance of these new properties is measured and compared to the market-wide performance. These results are used to test *Hypothesis* 6_2, which posits that new properties underperform the market.

The chapter is structured as follows. The next section will outline the methodologies to be used to test for first-sale price bias and subsequent returns underperformance. Section 6.3 describes the data to be used in applying the methodology, and Section 6.4 presents and discusses the results of the empirical tests. Finally, Section 6.5 summarises the empirical findings of this chapter and discusses the implications of these findings to future research in the application of hedonic modelling to residential real estate price data.

6.2 Research Design

This section outlines the methodology to be applied in this chapter. This methodology is designed to test *Hypotheses* 6_1 and 6_2. To test for bias in the price of properties at their first sale, stand-alone regression analysis and the combination of matched sampling and regression analysis are used. To empirically determine whether there is underperformance in the subsequent capital return of properties following their first sale, trade-pair analysis is used and compared to the market return.

6.2.1 First Price Bias

Hedonic-regression analysis is a process that incorporates characteristics data for each property in the estimation in order to control for heterogeneity in the housing stock. Using a dummy variable, *New*, indicating whether it is a property's first sale or not, the hedonic-regression model given by Equation 6.1 is estimated:

$$p_{i,t} = \alpha + \beta Size_i + \sum_{j=2}^{J} \gamma_j Suburb_{j,i} + \sum_{s=2}^{T} \delta_s Time_{s,t} + \Lambda New_{i,t} + \varepsilon_{i,t} \qquad 6.1$$

where,

$p_{i,t}$ is the natural logarithm of the price of property i in period t

$Size_i$ is the natural log of the land size (measured in hectares) of the i^{th} property

$Suburb_{j,i}$ is a dummy variable equal to 1 if the i^{th} property is located in $Suburb_j$, and 0 otherwise

$Time_{s,t}$ is a dummy variable representing the year and month of sale; equal to 1 if the sale of the i^{th} property occurred in the s^{th} time-period, and 0 otherwise

$New_{i,t}$ is a dummy variable equal to 1 if the sales observation of property i at period t is the first sale of property i, and 0 otherwise

α is the fitted regression intercept term

β is the estimated coefficient to $Size_i$, reflecting the expected percentage price change of a one unit change in the land size variable, *ceteris paribus*

γ_j is the estimated coefficient to the j^{th} suburb, reflecting the expected relative percentage price difference attributable to a location in $Suburb_j$ over $Suburb_1$, *ceteris paribus*

δ_s is the estimated coefficient to s^{th} time period, reflecting the expected percentage price difference attributable to a sale in $Time_s$ over $Time_1$, *ceteris paribus*

Λ is the estimated coefficient to $New_{,t}$, reflecting the expected percentage change in price of a property at its first sale, *ceteris paribus*

$\varepsilon_{i,t}$ is the regression error term to each observation, assumed i.i.d. $N[0,1]$.

A dummy variable for each suburb is included in the regression to account for variability in price attributable to location. Bedrooms and bathrooms are included to proxy for floor size, in the absence of information on actual living space. The inclusion of the ratio of bedrooms to bathrooms is intended to capture non-linearities in the relationship between property size and price. The inclusion of quarterly dummy variables accounts for differences in price arising from market drift.

The size and significance of the coefficient to the *New* dummy variable, Λ, captures the value of newness, all else being equal. *Hypothesis 6_1* is formally tested as:

This hypothesis is formally tested as:

$$H6_1^0 : \Lambda = 0$$

$$H6_1^A : \Lambda > 0$$

This hypothesis is rejected if the estimate of Λ is not statistically significant when the regression model given by Equation 6.1 is fitted. The finding of a statistically significant and positive estimate of Λ, however, will provide support for the alternative *Hypothesis 6₁*. This result would indicate the presence of a positive price bias to new properties over existing properties, *ceteris paribus*.

As demonstrated in the previous chapter, which considered the presence of a price premium to properties which sell by auction, if there is a systematic relationship between the properties of interest – in this chapter, new properties – and certain characteristics, the hedonic-regression method may not fully account for such variation. This manifests as a sample selectivity bias in the regression model, affecting the estimated Λ coefficient. For example, new properties may reflect more current architectural style, such as an open-plan layout (unobserved subjective factor), or appliances, such as dishwashers (unobserved objective factor). If consumers prefer these factors, they will be positively related to price, and consequently the estimated coefficient to newness will be positively biased in a hedonic-regression model which either omits them or does not account for such selectivity.

To control for this, the Heckman two-stage probit and regression procedure is applied. The probit model to be estimated is given by Equation 6.2:

$$New_i = \sum_{k=1}^{K} \beta_k^* Attribute_{k,i} + \sum_{j=2}^{J} \gamma_j^* Suburb_{j,i} + \sum_{s=2}^{T} \delta_s^* Time_{s,t} + \varepsilon_i \qquad 6.2$$

where

New_i is the binary dependent variable, equal to 1 if the i^{th} property listed for sale by auction, and 0 otherwise

ε is the error term of the probit model, assumed i.i.d. $N[0,1]$

the estimated coefficients – β^*, γ^*, δ^* – represent a measure of the predicted change in the probability of an auction of the i^{th} property in response to a given unit change in the respective independent variable, *ceteris paribus*

all other variables are defined as for Equation 6.1.

If new properties are more likely to exhibit certain features or be built in certain areas than existing properties, the coefficients to these variables will be positive. Due to increasing land scarcity, for example, new properties are expected to have smaller average land size than existing properties. The redevelopment of previously zoned agricultural or industrial areas, typically at the urban fringe, motivates the *a priori* expectation that the coefficients to suburbs meeting this criterion will be positive.

The structure of the probit model, required for application of the Heckman selectivity-corrected regression, follows the outline provided in the previous chapter, Section 5.3.2. Now, however, the probability of a property being sold for the first time, given a set of observed attributes, as given by Equation 6.2, now assumes,

$New_i = 1$ if $z_i^* > 0$

$New_i = 0$ if $z_i^* \leq 0$

$$P(New_i = 1) = \Phi\left[\sum_{k=1}^{K} \beta_k^* Attribute_{k,i} + \sum_{j=2}^{J} \gamma_j^* Suburb_{j,i} + \sum_{s=2}^{T} \delta_s^* Time_{s,t}\right]$$

$$P(New_i = 0) = 1 - \Phi\left[\sum_{k=1}^{K}\beta_k^* Attribute_{k,i} + \sum_{j=2}^{J}\gamma_j^* Suburb_{j,i} + \sum_{s=2}^{T}\delta_s^* Time_{s,t}\right]$$

where z_i^* is an unobserved selection variable determining the method by which a property is sold and Φ denotes the cumulative distribution function (CDF) of the Gaussian Normal distribution.

Estimation of the section-stage Heckman selectivity-corrected regression model requires estimation of the IMR, denoted by κ, using the estimates obtained from fitting the probit model given by Equation 6.2 for every observation. The calculation of κ is given by Equation 6.3:

$$\hat{\kappa}_i = \frac{\varphi\left[\sum_{k=1}^{K}\beta_k^* Attribute_{k,i} + \sum_{j=2}^{J}\gamma_j^* Suburb_{j,i} + \sum_{s=2}^{T}\delta_s^* Time_{s,t}\right]}{\Phi\left[\sum_{k=1}^{K}\beta_k^* Attribute_{k,i} + \sum_{j=2}^{J}\gamma_j^* Suburb_{j,i} + \sum_{s=2}^{T}\delta_s^* Time_{s,t}\right]} \qquad 6.3$$

where φ and Φ denote the probability density function (PDF) and the cumulative distribution function (CDF) of the Gaussian Normal distribution, respectively.

The selectivity-corrected regression is given by Equation 6.4:

$$p_{i,t} = \alpha + \sum_{k=1}^{K}\beta_k Attribute_{k,i} + \sum_{j=2}^{J}\gamma_j Suburb_{j,i} + \sum_{s=2}^{T}\delta_s Time_{s,t} + \Lambda New_{i,t} + \beta_\kappa \kappa_i + \varepsilon_{i,t} \qquad 6.4$$

where,

κ_i denotes the sample selectivity variable, as calculated from Equation 6.3

β_κ the coefficient of the sample selectivity variable

all other variables are defined as for Equation 6.1.

For robustness, the hedonic-regression model given by Equation 6.1 is re-estimated on a matched subsample of observations. Chapter 5 demonstrated that the observation of a premium to the auction sale mechanism arose when the regression model failed to completely account for heterogeneity in the presence of sample selectivity. While the previous chapter held that auctions were the more likely sale method for unique and higher-quality properties, in this chapter the concern is that new properties are more likely to exhibit contemporary architectural styles and features, which are not included in the model given by Equation 6.1.

To implement the matching process in this chapter, all first sales of residential properties in the sample are identified and separated – sample *A* – from the sales of existing properties – sample *B*. The matching requires that the sample B observation chosen has the same location and date as the given sample A observation, proxied by statistical subdivision and the quarter of sale, respectively. In the case where multiple matches are identified, the sample B observation with the closest size to the sample A observation is chosen as the match. This is proxied by land size for houses and the number of bedrooms and bathrooms for unit observations.

Significant and positive estimates of λ from the second-stage Heckman procedure (Equation 6.4) and the re-estimation of Equation 6.1 on the matched sample will provide further support for *Hypothesis* 6_1.

6.2.2 Relative Performance

To test for the presence of subsequent underperformance in a given segment of the residential real estate market, trade-pair analysis is undertaken. That is, the first-sale observations are paired with subsequent sales of the same property. A relationship between performance and new properties is modelled through Equation 6.5:

$$ACR_{i,t} = \alpha + \Lambda' New_{i,t} + \sum_{s=2}^{T} \delta_s Time_{s,t} + \varepsilon_{i,t} \qquad 6.5$$

where,

$ACR_{i,t}$ denotes the annual compound return to property i, calculated as

$$ACR = \left[\frac{P_t}{P_0}\right]^{\frac{365}{t}} - 1$$

Λ' is the coefficient of the *New* variable

all other variables are defined as for Equation 6.1.

The size and significance of the estimated coefficient to the dummy variable *New*, Λ', explains the performance of new properties relative to the market. *Hypothesis* 6_2 is formalised as:

$H6_2^0$: $\Lambda' = 0$

$H6_2^A$: $\Lambda' < 0$

The alternative hypothesis is supported if Λ' is found to be statistically significant and negative. Failure to reject this hypothesis indicates that new properties underperform the market following their first sales. That is, the average return to new properties is lower than the return to the market, *ceteris paribus*.

6.3 Data and Descriptive Statistics

The data used in this study is sourced from Rismark International, under licence from RPX. Virtually every sale of residential property in Australia is captured by RPX via data uploads with the VG in each state and territory. This data contains information on price, contract and settlement dates, full address and land size details. RPX augments this data with attributes information – bedrooms, bathrooms, pool, etc – collected from print and web media listings, real estate agents, and direct property viewing.

No information is currently available for off-the-plan or new sales. The year a property was built is captured by RPX. The first sale of a property – and, therefore, sales of new properties – can be inferred when the year of the contract date is equal to the year built. However, only the VG offices in South Australia and Western Australia collect and make widely available this data. Consequently, only the cities of Adelaide and Perth, as defined by the ASGC,[50] have sufficiently large sample sizes with all necessary variables for this study. This chapter focuses on the performance of the Perth residential real estate market as it covers a longer and larger sample than Adelaide. The results for Adelaide, presented in Appendix C, are consistent with those for the Perth market reported in this chapter.

Table 6-1 summarises the sample of sales data from the Perth market, covering the period January 1999 to June 2008 to be used in this chapter.

---

[50] *Standard Geography Volume I – Australian Standard Geographical Classification*, ABS, Catalogue 1216, July 2006.

Table 6-1:

Descriptive Statistics

This table provides the sample statistics of the sample of house and unit sales in Perth over the period January 1999 to June 2008. The information covers the sample periods and size, distribution measures of price, the average land size, bedrooms, bathrooms, and car spaces and the proportion of observations with given characteristics for binary variables: waterfrontage, scenic view, swimming pool, and air-conditioning.

	Houses		Units	
	All	New	All	New
Observations	236,259	3,255	60,951	583
Proportion (%)		1.378		0.957
Median Price ($)	248,000	355,000	191,000	295,000
Average Price ($)	302,244	376,105	230,308	310,651
Standard Deviation ($)	222,615	189,746	144,007	137,319
Land Size (m^2)	804.920	559.157	-	-
Bedrooms	3.359	3.768	2.411	3.151
Bathrooms	1.517	2.023	1.118	1.926
Car Spaces	1.025	1.581	0.779	1.655
Pool (%)	1.655	1.352	1.116	0.172
Air-Conditioning (%)	0.051	9.892	0.139	2.401
View (%)	11.681	1.720	4.169	0.343
Waterfront (%)	14.752	0.061	3.112	0

There are 236,259 house sale observations and 60,951 unit observations in Perth over the sample period January 1 1997 through 30 June 2008. The average and median prices for all houses in Perth during the sample period are \$302,244 and \$248,000, respectively. The observation of higher average prices than median prices indicates a negative skew in the distribution of house price in Perth. Unit prices exhibit a similar skew: the average unit price is \$230,308 while the median unit price in the sample is \$191,000.

New sales account for 1.38% of house sales and 0.96% of unit sales in the Perth market over the sample period. The negative skew in prices observed for the wider market exists among new properties, with average prices higher than median prices although to a smaller degree.

Of concern to the hypotheses tested in this chapter, the prices of new properties, by either the average or median measure, are higher than the full sample's. New houses and units in Perth sold during this sample, on average, for \$376,105 and \$310,651, respectively. On this measure, a price premium of between 7.4% and 8% is observed to new properties. This result, however, is not sufficient to conclude a price premium exists, given the heterogeneity in quality and attributes that is expected between new and resale properties.

Table 6-1 also presents details for the average size and representation of several hedonic attributes, including land size (for houses only), bedrooms, bathrooms, and binary variables such as pool, air-conditioning, scenic view, and waterfrontage.

The average land size for houses in the sample of sales is 805 m^2. This contrasts with the significantly lower average land size for new houses, 559 m^2. Despite this smaller average land size, new houses have more bedrooms and bathrooms than resale houses, on average. Average unit sizes have followed this trend: new units have 3.2 bedrooms and 1.9 bathrooms, on average, compared to 2.4 bedrooms and 1.1 bathrooms, on average, for resale units.

A higher proportion of new houses and units have air-conditioning, 9.9% and 2.4%, respectively, than resale houses and units, 0.05% and 0.14%. On average, however, new properties are less likely to have a pool than resale properties, which may be expected, given the smaller blocks of land on which the houses are being built. Furthermore, smaller proportions of new properties than resale properties have scenic views or waterfrontage. This may also be expected, given existing properties are more likely to have been built to take advantage of such features, limiting the ability of new property developments to access them.

These descriptive statistics indicate that significant differences exist between the subsample of new properties and the existing properties in the Perth market during the sample period January 1997 to June 2008. A comparison of median and average prices between these subsamples indicates a price premium to new properties. Such a measure, however, does not account for these observed hedonic differences. The next section presents the results of the regression-based methodologies outlined in Section 6.2.1 which statistically account for this heterogeneity to assess whether a systematic price difference exists between new and resale properties.

6.4 Results

This section reports the results of undertaking the methodology outlined in Section 6.2. Firstly, the fitted estimates of the hedonic regression are presented and discussed. From this an initial conclusion as to whether properties are fairly priced at their first sale can be made. This is followed by the results from probit modelling to determine if new properties are more likely to exhibit certain attributes. This represents the first stage of the Heckman selectivity correction method. The second-stage Heckman regression results are then reported. As an alternative selectivity correction method, the results from regression analysis on a matched sample of properties are reported. This examination of the pricing of new properties is followed by analysis of their subsequent investment performance relative to the broader market.

6.4.1 Hedonic Regression

The results of the first-sale price bias regression, run separately for house and units in Perth, are presented in Table 6-2.

The coefficient to the new dummy variable indicates that new houses sell at a premium of 10.2% over existing properties, *ceteris paribus*. This estimate is highly statistically significant. The price premium for new units is 7.5% which, while less than that estimated for houses, is still an economically and statistically significant value. On this result, it is not possible to reject *Hypothesis* 6_1, that the price of residential property at its first sale is greater than other properties.

Table 6-2:

New Property Pricing

This table reports the estimated OLS regression results from the hedonic model given by Equation 6.1 for the Perth residential real estate market. Observations for houses and units are modelled and reported separately. Given the large number of suburb and time parameters, the performance of these variables is summarised in this table. Full regression results are available from the author upon request.

	Houses		Units	
Variable	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic
Intercept	12.8398 *	1713.46	11.0764 *	427.81
New	0.1020 *	30.74	0.0745 *	7.56
Land Size	0.2429 *	162.33	-	-
Bedrooms	0.0917 *	65.52	0.0872 *	10.53
Bathrooms	0.0555 *	21.43	0.3439 *	27.27
Bed/Bath	-0.0569 *	-34.61	0.1672 *	19.55
Car Spaces	0.0443 *	78.85	0.0936 *	52.22
Pool	0.0470 *	42.45	-0.0189 *	-3.49
Air-Conditioning	0.0090 *	6.99	0.0201 *	4.23
Scenic View	0.0948 *	31.78	0.0631 *	7.15
Waterfront	0.0820 *	4.98	0.1445 *	5.89
	Significance at 1% (%)	Significance at 10% (%)	Significance at 1% (%)	Significance at 10% (%)
Suburbs (J_{HOUSE} = 311, J_{UNIT} = 249)	86.45	92.26	67.34	77.02
Months (T_{HOUSE} = T_{UNIT} = 45)	100.00	100.00	97.778	100.00
Observations	236,259		60,951	
Critical *t*	3.515		3.311	
Adjusted R^2	0.9086		0.8568	
F Statistic	6436.22 *		1208.16	

All hedonic variables included in the regression are highly significant with the *a priori* expected signs and relative size. In line with the results of a similar hedonic-regression model fitted to the Sydney housing market in Chapter 5, land size is found to be the most significant variable in explaining price. The coefficient estimates for bedrooms and bathrooms are also found to be highly significant. The ratio of these two variables is negative for houses. This supports the expectation of a non-linear relationship between price and property size. For units, however, a positive coefficient for this ratio is estimated. It is possible that theoretically a similar non-linear size-price relationship exists for units, but is not observed, given the relatively limited range of values for bedrooms and bathrooms.

For houses, unique property features are found to be significantly positively related to price. All else being equal, the presence of a swimming pool is expected to add 4.7% to a property's selling price, air-conditioning adds 1%, a scenic view adds 9.5%, and waterfrontage adds 8.2%.

Interestingly, swimming pools are found to have a negative relationship with the sales price of units in Perth. The *t* statistic for the estimated coefficient to the pool dummy variable (-0.019) is reasonably low. The result may be explained by higher strata and maintenance costs associated with pools in unit developments. Statistically significant positive coefficients are estimated for all other unique property features: a 2% price premium to air-conditioning, 6.3% price premium to a scenic view, and 14.5% premium to waterfrontage, *ceteris paribus*.

The model estimated explains a high degree of variance in price, achieving highly significant F statistics, and adjusted R^2 statistics of 0.9086 for houses and 0.8568 for units.

6.4.2 Probit Model

A probit model is estimated using the binary variable for first sale as the dependent variable against the set of hedonic explanatory variables. This form of model allows an examination of the qualities of new properties. That is, from the results of such analysis it may be determined if new properties are representative of the existing housing stock, and if not what features new properties are more likely to exhibit. Table 6-3 presents the results of fitting Equation 6.2 to the sample of house and unit sales from the Perth residential real estate market.

The results from probit modelling support the observation from the descriptive statistics regarding property size. That is, new houses are on average smaller in land size but have more bedrooms than existing houses, *ceteris paribus*. New units also have more bedrooms on average than exiting units, *ceteris paribus*.

Through this statistically controlled approach, it is observed that new houses are less likely to have a pool, air-conditioning, or waterfrontage. New units also are less likely than existing units to have these features, including a scenic view. The propensity for new houses to have a scenic view is, however, higher than for existing houses.

Table 6-3:

Probit Results

This table reports the results of MLE fitting of the probit model given by Equation 6.2. The coefficient estimates and related diagnostic statistics are presented. Due to the number of dummy variables used to account for the effect of suburb and month of sale in this model, a summarised description of the statistical performance of these variables is provided. Full results are available from the author upon request. Coefficient estimate significance at 1% level of significance is denoted by *.

	Houses		Units	
Variable	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic
Intercept	-8.0373 *	-30.98	-2.3746 *	-6963.64
Land Size	-1.7830 *	-39.27	-	-
Bedrooms	0.6559 *	16.20	0.0091 *	69.69
Bathrooms	-0.3963 *	-5.50	0.0903 *	336.94
Bed/Bath	-0.9919 *	-13.07	-0.0345 *	-233.11
Car Spaces	0.2292 *	15.28	0.1290 *	427.15
Pool	-0.7832 *	-12.82	-0.1308 *	-55.66
Air-Conditioning	-0.2952 *	-9.00	-0.0664 *	-35.70
Scenic View	0.2415 *	3.32	-0.0611 *	-15.01
Waterfront	-0.1572	-0.42	-0.1184 *	-10.03
	Significance at 1% (%)	Significance at 10% (%)	Significance at 1% (%)	Significance at 10% (%)
Suburbs (J_{HOUSE} = 311, J_{UNIT} = 249)	23.23	34.19	94.35	98.39
Months (T_{HOUSE} = T_{UNIT} = 45)	75.56	84.44	93.33	93.33
Observations	236,259		60,951	
Log-Likelihood	-10,628		-1,662	
AIC	21,984		6,504	
Schwartz Criterion	25,771		9,228	
Maximum Absolute Gradient	1.297		0.025	

The statistically significant estimated coefficients to this probit model indicate a large degree of attribute selectivity between new and existing properties. This presents a strong case for undertaking selectivity-corrected regression modelling in addition to the unadjusted hedonic regression fitted in Section 6.4.1. Following the methodological argument outlined in Section 6.2, if this selectivity is not fully accounted for, either through mis-specification of the hedonic function or omitted variables (including unobserved and subjective attributes), regression estimates of the coefficient to the dummy variable for first sale may be biased.

6.4.3 Selectivity-Corrected Regression

This section discusses the results from applying the Heckman selectivity correction procedure. The results from fitting the second-stage selectivity-corrected regression, given by Equation 6.4, are reported in Table 6-4.

The coefficient to the selectivity variable, β_K, is statistically significant and positive for the sample of house sales but not the sample of units sales. This supports the *a priori* expectation of unobserved positive-price factors to new houses over existing houses. The coefficient to the dummy variable representing these new houses remains statistically significant and positive. The results predict a price premium of 8.86% to new houses over existing houses, *ceteris paribus*. The estimate of the dummy variable for new units is also significant and positive, predicting an average price premium to new units over existing units of 7.91%, *ceteris paribus*. These results are consistent with the results from fitting the uncorrected regression model given by Equation 6.1 as reported in Section 6.4.1.

Table 6-4:

Selectivity-Corrected Regression

This table reports the results of OLS fitting of the second-stage Heckman regression model given by Equation 6.3 on the sample of sales for Perth houses and units separately. The estimated coefficients of the fitted model and their significance are reported with several regression diagnostic statistics. Due to the number of dummy variables used to account for the effect of suburb and month of sale, a summarised description of the statistical performance of these variables is provided. Full results are available from the author upon request.

	Houses		Units	
Variable	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic
Intercept	11.2785 *	112.88	12.2163 *	29.00
New	0.0886 *	25.88	0.0791 *	7.91
Sample Selectivity	0.1957 *	15.67	-0.4203	-2.71
Land Size	-0.0840 *	-4.01	-	-
Bedrooms	0.2133 *	27.04	0.0850 *	10.23
Bathrooms	-0.0202 *	-3.69	0.3085 *	16.98
Bed/Bath	-0.2416 *	-20.30	0.1789 *	18.68
Car Spaces	0.0859 *	31.69	0.0450	2.49
Pool	-0.0978 *	-10.51	0.0309	1.61
Air-Conditioning	-0.0448 *	-12.23	0.0451 *	4.34
Scenic View	0.1391 *	33.87	0.0867 *	7.00
Waterfront	0.0497	3.00	0.1899 *	6.39
	Significance at 1% (%)	Significance at 10% (%)	Significance at 1% (%)	Significance at 10% (%)
Suburbs ($J_{HOUSE} = 311$, $J_{UNIT} = 249$)	87.74	92.26	68.55	79.03
Months ($T_{HOUSE} = T_{UNIT} = 45$)	97.78	100.00	100.00	100.00
Observations	236,259		60,951	
Critical *t*	3.515		3.311	
Adjusted R^2	0.9087		0.8568	
F Statistic	6423.96		1204.33	

The estimated explanatory variable coefficients for the units sample are consistent with those from the unadjusted regression analysis. Several coefficient estimates from the sample of house sales that are positive and significant when Equation 6.1 is fitted, however, such as land size, bathrooms, pool and air-conditioning, are now negative.

From the results of the regression modelling performed in Sections 6.4.1 and 6.4.3, *Hypothesis* 6_1 – that the prices of new properties are higher than the prices of existing properties – cannot be rejected. A further examination of this issue is performed in the next section using a matched sampling procedure.

6.4.4 Matched Sampling

Rubin (1979) demonstrates that the matched sampling process significantly lowers bias in non-random allocation of treatment observations. In this scenario, the non-random allocation refers to the possibility that certain observations of price-determining attributes may be more highly associated with new than older properties – for example the push towards smaller properties driven by demographic changes. Details of the matched sample, created using the matching procedure algorithm described in Section 6.2.1, are presented in Table 6-4.

After controlling for several attributes in the matching process, the difference between the median price of new sales and resale properties is smaller. Median prices for new and resale Perth houses – \$355,000 and \$360,000, respectively – are almost equivalent following the matching process.

Table 6-5:

Matched Sample Descriptive Statistics

This table reports the descriptive statistics for the matched sample of Perth houses and units. Median and average prices are reported, as is the standard deviation of price. The average land size (houses only), number of bedrooms, bathrooms and car spaces are reported. For binary attributes – pool, air-conditioning, scenic view and waterfrontage – the proportion of properties with each attribute is reported. Statistics are reported for the total matched sample as well as the subsamples of new and existing sales.

	Houses			Units		
	Matched Sales	New	Existing	Matched Sales	New	Existing
Median Price ($)	357,750	355,000	360,000	265,000	295,000	231,000
Average Price ($)	386,001	376,034	395,969	289,919	310,651	269,187
Standard Deviation ($)	123,104	105,661	138,166	150,542	137,319	160,147
Land Size (m^2)	641.029	559,174	722.883	-	-	-
Bedrooms	3.38	3.82	3.01	3.16	3.15	3.18
Bathrooms	1.72	2.01	1.32	1.13	1.13	1.14
Car Spaces	1.21	1.59	0.84	1.22	1.66	0.79
Pool (%)	8.41	1.35	15.47	3.09	0.17	6.00
Air-Conditioning (%)	11.38	9.90	12.86	3.43	2.40	4.46
Scenic View (%)	2.21	1.69	2.74	1.03	0.34	1.72
Waterfront (%)	0.08	0.06	0.09	0.09	0.00	0.17

For units in Perth the median price for new properties ($295,000) is still higher than the price for resales ($231,000). In the matched sample, the average price for houses at their first sale ($376,034) is less than the average sales price for existing houses ($395,969). The average sales price for new units in Perth ($310,651) is higher than the average price for existing units ($269,187). These observations, however, are within the bounds of a single standard deviation, indicating that no significant price difference exists between new and resale properties in Perth following the matched sampling procedure.

The matching procedure controlled for location, size (proxied by land size for houses and the number of bedrooms for units) and the month of sale. These variables are significant determinants of price, as demonstrated in the hedonic-regression analysis presented in Section 6.4.1, and are significant determinants in observing new from existing properties. Large differences are observed, however, in the distribution of several hedonic attributes between the new and existing matched subsamples.

New houses in the matched sample have more bedrooms, bathrooms and car spaces than existing houses. A lower proportion of new houses have unique features such as a pool, air-conditioning, scenic view and waterfrontage. The pattern for units in the matched sample is similar, except for the average number of bedrooms, which is expected given this is a matching variable for units. Care must be taken in the interpretation of the unadjusted mean and median prices to new and existing properties in the matched sample as a result of this persistent heterogeneity. To further account for the differences across properties, the matched sampling procedure is combined with regression analysis, as prescribed by Rubin (1979).

Table 6-6 reports the results of the re-estimated regression given by Equation 6.1 using the matched sample. Using the sample-size adjusted critical *t* values (Connolly, 1989) to account for the smaller number of observations following matching, the estimated coefficient to the first-sale variable is found to be statistically significant: the results predict sales price premiums of 14.5 % to new houses and 19.2% to new units, *ceteris paribus*. This result is in line with the observation of a price premium to new properties obtained from the vanilla hedonic-regression analysis and sample selectivity-corrected regression performed in Sections 6.4.1 and 6.4.2, respectively.

The estimated coefficients to the set of hedonic variables from this matched sample are consistent with those estimated from the full sample and with *a priori* expectations. The high Adjusted R^2 and F statistics, even with this smaller sample, demonstrate that a high degree of variation in prices is explained by this model.

This section has demonstrated, using sample selectivity-corrected hedonic methods, that new properties sell at a premium to existing properties. The next section applies the methodology outlined in Section 6.4.3 to examine whether new properties subsequently underperform relative to the market.

Table 6-6:

Matched Sample Regression

This table reports the results of OLS fitting of the regression model given by Equation 6.1 on the full sales sample. The estimates of the explanatory variable coefficients, their *t* statistics and significance against the sample-size adjusted critical *t* value are reported. This table also reports the Adjusted R^2 and *F* statistic as well as the number of observations and the critical *t* value. Due to the number of dummy variables used to account for the effect of suburb and month of sale in the regression, the statistical performance of these coefficients is summarised. Full results are available from the author upon request.

	Houses		Units	
Variable	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic
Intercept	12.9702 *	221.22	11.4798 *	51.54
First Sale	0.1450 *	16.24	0.1924 *	7.12
Land Size	0.2069 *	20.84	-	-
Bedrooms	0.0606 *	6.91	0.0727	2.20
Bathrooms	0.0519 *	3.34	0.2414 *	4.50
Bed/Bath	-0.0348 *	-3.10	0.1348 *	3.40
Car Spaces	0.0277 *	7.71	0.0864 *	6.72
Pool	0.0379 *	4.39	0.1125 *	2.77
Air-Conditioning	0.0024	0.32	-0.0274	-0.75
Scenic View	0.0763 *	4.88	0.0897	1.40
Waterfront	0.1544	1.94	0.2199	1.06
	Significance at 1% (%)	Significance at 10% (%)	Significance at 1% (%)	Significance at 10% (%)
Suburbs (J_{HOUSE} = 248, J_{UNIT} = 156)	61.13	72.78	4.52	16.13
Months (T_{HOUSE} = T_{UNIT} = 45)	95.46	97.73	85.37	90.24
Observations	6,502		1,166	
Critical *t*	2.894		2.409	
Adjusted R^2	0.8837		0.8462	
F Statistic	165.10		32.26	

6.4.5 Underperformance Results

To estimate the ACR to residential real estate, all trade pairs in the sample are identified. A trade pair is defined as two consecutive normal sales of the same property. Consequently, the sample of trade pairs comprises properties that sold twice or more in the sample period. The ACR to new properties is calculated to their first subsequent sale only. That is, in the case of multiple trade pairs for a new property, only the first is considered the return to a new property; subsequent trade pairs are classified as returns to an existing property. Table 6-7 presents summary statistics for the performance and attributes of the trade pairs.

The estimates of ACR in the Perth residential real estate indicate an average ACR to houses and units of 17.70% and 18.32%, respectively. The return to units is observed to be higher than houses during this sample period. From their first sale to their second sale, new properties returned less than the overall market: the average ACR to new houses is 15.60% and 16.28% for units.

Differences exist between the attributes of properties that have sold only once and those that sell more than once in the sample. Wang and Zorn (1997) and Gatzlaff and Haurin (1997) argue that sample selection bias exists in the sample of properties that sell and that this is exaggerated in properties which sell more frequently. Several empirical papers suggest these more highly traded properties are 'winners' (Case et al., 1991; Abraham and Schaumann, 1991; Genesove and Mayer, 1991), having outperformed the market. Other research argues that these properties are more likely to be smaller and less expensive than the less frequently traded set of properties.

Table 6-7:

Trade Pairs

This table reports the results of descriptive statistics for trade pairs. This table reports the average and median ACR on new properties and the market, and the average holding period, calculated as the time between trade-pair sales observations. Descriptive statistics for the set of hedonic attributed is also reported. This includes the average land size; average number of bedrooms, bathrooms and car spaces; and proportion of properties with a pool, air-conditioning, scenic view, and waterfrontage.

	Houses		Units	
	All	New	All	New
Observations	54,818	704	14,558	114
Proportion		1.28%		0.78%
Average Holding Period (years)	4.15	3.71	4.31	3.68
Average ACR (%)	17.70	15.60	18.32	16.28
Median ACR (%)	16.35	13.50	16.71	13.65
Land Size (m^2)	783.02	578.83	-	-
Bedrooms	3.35	3.73	2.35	3.09
Bathrooms	1.51	2.04	1.09	1.91
Car Spaces	1.03	1.58	0.73	1.57
Pool (%)	12.82	0.71	3.06	0.00
Air-Conditioning (%)	6.14	4.55	2.26	1.75
Scenic View (%)	1.20	1.85	0.96	0.00
Waterfront (%)	0.06	0.00	0.16	0.00

A smaller proportion of new houses and units are resold in the sample period than they represent as individual sales in the descriptive statistics originally presented: 1.378% and 0.957% of all sales for houses and units, respectively, compared with 1.284% and 0.783% of paired sales. By land size, the properties that sell more than once are smaller on average (783 m^2 compared with 804 m^2), although the number of bedrooms and bathrooms are comparable to the total sales statistics. Interestingly, the proportion of properties selling repeatedly with unique property features, such as waterfrontage, is significantly less, supporting the finding of Englund, Quigley, and Redfearn (1998) that higher trade frequency properties are typically more modest.

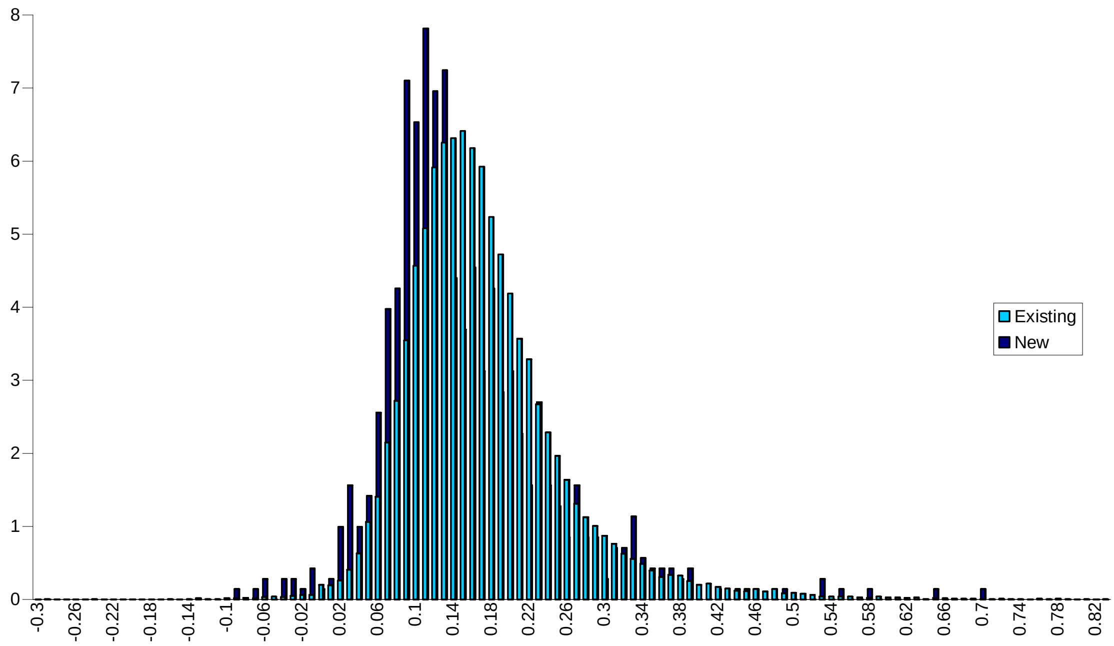
These differences are largely consistent across the subsamples of new and existing trade pairs. It is unlikely that this selectivity will impact upon the analysis of returns to new properties relative to the market. Figure 6-1 presents the distribution of returns, as estimated by the ACR, to new and existing properties.

From the histogram of returns to houses presented in Panel A it can be seen that the returns to existing properties form a slightly skewed, long-tailed bell-shaped figure distributed approximately around 16%. The distribution of returns to new properties from their first sale follows a roughly similar skewed pattern, but centred at a lower return, approximately 11%. New properties also have a higher proportion of extreme return observations than existing properties, although given the smaller overall representation of new sales in the trade-pair sample this observation must be interpreted with care.

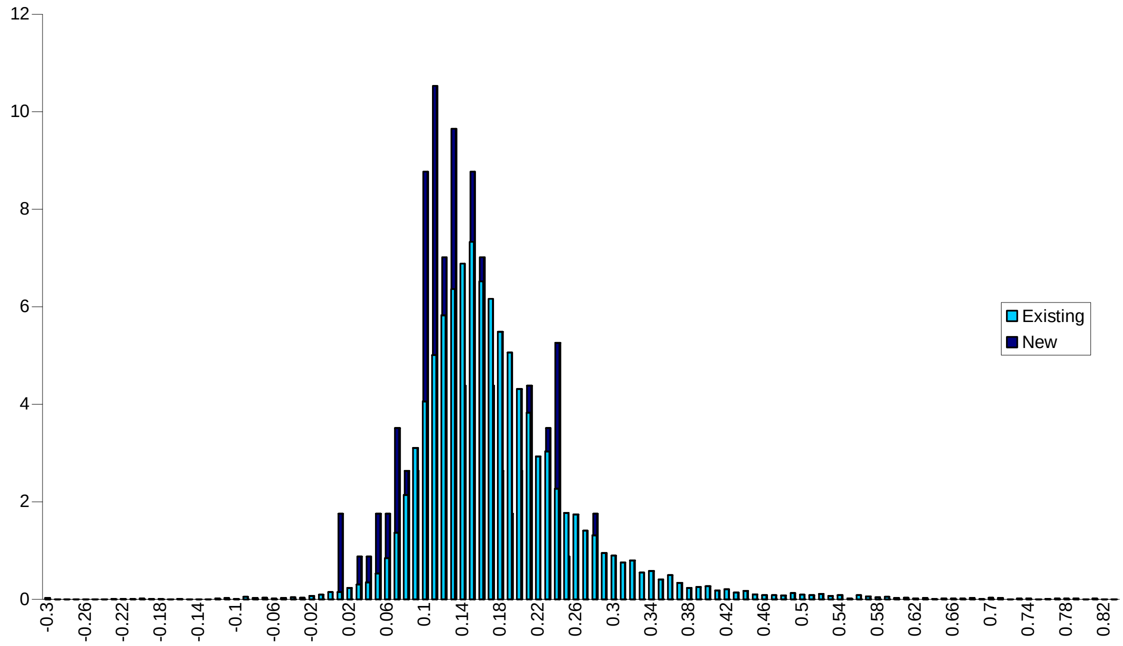
Figure 6-1:

Relative Performance

Panel A: Houses



Panel B: Units



The histogram of returns for units, presented in Panel B of Figure 6-1, shows a similar pattern of returns to units as for houses. The distribution is slightly skewed from the symmetric bell curve and with long tails. New units appear to underperform the rest of the market, with a similar-shaped left-shifted returns distribution.

Further analysis of the performance of investments in the Perth property market is undertaken to consider how returns change over time. Specifically, comparison in annual compound returns is also performed by segmenting the sample of trade pairs by holding period. Table 6-8 reports the average ACR estimates for new versus existing houses and units by holding period.

Two things are apparent from this analysis. Firstly, new houses in both samples tend to underperform the market regardless of the holding period. The pattern is similar but not as strong for new units, which may in part be attributable to the smaller sample size of new trade pairs.

Secondly, the return to new properties tends to decrease with each year of age. The results also indicate that the return to new houses and units by holding period is similar. However, the relationship between age and price observed from this analysis indicates a non-linear relationship.

Table 6-8:

Return Comparison

This table reports the average percentage ACR by holding period, as calculated by Equation 6.6, to the sample of sales from the Perth market by holding period. ACR is reported separately for houses and units, and separately for the sample of new properties from the total market.

Holding Period (years)	Houses		Units	
	New	All	New	All
2	19.51	25.52	23.25	18.93
3	16.38	23.92	15.53	19.10
4	15.27	21.77	16.15	19.71
5	14.01	19.96	15.02	18.77
6	11.50	17.46	12.36	17.30
7	12.23	18.07	10.34	14.33
8	11.88	17.09	12.33	15.94
9	11.19	18.30	12.44	16.75
10	12.29	17.31	11.62	14.62
11	10.22	16.01	-	-

These observations of underperformance are formally tested by estimating the regression model given by Equation 6.5. In these models, the return to properties for which trade pairs are identified, measured by the ACR, is regressed against a set of quarterly time dummy variables to control for the natural market growth pattern and a dummy variable equal to one if the trade pair includes the first sale of the property. The results from fitting this model are reported in Table 6-9.

Table 6-9:

Relative Performance

This table reports the results of OLS fitting of the regression model given by Equation 6.5 on the full sales sample.

	Houses		Units	
Variable	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic
First Sale	-0.0329 *	-11.82	-0.0314 *	-4.10
	Significance at 1% (%)	Significance at 10% (%)	Significance at 1% (%)	Significance at 10% (%)
Months ($T_{HOUSE} = T_{UNIT} = 42$)	97.62	100.00	97.62	97.62
Observations	54,818		14,558	
Critical *t*	3.302		3.092	
Adjusted R^2	0.8637[51]		0.8450	
F Statistic	8082.08		1846.44	

The results of this statistical analysis support the descriptive statistics: the return to both new houses and units for this sample is lower than the return to existing houses and units. These results provide insufficient evidence to reject *Hypothesis* 6_2. It is concluded that new properties sell at a significant price premium at their first sale only to subsequently underperform relative to the market.

[51] This model is estimated without an intercept term. Consequently, the interpretation of goodness of fit variables must be undertaken with care. The advantage to this approach is that an estimated annualised return to each quarter may be readily observed.

6.5 Summary

In this chapter the relationship between property age and sales price is investigated. A significant price premium to newly constructed residential properties is observed, *ceteris paribus*. New properties, however, subsequently underperform the market.

Using sales and hedonic attribute data for the Perth residential real estate market over the period January 1997 to June 2008, regression analysis estimates a price premium of 10% to new houses and 7% to new units. The Heckman two-stage procedure and a combination of matched sampling with regression analysis are applied in order to account for the expected non-random distribution of property traits across new and existing housing. The finding of a significant price premium to new properties holds in these sample selectivity-corrected models.

The return to new and existing properties is compared using an annualised compound return metric. New houses and units in Perth underperformed the market during this sample period on average by 6.32% and 2.1%, respectively.

The results presented in this chapter present insufficient evidence to reject *Hypothesis* 6_1 and *Hypothesis* 6_2. A number of significant implications arise from this with respect to residential real estate pricing and research methods. Property age and specifically the distinction between new and existing properties must be accounted for in pricing models and cross-sectional sampling research. This is a particularly important issue in repeat-sales based modelling, where property characteristics are assumed constant through time and age has no price effect.

7. Predicting Property Price Movements

7.1 Introduction

This chapter examines the value of information in the residential real estate market. Inefficiencies exist in the reporting and dissemination of house and unit sales information. As a result, the complete set of sales to have occurred by a given point in time is only known after a minimum of two months. This in turn creates a lag in the estimation and publication of house price indices.

Samples of sales data, however, are more readily available. These include the listings of properties for sale, newspaper reporting of recent sales, and the sales results of real estate agents. Indices are estimated from these timely data sources and their movements are compared with those of a market-wide index estimated from the population of sales. The results are used to test *Hypothesis* 7_1.

Using regression analysis it is found that a significant component of variation in returns to a market-wide index may be explained by indices constructed from newspaper reporting and agents' records. These timely data samples may consequently be used to predict the forthcoming market-wide index. This result should be a concern to the developers of residential real estate derivatives markets, whose products typically rely on population-based estimates.

This chapter is set out as follows. Section 7.2 outlines the index estimation procedure and comparative methodology to be used. Section 7.3 describes the data which will be used. Section 7.4 presents the results and the final section summarises and concludes.

7.2 Background

The motivation for the research presented in this chapter is driven largely by the development and refinement of derivatives markets for residential real estate products. In turn, the findings of this research have potentially significant implications for the design of such markets. This section provides a brief historical background and institutional setting of the current major global housing derivatives markets.

Derivatives over residential real estate enable investors to: diversify real estate holdings without physical ownership; achieve negative net exposure to the asset class (shorting); trade in the market with lower transaction costs and lower minimum levels of investment than are possible with direct investment; minimise maintenance fees on the investment;[52] and trade with faster execution and higher liquidity than are traditionally possible.

The first property derivatives market launched on the London Futures and Options Exchange in 1991 with little success. After a decade, and reflecting the boom in physical house prices, property derivatives re-emerged with several spread betting markets over housing launched in the UK between 2001 and 2004. These have largely failed and since been abandoned.

The CME launched futures and options in May 2006 over house prices in ten cities in the USA (Boston, Chicago, Denver, Las Vegas, Los Angeles, New York, San Diego,

---

[52] The EDHED European Real Estate Investment and Risk Management Survey undertaken in November 2007 estimates that the annual savings on physical ownership made possible by synthetic investments range between 248 and 295 basis points in the UK.

San Francisco, Washington D.C.) and a national composite. These indices are based on the weighted repeat-sales method of Case and Shiller (1987). The exchange reports average volume of 30-40 transactions per day, totalling a notional traded volume of $300-350 million in first 12 months.

Shiller (2008) describes a possible reason for the relatively low liquidity in this market as driven by the order imbalance towards investors wishing to sell real estate futures. Short interest is understandable in a new market for housing as the primary group of investors are looking to hedge their existing exposure: the long interest will apply to those looking to add real estate to their portfolios, and consequently will require more time.

Several over-the-counter residential real estate derivative products have also come and gone. In the UK these have been written over the Halifax house price index, which uses hedonic estimation techniques. The most successful OTC market in real estate derivatives has been written over the index developed by Radar Logic (also in the USA), which has been active since September 2007.[53]

[53] 'US Property Derivatives – High Time?', *Total Derivatives*, October 2007.

7.3 Research Design

This section outlines the index construction methodology to be applied to the samples of timely data and the population of sales and the method of comparing their results.

7.3.1 Index Estimation

Hedonic price indices are estimated from each of the 'predictive' data sources and the population of sales. Hedonic regression is chosen as the primary index construction technique in this analysis over competing index construction methodologies since the underlying index for the OTC and impending exchanged-based residential property derivatives market in Australia is likely to take some hedonic form.[54] Furthermore, the hedonic index is shown in Chapter 4 to be less prone to spurious seasonality and autocorrelation in returns than the median and repeat-sales indices.

The specific hedonic index approach taken is the adjacent-period hedonic model, which was also used in the estimation of the hedonic index in Chapter 4. The adjacent-period approach, first suggested in Triplett (2004) and applied empirically to residential real estate markets in Wright (2006) is a variation on the traditional pooling approach of hedonic models.

The advantage of the adjacent-period approach is that it allows for changes in the implicit values of attributes through time by pooling only a subset of adjacent months'

---

[54] 'Betting on the House', *Sydney Morning Herald*, May 14 2009.

data at a time and linking month-on-month growth by the chain method to form an index growth series.[55]

The hedonic-regression model to be estimated by OLS independently for the house and unit subsamples of each sales dataset is given by Equation 7.1:

$$p_{i,t} = \alpha + \sum_{j=1}^{K} \beta_j.Attribute_{i,j} + \sum_{h=1}^{N} \gamma_h.Location_{i,h} + \sum_{t=1}^{3} \delta_t Time_t + \varepsilon_{i,t} \qquad 7.1$$

where,

$p_{i,t}$ is the natural log of sales price of the i^{th} residential property at time t

α is the intercept term

β and γ are explanatory variable coefficients reflecting the implicit value of the set of property attributes and suburb, respectively

δ_t estimates the cumulative growth rate to time t

$Time_t$ is a set of dummy variables equal to 1 if the property sold in time-period t and zero otherwise

$\varepsilon_{i,t}$ is the random variation in price of the i^{th} residential property at time t not captured by the model.

$Attribute_{i,j}$ represents the matrix of hedonic attributes included in the model. Given the data available to this study, this includes: land size (houses only), the number of bedrooms and bathrooms, and the presence of a swimming pool or air-conditioning.

[55] Alternative hedonic specifications, such as the pooled model or imputation model, are not considered in this dissertation but are expected to generate similar results.

$Location_{i,k}$ is a dummy variable used to control for the variation in prices attributable to geography. Again determined by the available data, it may represent the city or region of a property, or something as granular as the individual suburb or street. In order to maintain full-rank in the explanatory variable matrix, the coefficient for one of the *k* locations is restricted to zero in every iteration.

Wright (2006) shows that the adjacent-period hedonic model is able to allow for the implicit prices of housing attributes to change through time, setting it apart from the traditional pooled data approaches. To estimate the three-month adjacent-period model, *T-2* subsets of data are created. The hedonic function detailed in Equation 7.1 is estimated independently for each subset of data by OLS.

To ensure a full-rank matrix (as required under the assumptions of OLS estimation), the coefficient of the first time period dummy variable in each subset is also set equal to zero. Taking the difference in estimated coefficients of the third and second time-period dummy variables, $\Lambda_3 - \Lambda_2$, thus produces the logarithmic growth attributable to the third month in the given adjacent-period subset.

A weighted market-wide adjacent-period hedonic index, $I_{t,HED}$, based at 100 in its first month, is created following Equation 7.2:

$$I_{t,HED} = I_{t-1,HED} e^{\lambda_t} \qquad 7.2$$

given,

$$\lambda_t = w_H(\Lambda_{H,2} - \Lambda_{H,1}) + w_u(\Lambda_{U,2} - \Lambda_{U,1})$$

where,

$I_{1,HED} = 100$

e is the base of the natural logarithm,

λ_t are the differenced time dummy variable coefficient estimates of $\Lambda_3 - \Lambda_2$

t is month and year, for all $t = 1, 2, \ldots, T$, and there are T months

w_H and w_U are the average capitalisation weights of houses and units equal to 0.7 and 0.3, respectively[56]

and all other variables are as defined for Equation 7.1.

In addition to the hedonic index methodology, median-price based indices are estimated for the timely datasets. Median indices are computationally simpler to estimate and significantly less data-intensive than hedonic indices.

Following the methodology introduced in Chapter 4 of this dissertation, the median-price series, m_t, is calculated and transformed into an index, $I_{t,MED}$, based at 100 through a chain linking process. This process follows the methodology set out in Section 4.3.

7.3.2 Prediction Comparison

This section presents the methodology by which the market-wide and predictive indices estimated using the hedonic and median methodologies outlined in Section

[56] These weightings are provided by RPX. They represent the relative capitalisation-weighted representations of houses and units in the ASGC-defined Sydney metropolitan region over the sample period considered in this chapter.

7.3.1, are to be compared. Specifically, the ability of the predictive indices to explain movements in the market-wide index is examined.

Regression analysis is used to assess the ability of the predictive indices to explain movements in the market-wide index. Using the alternative prediction indices as the independent variable, the simple regression model given by Equation 7.3 is estimated by OLS:

$$Index_t = \alpha_{SAMPLE,x} + \beta_{SAMPLE,x} I_{SAMPLE,t,x} + e_{SAMPLE,t,x} \qquad 7.3$$

where,

$Index_t$ is the return on the market-wide index estimated from the population of sales at time t

α is the regression intercept

e is the random error to the model assumed to be i.i.d. $N(0,1)$

Given the samples of timely data used to estimate the predictive indices are subsets of the total population of sales which underlie the market-wide index, it is expected that β will be positive. Following the conjecture developed in Chapter 3, it is expected that private information, such as the knowledge set held by agents, is more valuable than public information sets, such as the sample of sales reported in newspapers. This finding would not permit rejection of *Hypothesis* 7_1, which is formalised as

$H7_1^0 : \beta_{AGENT} = \beta_{NEWS}$

$H7_1^A : \beta_{AGENT} > \beta_{NEWS}$

7.4 Data

This study uses several databases, most of which have not been previously employed for research. These databases reflect: (1) the population of sales, (2) the sample of sales and auction results reported in the weekend newspaper, (3) the set of property listings, and (4) a sample of real estate agents' sales results. These are each described in turn.

7.4.1 Data Sources

Sales data are sourced from RPX, a publicly listed property database manager. This particular database is collated from each state's VG. As such, it holds the virtual population of sales of residential property in Australia.

RPX supplements the raw sales data with property attribute information, which is of importance to this study since hedonic-regression techniques are employed. Attribute information is originally sourced from advertisements, real estate agents and other clients of RPX, such as mortgage lenders.

This study focuses on the Sydney residential property market. Sydney is the largest city in Australia by population and aggregate property value. The metropolitan area of Sydney is defined using the 2007 ASGC Statistical Division (SD).[57] Sales and attribute information for the period 1 January 1999 to 31 December 2008 is used to estimate a market-wide sales-based index.

---

[57] Source: *Standard Geography Volume I – Australian Standard Geographical Classification*, ABS, Catalogue 1216, July 2006.

Three samples of timely sales data are used in this chapter. The three timely sales data samples represent: (1) the sample of sales and auction results reported in the weekend newspaper, (2) the set of property listings, and (3) a sample of real estate agents' sales results. These are each described in turn.

Residential property sales and auction results from the *Sydney Morning Herald* (SMH) are collated for the period 1 March 1999 to 31 December 2004.[58] This is a publicly available sample of sales and includes various attribute information. The supplied suburb and postcode of each sale is matched to the ASGC SD definition to ensure the sample refers solely to the city of Sydney. No property attribute data are added to this dataset: it is created solely from information available in newspapers.

Note that data in this sample is not available for the months of January and February, when special editions of the SMH are printed, and recent sales results are not published, in line with the lower sales volumes during the summer holiday period.

A database containing listings data for residential property is sourced from RPX. This database represents close to the population of public listings in the market as it is mainly compiled from online listing sites.

This data contains the listing type – generally private treaty or auction – as well as the asking price for private treaty listings and the auction date for auction listings. Given that the estimation of the hedonic index model outlined in Section 2.1 requires as the

[58] This database was collated and supplied by the Faculty of Economics and Business at the University of Sydney.

dependent variable a price, listings without asking prices (mainly the set of auction listings) are excluded in the index estimation.

Using property-identified codes supplied by RPX, hedonic attribute information is merged to this sample of data. ASGC codes are used to restrict the sample to the Sydney metropolitan area. The final sample covers the period 1 January 2005 to 31 December 2008.

A sample of 'insider information' is also sourced from RPX. Through its client network, RPX receives real time advice from agents relating to their recent sales. This database includes the price of the property and the date at which the agent reported it to RPX.

Using the RPX property identifiers again, hedonic attribute information is appended. The sample is restricted to the Sydney metropolitan region, as with the other samples using the ASGC definitions. The date range of this data is 1 March 2006 to 31 December 2008.

7.4.2 Descriptive Statistics

This section reports the descriptive statistics for the datasets used in this paper, followed by a summary of the estimated full sample and prediction indices.

Table 7-1 reports the descriptive statistics for the alternative timely-data samples and sales population datasets to which the methodology in Section 7.3 is applied.

There are 488,489 observations of Sydney house sales and 345,642 observations of Sydney unit sales in the total sample period, January 1999 to December 2008. The average house (unit) price in this cross section was $493,112 ($382,377), while the median price was $400,000 ($345,000). This is in line with previous analyses of the Sydney housing market that have found a negative skew in prices.

Considering the three predictive samples individually, it can be seen that the average (median) prices in these samples is considerably higher than the population average (median). Sales reported in the SMH total 61,854 for houses and 22,609 for units during the period for which data are available, January 1999 to December 2004 (excluding the months of January and February every year).

Thus, the SMH sample represents 22% and 11.5% of the total number of house and unit sales, 281,766 and 196,345, respectively, in this market over the same period. The average price of a house sale reported in the SMH, however, is $760,244; a premium of over 50% above the average price of all house sales at that time.

Table 7-1:

Sample Observations and Prices

This table presents descriptive statistics relating to the number of observations and prices of the alternative datasets used in this study: the sales population, and the predictive data sources comprising SMH reporting, listings and agents' data. Also presented are descriptive statistics for the full sample by the subperiods available for the predictive samples. The results for houses and units are presented separately in Panels A and B, respectively.

			Sales Price ($)		
Data Sample	Sales	Proportion	Average	Median	Standard Deviation
Panel A: Houses					
Full Sales	488,849		493,112	400,000	365,417
1999-2004 (ex Jan-Feb)	281,766		442,468	365,000	310,806
May 2005-08	151,339		609,838	483,000	446,713
March 2006-08	119,500		618,342	485,000	461,589
SMH	61,854	22.0%	760,244	612,000	592,819
Listings	92,888	61.4%	612,005	529,000	354,145
Agents	69,608	58.2%	696,85	605,000	402,273
Panel B: Units					
Full Sales	345,642		382,377	345,000	193,223
1999-2004 (ex Jan-Feb)	196,345		359,302	325,000	180,127
May 2005-08	113,445		429,680	385,000	210,146
March 2006-08	91,104		431,718	385,000	212,235
SMH	22,609	11.5%	501,142	402,000	386,547
Listings	53,140	46.8%	420,320	389,950	151,528
Agents	34,224	37.6%	422,762	392,500	169,987

A higher proportion of total house and unit sales, 61.4% and 46.8%, respectively, are captured in the listings data. The proportion of house and unit sales held in the agents' private information database is 58.2% and 37.6%, respectively.

The price bias between the prediction data and the population data continues, with house prices higher in the listings and agents' data than the population of sales over the same period by 24% and 41%, respectively. Unit prices for all predictive samples are similarly biased, although to a lower extent.

Table 7-2 presents a set of descriptive statistics relating to the hedonic attributes and location of properties. An immediate observation from this is the over-representation of properties located in the city and Eastern Suburbs in the SMH and agents' databases. Location alone is known to be a major determiner of prices (Sirmans, Macpherson and Zietz, 2005), and is likely to be driving the price discrepancy between the predictive samples and the population. The databases do not appear to skew towards other attributes in this systematic way. A bias in the 'type' (location, quality, etc) of properties in the predictive samples may affect the results of this study if patterns of price growth differ significantly across the tiers of housing stock.

The following section presents and discusses the hedonic and median indices estimated from these alternative data sources.

Table 7-2:

Hedonic Attributes and Location

This table presents descriptive statistics relating to several hedonic property attributes and their location in Sydney. The proportion of sales observed by location is also presented. Panel A contains the house statistics; statistics for units are contained in Panel B.

Attribute	Population				Timely Data Samples		
	All	1999 – 2004 (ex Jan-Feb)	May 2005-08	March 2006-08	SMH	Listings	Agents
Panel A: Houses							
Land Size (m^2)	606.67	609.0	598.8	597.8	-	1204.8	639.8
Bedrooms	3.40	3.4	3.4	3.4	3.2	3.9	3.4
Bathroom	1.84	1.9	1.8	1.8	-	1.8	1.9
Pool (%)	5.2	3.3	9.3	8.5	-	8.7	10.2
Scenic View (%)	7.8	9.2	6.3	4.8	-	3.8	5.0
Waterfront (%)	0.8	0.7	0.9	0.9	-	0.0	0.8
Air-Conditioning (%)	4.8	2.0	10.8	10.5	-	11.8	11.9
City-East (%)	9.0	8.7	10.3	10.4	30.6	5.6	16.7
South (%)	29.8	30.0	29.3	29.3	21.8	26.9	24.2
West (%)	26.8	27.1	25.8	25.9	14.7	24.4	20.1
North (%)	34.4	34.2	34.5	34.3	33.0	43.2	38.9
Panel B: Units							
Bedrooms	2.2	2.2	2.2	2.2	2.0	2.2	2.1
Bathroom	1.5	1.5	1.5	1.5	-	1.4	1.5
Pool (%)	2.3	1.5	3.9	3.7	-	3.9	4.8
Scenic View (%)	9.6	11.6	7.5	6.2	-	3.9	5.8
Waterfront (%)	1.2	1.2	1.2	1.1	-	0.0	1.4
Air-Conditioning (%)	2.3	1.0	5.0	5.1	-	6.3	6.3
City-East (%)	26.7	27.0	26.7	26.9	54.4	18.6	26.7
South (%)	24.0	24.6	22.6	22.6	11.6	26.2	22.9
West (%)	20.2	20.0	20.6	20.5	10.6	17.2	15.7
North (%)	29.2	28.4	30.1	29.9	23.4	38.0	34.7

7.4.3 House Price Indices

Figures 7-1, 7-2 and 7-3 chart the Sydney residential real estate market-wide index and returns against those estimated from each of the timely-data samples. Both the hedonic and median predictive indices are presented. For comparative purposes, the market-wide index is rebased to 100 in the first month of the predictive index.

The Sydney residential property market has experienced two major phases over the last decade. For the period 1999 to the first quarter of 2004, the market grew at a steady rate, reaching a peak at March 2004. Panel A of Figure 7-1 shows that from a base of 100, the market-wide index almost doubled in this period. From Panel B of this figure the strength of the market through this period can be observed: the return to almost every month is positive.

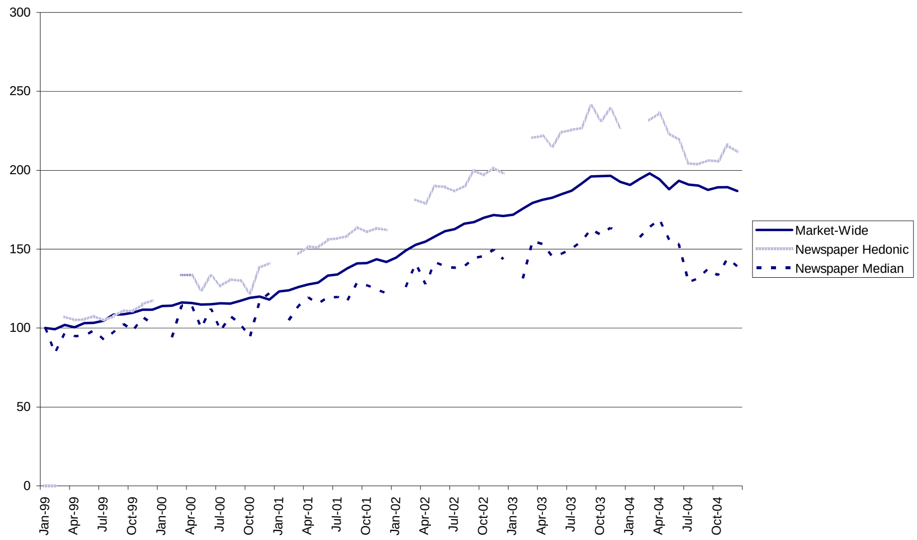
The following five years, approximately from the second quarter of 2004 to 2008 have seen the Sydney residential property market stagnate. The market-wide index portrayed in Figures 7-2 and 7-3 demonstrates this point. From a base of 100 in January 2005, the index has returned to approximately 100 in December 2008, having peaked during the period at 105 in mid-2007.

Visually, the general pattern of price rises and falls appears to be roughly matched by the predictive indices. The hedonic predictive indices, however, are more volatile than the market-wide index. In turn, the median-price based predictive indices are more volatile again. This is expected, given the smaller sales samples available to the timely datasets and supports the findings presented in Chapter 4 of higher volatility to median-price based indices.

Figure 7-1:
Index – SMH Data

This figure presents the results of fitting the sales data from the SMH sample following both the hedonic regression and median price methodologies described in Sections 7.2.1 and 7.2.2. Included for comparative purposes is the index estimated from the full sales sample. Panel A plots the indices, all based at 100 in January 1999, and Panel B plots the monthly index returns.

Panel A: Index



Panel B: Returns

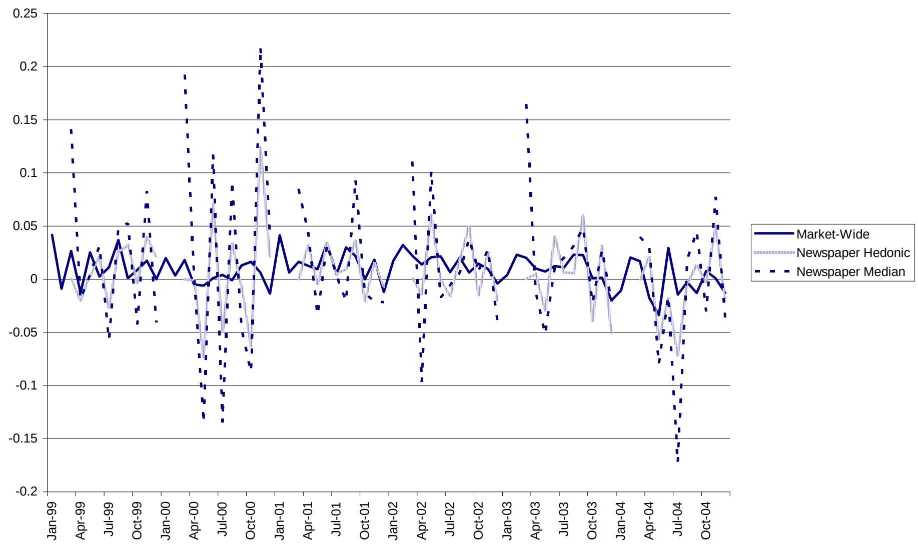
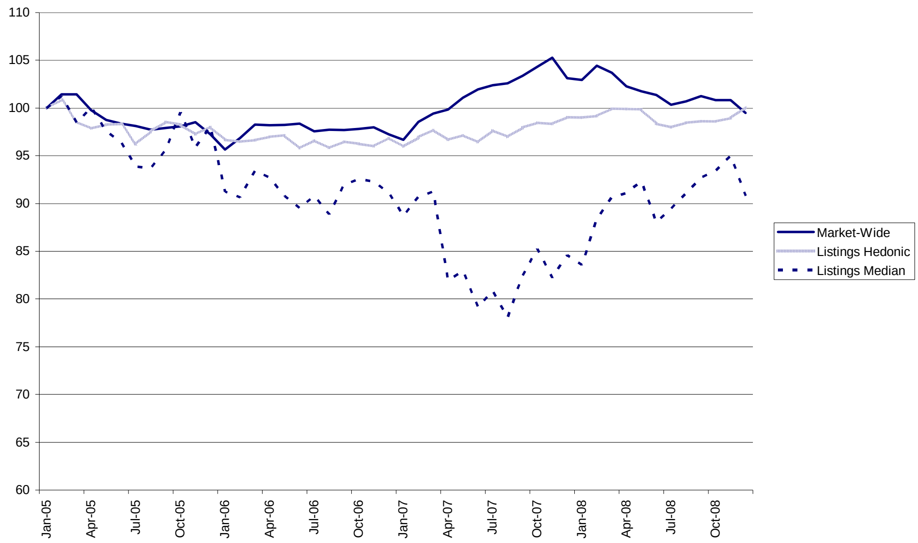


Figure 7-2:

Index – Listings

This figure presents the results of fitting the sales data from the advertisements (listings for sale) following both the hedonic regression and median price methodologies described in Sections 7.2.1 and 7.2.2. Included for comparative purposes is the index estimated from the full sales sample. Panel A plots the indices, all based at 100 in January 1999, and Panel B plots the monthly index returns.

Panel A: Index



Panel B: Returns

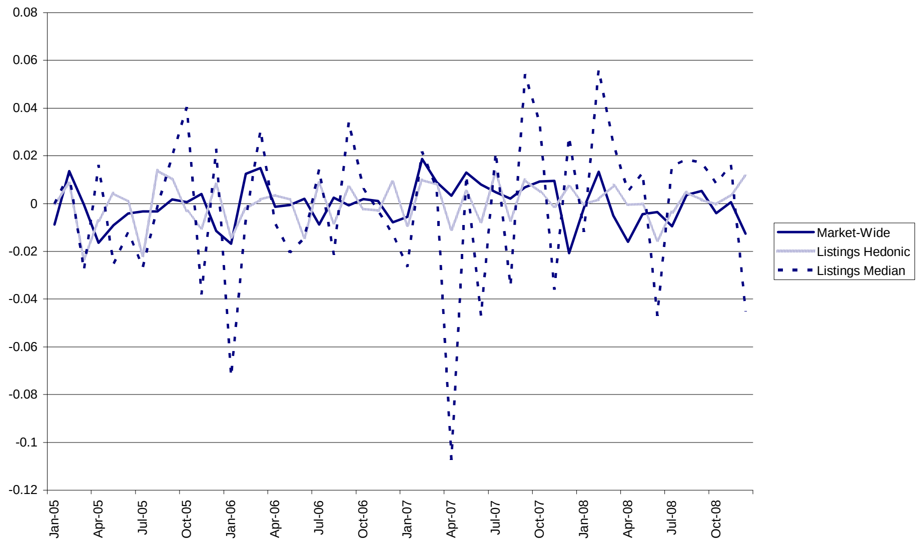
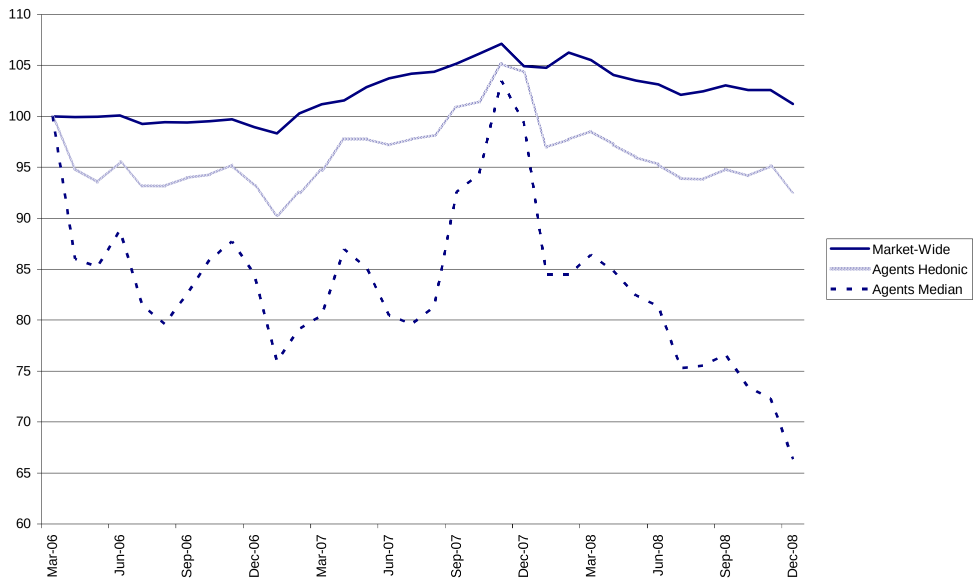


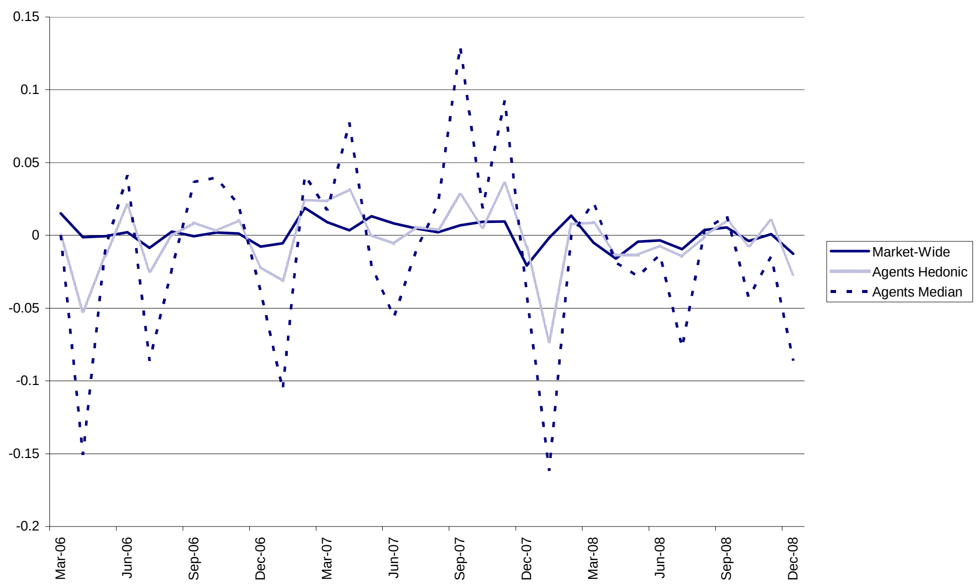
Figure 7-3:
Index – Agent's Information

This figure presents the results of fitting the sales data from the agents' set of sales information following both the hedonic regression and median price methodologies described in Sections 7.2.1 and 7.2.2. Included for comparative purposes is the index estimated from the full sales sample. Panel A plots the indices, all based at 100 in January 1999, and Panel B plots the monthly index returns.

Panel A: Index



Panel B: Returns



The dynamics of the prediction indices and the population index over the different subsamples are statistically analysed. Several statistics for the returns to the indices are presented in Table 7-3.

The statistics broadly support the visual observations made. For the entire sample period, 1999 to 2008, the Sydney residential real estate market grew at an average annual rate of 6.44%. This growth was not uniform. During the subperiod 1999 to 2004 it returned 7.70% on average, while for the subperiod 2005 to 2008 it returned -0.11%. Supporting the results made in Chapter 4, the kurtosis and skewness statistics indicate that residential property returns are not normally distributed.

The average return of the predictive hedonic indices over the given sample periods is close to the market-wide return. The average annualised return to the SMH-based hedonic index, for example, was 6.03%, while the return estimated from the market-wide index over the same period is 7.70%. This result provides early support for the value of timely data sources to predict market-wide index dynamics. On this measure, however, average return estimates from median-price indices are less valuable; the average annualised return to the SMH median-price index, by comparison, is 1.75%.

As observed in Figures 7-1, 7-2 and 7-3, the standard deviation of the predictive hedonic indices is greater than that of the market-wide indices, and the predictive medians in turn are substantially more volatile than the predictive hedonics.

The following section applied the methodology outlined in Section 7.3 to formally assess the predictive value of these timely data sources.

Table 7-3:

Index Dynamics

This table reports the dynamics of the returns to the population index and the predictive indices charted in Figures 7-1, 7-2 and 7-3. Statistics for the market-wide index for the entire sample period, 1999 to 2008, as well as the subperiods which match those of each of the timely datasets are reported in Panel A. Panels B and C report the statistics for the three alternative prediction indices as estimated using hedonic regression and median prices, respectively.

Variable	Average Annualised Return (%)	Annualised Standard Deviation (%)	Kurtosis	Skewness
Panel A: Market-Wide Index				
1999-2008	6.44	4.77	0.1211	0.2005
1999-2004	7.70	5.03	0.0925	-0.2507
2005-08	-0.11	3.13	-0.2181	-0.1668
2006-08	0.47	2.97	0.3412	-0.2906
Panel B: Hedonic-Predictive Indices				
SMH				
1999-2004	6.03	12.87	1.1181	0.2453
Listings				
2005-08	0.02	3.10	0.1034	-0.7395
Agents				
2006-08	-2.8	7.98	1.9206	-1.10019
Panel C: Median-Predictive Indices				
SMH				
1999-2004	1.75	22.95	1.8747	0.1788
Listings				
2005-08	-2.47	11.18	1.2891	-0.8394
Agents				
2006-08	-14.93	21.81	0.6829	-0.3384

7.5 Results

This section presents the results from applying the methodology outlined in Section 7.2.2 to test *Hypothesis* 7_1.

Firstly, a preliminary analysis into the relationship between the market-wide and predictive indices is undertaken. Table 7-4 reports the percentage of same-direction movements and the raw correlations between the returns in the market-wide index and the alternative prediction indices over the corresponding time periods.

From the results reported in Table 7-4 it can be seen that the predictive indices based on all samples of data and both index construction techniques – hedonic regression and median-price – move in the same direction as the market-wide index more than 50% of the time. The inside information, as expected, performs strongest, predicting the direction of the market 84.8% of the time when estimated using the hedonic method and 75.8% of the time when estimated using median prices. The publicly sourced SMH data also performs quite strongly on this measure, predicting the market in 66.7% of months using the hedonic index and 59.7% of months when using the median-price index.

This finding indicates that for the sample periods considered, these timely and public sales information sets may have been utilised to correctly predict the movement in the delayed-publication market-wide index most of the time. That the percentage of correct direction predictions is higher when the private sample of data is used is in line with the *a priori* expectations underlying *Hypothesis* 7_1.

Table 7-4:

Correlation Results

This table reports the results of correlation analysis between the predictive indices based on SMH, listings and agents' data and the market-wide index. Panel A presents the results for the predictive hedonic indices and Panel B presents the results for the predictive median-price indices. The statistics reported are the number of monthly return observations for the indices in their sample period, the percentage of months for which the indices move in the same direction, and the Pearson's Correlation statistic and the *p*-value of this statistic under the null hypothesis that the correlation equals zero. Rejection of this null is represented by *.

	SMH	Listings	Agents
Panel A: Hedonic			
Observations	54	47	33
Same Direction (%)	66.7	51.1	84.8
Pearson Correlation	0.3252 *	0.1021	0.5185
Correlation *p*-value	0.016	0.495	0.002
Panel B: Median			
Observations	54	47	33
Same Direction (%)	59.7	59.6	75.8
Pearson Correlation	0.3210 *	0.1837	0.4654
Correlation *p*-value	0.008	0.216	0.006

The Pearson's Correlation statistic supports the existence of a positive relationship between indices estimated from the newspaper and agents' advice data sources and a market-wide index. The correlation is highest for the agents' data – 0.5185 and 0.4654 from the hedonic index and the median index, respectively – followed by the SMH and the listings data. These results support the premises of *Hypothesis* 7_1 which are now formally tested using regression analysis.

Table 7-5 presents the regression results from the fitted models given by Equation 7.4. The data from sales reported in the SMH newspaper has a high degree of predictive power. The estimates of $\beta_{SMH,HED}$ (0.1270) and $\beta_{SMH,,MED}$ (0.0606) are statistically significant and positive. From this finding *Hypothesis* 7_1 may not be rejected. These estimates indicate that SMH reporting of sales may be used to predict with confidence the direction of the market prices and the relative magnitude of this movement.

In line with the *a priori* expectation that private information is of more value than public information, and consequently has higher predictive power, use of agents' data are found to explain a higher degree of variation in the market-wide index. Using this data alone achieves a regression statistic R^2 of 0.2452 in the hedonic index and 0.1913 in the median index. Estimates of $\beta_{AGENT,HED}$ (0.1931) and $\beta_{AGENT,,MED}$ (0.0634) are highly statistically significant and positive. As a result of this finding, *Hypothesis* 7_1 may not be rejected.

The sample of listing price data is not found to statistically explain any movement in the market-wide index. Some part of the poor result from the listings data may be attributable to the sample-period available to this data: as the index chart in Figure 7-2 and the return statistics reported in Table 7-3 demonstrate, the Sydney residential real estate market in the period 2005 through 2008 was stagnant, with low growth and volatility. Reassessment of this source of data for predictive content over another period needs to be undertaken as data becomes available.

Table 7-5:

Regression Results

This table reports the estimated parameters of the fitting the model given by Equation 7.3. Panel A presents results when the predictive hedonic indices are used and Panel B presents the results from using the predictive median-price index. The *t* statistic of the coefficient estimate, the number of observations used and the Adjusted R^2 statistic are also reported. Significance of the estimated intercept and independent variable coefficient at the 1% and 5% level of significance is denoted by *** and *, respectively.

	SMH	Listings	Agents
Panel A: Hedonic			
Intercept	0.0058 ***	-0.0001	0.0008
Coefficient Estimate	0.1270 *	0.1029	0.1931 ***
t Statistic	2.48	0.69	3.38
Observations	54	47	33
Adjusted R^2	0.0886	-0.0116	0.2452
Panel B: Median			
Intercept	0.0084 ***	0.0000	0.0012
Coefficient Estimate	0.0606 ***	0.0514	0.0634 ***
t Statistic	2.7328	1.2540	2.9274
Observations	54	47	33
Adjusted R^2	0.0893	0.0123	0.1913

7.6 Summary

This study demonstrates the effectiveness of publicly and privately available samples of sales data to predict movements in a population-based house price index. As expected, the inside information held by agents in this market performs the best on a number of tests of predictive ability. Publicly available data, however, from the SMH newspaper and even from advertisements of properties for sale is shown to perform reasonably well. Both sources of data predict movements in the market more than half the time. Furthermore, the data from the SMH is shown using regression analysis to have some power in predicting the relative size of the market-wide price movement.

This result suggests a profitable trading strategy exists where derivative products over a population-based housing index are available. Such products have been listed in the USA, with plans being made to introduce similar investments to markets in the UK, Asia and Australia. By accessing the ‘predictive’ information investors are able to estimate the latest market movements. This will be known up to a month prior to the publication of and settlement against a population-based index.

8. Conclusion

This chapter concludes the dissertation. The following sections provide a summary of the findings from each of the empirical chapters; outline the key contributions of this research and the implications of the results; and suggests several directions for future research in this area.

8.1 Summary of Findings

In each chapter, methodology – and, specifically, the way in which heterogeneity across property assets is controlled – is found to have an impact on the results of the research.

Chapter 4 examined the efficiency of the Sydney and Melbourne residential real estate markets. Through the estimation of ARIMA models, the hypothesis of market efficiency is rejected in its weakest form. Regression analysis, however, indicates that monthly seasonality is not a determinant of returns to the housing market. The results demonstrate that the conclusions to such research are not consistent across the returns from alternative index estimation methods. Returns to the median, for example, indicate a significant negative first- and second-order autocorrelation and a significant seasonal component. Returns from repeat-sales index, on the other hand, indicate significant positive short-order inertia in returns.

The findings of the empirical research presented in Chapter 5 indicate that the auction sale mechanism has no significant price impact. This is a major finding, as it opposes the conclusions of all previous research in this area. This conclusion is made using a detailed database of Sydney house sales to estimate sample selectivity-corrected hedonic regressions. The methods undertaken include a Heckman correction and a matched sampling procedure (Rubin, 1979). Uncorrected hedonic regressions, following the Dotzour et al. (1998) methodology, however, support the existence of a price premium to auctions over private treaty sales in the Sydney residential real estate market. These findings suggest that the auction premium that has previously been documented in the residential real estate markets of Australia and New Zealand is a function of the methodology used rather than a result of the sale mechanism.

Regression analysis is applied in Chapter 6 to examine the pricing of new properties in Perth. The results predict a price premium of 10% (7%) in sales of new houses (units) over existing properties, *ceteris paribus*. The robustness of these results is supported by the use of Heckman two-stage estimation technique and use of a matched sampling procedure. The relative price performance of new properties with the whole market is also compared using an annualised compound return metric. It is found that houses (units) in Perth underperform the market by 6.32% (2.1%).

Chapter 7 considers the value of private and public information in a hypothetical residential real estate derivatives market. Using an underlying house price index for the population of sales in Sydney, it is shown that price movements based on the sample of sales from public newspaper reporting and agents' inside information are potentially strong predictors of index movements. This result holds using both advanced hedonic techniques and the less intensive method of tracking changes in the median sales price.

8.2 Contributions and Implications

The research presented in this dissertation adds significantly to the existing literature to have considered pricing in the residential real estate market. Much of this work has not been previously applied to Australian data, or with as comprehensive a database. As a result several completely innovative research areas were explored and methods beyond the ability of past research were used.

The research presented in Chapter 4 represents the first robust analysis of weak-form efficiency in the returns to the residential property markets of Australia's two largest cities, Sydney and Melbourne. The results presented in Chapter 5 conflict with those of previous research which have suggested the existence of a price premium attributable to auctions. This implies that the observation of an 'auction premium' in previous work may be a result of methodological shortcomings. Chapter 6 contains the first academic research to explicitly examine both the pricing and performance of new residential properties. The theoretical context for the existence of an information asymmetry in sales of new properties, as developed through Sections 2.3.3 and 3.3, is also the first of its kind. Finally, in considering the design of residential real estate derivative products, the research presented in Chapter 7 is the first research to examine the value of information in this market, particularly given the opportunity for higher-frequency trading through such products.

These contributions are enabled by access to an incredibly extensive database of the virtual population of property sales in Australia. A further advantage is the ability to use and consequently empirically compare the results from alternative methods in residential real estate pricing research

Using competing methodologies concurrently, this thesis makes two further contributions to the residential real estate pricing literature. Firstly, it is shown that the choice of method has a significant impact on the results of such research. Secondly, the direction and cause of this impact is found.

This finding has a number of implications for the interpretation of past and the design of future research. The results of past research must be viewed in the context of the author's methodological choices. Future research must take care in its choice of methodology and be aware of the limitations.

8.3 Further Research Directions

This dissertation makes several important contributions to the literature. It is worth noting, however, that these findings represent a single case study. The extension of this work to other markets is essential for the external validity of these conclusions.

It is inevitable that sales and attributes data for transactions of residential real estate will become more accessible into the future. As this takes place, future research can extend the comparisons made in this dissertation to other markets. Of particular interest will be the application of alternative methods to markets where the bulk of past research has taken place, particularly the USA.

With broader understanding of the biases and effects peculiar to individual methodologies and access to improved data, future research may apply more robust techniques in studies of residential real estate market structure and price dynamics. The research presented in this dissertation, for example, did not incorporate the yields to residential real estate investment.

Future research may also build upon the new areas of research developed in this dissertation, such as the pricing of new properties and the structure of housing derivatives markets. Exploration of other leading information factors in predicting derivatives movements presents a possible avenue.

Like any developing area of academic interest, the future research options in the residential real estate market are many and varied. Though the availability of data and technology may determine the direction of subsequent research, subjects worthy of exploration include the efficiency of the market, asset valuation methods, and the strategies and behaviours of participants. Ultimately, as understanding of residential real estate markets improves and the ability to invest in property assets and associated strategies are redefined through emerging real estate derivatives markets,[59] the importance of research such as this will continue to grow.

---

[59] For example, the CME-listed futures and options written over repeat-sales indices for 10 USA cities in 2006. Recent media has also reported the ASX is planning to list derivatives over hedonic indices for Australian cities (see, for example, 'Betting on the House', *Sydney Morning Herald*, May 14 2009).

References

Abraham, J., and W. Schaumann, 1991, New Evidence on Home Prices from Freddie Mac Repeat Sales, *Real Estate Economics* 19(3), 333-352.

Adams, P., B. Kruger and S. Wyatt, 1992, Integrating Auction and Search Markets: The Slow Dutch Auction, *Journal of Real Estate Finance and Economics* 5(3), 239-254.

Aggarwal, R., and P. Rivoli, 1989, Seasonal and Day of the Week Effects in Four Emerging Stock Markets, *Financial Review* 24(4), 541-550.

Akerlof, G., 1970, The Market for Lemons: Quality Uncertainty and the Market Mechanism, *Quarterly Journal of Economics* 84(3), 488-500.

Allen, M., and J. Swisher, 2000, An Analysis of the Price Formation Process at a HUD Auction, *Journal of Real Estate Research* 20(3), 279-298.

Amidu, A., and A. Agboola, 2009, Empirical Evidence of The Influences On First-Price Bid Auction Premiums, *International Real Estate Review* 12(2), 157-170.

Ashenfelter, O., and D. Genesove, 1992, Testing for Price Anomalies in Real-Estate Auctions, *American Economic Review* 82(2), 501-505.

Bailey, M., R. Muth and H. Nourse, 1963, A Regression Method for Real Estate Price Index Construction, *Journal of the American Statistical Association* 58(304), 933-942.

Baron, D., 1979, The Incentive Problem and the Design of Investment Banking Contracts, *Journal of Banking and Finance* 3(2), 157-175.

Baron, D., 1982, A Model of the Demand for Investment Banking, Advising, and Distribution Services for New Issues, *Journal of Finance* 37(4), 955-976.

Baron, D., and B. Holmström, 1980, The Investment Banking Contract for New Issues Under Asymmetric Information: Delegation and the Incentive Problem, *Journal of Finance* 35(5), 1115-1138.

Baye, M., 2009, *Managerial Economics and Business Strategy*, McGraw-Hill: Boston.

Beatty, R., and J. Ritter, 1986, Investment Banking, Reputation, and the Underpricing of Initial Public Offerings, *Journal of Financial Economics* 15(1), 213-232.

Beggs, A., and K. Grady, 1997, Declining Values and the Afternoon Effect: Evidence from Art Auctions, *RAND Journal of Economics* 28(3), 544-565.

Ben-Shahar, D., 2003, A Performance Comparison between Dwellings and Financial Assets in Israel, *Journal of Real Estate Literature* 11(2), 179-194.

Berkovec, J., and J. Goodman, 1996, Turnover as a Measure of Demand for Existing Homes, *Real Estate Economics* 24(4), 421-440.

Bodie, Z., A. Kane and A. Marcus, 2002, *Investments,* McGraw Hill: New York.

Bond, M., and M. Seiler, 1998, Real Estate Returns and Inflation: An Added Variable Approach, *Journal of Real Estate Economics* 15(3), 327-338.

Booth, J., and L. Chua, 1996, Ownership Dispersion, Costly Information and IPO Underpricing, *Journal of Financial Economics* 41(2), 291-310.

Brennan, M., and J. Franks, 1997, Underpricing, Ownership and Control in Initial Public Offerings of Equity Securities in the UK, *Journal of Financial Economics* 45(3), 391-413.

Brown, R., L. Li and K. Lusht, 2000, A Note on Intracity Diversification of Real Estate Portfolios: Evidence from Hong Kong, *Journal of Real Estate Portfolio Management* 6(1), 131-140.

Campa, J., and N. Fernandes, 2006, Sources of Gains from International Portfolio Diversification, *Journal of Empirical Finance* 13(4), 417-443.

Campbell, J., A. Lo and A. MacKinlay, 1997, *The Econometrics of Financial Markets,* Princeton University Press: New Jersey.

Campbell, J., and J. Cocco, 2007, How Do House Prices Affect Consumption: Evidence from Micro Data, *Journal of Monetary Economics* 54(3), 591-621.

Can, A., 1992, Specification and Estimation of Hedonic Housing Price Models, *Regional Science and Urban Economics* 22(3), 453-474.

Capozza, D., P. Hendershott, C. Mack and C. Mayer, 2002, *Determinants of Real House Price Dynamics*, Working Paper, NBER.

Case, B., H. Pollakowski and S. Wachter, 1991, On Choosing Among House Price Index Methodologies, *Real Estate Economics* 19(3), 286-307.

Case, K., J. Quigley and R. Shiller, 2005, Comparing Wealth Effects: The Stock Market Versus the Housing Market, *Advances in Macroeconomics* 5(1), 1-32.

Case, K., and R. Shiller, 1987, Price of Single Family Homes Since 1970: New Indexes for Four Cities, *New England Economic Review* 19(5), 45-56.

Case, K., and R. Shiller, 1989, The Efficiency of the Market of Single Family Homes, *American Economic Review* 79(1), 125-137.

Catte, P., N. Girouard, R. Price and C. André, 2004, *Housing Markets, Wealth and the Business Cycle*, Working Paper, Economics Department, OECD.

Chang, C., and C. Ward, 1993, Forward Pricing and the Housing Market: The Pre-Sales Housing System in Taiwan, *Journal of Property Research* 10(3), 217-227.

Chua, A., 1999, The Role of International Real Estate in Global Mixed-Asset Investment Portfolios, *Journal of Real Estate Portfolio Management* 5(2), 129-137.

Clapp, J., and C. Giacotto, 1992, Estimating Price Indices for Residential Property: A Comparison of Repeat-Sales and Assessed Value Methods, *Journal of the American Statistical Association* 5(4), 357-374.

Clayton, J., 1998, Further Evidence on Real Estate Market Efficiency, *Journal of Real Estate Research* 15(1), 41-57.

Cocco, J., 2005, Portfolio Choice in the Presence of Housing, *Review of Financial Studies* 18(2), 535-567.

Coniffe, D., and D. Duffy, 1999, Irish House Price Indices – Methodological Issues, *Economic and Social Review* 30(4), 403-423.

Connolly, R., 1989, An Examination of the Robustness of the Weekend Effect, *Journal of Financial and Quantitative Analysis* 24(2), 133-169.

Copeland, T., J. Weston and K. Shastri, 2005, *Financial Theory and Corporate Policy: International Edition*, Pearson Addison Wesley: Sydney.

Costello, G., 2001, Seasonal Influences in Australian Housing Markets, *Pacific Rim Property Research Journal* 7(1), 47-60.

Crone, T., and R. Voith, 1992, Estimating House Price Appreciation: A Comparison of Methods, *Journal of Housing Economics* 2(4), 324-338.

Davidson, S., and R. Faff, 1999, Some Additional Australian Evidence on the Day-of-the-Week Effect, *Applied Economics Letters* 6(4), 247-249.

Davis, M., and M. Palumbo, 2001, *A Primer on the Economics and Time Series Econometrics of Wealth Effects*, Finance and Economics Discussion Series, Federal Reserve Board, Washington.

Deaton, A., and J. Muellbauer, 1983, *Economics and Consumer Behaviour,* Cambridge University Press: Cambridge.

DiPasquale, D., and W. Wheaton, 1996, *Urban Economics and Real Estate Markets,* Prentice-Hall: New Jersey.

Dickey, D., D. Hasza and W. Fuller, 1984, Testing for Unit Roots in Seasonal Time Series, *Journal of the American Statistical Soceity* 79(386), 355-367.

Dotzour, M., E. Moorhead and D. Winkler, 1998, The Impact of Auctions on Residential Sales Prices in New Zealand, *Journal of Real Estate Research* 16(1), 57-71.

Driessen, J., and L. Laeven, 2007, International Portfolio Diversification Benefits: Cross-Country Evidence from a Local Perspective *Journal of Banking and Finance* 31(6), 1693-1712.

Eichholtz, P., M. Hoesli, M. MacGregor and N. Nanthakumaran, 1995, Real Estate Portfolio Diversification by Property Type and Region, *Journal of Property Finance* 6(3), 39-59.

Englund, P., M. Hwang and J. Quigley, 2002, Hedging Housing Risk, *Journal of Real Estate Finance and Economics* 24(1), 167-200.

Englund, P., J. Quigley and C. Redfearn, 1998, Improved Price Indexes for Real Estate: Measuring the Course of Swedish House Prices, *Journal of Urban Economics* 44(2), 171-196.

Fama, E., 1965, The Behaviour of Stock Market Prices, *Journal of Business* 38(1), 34-105.

Fama, E., 1970, Efficient Capital Markets: A Review of Theory and Empirical Work, *Journal of Finance* 25(2), 383-417.

Fama, E., 1991, Efficient Capital Markets: II, *Journal of Finance* 46(5), 1575-1617.

Fisher, J., and C. Sirmans, 1994, The Role of Commercial Real Estate in a Multi-Asset Portfolio, *Journal of Property Management* 59(1), 54-59.

Flavin, M., and T. Yamashita, 2002, Owner-Occupied Housing and the Composition of the Household Portfolio, *American Economic Review* 92(1), 345-362.

French, K., and R. McCormick, 1984, Sealed Bids, Sunk Costs, and the Process of Competition, *Journal of Business* 57(4), 417-441.

Friedman, H., 1971, Real Estate Investment and Portfolio Theory, *Journal of Financial and Quantitative Analysis* 6(2), 861-874.

Gatzlaff, D., and D. Haurin, 1997, Sample Selection Bias and Repeat-Sales Index Estimates, *Journal of Real Estate Finance and Economics* 14(1), 33-50.

Gatzlaff, D., and D. Ling, 1994, Measuring Changes in Local House Prices: An Empirical Investigation of Alternative Methodologies, *Journal of Urban Economics* 35(2), 221-244.

Gau, G., 1984, Weak Form Tests of the Efficiency of Real Estate Investment Markets, *Financial Review* 19(4), 301-320.

Gau, G., 1985, Public Information and Abnormal Returns in Real Estate Investments, *Real Estate Economics* 13(1), 15-31

Gau, G., 1987, Efficient Real Estate Markets: Paradox or Paradigm? *Real Estate Economics* 15(1), 1-15.

Gau, G., and D. Quan, 1992, *Market Mechanism Choice and Real Estate Disposition: Negotiated Sale Versus Action*, Working Paper, Anderson Graduate School of Management, UCLA.

Gay, G., and T. Kim, 1987, An Investigation into Seasonality in the Futures Market, *Journal of Futures Markets* 7(2), 169-181.

Geltner, D., 1993, Estimating Market Values from Appraised Values without Assuming an Efficient Market, *Journal of Real Estate Research* 8(3), 325-245.

Geltner, D., 1997, Bias and Precision of Estimates of Housing Investment Risk Based on Repeat-Sales Indices: A Simulation Analysis, *Journal of Real Estate Finance and Economics* 14(1), 155-171.

Genesove, D., and C. Mayer, 1991, Loss Aversion and Seller Behaviour: Evidence from the Housing Market, *Quarterly Journal of Economics* 116(4), 1233-1260.

Goetzmann, W., 1993, Accounting for Taste: Art and Financial Markets Over Three Centuries, *American Economic Review* 83(5), 1370-1376.

Goetzmann, W., and R. Ibbotson, 1990, The Performance of Real Estate as an Asset Class, *Journal of Applied Corporate Finance* 3(1), 65-76.

Goetzmann, W., and M. Spiegel, 1995, Non-Temporal Components of Real Estate Appreciation, *Review of Economics and Statistics* 77(1), 199-206.

Goodman, J., 1993, A Housing Market Matching Model of the Seasonality in Geographic Mobility, *Journal of Real Estate Research* 8(1), 117-137.

Granger, C., and P. Newbold, 1974, Spurious Regressions in Econometrics, *Journal of Econometrics* 2(2), 111-120.

Grauer, R., and N. Hakansson, 1995, Gains from Diversifying into Real Estate: Three Decades of Portfolio Returns based on the Dynamic Investment Model, *Real Estate Economics* 23(2), 117-159.

Greene, W., 2003, *Econometric Analysis,* Pearson Education: New Jersey.

Greenwood, J., and Z. Hercoqitz, 1991, The Allocation of Capital and Time Over the Business Cycle, *Journal of Political Economy* 99(61), 1188-1214.

Griliches, Z., 1971, *Hedonic Price Indices for Automobiles: An Econometric Analysis of Quality Change,* Harvard University Press: Cambridge.

Groenewold, N., and K. Kang, 1993, The Semi-Strong Efficiency of the Australian Share Market, *The Economic Record* 69(207), 405-410.

Gultekin, M., and N. Gultekin, 1983, Stock Market Seasonality: International Evidence, *Journal of Financial Economics* 12(4), 469-481.

Gurdgiev, C., S. Stevenson and J. Young, 2010, A Comparison of the Appraisal Process for Auction and Private Treaty Residential Sales, *Journal of Housing Economics* forthcoming.

Heckman, J., 1978, A Partial Survey of Recent Research on the Labour Supply of Women, *American Economic Review* 68(2), 200-207.

Heckman, J., 1979, Sample Selection Bias as a Specification Error, *Econometrica* 47(1), 153-61.

Hill R., C. Sirmans and J. Knight, 1999, A Random Walk Down Main Street? *Regional Science and Economics* 29(1), 89-103.

Himmelberg, C., C. Mayer and T. Sinai, 2005, Assessing High House Prices: Bubbles, Fundamentals, and Misperceptions, *Journal of Economic Perspectives* 19(4), 67-92.

Hoesli, M., and F. Hamelink, 1997, An Examination of the Role of Geneva and Zurich Housing in Swiss Institutional Portfolios, *Journal of Property Valuation and Investment* 15(4), 354-71.

Hosios, A., and J. Pesando, 1991, Measuring Prices in Resale Housing Markets in Canada: Evidence and Implications, *Journal of Housing Economics* 1(4), 303-317.

Hutchison, N., 1994, Housing as an Investment? A Comparison of Returns from Housing with Other Types of Investment, *Journal of Property Finance* 5(2), 47-61.

Iacoviello, M., 2005, House Prices, Borrowing Constraints, and Monetary Policy in the Business Cycle, *American Economic Review* 95(3), 739-764.

Ibbotson, R., 1975, Price Performance of Common Stock New Issues, *Journal of Financial Economics* 2(3), 235-272.

Ibbotson, R., and L. Siegel, 1984, Real Estate Returns: A Comparison with Other Investments, *Real Estate Economics* 12(3), 219-242.

Ito, T., and K. Hirono, 1993, *Efficiency of the Tokyo Housing Market*, Working Paper, NBER.

Kallberg, J., C. Liu and W. Grieg, 1996, The Role of Real Estate in the Portfolio Allocation Process, *Real Estate Economics* 24(3), 359-378.

Kiyotaki, N., and J. Moore, 1997, Credit Cycles, *Journal of Political Economy* 105(2), 211-248.

Kuo, C, 1996, Serial Correlation and Seasonality in the Real Estate Market, *Journal of Real Estate Finance and Economics* 12(2), 139-162.

Leamer, E., 2007, *Housing is the Business Cycle*, Working Paper, NBER.

Lee, C., 2008, Housing in Australia as a Portfolio Investment, *International Journal of Housing Markets and Analysis* 1(4), 352-361.

Lin, V., and C. Huang, 2005, *The Comparison between Semi-parametric and Parametric CAMA Modelling of Court Auction Residential Housing Market in the Taipei Metropolitan Area*, presented at the 10th Asian Real Estate Society International Conference (Sydney).

Lin, V., F. Tsai and C. Chang, 1997, *The Study of Price Factors on the Auction Housing Market in Taipei City*, presented at the Conference of Chinese Society of Housing Studies (Taipei).

Lo, A., and A. MacKinlay, 1990, An Econometric Analysis of Nonsynchronous-Trading, *Journal of Econometrics* 45(1), 181-212.

Loughran, T., and J. Ritter, 1995, The New Issues Puzzle, *Journal of Finance* 50(1), 23-51.

Loughran, T., and J. Ritter, 2002, Why Don't Issuers Get Upset About Leaving Money on the Table in IPOs?, *Review of Financial Studies* 15(2), 413-444.

Lusht, K., 1996, A Comparison of Prices Brought by English Auction and Private Negotiations, *Real Estate Economics* 24(4), 517-530.

MacGregor, B., and N. Nanthakumaran, 1992, The Allocation to Property in the Multi-Asset Portfolio: The Evidence and Theory Reconsidered, *Journal of Property Research* 9(1), 5-32.

Maher, C., 1989, Information, Intermediaries and Sales Strategy in an Urban Housing Market: The Implications of Real Estate Auctions in Melbourne, *Urban Studies* 26(5), 495-509.

Mahieu, R., and A. van Bussel, 1996, *A Repeat-Sales Index for Residential Property in the Netherlands*, Working Paper, University of Maastricht.

Mandelker, G., and A. Raviv, 1977, Investment Banking: An Economic Analysis of Optimal Underwriting Contracts, *Journal of Finance* 32(3), 683-694.

Mark, J., and M. Goldberg, 1984, Alternative Housing Price Indices: An Evaluation, *Real Estate Economics* 12(1), 30-49.

Markowitz, H., 1952, Portfolio Selection, *Journal of Finance* 7(1), 77-91.

Matysiak, G., and P. Wang, 1995, Commercial Property Market Prices and Valuations: Analysing the Correspondence, *Journal of Property Research* 12(3), 181-202.

Mayer, C., 1995, A Model of Negotiated Sales Applied to Real Estate Auctions, *Journal of Urban Economics* 38(1), 1-22.

Mayer, C., 1998, Assessing the Performance of Real Estate Auctions, *Real Estate Economics* 26(1), 41-66.

Milgrom, P., 1987, *Auction Theory* Cambridge University Press: Cambridge.

Milgrom, P., 1989, Auctions and Bidding: A Primer, *Journal of Economic Perspectives* 3(1), 3-22.

Montezuma, J., and K. Gibb, 2006, Residential Property as an Institutional Asset: The Swiss and Dutch Cases, *Journal of Property Research* 23(4), 323-345.

Moran, A., 2006, *The Tragedy of Planning: Losing the Great Australian Dream* Pinnacle Print Group: Melbourne.

Muscarella, C., and M. Vetsuypens, 1989, A Simple Test of Baron's Model of IPO Underpricing, *Journal of Financial Economics* 24(1), 125-135.

Newell, G., J. MacFarlane, K. Lusht and S. Bulloch, 1993, *Empirical Analysis of Real Estate Auction Versus Private Sale Performance*, Working Paper, University of Western Sydney.

Ngai, L., and S. Tenreyro, 2009, *Hot and Cold Seasons in the Housing Market*, Working Paper, Financial Markets Group Research Centre, London School of Economics.

Officer, R., 1975, Seasonality in Australian Capital Markets: Market Efficiency and Empirical Issues, *Journal of Financial Economics* 2(1), 29-51.

Oikarinen, E., M. Hoesli and C. Serrano, 2009, Linkages between Direct and Securitized Real Estate, *Journal of Real Estate Finance and Economics*, forthcoming.

Ong, S., 2006, Price Discovery in Real Estate Auctions: The Story of Unsuccessful Attempts, *Journal of Real Estate Research* 28(1), 39-59.

Ong, S., and K. Ng, 2009, Developing the Real Estate Derivative Market for Singapore: Issues and Challenges, *Journal of Property Investment and Finance* 27(4), 425-432.

Praetz, P., 1973, A Spectral Analysis of Australian Share Prices, *Australian Economic Papers* 12(20), 70-78.

Prasad, N., and A. Richards, 2006, *Measuring House Price Growth – Using Stratification to Improve Median-Based Measures*, Research Discussion Paper, Reserve Bank of Australia.

Quan, D., 2002, Market Mechanism Choice and Real Estate Disposition: Search Versus Auction, *Real Estate Economics* 30(3), 365-384.

Quigley, J., 2001, Real Estate and The Asian Crisis, *Journal of Housing Economics* 10(2), 129-161.

Rayburn, W., M. Devaney and R. Evans, 1987, A Test of Weak-Form Efficiency in Residential Real Estate Returns, *Real Estate Economics* 15(3), 220-233.

Reed, R., J. Robinson and P. Williams, 2002, *Does an Auction Represent Fair Market Value*, presented at the University of Queensland Property Conference (Brisbane).

Riley, J., and W. Samuelson, 1981, Optimal Auctions, *American Economic Review* 71(3), 381-392.

Ritter, J., 2003, Differences between American and European IPO Markets, *European Financial Management* 9(4), 421-434.

Rosen, S., 1974, Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition, *Journal of Political Economy* 82(1), 34-55.

Rosenthal, L., 2006, Efficiency and Seasonality in the UK Housing Market, *Oxford Bulletin of Economics and Statistics* 68(3), 289-317.

Rossini, P., 2000, *Estimating the Seasonal Effects of Residential Property Markets – A Case Study of Adelaide*, presented at the Sixth Annual Pacific-Rim Real Estate Society Conference (Sydney).

Rossini, P., 2002, *Calculating Stratified Residential Property Price Indices to Test for Differences in Trend, Seasonality and Cycle*, presented at the Eighth Annual Pacific-Rim Real Estate Society Conference (Christchurch).

Roulac, S., 1976, Can Real Estate Returns Outperform Common Stocks, *Journal of Portfolio Management* 2(1), 26-43.

Rozeff, M., and W. Kinney, 1976, Capital Market Seasonality: The Case of Stock Returns, *Journal of Financial Economics* 3(4), 379-402.

Rubin, D., 1979, Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies, *Journal of the American Statistical Society* 74(366), 318-328.

Schneeweis, T., and J. Woolridge, 1979, Capital Market Seasonality: The Case of Bond Returns, *Journal of Financial and Quantitative Analysis* 14(5), 939-958.

Scott, L., 1990, Do Prices Reflect Market Fundamentals in Real Estate Markets? *Journal of Real Estate Finance and Economics* 3(1), 5-23.

Shiller, R., 2008, *Derivatives Markets for Home Prices*, Working Paper, NBER/Cowles Foundation, Yale University.

Shiller, R., 2009, *The Subprime Solution: How Today's Global Financial Crisis Happened and What to do About It,* Princeton University Press: New Jersey.

Sirmans, G., D. Macpherson and E. Zietz, 2005, The Composition of Hedonic Pricing Models, *Journal of Real Estate Literature* 13(1), 3-43.

Solnik, B., 1974, Why Not Diversify Internationally Rather Than Domestically?, *Financial Analysts Journal* 51(1), 48-54.

Steele, M., and R. Goy, 1992, Short Holds, The Distribution of First and Second Sales, and Bias in the Repeat-Sales Price Index, *Journal of Real Estate Finance and Economics* 14(1), 133-154.

Stevenson, S., and J. Young, 2004, Valuation Accuracy: A Comparison of Residential Guide Prices and Auction Results, *Property Management* 22(1), 45-54.

Susilawati, C., and V. Lin, 2006, *Case Analysis of Auction Market in Brisbane Housing System*, presented at the 12th Annual Pacific Rim Real Estate Conference (Auckland).

Triplett, J., 2004, *Handbook on Hedonic Indexes and Quality Adjustment in Price Indexes: Special Application to Information Technology Products*, Working Paper, Science, Technology and Industry Department, OECD.

Vickrey, W., 1961, Counterspeculation, Auctions, and Competitive Sealed Tenders, *Journal of Finance* 16(1), 8-37.

Wallace, N., 1996, Hedonic Based Price Indexes for Housing: Theory, Estimation and Index Construction, *Economic Review* 2(3), 34-48.

Wang, F., and P. Zorn, 1997, Estimating House Price Growth with Repeat Sales Data: What's the Aim of the Game?, *Journal of Housing Economics* 6(2), 93-118.

Webb, C., 1988. *A Probabilistic Model for Price Levels in Discontinuous Markets* Physica-Verla: Heidelberg.

Webb, J., R. Curcio and J. Rubens, 1988, Diversification Gains From Including Real Estate in Mixed-Asset Portfolios, *Decision Sciences* 19(2), 434-452.

White, H., 1980, A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica* 48(4), 817-838.

Wilson, P., and R. Zurbruegg, 2003, *International Diversification of Real Estate Assets - Is It Worth It? Evidence from the Literature*, Working Paper, University of Technology, Sydney.

Winship, C., and R. Mare, 1992, Models for Sample Selection Bias, *Annual Review of Sociology* 18(1), 327-350.

Worthington, A., and H. Higgs, 2009, Efficiency in the Australian Stock Market, 1875-2006: A Note on Extreme Long-Run Random Walk Behaviour, *Applied Economics Letters* 16(3), 301-306.

Wright, D., 2006, *Bias and Precision in the Measurement of Residential Real Estate Performance*, Working Paper, University of Sydney.

Appendices

A. Additional Tables to Chapter 4

Appendix A presents a set of additional empirical results to further support the analysis carried out in Chapter 4.

The results from the ARIMA Identification and Estimation stages for the Sydney unit market, and Melbourne house and unit market are first presented. Tables A-1, A-2 and A-3 present the ACF and PACF for the three Sydney unit indices, while Table A-4 summarises the estimated ARIMA models for this market. Similarly, Tables A-5 through A-8 report the ACF, PACF and estimated ARIMA model for the Melbourne house indices, and Tables A-9 through A-12 report the results for Melbourne units.

Table A-1:

Correlogram – Sydney Unit Median-Price Index

This table presents the ACF and PACF for returns to the median-price index for Sydney units. Correlations are calculated over first differences in the log median-price index levels, with significance at the 5% level denoted by *.

	Autocorrelation Function			Partial Autocorrelation Function		
Lag	Correlation	Correlogram		Correlation	Correlogram	
1	-0.3032 *	******		-0.3032 *	******	
2	-0.1888 *	****		-0.3092 *	******	
3	0.2520		*****	0.1014		**
4	-0.1199	**		-0.0588	*	
5	-0.0804	**		-0.0674	*	
6	0.2370 *		*****	0.1524 *		***
7	-0.0631	*		0.0755		**
8	-0.0807	**		0.0165		
9	0.1479		***	0.0767		**
10	-0.1005	**		-0.0347	*	
11	-0.1171	**		-0.1226	**	
12	0.3995 *		********	0.3043 *		******
13	-0.0174			0.2577 *		*****
14	-0.1769	****		0.0822		**
15	0.1168		**	-0.0244		
16	-0.0478	*		-0.0517	*	
17	0.0067			0.0944		**
18	0.0789		*	-0.0160		

Table A-2:

Correlogram – Sydney Unit Hedonic Price Index

This table presents the ACF and PACF for returns to the hedonic index for Sydney units. Correlations are calculated over first differences in the log hedonic index levels, with significance at the 5% level denoted by *.

	Autocorrelation Function		Partial Autocorrelation Function	
Lag	Correlation	Correlogram	Correlation	Correlogram
1	-0.1454 *	***	-0.1454 *	***
2	0.0940	**	0.0744	*
3	-0.0447	*	-0.0218	
4	0.0429	*	0.0278	
5	-0.1044	**	-0.0924	**
6	0.0121		-0.0205	
7	-0.0033		0.0131	
8	-0.0056		-0.0098	
9	-0.0907	**	-0.0918	**
10	0.1263	***	0.0995	**
11	-0.0535	*	-0.0136	
12	0.0890	**	0.0642	*
13	-0.1256	***	-0.1012	**
14	0.1184	**	0.0627	*
15	-0.1699	***	-0.1178	**
16	0.0482	*	-0.0032	
17	-0.0033		0.0341	*
18	0.0912	**	0.0737	*

Table A-3:

Correlogram – Sydney Unit Repeat-Sales Index

This table presents the ACF and PACF for returns to the repeat-sales index for Sydney units. Correlations are calculated over first differences in the log repeat-sales index levels, with significance at the 5% level denoted by *.

	Autocorrelation Function		Partial Autocorrelation Function	
Lag	Correlation	Correlogram	Correlation	Correlogram
1	-0.4921 *	**********	-0.4921 *	**********
2	-0.0026		-0.3230 *	******
3	0.0005		-0.2360*	*****
4	-0.0062		-0.1924 *	****
5	0.0039		-0.1561 *	***
6	-0.0003		-0.1304 *	***
7	0.0097		-0.0945	**
8	-0.0108		-0.0858	**
9	0.0035		-0.0719	*
10	0.0055		-0.0512	*
11	-0.0096		-0.0516	*
12	0.0143		-0.0263	*
13	-0.0103		-0.0250	
14	0.0099		-0.0063	
15	-0.0115		-0.0124	
16	0.0040		-0.0096	
17	0.0094		0.0085	
18	-0.0110		0.00272	

Table A-4:

ARIMA Models – Sydney Unit Price Indices

This table reports the estimated parameters and their *t* statistics for the following ARIMA models that best fit the returns from the median-price, hedonic and repeat-sales indices for Sydney units given by:

$$Y_t^{MED} - Y_{t-1}^{MED} = \mu + \frac{(1-\theta_1 B)(1-\theta_{12} B^{12})}{(1-\phi_1 B)(1-\phi_{12} B^2 - \phi_{12} B^{12})}\alpha_t$$

$$Y_t^{HED} - Y_{t-1}^{HED} = \mu + \frac{(1-\theta_1 B)}{(1-\phi_1 B)}\alpha_t$$

$$Y_t^{REP} - Y_{t-1}^{REP} = \mu + (1-\theta_1 B)\alpha_t$$

Statistical significance of the parameters at the 1% and 10% level of significance is represented by *** and *, respectively. The number of observations, AIC and SBC statistics are also reported.

	Median Price		Hedonic		Repeat-Sales	
Parameter	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic
μ	0.0040	0.98	0.0033 *	1.77	0.0042 *	2.27
Autoregressive Terms						
Φ_1	0.5288 ***	3.81	-0.9687 ***	-12.04		
Φ_2	0.4328 ***	4.88				
Φ_{12}	0.9999 ***	13.09				
Moving-Average Terms						
θ_1	0.9497 ***	7.60	-0.9174 ***	-20.40	0.8459 ***	17.94
θ_{12}	0.7382 ***	5.23				
Observations	131		131		131	
AIC	-642.550		-625.046		-157.145	
SBC	-625.299		-616.420		-151.394	

Table A-5:

Correlogram – Melbourne House Median-Price Index

This table presents the ACF and PACF for returns to the median-price index for Melbourne houses. Correlations are calculated over first differences in the log median-price index levels, with significance at the 5% level denoted by *.

	Autocorrelation Function			Partial Autocorrelation Function		
Lag	Correlation	Correlogram		Correlation	Correlogram	
1	-0.3401 *	*******		-0.3401 *	*******	
2	-0.0930	**		-0.2359 *	*****	
3	-0.0110			-0.1560 *	***	
4	-0.1328	***		-0.2686 *	*****	
5	-0.0638	*		-0.3327 *	*******	
6	0.4154 *		********	0.2331 *		*****
7	-0.0978	**		0.1679 *		***
8	-0.0991	**		0.0349		*
9	-0.0235	*		-0.0137		
10	-0.0506	*		0.0363		*
11	-0.2439*	*****		-0.3588 *	*******	
12	0.6259 *		************	0.3517 *		*******
13	-0.1757	****		0.1611 *		***
14	-0.1731	***		-0.0974	**	
15	0.0162			-0.0911	**	
16	-0.0648	*		0.0344		*
17	-0.0361	*		0.0971		**
18	0.2746		*****	-0.1165	**	

Table A-6:

Correlogram – Melbourne House Hedonic Index

This table presents the ACF and PACF for returns to hedonic index for Melbourne houses. Correlations are calculated over first differences in the log hedonic index levels, with significance at the 5% level denoted by *.

	Autocorrelation Function		Partial Autocorrelation Function	
Lag	Correlation	Correlogram	Correlation	Correlogram
1	-0.4080 *	********	-0.4080 *	********
2	-0.0292	*	-0.2347 *	*****
3	0.0015		-0.1383 *	***
4	-0.0500	*	-0.1482 *	***
5	0.2971 *	******	0.2685 *	*****
6	-0.1441	***	0.1494 *	***
7	0.0461	*	0.1722 *	***
8	-0.0340	*	0.0840	**
9	0.0096		0.0437	*
10	-0.0558	*	-0.2046 *	****
11	0.1108	**	-0.0295	*
12	-0.0468	*	-0.0852	**
13	0.0471	*	0.0620	*
14	-0.0420	*	0.0401	*
15	-0.0518	*	0.0455	*
16	0.1018	**	0.0413	*
17	-0.0591	*	0.0062	
18	0.0438	*	-0.0354	*

Table A-7:

Correlogram – Melbourne House Repeat-Sales Index

This table presents the ACF and PACF for returns to the repeat-sales index for Melbourne houses. Correlations are calculated over first differences in the log repeat-sales index levels, with significance at the 5% level denoted by *.

	Autocorrelation Function			Partial Autocorrelation Function		
Lag	Correlation	Correlogram		Correlation	Correlogram	
1	-0.0134			-0.0134		
2	0.2937 *		******	0.2936 *		******
3	0.1103		**	0.1281 *		***
4	0.0891		**	0.0105		
5	0.1467		***	0.0904		**
6	0.1355		***	0.1160		**
7	0.1013		**	0.0407		*
8	-0.0589	*		-0.1648 *	***	
9	0.0223			-0.0697	*	
10	-0.0683	*		-0.0533	*	
11	-0.0826	**		-0.1090	**	
12	0.1271		***	0.1502 *		***
13	-0.0194			0.0891		**
14	-0.1241	**		-0.1728 *	***	
15	0.0131			0.0024		
16	-0.0255			0.1073		**
17	0.0750		*	0.1164		**
18	0.0491		*	-0.0005		

Table A-8:

ARIMA Models – Melbourne House Price Indices

This table reports the estimated parameters and their t statistics for the following ARIMA models that best fit the returns from the median, hedonic and repeat-sales indices for Melbourne houses given by

$$Y_t^{MED} - Y_{t-1}^{MED} = \mu + \frac{(1-\theta_1 B)(1-\theta_6 B^6 - \theta_{12} B^{12})}{(1-\phi_{12} B^{12})}\alpha_t$$

$$Y_t^{HED} - Y_{t-1}^{HED} = \mu + (1-\theta_1 B - \theta_5 B^5)\alpha_t$$

$$Y_t^{REP} - Y_{t-1}^{REP} = \mu + (1-\theta_2 B^2)(1-\theta_{12} B^{12})\alpha_t$$

Statistical significance of the parameters at the 1% and 10% level of significance is represented by *** and *, respectively. The number of observations, AIC and SBC statistics are also reported.

	Median		Hedonic		Repeat-Sales	
Parameter	Coefficient Estimate	t Statistic	Coefficient Estimate	t Statistic	Coefficient Estimate	t Statistic
μ	0.0074	1.56	0.0083 ***	3.35	0.0080 ***	5.24
Autoregressive Terms						
Φ_{12}	0.9999 ***	18.52				
Moving-Average Terms						
θ_1	0.3895 ***	4.62	0.4976 ***	6.87		
θ_2					-0.3662 ***	-4.40
θ_5			-0.3035 ***	-4.18		
θ_6	-0.1390 *	-2.02				
θ_{12}	0.6374 ***	6.17			-0.2222 *	-2.36
Observations	131		131		131	
AIC	-551.917		-497.944		-812.164	
SBC	-537.541		-489.319		-803.538	

Table A-9:

Correlogram – Melbourne Unit Median-Price Index

This table presents the ACF and PACF for returns to the median-price index for Melbourne units. Correlations are calculated over first differences in the log median-price index levels, with significance at the 5% level denoted by *.

	Autocorrelation Function		Partial Autocorrelation Function	
Lag	Correlation	Correlogram	Correlation	Correlogram
1	-0.3520 *	******* |	-0.3520 *	******* |
2	0.0490	| *	-0.0856	** |
3	0.0064	|	-0.0058	|
4	-0.0595	* |	-0.0630	* |
5	0.0054	|	-0.0433	* |
6	0.0523	| *	0.0435	| *
7	-0.0002	|	0.0412	| *
8	-0.0624	* |	-0.0583	* |
9	0.1064	| **	0.0714	| *
10	0.0557	| *	0.1458 *	| ***
11	-0.1356	*** |	-0.0717	* |
12	0.1447	| ***	0.0720	* |
13	-0.0676	* |	0.0226	|
14	0.0281	| *	0.0353	| *
15	0.1011	| **	0.1180	| **
16	-0.0334	* |	0.0503	| *
17	-0.0157	|	0.0112	|
18	-0.0645	* |	-0.0909	|

Table A-10:

Correlogram – Melbourne Unit Hedonic Index

This table presents the ACF and PACF for returns to the hedonic index for Melbourne units. Correlations are calculated over first differences in the log hedonic index levels, with significance at the 5% level denoted by *.

	Autocorrelation Function		Partial Autocorrelation Function	
Lag	Correlation	Correlogram	Correlation	Correlogram
1	-0.2729 *	*****	-0.2729 *	*****
2	-0.0817	**	-0.1688 *	***
3	0.0079		-0.0714	*
4	-0.0709	*	-0.1157	**
5	0.0706	*	0.0087	
6	-0.0378	*	-0.0397	*
7	0.0907	**	0.0851	**
8	-0.1440	***	-0.1147	**
9	0.0320	*	-0.0197	
10	0.0762	**	0.0481	*
11	-0.1971 *	****	-0.1742 *	***
12	0.1781 *	****	0.0714	*
13	-0.0359	*	0.0062	
14	0.0755	**	0.1048	**
15	-0.0050		0.0507	*
16	-0.1226	**	-0.0836	**
17	0.1616	***	0.1153	**
18	-0.1877 *	****	-0.1199	**

Table A-11:

Correlogram – Melbourne Unit Repeat-Sales Index

This table presents the ACF and PACF for returns to the repeat-sales index for Melbourne units. Correlations are calculated over first differences in the log price repeat-sales index levels, with significance at the 5% level denoted by *.

	Autocorrelation Function			Partial Autocorrelation Function		
Lag	Correlation	Correlogram		Correlation	Correlogram	
1	-0.2120 *	****		-0.2120 *	****	
2	0.1170		**	0.0755		**
3	0.2174 *		****	0.2700 *		*****
4	-0.1954	****		-0.1206	**	
5	0.2044 *		****	0.0957		**
6	0.0109			0.0605		*
7	-0.0061			0.0427		*
8	0.1016		**	0.0116		
9	-0.1087	**		-0.0771		
10	0.0477		*	-0.0138		
11	0.0116			0.0120		
12	0.0368		*	0.0909		**
13	0.0550		*	0.0336		*
14	-0.0365	*		-0.0262	*	
15	0.00325			-0.0426	*	
16	-0.0070			-0.0082		
17	-0.0211			-0.0111		
18	0.1030		**	0.0880		**

Table A-12:

ARIMA Models – Melbourne Unit Price Indices

This table reports the estimated parameters and their *t* statistics for the following ARIMA models that best fit the returns from the median-price, hedonic and repeat-sales indices for Melbourne units as given by,

$$Y_t^{MED} - Y_{t-1}^{MED} = \mu + (1 - \theta_1 B)\alpha_t$$

$$Y_t^{HED} - Y_{t-1}^{HED} = \mu + (1 - \theta_1 B)\alpha_t$$

$$Y_t^{REP} - Y_{t-1}^{REP} = \mu + (1 - \theta_1 B^1 - \theta_3 B^3)\alpha_t$$

Statistical significance of the parameters at the 1% and 10% level of significance is represented by *** and *, respectively. The number of observations, AIC and SBC statistics are also reported.

Parameter	Median Price		Hedonic		Repeat-Sales	
	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic
μ	0.0073 ***	4.27	0.0071 ***	3.62	0.0069 ***	5.09
Moving-Average Terms						
θ_1	0.3720 ***	4.53	0.3831 ***	4.69	0.1606 *	1.85
θ_3					-0.1839 *	-2.12
Observations	131		131		131	
AIC	-536.301		-496.928		-723.180	
SBC	-530.551		-491.177		-714.554	

B. Additional Tables to Chapter 5

Appendix B presents a series of additional Tables to support the empirical work undertaken in Chapter 5 of this dissertation.

Descriptive statistics and regressions results from the sample of matched sales, matched using a without replacement procedure, are reported in Tables B-1 and B2, respectively.

Two alternative matching procedures to the one reported in Chapter 5 are also considered. Table B-3 presents the results of the re-estimated hedonic-regression model following matching by land size and Table B-4 presents the results following matching by month of sale, location, and number of bedrooms.

Finally, a small sample of house sales is obtained for the Christchurch area of New Zealand – the same housing market considered by Dotzour et al. (1998) – for the period January 2005 to December 2008. The sale details and attributes for 3,754 house sales are sourced from RPX. Auctions represent approximately 3% of the sales in this sample.

Table B-5 presents the results of the hedonic-regression models, given by Equations 5.1 and 5.3, fitted to the sample of sales from Christchurch (subject to attribute data availability). Table B-6 presents the results of the re-estimated hedonic-regression models, given by Equations 5.1 and 5.3, to the matched sample of sales from Christchurch, matched following the procedure outlined in Section 5.3.3.

Table B-1:

Matched Without Replacement Descriptive Statistics

This table presents the descriptive statistics of the without replacement matched sample. Panel A reports the number of matched observations as well as several price statistics. These include the average, median, and standard deviation of price in the matched sample, as well as between the post-matched auction and private treaty sales sub-samples. Panel B provides information on the average composition of hedonic attributes in the matched-sample and the post-matching auction and private treaty sales sub-samples.

Panel A: Matched Sample Observations and Prices

		Price		
Sample	Observations	Average	Median	Standard Deviation
All Sales	7,902	752,780	645,000	443,101
Private Treaty	3,951	713,688	615,000	412,799
Sold at Auction	3,951	791,871	680,000	468,259

Panel B: Matched Sample Hedonic Attributes

Variable	All Sales	Private Treaty	Auctions
Land Size (m^2)	552.46	546.60	558.31
Bedrooms	3.22	3.22	3.22
Bathrooms	1.65	1.66	1.64
Other Rooms	1.01	1.01	1.02
Car Spaces	0.90	0.90	0.90
Pool (%)	10.96	10.98	10.93
Air-Conditioning (%)	9.49	8.93	10.05
Scenic View (%)	7.05	5.69	8.40
Waterfront (%)	0.85	0.51	1.19

Table B-2:

Replication Regression Results

This table reports the results of the re-estimated regression models given by Equations 5.1 and 5.3 using the matched sample obtained without replacement. Coefficient estimates of the fitted model and several regression diagnostic statistics are presented. Coefficient significance at the sample-size adjusted critical *t* value is denoted by *. Due to the number of dummy variables used to account for the effect of suburb and month of sale in the regression, these individual estimated coefficients are not reported. Instead, a summarised description of the statistical performance of the suburb and month of sale explanatory variables is provided. Full results are available from the author upon request.

	Restricted OLS Model		Hedonic OLS Model	
Variable	Coefficient	*t* Statistic	Coefficient	*t* Statistic
Intercept	13.5339 *	25.22	12.8658 *	28.82
Auction	0.0098	1.91	0.0024	0.56
Land Size	0.3764 *	41.57	0.2922 *	37.71
Interest Rate	0.0858	1.08	0.0710	1.08
Bedrooms			0.0707 *	13.4
Bathrooms			0.0919 *	11.43
Bed/Bath			0.0142	2.12
Other Rooms			0.0337 *	4.80
Car Spaces			0.0284 *	10.30
Pool			0.0476 *	6.87
Air-Conditioning			0.0459 *	6.40
Scenic View			0.0416 *	4.87
Waterfront			0.2314 *	9.80
	Significance at 1% (%)	Significance at 10% (%)	Significance at 1% (%)	Significance at 10% (%)
Suburbs ($J = 534$)	57.87	73.03	60.67	75.09
Months ($T = 47$)	2.13	6.38	6.38	8.51
Observations	7,902		7,902	
Critical *t*	2.884		2.882	
Adjusted R^2	0.8164		0.8733	
F Statistic	61.16 *		92.83 *	

Table B-3:

Matched (Land Size) Sample Regression

This table reports the results of the re-estimated regression models given by Equations 5.1 and 5.3 using the sample obtained by an alternative matching procedure. Specifically, properties are matched solely on nearest land size. Estimated coefficients and several regression diagnostic statistics are presented. Due to the number of dummy variables used to account for the effect of suburb and month of sale in the regression, these individual estimated coefficients are not reported. Instead, a summarised description of the statistical performance of the suburb and month of sale explanatory variables is reported. Full results are available from the author upon request. Statistical significance measured by the sample size adjusted critical *t* value is indicated by *.

	Restricted OLS Model		Hedonic OLS Model	
Variable	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic
Intercept	12.6104 *	39.85	11.798 *	44.54
Auction	0.0025	0.46	0.012	2.59
Land Size	0.3869 *	79.07	0.294 *	68.21
Interest Rate	0.2504 *	5.22	0.257 *	6.42
Bedrooms			0.079 *	23.2
Bathrooms			0.085 *	16.51
Bed/Bath			0.001	0.26
Other Rooms			0.018 *	4.07
Car Spaces			0.026 *	14.7
Pool			0.023 *	4.85
Air-Conditioning			0.051 *	10.54
Scenic View			0.007	1.43
Waterfront			0.216 *	14.13
	Significance at 1% (%)	Significance at 10% (%)	Significance at 1% (%)	Significance at 10% (%)
Suburbs (J = 532)	80.26	86.64	83.08	88.91
Months (T = 50)	62.00	86.00	84.00	88.00
Observations	19,358		19,358	
Critical *t*	3.094		3.094	
Adjusted R^2	0.8299		0.8815	
F Statistic	162.39		243.48	

Table B-4:

Matched (Time, Location, Bedrooms) Sample Regression

This table reports the results of the re-estimated regression models given by Equations 5.1 and 5.3 using the sample obtained by an alternative matching procedure. Specifically, the matching procedure used to obtain this sample uses nearest bedrooms as an alternative to land size as a secondary filter. Coefficient estimates of the estimated model and several regression diagnostic statistics are presented. Due to the number of dummy variables used to account for the effect of suburb and month of sale in the regression, these individual estimated coefficients are not reported. Instead, a summarised description of the statistical performance of the suburb and month of sale explanatory variables is reported. Full results are available from the author upon request.

	Restricted OLS Model		Hedonic OLS Model	
Variable	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic
Intercept	13.5022 *	39.98	12.966 *	46.09
Auction	0.0058	1.62	0.001	0.47
Land Size	0.3823 *	67.21	0.296 *	60.17
Interest Rate	0.0937	1.90	0.056	1.37
Bedrooms			0.081 *	19.95
Bathrooms			0.088 *	14.44
Bed/Bath			0.005	1.02
Other Rooms			0.038 *	7.67
Car Spaces			0.024 *	12.28
Pool			0.046 *	8.66
Air-Conditioning			0.047 *	8.9
Scenic View			0.044 *	7.22
Waterfront			0.193 *	11.65
	Significance at 1% (%)	Significance at 10% (%)	Significance at 1% (%)	Significance at 10% (%)
Suburbs ($J = 544$)	66.91	77.21	67.65	79.41
Months ($T = 47$)	6.38	27.66	14.89	29.79
Observations	15,866		15,866	
Critical *t*	3.052		3.051	
Adjusted R^2	0.8032		0.8640	
F Statistic	109.99		168.18	

Table B-5:

Regression Results

This table reports the results of fitting the models given by Equations 5.1 and 5.3 to a sample of Christchurch, New Zealand, house sales. The estimated coefficients and several diagnostic statistics are presented. Coefficient significance against the sample size adjusted critical t value is denoted by * Due to the number of suburb and monthly time dummy variables the estimated coefficients for each of these variables are not reported. Instead, a summarised description of the statistical performance of these variables is provided. Full results are available from the author upon request.

	Restricted OLS Model		Hedonic OLS Model	
Variable	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic
Intercept	13.1845 *	127.95	12.5559 *	122.17
Auction	0.0769 *	4.53	0.0686 *	4.46
Land Size	0.1297 *	10.04	0.0914 *	7.52
Bedrooms			0.1010 *	9.65
Bathrooms			0.1237 *	6.94
Bed/Bath			-0.0243	-1.40
Car Spaces			0.0069	1.94
Pool			0.2025	1.19
	Significance at 1%	Significance at 10%	Significance at 1%	Significance at 10%
Suburbs (J = 80)	27.50	43.75	25.00	41.25
Months (T = 47)	27.66	48.94	42.55	53.19
Observations	3,754		3,754	
Critical t value	2.820		2.820	
Adjusted R^2	0.4387		0.5409	
F statistic	23.74		34.00	

Table B-6:

Matched Sample Regression

This table reports the results of the re-estimated regression models, given by Equations 5.1 and 5.3, using the matched sample of Christchurch house sales. The estimated coefficients and several regression diagnostic statistics are presented. Coefficient significance as measured against the sample size adjusted critical *t* value is denoted by *. Due to the number of dummy variables used to account for the effect of suburb and month of sale in the regression, these individual estimated coefficients are not reported. Instead, a summarised description of the statistical performance of the suburb and month of sale explanatory variables is reported. Full results are available from the author upon request.

	Restricted OLS Model		Hedonic OLS Model	
Variable	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic
Intercept	13.7256 *	52.71	13.3184 *	37.82
Auction	0.0654	1.81	0.0603	1.79
Land Size	0.2533 *	3.14	0.2715 *	3.65
Bedrooms			0.1008	1.62
Bathrooms			0.1208	1.19
Bed/Bath			-0.0625	-0.58
Car Spaces			-0.0105	-0.44
	Significance at 1%	Significance at 10%	Significance at 1%	Significance at 10%
Suburbs ($J = 51$)	1.96	25.49	5.88	27.45
Months ($T = 45$)	0.00	2.22	0.00	6.67
Observations	254		254	
Critical *t*	1.848		1.824	
Adjusted R^2	0.3172		0.4382	
F Statistic	2.20		2.94	

C. Additional Tables to Chapter 6

Appendix C contains a set of additional tables to support the empirical work presented in Table 6. These specifically represent a replication of the new property pricing analysis using a sample of sales for the Adelaide residential property market.

Table C-1 reports the descriptive statistics for this sample of sales. Table C-2 reports the results of fitting the regression model given by Equation 6.1 to the Adelaide sample. A probit model is fitted, as given by Equation 6.2, the results of which are reported in Table C-3. Finally, the sample-selectivity corrected regression model, following the two-stage Heckman procedure, is estimated. The results of this model, given by Equation 6.4, are reported in Table C-4.

Table C-1:

Adelaide Summary Statistics

This table provides the sample statistics of the sample data. The information covers the sample periods and size, distribution measures of price, the average land size, bedrooms, bathrooms, and car spaces and the proportion of observations with given characteristics for binary variables: waterfrontage, scenic view, swimming pool, and air-conditioning.

	Houses		Units	
	All	New	All	New
Observations	64,784	753	17,583	224
Proportion (%)		1.16		1.27
Median Price ($)	285,000	312,500	227,000	320,000
Average Price ($)	315,412	330,954	245,450	325,703
Standard Deviation ($)	132,965	105,755	99,428	97,403
Land Size (m^2)	718.67	409.14		
Bedrooms	3.21	3.13	2.28	2.75
Baths	1.51	1.60	1.31	1.64
Car Spaces	0.44	0.31	0.32	0.44
Pool (%)	5.56	0.13	0.47	0.45
Air-Conditioning (%)	37.76	45.02	33.71	37.05
Scenic View (%)	3.42	4.12	1.80	2.68
Waterfront (%)	0.13	0.53	0.57	0.89

Table C-2:

New Property Pricing

This table reports the results of OLS fitting of the regression model given by Equation 6.1 on the sample of sales from Adelaide. Coefficient significance at the sample size adjusted 5% level of significance is denoted by *. Due to the number of dummy variables used to account for the effect of suburb and month of sale in the regression coefficients have not been presented here. Instead, a summarised description of the statistical performance of these variables is reported. Full results are available from the author upon request.

	Houses		Units	
Variable	Coefficient Estimate	t Statistic	Coefficient Estimate	T Statistic
Intercept	12.4699 *	1176.61	11.4901 *	297.91
First Sale	0.1143 *	18.55	0.1923 *	13.15
Land Size	0.1301 *	58.55	-	-
Bedrooms	0.1145 *	57.24	0.2643 *	30.49
Bathrooms	0.0321 *	10.63	0.0440 *	4.26
Bed/Bath	-0.0124 *	-5.01	-0.0142	-1.41
Car Spaces	0.0119 *	11.66	0.0328 *	10.35
Pool	0.0532 *	18.42	0.0224	0.92
Air-Conditioning	0.0071 *	5.36	0.0267 *	7.74
Scenic View	0.0460 *	12.66	0.0943 *	7.57
Waterfront	0.1069 *	6.06	0.0819 *	3.79
	Significance at 1% (%)	Significance at 10% (%)	Significance at 1% (%)	Significance at 10% (%)
Suburbs (J_{HOUSE} = 374, J_{UNIT} = 328)	87.70	91.98	55.49	6.77
Months (T_{HOUSE} = 99, T_{UNIT} = 93)	93.94	100.00	78.13	84.38
Observations	64,784		17,583	
Critical t value	3.316		3.088	
Adjusted R^2	0.8139		0.7211	
F-Statistic	680.43 *		124.21 *	

Table C-3:

Probit Results

This table reports the results of MLE fitting of the probit model given by Equation 6.2 to the sample of sales from Adelaide. The coefficient estimates and related diagnostic statistics are presented. Coefficient significance at the 1% and 10% levels of significance is donated by *** and *, respectively Due to the number of dummy variables used to account for the effect of suburb and month of sale in this model, these estimated coefficients are not presented. Instead, a summarised description of the statistical performance of these variables is provided. Full results are available from the author upon request.

	Houses		Units	
Variable	Coefficient Estimate	*t* Statistic	Coefficient Estimate	*t* Statistic
Intercept	-10.7634 ***	-21.93	-9.8007 ***	-15.54
Land Size	-2.1982 ***	-28.42	-	-
Bedrooms	0.3397 ***	5.30	0.9776 ***	6.81
Bathrooms	0.0681	0.74	-0.1759	-0.95
Bed/Bath	-0.0176	-0.25	-0.3387 *	-2.04
Car Spaces	-0.0964 ***	-2.58	0.1521	0.31
Pool	-1.1866 ***	-3.14	-0.6486	-1.03
Air-Conditioning	0.0172	0.42	-0.0440	0.029
Scenic View	0.1664	1.45	-0.5094 *	-1.99
Waterfront	0.7996 *	2.34	0.5517	0.90
	Significance at 1% (%)	Significance at 10% (%)	Significance at 1% (%)	Significance at 10% (%)
Suburbs (J_{HOUSE} = 367, J_{UNIT} = 82)	10.63	22.61	100.00	100.00
Months (T_{HOUSE} = 95, T_{UNIT} = 67)	17.35	41.84	13.43	44.78
Observations	64,784		17,583	
Log-Likelihood	-2350.16		-696.04	
Maximum Absolute Gradient	3.738		0.012	

Table C-4:

Selectivity-Corrected Regression

This table reports the results of OLS fitting of the second-stage Heckman regression model and their significance at the sample size adjusted critical t value, denoted by *, are reported with several regression diagnostic statistics. Due to the number of dummy variables used to account for the effect of suburb and month of sale in the regression these coefficients have not been presented here. Instead, a summarised description of the statistical performance of these variables is provided. Full results are available from the author upon request.

	Houses		Units	
Variable	Coefficient Estimate	t Statistic	Coefficient Estimate	t Statistic
Intercept	9.8541 *	51.91	11.8717 *	48.08
New	0.0897 *	14.02	0.1980 *	13.19
Sample Selectivity	0.2474 *	13.83	-0.0358	-1.45
Land Size	-0.3813 *	-10.29	-	-
Bedrooms	0.1944 *	31.85	0.2305 *	9.46
Bathrooms	0.0482 *	14.97	0.0504 *	4.59
Bed/Bath	-0.0156 *	-6.71	-0.0019	-0.16
Car Spaces	-0.0102 *	-5.34	0.0285 *	5.87
Pool	-0.2293 *	-11.13	0.0420	1.50
Air-Conditioning	0.0113 *	8.17	0.0289 *	7.85
Scenic View	0.0846 *	18.44	0.1129 *	6.53
Waterfront	0.2902 *	13.22	0.0683	2.64
	Significance at 1% (%)	Significance at 10% (%)	Significance at 1% (%)	Significance at 10% (%)
Suburbs (J_{HOUSE} = 374, J_{UNIT} = 328)	10.63	22.61	100.00	100.00
Months (T_{HOUSE} = 99, T_{UNIT} = 93)	17.35	41.84	13.43	44.78
Observations	64,784		17,582	
Critical t	3.316		3.088	
Adjusted R^2	0.8163		0.7214	
F Statistic	590.48 *		106.62 *	