Evidence on the determinants and variation of idiosyncratic risk in housing markets*

Lydia Cheung[1], Jaqueson K. Galimberti[2,4,5], and Philip Vermeulen[3,4]

[1]Auckland University of Technology
[2]Asian Development Bank
[3]University of Canterbury
[4]Centre for Applied Macroeconomic Analysis, Australian National University
[5]KOF Swiss Economic Institute, ETH Zurich

December 19, 2024

Abstract

Using around one million repeat sales, we show that idiosyncratic risk in real house price appreciation varies considerably and systematically across houses. First, we find that idiosyncratic risk is time-varying, depends negatively on the initial house price, varies across locations, and reduces as the holding period of the house increases. Second, these systematic movements in idiosyncratic risk can be explained by time and space variations in market thinness and differences in information quality across markets. We find that borrowing costs and deposit requirements have offsetting effects on risk. Higher interest rates are associated with lower idiosyncratic pricing, while tighter deposit requirements are associated with shorter holding periods, which are subject to a higher risk. Third, we find that the systematic variations in idiosyncratic housing risk are mostly priced in expected returns, except across locations, while the higher risk in shorter holding periods is rewarded with only slightly higher returns. The risk-return relationship across initial house prices depends on the current state of the market. During busts of the housing cycle, the distribution of house prices widens and cheaper houses depreciate faster than more expensive houses, leading to an inverted risk-return relationship.

Keywords: idiosyncratic risk, house prices, housing markets
JEL classifications: G1, R1

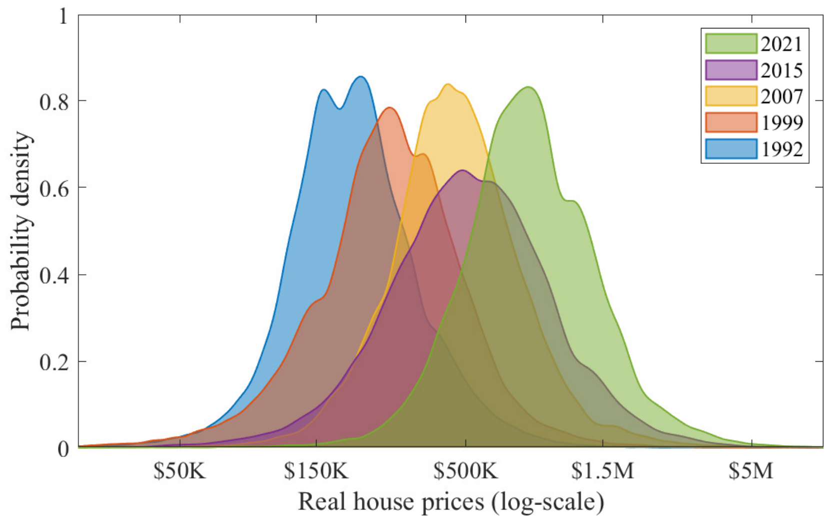
---

*Earlier versions of this paper were presented at seminars at the Reserve Bank of New Zealand, Auckland University of Technology, the Asian Development Bank, and at the Asia Meeting of the Econometric Society, East & Southeast Asia. This project received funds from the Faculty of Business, Economics and Law, at Auckland University of Technology, for which we gratefully acknowledge the support. The views expressed in this paper are those of the authors, and do not necessarily represent the views of their corresponding institutional affiliations. Corresponding author: Jaqueson K. Galimberti. Contact details: Asian Development Bank, 6 Adb Avenue, 1550 Mandaluyong, Metro Manila, Philippines. Emails: jgalimberti@adb.org, lydia.p.cheung@aut.ac.nz, philip.vermeulen@canterbury.ac.nz.

1 Introduction

House price appreciation is not homogeneous. In fact, long-established evidence conveys very heterogeneous appreciation rates across houses. For example, Case and Shiller (1989) report that annual house price appreciation has a standard deviation of close to 15 percent for individual houses and go on to argue that home-owners should not expect to realize capital gains measured by house price indices. Figure 1 depicts how the distribution of house prices has evolved since 1992 in New Zealand. It is clear that house prices are widely dispersed at any point in time and their distribution dynamics are nontrivial. Homeowners typically own only one house, the family home, and therefore carry a considerable idiosyncratic risk on housing. Although the family home has by far the largest weight in many households' investment portfolios (OECD, 2021), surprisingly little is known about how idiosyncratic risk in real house price appreciation varies across individual houses, whether the risk is priced and what are its determinants. This is unfortunate because a better understanding of the nature of idiosyncratic risk of housing investment is definitely important as it cannot be easily diversified away, contrary to, say, a stock portfolio that diversifies away the idiosyncratic risk of individual stocks.

Figure 1: Evolution of the distribution of house prices in New Zealand.



Notes: Real house prices are based on the New Zealand Consumer Price Index, taking 2021 as the base year.

Our paper provides answers to three important questions on the nature of idiosyncratic risk in housing markets. First, is there any systematic variation in idiosyncratic risk? If certain houses carry higher idiosyncratic risk, households clearly would want to know, since households *choose* which house to buy.[1] Households decide *when* to buy, *where* to buy, *what* house to buy, and *how long* to hold the house before resale. Does idiosyncratic risk vary along those choices? We find that idiosyncratic risk varies systematically across the four dimensions of time of purchase, location, the initial price of the house, and the holding period. Second, why does idiosyncratic risk vary along these dimensions? We find that the systematic variation in idiosyncratic risk identified in our model can be associated with credit conditions and valuation uncertainty, which are important factors determining search and matching frictions in the house trading process. Third, is systematically higher idiosyncratic risk rewarded with higher returns? The risk-return trade-off is a perennial question in finance. We show that existing answers in the literature based on house price indexes or raw repeat sales data can be seriously misleading. Rather, we answer this question using individual house price data and estimates of idiosyncratic risk that take full advantage of the information provided by overlapping repeat sales data. Importantly, we find that some but not all systematic variation in idiosyncratic risk is rewarded with higher returns.

To obtain answers to our three questions, we use a novel large dataset of around one million repeat sales observations in the entire New Zealand housing market over a long period (1992-2021). To answer our first question, we estimate how idiosyncratic risk varies across four dimensions in an extension of the repeat sales model of house price changes of Landvoigt et al. (2015). Landvoigt et al. (2015) considers a log-linear regression model of annual price changes of individual houses. In the model, annual house price appreciation is a function of annual time dummies (measuring the annual average increase in house prices), the current price of the house, and an idiosyncratic shock $\varepsilon_{i,t}$. Idiosyncratic risk is measured by the standard deviation of the shock. It measures how much individual house annual price appreciation deviates from the average annual appreciation, also accounting for a time-dependent house price effect intended to capture segmentation in housing markets. Landvoigt et al. (2015) allow this standard deviation, i.e. idiosyncratic risk, to vary across time. They estimate the model using repeat sales data from San Diego County and thus obtain a yearly estimate of idiosyncratic risk.

[1] Others, such as mortgage lenders and investment funds, should also be interested: in case of default, the recovery amount would depend on the idiosyncratic risk, while portfolio optimization requires a proper quantification of risks.

They show that idiosyncratic risk in the San Diego housing market is indeed time-varying and quite large. It varies between 8 percent and 13.8 percent depending on the year.

We extend the model in several directions. First, in contrast to Landvoigt et al. (2015), we have information on house features (such as number of bedrooms, bathrooms, garages, and floor area). This allows us to control for house remodeling changes in the regression. This is important as variation in house price appreciation due to remodeling of the house could affect the estimated magnitude of idiosyncratic risk. This argument has been made recently by Giacoletti (2021) who controls for major remodeling investments when estimating idiosyncratic risk in the California housing market. We find that indeed remodeling increases the average rate of appreciation of house prices but that controlling for remodeling has only limited effects on our estimates of idiosyncratic risk. Second, we allow the average annual house price appreciation to vary across 16 different regional markets in New Zealand, as well as pinning down the house price segmentation effect to the median house in each regional market. Third, we allow idiosyncratic risk to vary in the four dimensions of time, location, initial house price, and holding period. The last three dimensions are an innovation to the previous model.

We find considerable variation in idiosyncratic risk across houses in New Zealand in all four dimensions. First, we find that idiosyncratic risk varies across time. For example, a median-priced house in Auckland (held for a period of 5 years) has an idiosyncratic risk that varies between 7.90 percent (in 2018) to 10.55 percent (in 2002). Second, we show large regional differences in idiosyncratic risk across New Zealand. Moving away from the Auckland region (which includes the largest city) to Nelson (situated in the North of the South Island) households are able to reduce idiosyncratic risk by as much as 26 percent. Equally, moving from Auckland to the West Coast of the South Island increases idiosyncratic risk by 25 percent. These results show that location is an important factor driving differences in the level of idiosyncratic risk in housing. Third, we find that households can reduce idiosyncratic risk by buying a house higher up in the price distribution. Landvoigt et al. (2015) argue that the price of the house is a reasonable summary indicator of quality. Our results therefore indicate that cheaper houses or houses of lower quality are more risky. A house at the 10th (90th) percentile of the house price distribution has an 8 percent higher (lower) idiosyncratic risk than a house at the median. These estimates corroborate the findings of Guerrieri et al. (2013) who documented a higher variance in appreciation rates for initially low-priced neighborhoods during city-wide housing booms in U.S. cities.[2] Fourth, we find that holding the house for a longer period reduces idiosyncratic risk. The housing market in New Zealand is very active and houses are typically held for only a short period. The median holding period is 5 years. Each additional year the house is kept reduces the idiosyncratic risk by around 7.8 percent. This is consistent with recent findings in Giacoletti (2021) for the Californian housing market.[3]

We believe our paper to be the first that considers simultaneously the four important ways in which idiosyncratic risk varies systematically. This enables us to determine the relative importance of these four factors (time, location, price, and holding period). Particularly, comparing a house at the 10th percentile of idiosyncratic risk with a house at the 90th percentile,[4] we find that: 58 percent of the systematic variation in idiosyncratic risk is due to the holding period factor, 21 percent is due to the regional factor, 11 percent is due to the time variation and 10 percent is due to the initial price variation. The holding period is the most important factor driving the systematic variation in idiosyncratic risk, followed by location. The answer to our first question broadens our understanding of the nature and importance of idiosyncratic risk and the potential of households to mitigate such risk.

With respect to our second question, we provide an empirical analysis of why idiosyncratic risk varies systematically along the dimensions we investigate. We approach that question by separately regressing the different dimensions of variation in idiosyncratic risk, while keeping the other dimensions constant, on explanatory variables that cover different channels. Particularly, and guided by the previous literature, we consider how credit conditions and informational asymmetries can be associated with the systematic variation in risk elicited by our estimates. For credit conditions, we find that borrowing costs and deposit requirements have offsetting effects on risk. Although higher interest rates are associated with lower idiosyncratic risk, both across time and through higher holding periods, higher deposit requirements

[2] Guerrieri et al. (2013) finds that the standard deviation of housing price growth between 2000 and 2006 is 61% (or 10.17% in annual terms) for neighborhoods in the lowest initial house price quartile and 46% (or 7.67% in annual terms) for neighborhoods in the top initial house price quartile.

[3] Giacoletti (2021) finds that idiosyncratic risk declines non-linearly with the holding period, following a quadratic function: from 12% for a 2-year hold, to 8.6% for a 5-year hold, to 7% for a 15-year hold.

[4] In absolute terms this corresponds to an increase in idiosyncratic risk from 2.8 percent to 13.8 percent.

are associated with shorter holding periods, which increases risk. These effects can be interpreted through the lens of a "market thinness" mechanism. Namely, the sizes of the pools of buyers and sellers in the market impact idiosyncratic pricing through their effect on matching frictions. While tighter access to credit increases risk by decreasing the pool of potential buyers in the market, higher borrowing costs decrease risk by constraining mortgage rollover and leading to forced sales, which increases the pool of houses available in the market. Moreover, the impact of borrowing costs is also consistent with a risk-taking behavioral channel, where higher interest rates may restrain high-risk speculative investments that would otherwise increase volatility in housing prices. We believe we are the first to document such a relationship between borrowing costs and idiosyncratic risk, the identification of which benefited from the risk decomposition provided by our model.[5]

We also find that informational asymmetries are relevant in determining idiosyncratic risk through a "valuation uncertainty" channel in the house search process. Poorer information and difficulties in valuing houses lead to more heterogeneous house valuations and increase the potential variation of transaction prices. To assess this channel we consider two variables related to benchmarking difficulties and information access. First, for the benchmarking issues, we construct regional measures of house atypicality, which compare how the characteristics of individual houses deviate from typical characteristics of houses in their corresponding suburbs. We find that more atypical houses tend to be located in regions with higher risk and also tend to be held for a shorter period. Hence, part of the regional and holding period differences in risk can be attributed to the varying degrees of relative heterogeneity of housing characteristics and the underlying valuation uncertainty. Second, we consider how access to information impacts risk by looking at (sub-)regional differences in internet access. We find that regions with lower internet access have higher risk estimates. In addition, we also find that cheaper houses tend to be located in suburbs with lower internet access. Considering the relative house prices as a proxy for house quality, this evidence suggests that lower-quality houses are subject to higher

[5]There is substantial research on the relationship between interest rates and *average* house prices. While the workhorse present value user cost model predicts a negative relationship between interest rates and house prices, empirical evidence suggests a limited role for borrowing costs in explaining fluctuations in average house price (see, e.g., Glaeser et al., 2012). For New Zealand, Shi et al. (2014) find that real interest rates are positively related to real housing prices.

valuation uncertainty due to limited access to information about the house. All in all, our empirical analysis indicates that credit conditions and informational asymmetries are key drivers of systematic variation in idiosyncratic risk.

Our last question asks whether idiosyncratic risk is priced, i.e., do houses with higher idiosyncratic risk have higher returns? A large part of the literature on housing return and risk ignores the idiosyncratic risk of an individual homeowner. Indices of house prices, not individual house prices, are typically used to calculate historical risk and return in housing markets. For instance, Jorda et al. (2019) ignores idiosyncratic risk when arguing that over long periods of time, residential real estate has been the best long-run investment with returns around the same level of equity, but with much lower volatility. However households buy a house not a house price index. It is well known that house price capital gains vary across locations; however, idiosyncratic risk is often omitted from studies that investigate geographical variation in risk and return.[6] In contrast to this literature, our paper uses individual house price data to investigate the risk-return relationship. We find that idiosyncratic risk is mostly priced with one important exception, namely, across regions. That is, households moving to a region with higher idiosyncratic risk are not rewarded with higher returns. Moreover, although the higher risk associated with a shorter holding period is rewarded with a higher return, the sensitivity of returns to risk across holding periods is much smaller than what raw repeat sales data imply. Finally, we show that the risk-return trade-off also depends on the relative price of the house, and this relationship varies with the housing cycle. While during normal times the higher risk of idiosyncratic price changes in cheaper houses is priced in higher average returns, during housing market busts the risk-return trade-off by relative house price inverts as cheaper houses

[6] Sinai (2009) documents widely varying volatility of housing markets across metropolitan areas in the US, using house price indices at the metropolitan level but staying silent on the idiosyncratic risk of an individual home-owner within a metropolitan area. Peng and Thibodeau (2017) analyze idiosyncratic risk only at the zip-code level in the U.S. and find evidence that risk follows a U-shaped relationship with the neighborhood median household income, being higher for low-income and to a lesser degree for high-income markets. Han (2013) uses metropolitan statistical level repeat sales house price indices to investigate the relationship between expected return and risk in housing markets in the US. Cannon et al. (2006) argue that broad metropolitan area indices may be misleading for investors as an indicator of capital appreciation or risk and investigate house price risk and return at the ZIP code level but falls short of using individual house price data. Guerrieri et al. (2013) document empirical facts on the variation in house prices within U.S. cities, including a relationship between initial house prices and the variance of appreciation rates during city-wide housing booms, but only at the ZIP code and census tract levels.

end up depreciating faster.[7]

The rest of the paper is structured as follows. In section 2 we describe the data. In section 3 we present our model followed by estimation results in section 4. In section 5 we discuss the determinants for the variation in idiosyncratic risk. In section 6 we investigate whether higher idiosyncratic risk is rewarded with higher returns. In section 7 we conclude.

2 Data

2.1 Data source

We use individual house sale transactions data from the Real Estate Institute of New Zealand (REINZ). REINZ is a membership organisation representing more than 14,000 real estate professionals in New Zealand. According to REINZ, more than 90 percent of New Zealand's real estate agents are a member. It collects transactions data from its real estate agent members and it is one of the leading sources of real estate transaction data in New Zealand. Our original dataset contains over 2.4 million home sales transactions and covers the years from 1992 to 2021. Geographically, our dataset covers the entire New Zealand. Individual house sales transactions variables include the basic characteristics of the property, such as the address, number of bedrooms, bathrooms, garages, floor area as well as the sale price and sale date. Each individual property has a unique ID, which we use to identify repeat sales.

New Zealand has a surface area of 268,021 km^2 (about the size of the state of Colorado in the US or a bit larger than the United Kingdom) with a small population of around 5.1 million (estimate 2022) and most people live close to regional city centres. The New Zealand housing market can be best described as a set of geographically separated markets each centred around a major city. Major cities in New Zealand are quite small in population in international comparison. Geographically, the New Zealand housing market can be divided up into 16 regional markets. The largest housing market in terms of population is the Auckland region (with Auck-

[7]Two important caveats to our results are in order. First, the estimates of risk and return examined in this paper are based on *ex post* realized capital gains. This is important because investment decisions are based on expected returns. Second, we do not account for rents, which are important for both non-owner occupier investment returns and the shadow cost of housing for owner occupiers. Nevertheless, the historical analysis of realized capital gains offers lessons about the sources of risk that need to be priced in housing investment decisions.

land city as the main centre) with around 1.4 million inhabitants for the entire region (according to 2013 census). The second largest region is Canterbury with less than 600,000 inhabitants (and Christchurch as the main centre). The third largest region is Wellington, centred around New Zealand's capital (with the same name as the region) with less than 500,000 inhabitants. Six regions in New Zealand have each less than 100,000 inhabitants. For comparison, when estimating idiosyncratic risk, Landvoigt et al. (2015) consider San Diego county as one market. San Diego county has around 3.3 million inhabitants which is twice the population size of the Auckland region in New Zealand.

2.2 Sample

The raw dataset contains 2,445,989 individual house sales observations. We exclude some clear outlying observations. This guards against potential input error in the information about house features. We remove a house sale observation if the house has either nine or more bedrooms, six or more bathrooms, seven or more garages, or a floor area larger than 885 m^2.[8] A total of 8,266 sales observations are removed by this filter.

From this dataset we compile the repeat sales, i.e., observations of the same house sold at least twice over different years. Repeated sales that occur within the same year are averaged into a single observation. This leaves us with a total of 1,065,782 repeated sales transactions on 585,242 houses. We then calculate the house price real appreciation between repeat sales. All house prices used in our calculations are in real terms. We convert nominal prices to real prices using the New Zealand consumer price index, taking 2021 as the base year. Finally, we exclude observations with absolute annualized capital gain greater than 50 percent. This leaves us with a final sample of 1,058,391 repeat sales observations on 583,561 houses. Table 1 presents the summary statistics for this sample.[9] The average annualized capital gain is 5.52 percent. The standard deviation is 8.20 percent, showing indeed a large heterogeneity in the capital gains among home-owners.

We use five variables to control for remodelling changes: number of bedrooms, number of bathrooms, number of garages, floor area, and whether there was a construction of a new

[8]These cut-off values are obtained as the top 0.1 percent available values for each variable.
[9]Table B.1 in the Appendix presents the number of observations by region.

dwelling on the property. Unfortunately, for a substantial number of sale observations, there is some missing information on the house features. As we do not want to throw away data, we construct four different samples for our analysis. In our first sample we keep all 1,058,391 repeat sales observations on 583,561 houses. This sample provides our baseline estimates. In these baseline estimates, we treat missing values in the house features as no remodelling change for those features. The effect of this assumption is that some actual remodelling changes are not captured in our baseline estimates. This does have some small upward effect on the idiosyncratic risk estimates. In a second sample we exclude all house sale observations where the record shows that a new dwelling was built on the property. We do this to avoid comparing an older building with an entire new building. This sample has 1,026,613 repeat sales, somewhat reduced from our first sample, where we simply control for new dwellings in the regression with a dummy variable. In our third sample we exclude all observations where some features of the house is missing. This greatly reduces our sample size, which still remains at a substantial 522,616 observations. In our fourth sample, we exclude all repeat sales that are less than 2 years apart. This sample therefore excludes houses that are bought for quick ‘flipping’ purposes.

Table 1: Full Sample Summary Statistics.

			Percentiles					
Variable	Mean	Std. dev.	1^{st}	25^{th}	50^{th}	75^{th}	99^{th}	N. Obs.
House characteristics:								
Bedrooms	3.06	0.82	1	3	3	3	5	1,054,105
Bathrooms	1.43	0.69	0	1	1	2	4	541,315
Garages	1.46	0.87	0	1	1	2	4	1,026,847
Floor area (m^2)	143.11	65.16	50	100	126	174	351	1,029,021
Repeat sales:								
Initial price deviation* (%)	12.67	66.63	-71.43	-25.64	-1.27	32.90	253.69	1,058,391
Price change (% ann.)	5.52	8.20	-12.27	1.20	4.28	8.30	34.96	1,058,391
Holding years	6.70	5.23	1	3	5	9	24	1,058,391

Notes: *Price of purchase relative to the regional median.

3 Model and Estimation

To analyse idiosyncratic risk, we estimate the model of annual house price changes of Landvoigt et al. (2015). In this model, idiosyncratic risk is time-varying. We extend the model in three dimensions. First, we account for regional differences in annual average price changes. This extension accounts for the fact that location factors determine average price appreciation. Second, we account for the effect of remodelling on house price appreciation. This extension allows us to control for changes in house features that will affect realized capital gains. Third, inspired by findings in the literature we allow the variance of the idiosyncratic shocks to depend on location, the initial house price (measured as a deviation from the regional median price) and the holding period.[10]

The model is estimated in two stages. The first stage focuses on house price changes. The residuals from the first stage are then used in the second stage to obtain estimates of the variance of idiosyncratic shocks.

3.1 Model of Price Changes

Letting $p_{i,t}$ denote the (log) price of a house i in year t, and $X_{i,t} = [x_{1,i,t}, ..., x_{F,i,t}]'$ denote the change in a set of F house features for house i between periods t and $t+1$, the price change of house i between year t and $t+1$ is given by

$$p_{i,t+1} - p_{i,t} = \alpha_{r,t} + \beta_t (p_{i,t} - \widetilde{p}_{r,t}) + \Psi X_{i,t} + \varepsilon_{i,t}, \tag{1}$$

where $\varepsilon_{i,t}$ represents idiosyncratic shocks with mean zero, $\alpha_{r,t}$ captures the expected price change that is common for all houses in region r, β_t captures the effect of the house price deviation from the regional median, $\widetilde{p}_{r,t}$, on its expected price change, and $\Psi = [\psi_1, ..., \psi_F]$ is a set of coefficients that capture the (average) effect of changes in the corresponding house fea-

[10] Sinai (2009), Peng and Thibodeau (2017) and Han (2013) provide evidence on varying returns and risk across location. Guerrieri et al. (2013) provides evidence on the variation of the variance of appreciation rates according to initial house price and Giacoletti (2021) shows that idiosyncratic risk varies as a function of the holding period. None of these papers however combine all these factors.

tures on the expected price appreciation of the house.[11] The coefficients β_t affect the evolution of the distribution of price changes across house quality. Namely, $\beta_t < 0$ implies that prices of initially cheaper houses will, on average, have higher price appreciation than more expensive houses, i.e., there is convergence of house prices towards the regional median between periods t and $t+1$. In contrast, if $\beta_t > 0$ the distribution of house prices is diverging as initially cheaper houses appreciate less than more expensive houses.

In order to capture the dynamics of the distribution of house prices the relevant coefficients of model (1) are time dependent. Therefore, the model identification comes from the cross section variation of house prices and capital gains observed from the repeat sales data. To avoid selection bias, estimates of these coefficients are obtained from all repeat sales simultaneously. Namely, the estimated coefficients $\alpha_{r,t}$ and β_t reflect any repeat sale that brackets the year t. In contrast, the effect of changes in house features, captured by Ψ, is assumed constant over time. In spite of having assumed a time independent effect on price appreciation, the effect of changes in house features for repeat sales extending over multiple periods need to be compounded with the differential effects of initial house price. Given that the state of house features is only observed when the house is sold, an assumption regarding the evolution of changes between repeat sales is required. Here we assume that such changing house features evolved according to a simple average in the (unobserved) periods between repeat sales.

Under these assumptions, estimation of model (1) with repeat sales data is achieved by extending the model to a system of non-linear equations covering all possible pairs of repeat sales observed in the sample. Details about this system of equations and the mappings to the structural parameters $\alpha_{r,t}$, β_t and Ψ are provided in Appendix A. The parameters of the model of price changes are estimated using non-linear least squares in two steps. In the first step, every equation in the system is weighted equally. In the second step, each equation is weighted by the inverse of the variance of their residuals from the first step of estimation.

[11] Landvoigt et al. (2015) do not subtract the regional median $\widetilde{p}_{r,t}$ from the price of the house as they only estimate their model on one region. We introduce this useful re-scaling to aid in interpreting the regional coefficients $\alpha_{r,t}$. These represent the average price change in region r of the median priced house with no remodelling changes.

3.2 Variances of Idiosyncratic Shocks

We are interested in systematic variation of the idiosyncratic risk that is associated with housing investment. In our repeat sales model, idiosyncratic risk is captured by the standard deviation of the idiosyncratic shocks $\varepsilon_{i,t}$. We assume the idiosyncratic shocks $\varepsilon_{i,t}$ are independently normally distributed with variance $\sigma^2_{\varepsilon,i,t}$. As in Landvoigt et al. (2015), we allow this variance to change over time. However, we also allow the variance to depend on three additional factors: (i) the region where the house is located; (ii) the initial price of the house; and, (iii) the holding period between the house's repeat sales.

To be precise, the idiosyncratic shocks are assumed to behave as a zero mean normal random variable with variance given by

$$\sigma^2_{\varepsilon,i,t} = \exp\left(\phi_{r_i} + \delta \ddot{p}_i + \rho\left(k_i - 1\right)\right)\sigma^2_t, \tag{2}$$

where $\ddot{p}_i$ stands for the initial house price deviation from the regional median (in logs), ϕ_{r_i} captures regional differences in idiosyncratic risk, δ regulates the sensitivity of idiosyncratic risk to the initial house price deviation from the regional median, and ρ captures the effect of the holding period, given by k_i as the number of years between repeat sales of the same house. The set of parameters $\phi_r, \delta, \rho, \sigma^2_t$ determine the "expected" idiosyncratic risk of an individual house, at a particular location, initial price and holding period, in a particular year.[12] In econometric terms, this is equivalent to heteroskedastic errors in the determination of housing returns.

Under these assumptions, the parameters underlying the variances of the idiosyncratic shocks are estimated using maximum likelihood on the residuals from the system of multi-period repeat sales equations defined above. Derivations of the likelihood function are provided in Appendix A.

[12]This quantity is "expected" in the statistical sense of the expected value of the second moment of a random variable. As discussed in the text, our estimates are based on a full sample of realized housing returns, hence not a direct estimate of *ex ante* returns.

4 Estimation Results

In this section, we present the estimation results of our extended model. We estimate the model on our main sample of 1.06 million observations and re-estimate it on three other samples to check for robustness. We first discuss the resulting estimates for the model of price changes. We then focus on the estimates of systematic variation in idiosyncratic risk.

4.1 Price Changes Model Estimates

Table 2 shows the estimates of Ψ, which determine the effect of remodeling on annual price appreciation. The positive and significant coefficient estimates show, unsurprisingly, that adding bedrooms, bathrooms, or garages increases the value of the house. So does adding floor space or a new dwelling. For instance, adding one bedroom increases the value of the house by around 6.5 percent, while adding a bathroom adds around 1.5 percent to the value of the house. The results are quite robust across the different samples.

Figure 2 shows the time-varying parameter estimates in this model. Panel (a) shows the coefficient estimates of the regional time dummies $\alpha_{r,t}$, which determine the time varying annual average price appreciation. These estimates are consistent with the price trends observed over the period. Namely, a sharp and broad-based increase in house prices prior to the GFC, then followed by another period with a positive trend over the last decade. Panel (b) shows the coefficient estimates of β_t in Equation 1. Recall that these estimates give us an indication of what is happening with the distribution of house prices. Namely, a negative β indicates that cheaper houses have been appreciating faster than more expensive ones, hence house prices are converging (and vice versa if β is positive). As depicted in Figure 2 these estimates are significantly cyclical, and mostly negative, hence suggesting the distribution of house prices has tended to shrink over time. But there were some periods of divergence too, particularly between 1997-2000 and 2007-09, the bust phases of the housing cycles. One potential interpretation for this effect is that owners of cheaper houses are more likely to be forced to sell during the busts, which makes cheaper homes depreciate faster. This evidence is consistent with Landvoigt et al. (2015) findings for San Diego county who argue that cheaper credit for poor households was a

Table 2: Model estimates.

	(1)	(2)	(3)	(4)
Elasticity estimates:				
Bedrooms	6.515	6.446	6.092	7.331
	(0.061)	(0.062)	(0.081)	(0.069)
Bathrooms	1.467	1.477	1.284	1.574
	(0.082)	(0.082)	(0.085)	(0.097)
Garages	0.637	0.645	0.331	0.666
	(0.050)	(0.049)	(0.058)	(0.053)
Floor area	0.539	0.531	0.545	0.587
($25m^2$)	(0.029)	(0.030)	(0.035)	(0.032)
New dwelling	2.402		2.834	1.925
	(0.125)		(0.172)	(0.157)
Variance parameter estimates:				
Initial price deviation, $\hat{\delta}$	-0.278	-0.286	-0.332	-0.275
	(0.008)	(0.008)	(0.011)	(0.008)
Holding period, $\hat{\rho}$	-0.156	-0.152	-0.192	-0.117
	(0.003)	(0.003)	(0.005)	(0.003)
Sample	All	Excluding new dwelling	Excluding missing	Excluding hold < 2 years
R-squared (equation 1)	0.760	0.761	0.776	0.773
Avg. log likelihood (equation 2)	0.192	0.195	0.261	0.143
N. Obs.	1,058,391	1,026,613	522,616	947,089

Notes: Elasticity estimates are estimates of Ψ from equation 1. These can be interpreted in terms of annualized (log) capital returns. Variance parameter estimates are estimates of δ and ρ from equation 2. The complete set of estimates of $\alpha_{r,t}$ and β_t, and estimates of ϕ_r and σ_t^2, are reported in the Appendix. Robust standard errors are presented between brackets. All estimates are statistically significant at the 1% significance level.

major driver of house prices at the low end of the market.[13] Of course, these predictions do not account for idiosyncratic variation in house prices, which we now focus on.

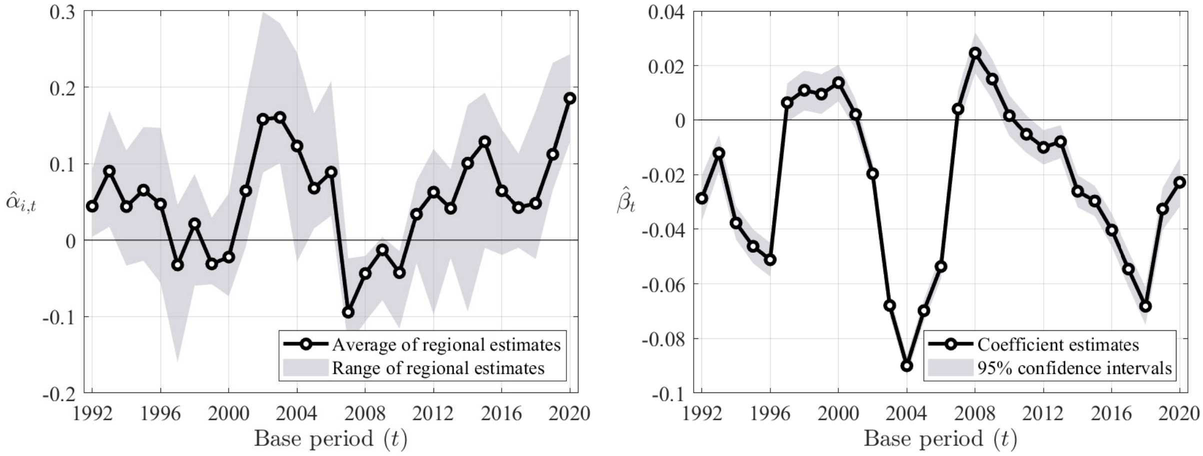
4.2 Idiosyncratic Risk Estimates

Idiosyncratic risk is measured by the standard deviation of the shocks to house prices not captured by the model of price changes. Table 2 shows the estimates of the parameters determining how the variance of idiosyncratic shocks is affected by the initial price of the house relative to the regional median (δ) and by the holding period between purchase and resale of the house (ρ).

[13] The impact of the initial house price on the house price appreciation in our sample is nevertheless limited, as indicated by the magnitudes of $\hat{\beta}_t$. Figure D.2 in the Appendix shows that the time variation in returns fitted by the model is mostly determined by shifts in $\hat{\alpha}_t$.

Figure 2: Time-varying model parameter estimates.

(a) Regional capital gains estimates. (b) Effects of initial house prices on capital gains.



Notes: Panel (a) shows estimates of $\alpha_{r,t}$ from Equation 1. These represent estimates of the average (log) annual capital gains for a median house in each region. The regional average is weighted by the number of sales in the region/year. The shaded areas indicate the minimum and maximum $\hat{\alpha}_{r,t}$ values across the regions for each year. Panel (b) shows estimates of β_t from Equation 1. These capture the effect that the initial price of the house, relative to the regional median house price, has on the average (log) annual capital gains. Negative β means a house cheaper than the median regional house gained more value than the houses that are more expensive than the median. The confidence band is based on robust standard errors.

The estimates of δ in Table 2 are consistently negative, indicating that a house purchased at a price below the regional median is subject to a higher risk of idiosyncratic price changes. In terms of magnitude, the baseline estimate, $\hat{\delta} = -0.278$, indicates that a house purchased at a price 25% below the regional median has an idiosyncratic risk about 4.1% higher than the median house.[14] Cheap houses are therefore more risky.

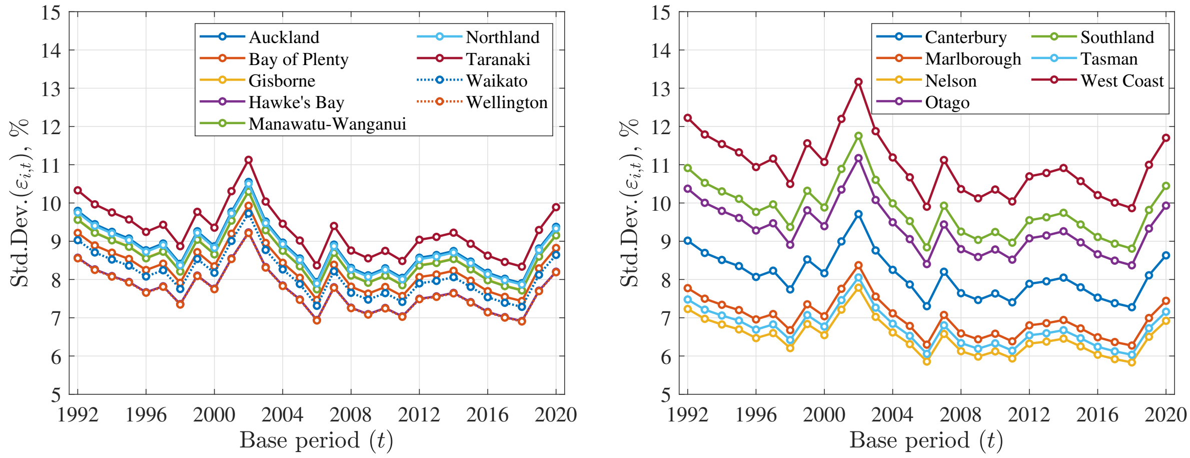
The estimates of ρ in Table 2 are also negative, indicating that holding the house for longer reduces the risk of idiosyncratic price changes. The baseline estimate, $\hat{\rho} = -0.156$, indicates that for every additional year the house is held idiosyncratic risk declines by about 7.8%. Giacoletti (2021) shows that idiosyncratic risk varies across holding period. Houses that are held over longer periods have less idiosyncratic risk. We find this to be also true in New Zealand.

The absolute level of idiosyncratic risk is determined by considering these factors jointly with the regional and time factors in the model, Equation 2. Figure 3 shows how idiosyncratic risk varies over time and across regions for a median-priced house that is held for five years (so $\ddot{p} = 0$ and $k = 5$). We use a holding period of five years as this is the median holding

[14] Noting that the price deviation in Equation 2 is transformed to logarithms, the relative effect can be calculated as $\sqrt{(1-0.25)^{-0.278}} - 1$.

Figure 3: Idiosyncratic risk estimates by region.

(a) North island. (b) South island.



Notes: These are obtained by maximum likelihood from the residuals of the repeat sales model, assuming the idiosyncratic shocks are distributed as a Normal and iid with variance given by equation 2. The estimates are for a median priced house in the corresponding regions and holding period of five years between repeat sales. Estimates shown are those obtained using the baseline sample from Table 2.

period for houses in New Zealand. Estimates of σ_t and ϕ_r which determine the regional shifts in idiosyncratic risk are in the Appendix (see Table C.2 and Table C.3).

Figure 3 shows that idiosyncratic risk varies over time. It was relatively high in 1992 at the start of our sample, then gradually dropped until 1998. It rose again to reach a high in 2002, after which it dropped again until 2006. It rose in 2007, however between 2008 and 2018 it was relatively stable.[15] Interestingly, during the Covid years idiosyncratic risk rose significantly (by about 0.9 and 0.6 percentage points in 2019-20 and 2020-21, respectively). All in all however, the time-variation of idiosyncratic risk is within limits. For instance, for a median-priced house in Auckland and a holding period of five years, idiosyncratic risk varied from a low of 7.90 percent (2018) to a high of 10.55 percent (2002).

Figure 3 also shows the wide regional variation in risk. Our specification assumes that

[15]This evolution of risk in housing parallels with the developments of the banking sector and financial regulation in New Zealand over the period (see, e.g., Murphy, 2011). Following a period of financial liberalization in the 1990s, whereby private banks replaced the state as the primary suppliers of mortgage finance, the early 2000s were characterized by increased price competition and financial product innovations in the banking sector. At the same time a substantial portion of the New Zealand mortgage market at the time was funded by overseas carry-trades, which tied the domestic credit conditions to a roughly unregulated and less conservative global market. Prompted by the GFC turmoil, the RBNZ introduced a series of new prudential liquidity policies that sought to reduce the retail banks' reliance on external sources of funding (see, e.g., Nield, 2008; Hoskin et al., 2009). Moreover, in October 2013, the RBNZ introduced loan-to-value ratio (LVR) restrictions to curb risky mortgage lending, while the LVR values have been revised over time (see Lu, 2019, for an early review, and McDonald and Markham, 2023 for a recent account of LVR's evolution).

regions vary simultaneously in risk across time. Essentially we have only allowed regions to differ in risk with a shift factor, $exp(\phi_r)$, which is constant over time. Otherwise said, we estimate 29 parameters σ_t^2 (one for each t) and 15 parameters ϕ_r (one for each r). (Shift factors are relative to Auckland, which is the base region with $\phi_r = 0$.) In the Appendix we show results where we estimate regional time varying risk, i.e. 464 estimates $\sigma_{r,t}^2$ (see Figure D.5 and Figure D.6). The shift factor however explains most of the variation across regions in idiosyncratic risk.

To determine the relative importance of each of the factors in the systematic variation of idiosyncratic risk, we do the following calculation. We calculate the idiosyncratic risk of a house that is at the 10th percentile in all dimensions (of time, region, holding period and initial price) and compare it with a house at the 90th percentile in all dimensions. Due to the multiplicative nature of the relative risk factors we can calculate the relative contribution of each factor in the increase of risk of buying a house at the 90th percentile versus the 10th percentile.[16] A house at the 10th percentile has an absolute idiosyncratic risk of 2.8%. A house at the 90th percentile has an absolute idiosyncratic risk of 13.8%.[17] Calculating the relative contributions as explained above, this almost five-fold increase of risk is 58% due to the holding period factor, 21% percent due to the regional factor, 11% due to the time variation, and 10% due to the initial price variation. The holding period is the most important factor driving the systematic variation in idiosyncratic risk, followed by location.

5 Why does idiosyncratic risk vary?

We find that idiosyncratic risk varies systematically in the dimensions of region, time, holding period and initial price. What can explain this variation? In this section, we investigate its potential determinants by regressing the idiosyncratic risk estimates on explanatory variables that cover different channels.

[16] This calculation boils down to calculating the ratio of each of the bracketed terms in $0.5\left[\hat{\phi}_{r90} - \hat{\phi}_{r10}\right] + 0.5\left[\hat{\delta}\ddot{p}_{90} - \hat{\delta}\ddot{p}_{10}\right] + 0.5\left[\hat{\rho}(k_{90}-1) - \hat{\rho}(k_{10}-1)\right] + \left[ln(\sigma_{t90}) - ln(\sigma_{t10})\right]$ to the total sum of bracketed terms, where the subscripts 90 and 10 indicate the 90th and 10th percentile.

[17] A house at the 10th percentile of risk is located in Tasman, was bought in 2017 at a price 80% above the median, and is held for 14 years. A house at the 90th percentile of risk was located in Otago, was bought in 2003 at a price 43% below the median, and was held for 2 years.

5.1 Potential channels

Before describing the empirical framework of this section, we discuss the potential mechanisms through which variation in idiosyncratic risk may be determined. Generally, idiosyncratic pricing is associated with "matching uncertainty" in the house trading process, which, in turn, can emerge through two main channels: (i) market thinness, and (ii) informational asymmetries.

Market thinness is a leading potential driver of idiosyncratic pricing, as smaller pools of potential buyers and sellers decrease the probability of matching. Given that housing investments are highly dependent on mortgage financing, credit conditions are a key determinant of market thinness. Tighter credit conditions reduce the availability of funds for borrowers, thereby reducing the pool of potential house buyers. However, in markets where mortgage rates are re-negotiated throughout the loan tenure, credit conditions can also affect the supply side of the housing market. Higher borrowing costs can constrain mortgage rollover and lead to forced sales, increasing the pool of houses available in the market as some of the distressed households switch to renting. Credit conditions can also have a behavioral impact on risk-taking. Particularly, lower borrowing costs may lead to increased speculative and risk-taking behavior, consequently increasing idiosyncratic volatility in housing prices. Hence, in theory, the impact of credit conditions on idiosyncratic risk is ambiguous. In our empirical analysis, we investigate how credit conditions are related to the estimated variation in risk across time, initial house prices, and holding period. In addition, housing market thinness is likely to depend on regional market sizes, hence we also control for variation in market size and liquidity across time and regions.[18]

Another key potential driver of idiosyncratic risk relates to informational asymmetries. According to the theory of search and matching, the poorer the information on house valuations the larger the heterogeneity of buyer's valuations, which increases the potential variation of transaction prices. Giacoletti (2021) provides empirical evidence that houses with higher valuation uncertainty have both a larger idiosyncratic risk and a more steeply declining term structure of

[18] In New Zealand, cities are relatively small with significant variation in population density across regions. For example, West Coast, which has the largest estimated idiosyncratic risk, is the region with the lowest population (below 50,000) and the lowest population density (below 2 persons per square km), while Nelson, the lowest risk region, has the second highest population density (below that of Auckland).

risk over an increasing holding period.[19] This uncertainty can further interact with the supply of credit at the individual level, as poorer information about the house may lead to collateral value uncertainty and reduced mortgage credit.[20] Information quality may also be associated with the quality of the house, which may be proxied by the house price. Hence, the effect of initial house price on risk may also be associated with the varying degree of information uncertainty across house prices. Finally, heterogeneity of information access may be another dimension through which this channel operates, especially across regions. As housing markets rely heavily on internet advertising, regions with less internet access will tend to face higher valuation uncertainty. The remoteness of certain areas in New Zealand indeed implies some areas have no internet access. Although the data does not allow a proper disentangling of these channels, we test their general implications for estimated risk variation across regions, initial prices, and holding periods, by looking at a measure of house atypicality and data on internet access across different levels of geographical granularity.[21]

5.2 Regression framework

In order to identify the determinants of risk variation across each dimension, we first calculate risk estimates controlling for the variation in the remaining dimensions. For example, when looking at risk variation across time, we subtract from the risk estimate the variation that is determined by the other three dimensions. The general regression specification is given by

$$\ln \hat{\sigma}_{\varepsilon,i,t} - \ln \widetilde{\sigma}_{\varepsilon,i,t} = \alpha + \Lambda \mathbf{X}_{i,t} + u_{i,t}, \tag{3}$$

where $\hat{\sigma}_{\varepsilon,i,t}$ is calculated according to Equation 2 for each repeat sales i and sequence of years t between the repeat sale initial and end transactions, $\widetilde{\sigma}_{\varepsilon,i,t}$ is calculated similarly but keeping one (or all) of the dimensions of risk variation constant, and $\mathbf{X}_{i,t}$ is a set of explanatory variables.[22]

[19] Similarly, investigating house price changes in New Zealand, Bourassa et al. (2009) find that atypical houses tend to have more volatile prices than those of standard properties.

[20] For example, Jiang and Zhang (2023) estimate a hedonic price model and find that less standardized houses tend to have greater value uncertainty, which ends up affecting the amount of credit offered in mortgages.

[21] Bargaining power asymmetry is another potential channel through which market thinness and informational quality can lead to higher price variability (see, e.g., Harding et al., 2003).

[22] For example, when we look at year effects, $\widetilde{\sigma}_{\varepsilon,i,t} = \exp\left(\frac{1}{2}\hat{\phi}_{r_i} + \frac{\hat{\delta}}{2}\ddot{p}_{i,t} + \frac{\hat{\rho}}{2}(k_{i,t}-1)\right)\hat{\sigma}_{1992}$, where 1992 is taken as baseline; when we look at regional effects we set $\hat{\phi}_{r_i} = 0$ (i.e., Auckland is the baseline) and calculate

Note that a logarithmic transformation is adopted due to the multiplicative formulation of risk estimates. Also note that our dataset is expanded to cover all periods between repeat sales in order to allow identification of time-specific effects from time-varying explanatory variables, such as interest rates. Finally, there is only time and (sub-)regional variation in the explanatory variables.[23]

As explanatory variables we use measures associated with the two broad channels discussed above: (i) for credit conditions we use the central bank's policy rate to capture borrowing costs and loan-to-value ratios (LVR) to capture borrowing constraints; (ii) for informational asymmetries and uncertainty we construct a measure of house atypicality based on a hedonic regression and use census data on the share of households with internet access.[24] As control variables for market activity and size we use the number of house sales, total population, and new building consented area.[25]

5.3 Regression estimates

Table 3 presents the results of five regressions. Columns (1) to (4) show how the explanatory variables correlate with the estimated idiosyncratic risk variation across time, region, initial house price, and holding period, respectively, each at a time while controlling for the variation in the other dimensions. In each of these regressions, as much as possible we focus on variables presenting variation along the risk dimension of interest. Column (5) shows how the same variables correlate with the total variation in the estimated idiosyncratic risk, i.e., without controlling for any dimension.

Credit conditions are found to be particularly relevant for risk across time and holding pe-

$\widetilde{\sigma}_{\varepsilon,i,t} = \exp\left(\frac{\hat{\delta}}{2}\ddot{p}_{i,t} + \frac{\hat{\rho}}{2}(k_{i,t}-1)\right)\hat{\sigma}_t$; and so on for the initial price and holding period dimensions.

[23]Because relative prices and holding periods vary both across time and across houses, we consider variables that vary across time and across statistical areas (level 2, SA2), which is the lowest level of geographical aggregation we managed to match with the house sales data. There are a total of 2,104 SA2s with over 2 million matched house sales in our data. The median area and population of SA2s are 1.93km^2 and 1,942 people, respectively.

[24]Our measure of house atypicality is similar to that adopted in the literature (e.g., Bourassa et al., 2009): first, we estimate a hedonic regression of house (log) prices on house characteristics, which include the number of bedrooms, bathrooms, and garages, floor area, decade built, and fixed effects for region, year, and type of property; second, we calculate the deviations of each house's characteristics from their corresponding neighborhood (statistical area level 2) mode values (except for floor area where we take the median); third, we use the estimated hedonic regression to calculate the total implicit price of each house's characteristics deviations from their neighborhood's typical values; fourth, we average the absolute standardized valuation differences by region.

[25]Descriptive statistics and sources for these variables are provided in Table B.3 in the Appendix

Table 3: Regression results on idiosyncratic risk estimates.

Explanatory variables	(1)	(2)	(3)	(4)	(5)
(i) Credit conditions:					
• Interest rate	-2.080**		0.017	-2.372***	-2.262***
(%, national)	(0.773)		(0.026)	(0.532)	(0.616)
• Loan-to-value ratio	0.156		-0.012*	-0.536**	-0.443**
(%, national)	(0.105)		(0.006)	(0.198)	(0.153)
(ii) Informational asymmetries/uncertainty:					
• Average house atypicality		0.276**	0.009	1.425***	0.982**
(regional)		(0.116)	(0.036)	(0.466)	(0.399)
• Internet access, 2006		-1.363**	-0.266***	0.013**	-0.328***
(%, regional/SA2)		(0.511)	(0.071)	(0.005)	(0.045)
Variation in risk	Year	Region	Price	Hold	All
Explan. vars. aggregation	Country	Region	SA2	SA2	SA2
N. Obs.	7,088,089	6,834,835	6,535,077	6,535,077	6,535,077
R-squared	0.575	0.275	0.282	0.137	0.101

Notes: Each column represents a regression with dependent variable equal to the repeat sales (log) estimate of idiosyncratic risk excluding the controlled variation, $\ln\hat{\sigma}_{\varepsilon,i,t} - \ln\widetilde{\sigma}_{\varepsilon,i,t}$, where applicable. All regressions include additional controls for market activity/size given by year/region/SA2 number of sales, total population, and new building consents area. Cluster robust standard errors are presented between brackets, where the clusters are defined according to the variation in risk for regressions (1) to (4) and two-way clusters by region and year for regression (5). ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels of significance, respectively. SA2 stands for statistical area (level 2) definition from the New Zealand's 2018 Census of Population and Dwellings.

riods. Across time, higher interest rates and lower LVR are associated with lower risk, though the latter effect is statistically insignificant (panel (i), column 1). Hence, tighter credit conditions are associated with periods of lower idiosyncratic risk, which is consistent with the costly re-financing and risk-taking behavioral channels. Interest rates have a similar effect on risk across holding periods, but the impact of macro-prudential policy inverts (panel (i), column 4). Namely, a more restrictive LVR (equivalent to a lower LVR) is associated with higher risk across holding periods. As the risk estimates decrease with the holding period, these estimates indicate that while higher borrowing costs incentivize holding onto a property for longer, which is consistent with the costly re-financing and risk-taking behavioral channels, periods of higher deposit requirements are associated with shorter holding periods, leading to higher risk. Hence, the impact of LVR on idiosyncratic housing risk is consistent with the market thinness channel.[26] The effects of credit conditions on risk variation across holding periods dominate the

[26]This results is consistent with one of Giacoletti (2021)'s findings that contractions in local credit supply are associated with a higher level of idiosyncratic risk.

impact on total risk variation (panel (i), column 5). Quantitatively, an interest rate decrease of 1 percentage point is associated with an increase in idiosyncratic risk of 2.26%, while an LVR tightening of 10 percentage points increases risk by 4.43%.

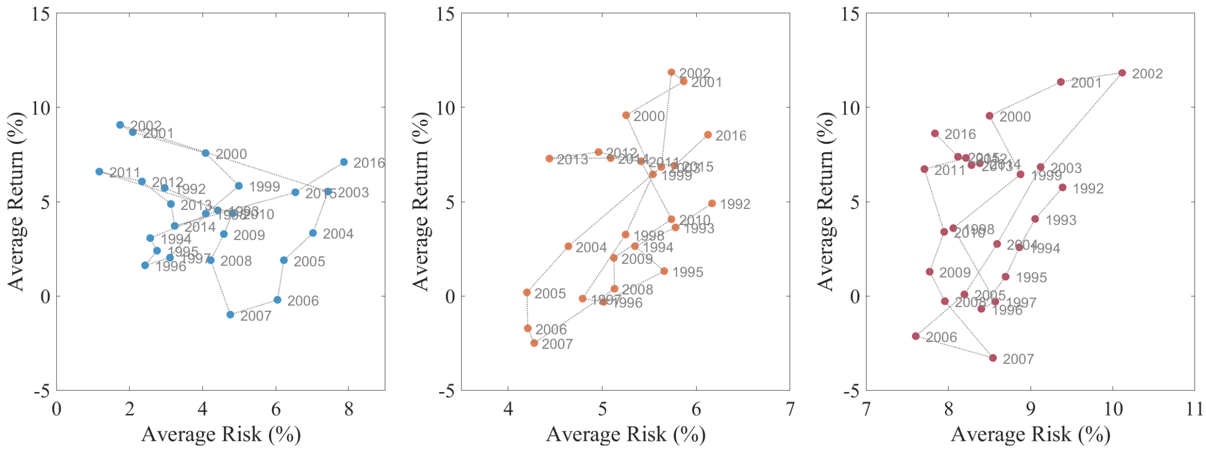
Information uncertainty effects are found to be mostly consistent in determining idiosyncratic risk across regions, relative house prices, and holding periods. First, regions with more atypical houses tend to have higher risk estimates (panel (ii), column 2) and shorter holding periods (panel (ii), column 4). Second, regions with lower internet access also tend to have higher risk estimates. Third, differences in internet access are also associated with the variation in risk across initial house prices (panel (ii), column 3). Recall risk is found to decrease with the relative house price. Hence, the effects here may be interpreted in relation to how the variables impact relative house prices. Lower internet access is associated with higher risk, hence cheaper houses. Although it is not possible to determine the direction of causality, the fact that internet access is a significant determinant of regional risk, where price effects are held constant, suggests that access to information is an important factor in the determination of idiosyncratic risk. The small (though positive and statistically significant) coefficient on internet access for risk variation across holding periods (panel (ii), column 4) suggests this specific channel is not a major determinant of housing turnover. Finally, the impact of information uncertainty on total risk variation is consistent with the expected signs of this channel (panel (ii), column 5). Thus, by affecting the house valuation process, informational asymmetries also have a bearing on idiosyncratic housing risk, with effects operating mainly through differences in access to information across regions and house prices, and varying degrees of heterogeneity of housing characteristics across regions and its impacts on holding periods.

6 Risk-return Relationships

Is the systematic variation in idiosyncratic risk priced in housing returns? In this section, we analyze the risk-return relationship in housing investments implied by our model estimates. Our model allows us to draw risk-return relationships across four dimensions of idiosyncratic risk variation: time, location, holding period, and initial house price. Throughout this section,

Figure 4: Risk-return across time and estimation methods.

(a) Index-based. (b) Raw repeat sales. (c) Model-based.



Notes: All risk and return statistics are annualized rates for a 5-year holding period starting at the indicated year. Index-based returns are the average annualized 5-year (log) returns computed from the national average of prices. Index-based risk is calculated as the standard deviation of year-on-year (log) returns for each 5-year holding period. For the raw repeat sales we average the annualized returns between repeat sales with a 5-year holding period, and take the standard deviation of the same annualized returns for average risk estimates. Model-based estimates are calculated using our model's predictions.

we also compare estimates of risk and return obtained from house price indices and the raw repeat sales data. The methods used for the calculation of these estimates of risk and return are described in Appendix A.

6.1 Risk-return across Time

Figure 4 presents the estimates of risk and return across time and different estimation methods, focusing on investments over a 5-year holding period. First, it is interesting to note that the level of risk is different across the measures (note the different horizontal axis ranges across the panels). The index-based estimates (panel a) generally underestimate risk because they disregard the cross-sectional variation in house price changes. The raw repeat sales estimates (panel b) account for that idiosyncratic component, but still underestimate risk because they are based only on repeat sales for the specific holding period. Those estimates also do not control for the effects of remodeling and the initial house price. Accounting for all these factors, the model-based estimates (panel c) show that the level of risk is about 4 percentage points higher than what would be inferred from repeat sales data alone.

Next, we consider whether risk is priced in housing returns over time. Using index-based estimates (panel a) one would be misled to think that there is no trade-off between risk and return over time. Using the raw repeat sales data (panel b) and the model-based estimates derived from that data (panel c) we find that there is a positive relationship between idiosyncratic risk and returns associated with the time the house is purchased. The relationship is rather steep, as a small range of variation in risk (about 2%) is associated with large changes in returns (about 15%). That means the time variation in risk is mostly priced in the time variation of returns.

6.2 Risk-return across Locations

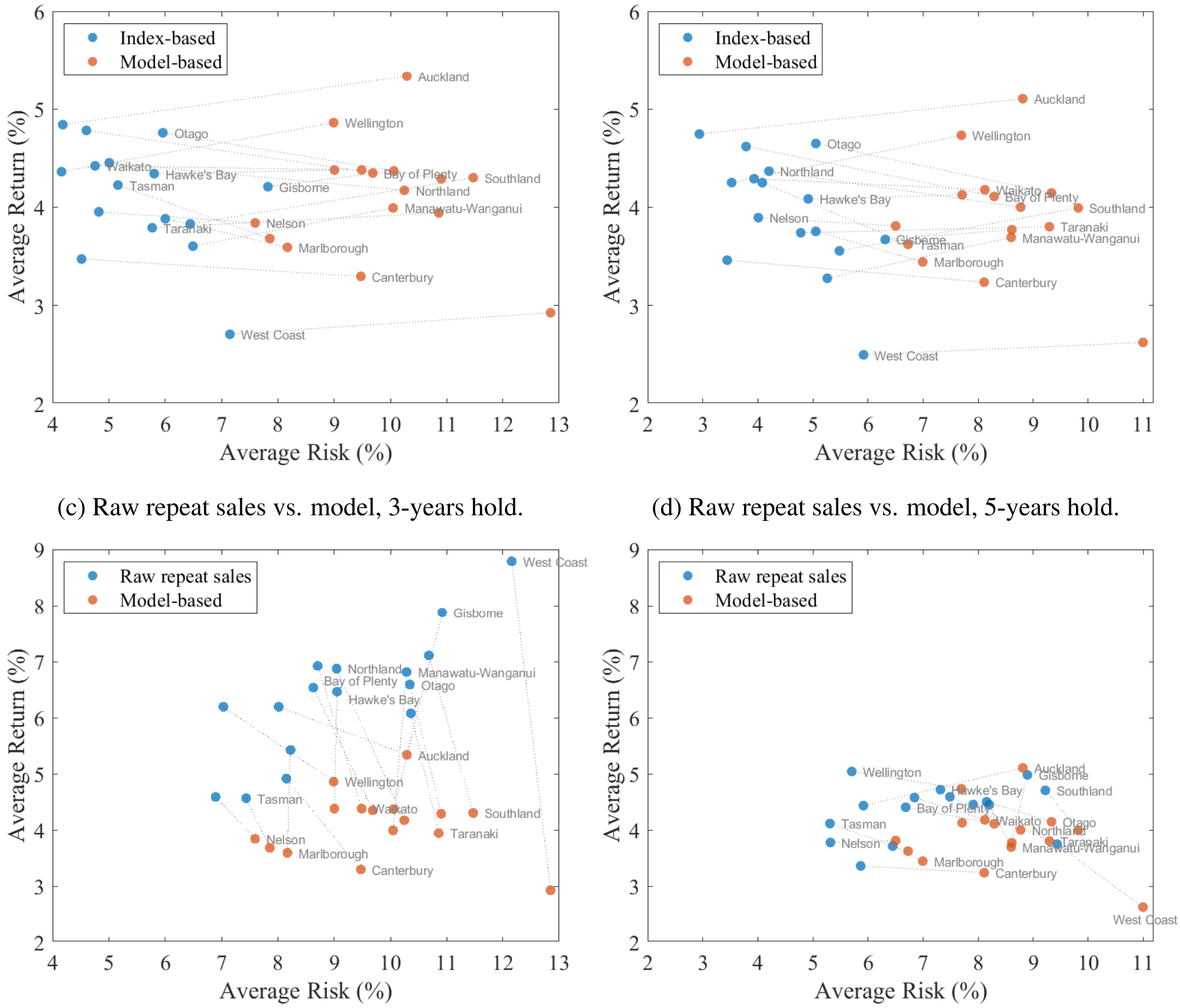
Figure 5 presents the estimates of risk and returns across regions. Panels (a) and (b) show that model-based estimates of risk are higher than index-based estimates. This is because the latter does not account for the large within-region variation in house appreciation. Panels (c) and (d) show that model-based estimates of risk tend to be higher than estimates based on raw repeat sales as well. This is because the raw repeat sales only account for the price changes of specific holding periods. This also leads to a mismeasurement of returns for shorter holding periods (panel c).

Overall, the regional variation in idiosyncratic risk is not priced in regional returns. Moving from a low-risk region to a high-risk one does not increase housing returns. In other words, when households move across regions and move to regions with higher idiosyncratic risk they are not rewarded with higher returns on their housing. This is in contrast with what we found about risk-return across time. Such a dichotomy between pricing of housing risk across time and location implies that home buyers exert arbitrage mainly across time, i.e., timing when to buy but not where to buy. We believe this is reasonable as the decision of location is likely determined by factors such as family origins, and employment and education opportunities, which are, arguably, mostly exogenous to the decision of where to buy.

6.3 Risk-return across Holding Periods

Although our model allows idiosyncratic risk to depend on the holding period, that is not the case for the specification of price changes. Hence, it is interesting to see whether the declines in

Figure 5: Risk-return across regions and estimation methods.



(a) Index vs. model, 3-years hold.

(b) Index vs. model, 5-years hold.

(c) Raw repeat sales vs. model, 3-years hold.

(d) Raw repeat sales vs. model, 5-years hold.

Notes: Index-based return and risk are calculated by (i) averaging prices by year and region, (ii) calculating annualized k-year (log) returns based on each region time series of average prices, where k stands for the holding period, and then (iii) averaging the resulting time series of returns and taking the standard deviation of the same series for average risk estimates. For the raw repeat sales we average the annualized returns between repeat sales with a holding period of k years, and take the standard deviation of the same annualized returns for average risk estimates. Model-based estimates are calculated using our model's predictions.

risk associated with longer holding periods are also priced in the returns of holding for longer.

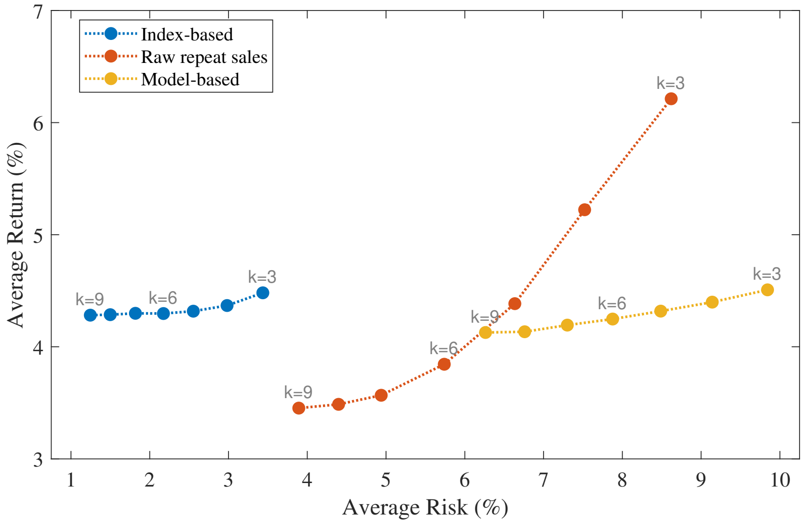
Figure 6 presents a comparative of risk and return estimates by holding period and across estimation methods. Although all methods render increasing risk and return as the holding period decreases, the degree to which risk is priced in returns by varying holding periods is limited according to our model-based estimates. Between holding periods of 3 and 9 years (25th and 75th percentiles of the sample, respectively), model-based estimates of risk decrease by about 4 percentage points while estimated returns decrease by less than a half percentage point. Index-based estimates indicate a similarly lower sensitivity of returns to risk.

In contrast, the raw repeat sales data misleadingly suggest a larger role for holding periods, with steeper risk pricing into returns (see red dots in Figure 6). The raw repeat sales lead to biased estimates of risk and return because these do not account for unobserved price changes of repeat sales of different but overlapping holding periods. For example, houses bought in 2000 and re-sold in 2003, which would enter the raw repeat sales calculations for a holding period of 3 years (k=3), overlapped in time with houses bought in 1999 and re-sold in 2004 (k=5); because part of the price changes are time-specific (recall Figure 4) the time overlap between repeat sales of different holding periods makes them jointly informative about the risk and return specific to each holding period. Our model accounts for these overlapping effects while the raw repeat sales miss them and tend to over(under)estimate returns at lower(higher) holding periods.

6.4 Risk-return across Initial Prices

Finally, our model estimates also allow us to draw the risk-return relationship across initial house prices relative to the regional median. This is one of the features that is new in our model, particularly with respect to the dependence of risk on the initial price of the house. But since the relation between price change and the initial price is allowed to vary over time we also obtain time-varying risk-return trade-offs by initial price. Figure 7 shows that the slope of such a relationship indeed depends on the time of purchase. Recall that in our model of price changes, Equation 1, the relationship between initial house price and returns, β_t, is allowed to vary over time in order to capture dynamics on the distribution of house prices.

Figure 6: Risk-return across holding periods and estimation methods.



Notes: Index-based returns are the average annualized k-year (log) returns computed from the national average of prices. Index-based risk is calculated as the standard deviation of the annualized k-year (log) returns. For the raw repeat sales we average the annualized returns between repeat sales with a holding period of k years, and take the standard deviation of the same annualized returns for average risk estimates. Model-based estimates are calculated using our model's predictions using baseline estimates of Equation 1 and Equation 2 for a median-priced house, then averaged across regions and time for every possible combination of a holding period of k years.

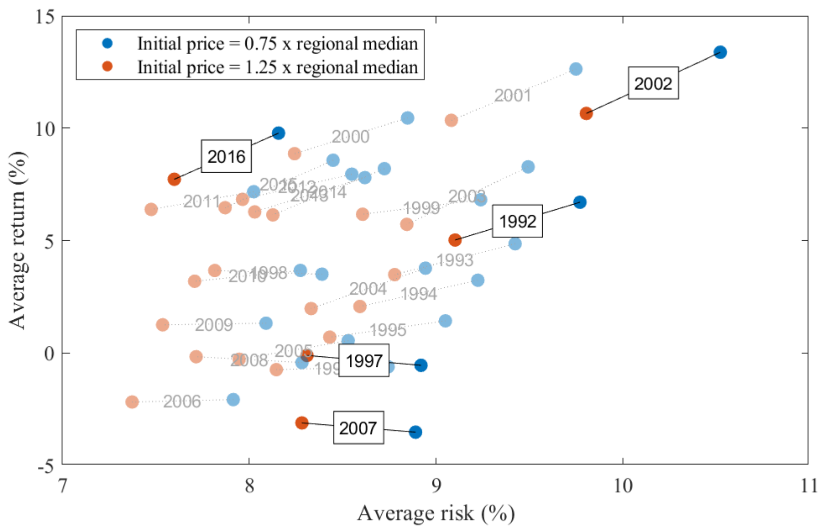
Particularly, during busts of the housing cycle our estimates indicated that house prices are diverging. As Figure 7 shows, housing investment over these periods, e.g., buying in 1997 or 2007 and holding for five years, also resulted in an inverted risk-return relationship across house prices. Namely, more expensive houses not only presented a lower risk but also yielded higher returns (or lower depreciation as these periods also tend to yield negative returns). Across the remaining periods, and for most of the sample period, $\hat{\beta}_t < 0$ and a positive risk-return relationship is observed across initial house prices.

6.5 Discussion

In this section, we presented what our model and other estimates based on the same underlying data imply for the risk-return relationship in housing investments.

The comparison to estimates based on raw repeat sales data and average indices points to the advantages of using model-based estimates. First, although it is quite common in the discussion of housing returns and risk to use index-based time series averages and standard deviations (see, e.g., Jorda et al., 2019), this practice ignores the fact that people don't buy house price indexes

Figure 7: Risk-return across initial price and estimation methods.



Notes: These are model-based estimates for house purchases at the indicated year and a 5-year holding period. Estimates are calculated at the regional level and then averaged. Selected years are highlighted to show how the risk-return relationship by initial house price varies over time.

but individual houses. In other words, it ignores the existence of idiosyncratic risk. Although there is no systematic difference in the average returns, the index-based risk estimates are all lower than the risk estimates based on our model estimates. Quantitatively, index-based risk estimates can be about 5 percentage points lower than model-based estimates.

Estimates based on the raw repeat sales data also result in lower risk estimates than estimates obtained using our model. In addition, the raw repeat sales point to a larger dispersion of returns averaged across holding periods. These biases emerge mainly because looking only at the raw repeat sales data ignores the impacts of unobserved price changes of unsold houses. Our model incorporates those effects by jointly accounting for the repeat sales of different holding periods.[27] Quantitatively, the raw repeat sales risk estimates can be about 2 to 3 percentage points lower than model-based estimates. Hence, both index-based and raw repeat sales estimates of housing risk generally underestimate the risk that households are exposed to.

Regarding the pricing of risk on housing returns, our model estimates point to interesting

[27]Specifically, the differences between model-based and raw repeat sales risk estimates can be traced to (i) the time variation in β_t, which weight the averaging of idiosyncratic shocks between initial purchase and resale, (ii) the time variation in σ_t that is identified through the overlapping repeat sales, and (iii) the controlled effects of regional and relative price differences in risk. In applications that use raw repeat sales data directly on regressions, effect (i) is absent because the capital returns are simple averages over the holding period, while effects (ii) and (iii) could be partially controlled through time and regional fixed effects and the inclusion of the initial house price in the regression.

findings. First, idiosyncratic risk is mostly priced in returns over time, and to some extent with respect to the holding period, but not across regions. Hence, timing when to buy and for how long to hold on to a house before re-selling are critical determinants of performance in housing investments. In contrast, while moving across regions can significantly impact a household's exposure to idiosyncratic price changes, the regional differences in housing risk are not compensated by proportional changes in average returns. Finally, our model provides evidence of a new channel of arbitrage in housing investments. Namely, the risk-return trade-off also depends on the initial price of the house. Our estimates show that during normal times cheaper houses are exposed to a higher risk of idiosyncratic price changes, and this risk is priced in higher average returns. However, this risk-return relationship is also dependent on the housing cycle. Particularly, during housing market busts the risk-return trade-off by initial price inverts as cheaper houses end up depreciating faster than more expensive ones.

7 Conclusions

We estimate idiosyncratic risk of annual house price appreciation using around 1.06 million repeat sales observations in New Zealand from 1992 to 2021. Our estimates show that idiosyncratic risk varies considerably and systematically along four dimensions. Time, location, initial house price and holding period are all important factors determining the idiosyncratic risk of the house, and the holding period is the most important, followed by location.

Our results have important implications for housing demand and housing portfolio choice. As households choose which house to buy, and therefore have a choice on the determinants of idiosyncratic risk, our results suggest that households are able to affect the risk they are facing. First, we show that location matters. The region where the house is located affects the magnitude of idiosyncratic risk. This is due to varying degrees of market thinness and information quality across regional housing markets. By buying in regions that have thicker markets and better information, households can reduce idiosyncratic risk. However since we find that higher risk regions are not rewarded with higher returns this implies that households location choice is not perfectly priced. Likely other determinants such as good schools or high paying jobs attract people to also live in locations where risk is high but average returns are not. Second, holding the house for longer reduces the risk, a result that coincides with findings for the Californian market in Giacoletti (2021). Households can affect this risk by waiting to buy when family size is stable and when job certainty is higher (so that labor market shocks do not necessitate a move out of the area). Moreover, although we find that this risk is priced, holding houses for longer reduces their return only marginally. Third, buying a higher-priced house (i.e., a better-quality house) reduces risk. Interestingly this suggests that the households who need to be least shielded from idiosyncratic risk, i.e. the "rich", are likely most protected from it. Poorer households will naturally (have to) buy cheaper houses, with bigger risks.

Our analysis of the determinants of these systematic variations in idiosyncratic risk in housing also provides some useful lessons for policymaking. First, credit conditions are important drivers of time and holding period variation in risk. Interestingly, we find that borrowing costs and deposit requirements have offsetting effects on risk. Higher interest rates are associated with lower risk, consistent with a decrease in risk-taking behavior as well as the potential impact that constrained mortgage rollover can have on market thinness. At the same time, tighter deposit requirements increase risk as the reduced availability of credit decreases the pool of potential buyers in the market. Hence, the New Zealand experience shows that macroprudential regulation can complement, if not substitute, interest rate policy in the control of housing risk. Second, information uncertainty is another important determinant of risk through its effects on the housing search and valuation process. We find that these valuation uncertainty effects operate mainly through differences in access to information across regions and relative house prices, as well as varying degrees of benchmarking difficulties across regions. Hence, houses of lower quality may be harder to value and also lead to lower mortgage credit provisions due to collateral uncertainty. Moreover, housing atypicality is also associated with shorter holding periods and the implied higher risk of idiosyncratic pricing. These findings provide useful directions for the design of regulatory policies in housing markets. Namely, improving access to information and public transparency in the valuation of houses can help mitigate the large and arguably undiversifiable idiosyncratic risks faced by homeowners.

Finally, this paper is primarily focused on the documentation of empirical evidence about systematic variation of idiosyncratic risk in housing markets. We hope the evidence that this type of risk varies simultaneously across the four dimensions of time, location, relative house price, and holding period motivates the development of richer structural models of housing markets.

Data availability

Due to license restrictions, the repeat sales data used in this project are not publicly available. We obtained these data through a data supply agreement with the Real Estate Institute of New Zealand (REINZ). The data can be acquired upon application directly to REINZ.

References

Bourassa, S. C., D. R. Haurin, J. L. Haurin, M. Hoesli, and J. Sun (2009). House price changes and idiosyncratic risk: The impact of property characteristics. *Real Estate Economics 37*(2), 259–278.

Cannon, S., N. G. Miller, and G. Pandher (2006). Risk and return in the U.S. Housing Market: A cross-sectional Asset-Pricing Approach. *Real Estate Economics 34*, 519–552.

Case, K. E. and R. J. Shiller (1989). The Efficiency of the Market for Single-Family Homes. *The American Economic Review 79*(1), 125–137.

Giacoletti, M. (2021). Idiosyncratic Risk in Housing Markets. *The Review of Financial Studies 34*(8), 3695–3741.

Glaeser, E. L., J. D. Gottlieb, and J. Gyourko (2012). Can Cheap Credit Explain the Housing Boom? In *Housing and the Financial Crisis*, NBER Chapters, pp. 301–359. National Bureau of Economic Research, Inc.

Guerrieri, V., D. Hartley, and E. Hurst (2013). Endogenous gentrification and housing price dynamics. *Journal of Public Economics 100*, 45–60.

Han, L. (2013). Understanding the Puzzling Risk-Return Relationship for Housing. *The Review of Financial Studies 26*(4), 877–928.

Harding, J. P., S. S. Rosenthal, and C. F. Sirmans (2003). Estimating bargaining power in the market for existing homes. *The Review of Economics and Statistics 85*(1), 178–188.

Hoskin, K., I. Nield, and J. Richardson (2009). The Reserve Bank's new liquidity policy for banks. *Reserve Bank of New Zealand Bulletin 72*, 5–18.

Jiang, E. X. and A. L. Zhang (2023). Collateral value uncertainty and mortgage credit provision. In *Allied Social Science Associations 2023 Annual Meeting*, New Orleans, LA. American Finance Association.

Jorda, O., K. Knoll, K. Dmitry, M. Schularick, and A. M. Taylor (2019). The rate of return on everything, 1870-2015. *The Quarterly Journal of Economics 134*(3), 1225–1298.

Landvoigt, T., M. Piazzesi, and M. Schneider (2015). The housing market(s) of San Diego. *American Economic Review 105*(4), 1371–1407.

Lu, B. (2019). Review of the Reserve Bank's loan-to-value ratio policy. *Reserve Bank of New Zealand Bulletin 82*(6), 1–24.

McDonald, C. and S. Markham (2023). Reflections on a decade of using macroprudential policy. *Reserve Bank of New Zealand Bulletin 86*, 1–19.

Murphy, L. (2011). The global financial crisis and the australian and new zealand housing markets. *Journal of Housing and the Built Environment 26*(3), 335–351.

Nield, I. (2008). Evolution of the Reserve Bank's liquidity facilities. *Reserve Bank of New Zealand Bulletin 71*.

OECD (2021). *Brick by brick: building better housing policies*. OECD Publishing, Paris.

Peng, L. and T. G. Thibodeau (2017). Idiosyncratic Risk of House Prices: Evidence from 26 Million Home Sales. *Real Estate Economics 45*(2), 340–375.

Shi, S., J.-B. Jou, and D. Tripe (2014). Can interest rates really control house prices? effectiveness and implications for macroprudential policy. *Journal of Banking Finance 47*, 15–28.

Sinai, T. (2009). Spatial Variation in the Risk of Home Owning. In E. L. Glaeser and J. M. Quigley (Eds.), *Housing Markets and the Economy: Risk, Regulation, and Policy*, Chapter 4, pp. 83–112. Lincoln Institute of Land Policy.

A Technical Appendix

A.1 Repeat sales model

Compounding of (1) provides a model for the capital gain over k periods given by

$$p_{i,t+k} - p_{i,t} = a_{r,t,k} + b_{t,k} p_{i,t} - c_{t,k} + d_{t,k} \Psi X_{i,t} + e_{i,t,k}, \tag{A.1}$$

where $a_{r,t,k}$, $b_{t,k}$ and $d_{t,k}$ are functions of the parameters of equation (1), $c_{t,k}$ is a composite of median regional prices, $\widetilde{p}_{r,t}$, and $e_{i,t,k}$ is also a composite of the original idiosyncratic shocks, $\varepsilon_{i,t}$. Specifically, it can be shown that the parameters and error term of equation (A.1) evolve with the repeat sale interval ($k > 1$) according to

$$a_{r,t,k} = a_{r,t,k-1}\left(1 + \beta_{t+k-1}\right) + \alpha_{r,t+k-1}, \tag{A.2}$$
$$b_{t,k} = b_{t,k-1}\left(1 + \beta_{t+k-1}\right) + \beta_{t+k-1}, \tag{A.3}$$
$$c_{t,k} = c_{t,k-1}\left(1 + \beta_{t+k-1}\right) + \beta_{t+k-1} \widetilde{p}_{r,t+k-1}, \tag{A.4}$$
$$d_{t,k} = d_{t,k-1}\left(1 + \beta_{t+k-1}\right) + 1, \tag{A.5}$$
$$e_{i,t,k} = e_{i,t,k-1}\left(1 + \beta_{t+k-1}\right) + \varepsilon_{i,t+k-1}, \tag{A.6}$$

where the recursions depart from $a_{r,t,1} = \alpha_{r,t}$, $b_{t,1} = \beta_t$, $c_{t,1} = \beta_t \widetilde{p}_{r,t}$, $d_{t,1} = 1$, and $e_{i,t,1} = \varepsilon_{i,t}$. Note $d_{t,k}$ is derived under the assumption that changing house features evolved according to an arithmetic average during the (unobserved) periods between repeat sales.

The coefficients α_t and β_t can be estimated from repeat sales data using non-linear estimation methods on the system of equations characterized by one equation (A.1) for each possible pair of repeat sale periods, and the restrictions imposed by equations (A.2)-(A.5). Note the number of equations in this system grows with the number of periods in the sample (T) according to the binomial coefficient formula, $T!/(T-2)!2$.

A.2 Maximum likelihood estimation

The variances of the idiosyncratic shocks in equation (1) can be estimated by maximum likelihood. Assuming $\varepsilon_{i,t} \sim N\left(0, \sigma^2_{\varepsilon,i,t}\right)$, it follows from (A.6) that $e_{i,t,k} \sim N\left(0, \sigma^2_{e,i,t,k}\right)$, where

$$\sigma^2_{e,i,t,k} = \frac{1}{\left(1 + \beta_{t+k}\right)^2} \sum_{h=0}^{k-1} \sigma^2_{\varepsilon,i,t+h} \prod_{l=h+1}^{k} \left(1 + \beta_{t+l}\right)^2. \tag{A.7}$$

Under normality, the probability density function of an estimated residual, from equation (A.1), is given by

$$f\left(e_{i,t,k}\right) = \frac{1}{\sigma_{e,i,t,k}\sqrt{2\pi}} \exp\left(-\frac{e^2_{i,t,k}}{2\sigma^2_{e,i,t,k}}\right), \tag{A.8}$$

and the corresponding log likelihood function for the sample of N estimated residuals is given by[28]

$$\ln LF\left(\left\{\sigma^2_{e,i,t,k}\right\}\right) = -\frac{N}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{N} \ln \sigma^2_{e,i,t_i,k_i} - \frac{1}{2}\sum_{i=1}^{N} \frac{e^2_{i,t,k}}{\sigma^2_{e,i,t_i,k_i}}. \tag{A.9}$$

[28] A slight abuse of notation here, where we let i index for the repeat sale observation, and t_i and k_i stand for observation i's period t and repeat sale interval k, respectively.

Estimates of $\sigma^2_{\varepsilon,t}$ can be obtained by maximizing (A.9), using (A.7) to expand $\left\{\sigma^2_{e,i,t,k}\right\}$ in terms of $\left\{\sigma^2_{\varepsilon,i,t}\right\}$. To account for potential heteroskedasticity coming from regional variation, the initial house price, and holding period, the variance of the idiosyncratic shocks can be estimated by setting $\sigma^2_{\varepsilon,i,t}$ according to (2). This only adds 2 + number of regions parameters (δ, ρ, and $\{\phi_r\}$) to the estimation of $\sigma^2_{\varepsilon,i,t}$, and a re-scaling of $\sigma^2_{e,i,t,k}$ in equation (A.7) by $\exp\left(\phi_{r_i} + \delta\ddot{p}_i + \rho\left(k_i - 1\right)\right)$.

A.3 Risk-return Calculations

A.3.1 Model-based

Model-based return estimates can be obtained from estimates of Equation 1. Focusing on predictions for median-priced houses in each region and assuming no improvement is made, the returns are given by $\hat{\alpha}_{r,t}$. To calculate annualized k-holding years returns across *time* we first accumulate the regional returns, $\hat{\alpha}_{r,t,k} = \frac{1}{k}\sum_{\tau=t}^{t+k-1}\hat{\alpha}_{r,\tau}$, and then average the regional estimates, $\hat{\alpha}_{t,k} = \sum_r w_r\hat{\alpha}_{r,t,k}$, using the share of repeat sales of each region to the total, w_r, as weights. For the annualized k-holding years returns across *locations* we average the regional returns over the years, $\hat{\alpha}_{r,k} = \frac{1}{T-k}\sum_{t=1}^{T-k}\hat{\alpha}_{r,t,k}$. For the annualized returns across *holding periods* we average the yearly averaged returns by holding period, $\hat{\alpha}_k = \frac{1}{T-k}\sum_{t=1}^{T-k}\hat{\alpha}_{t,k}$. For the annualized k-holding years returns across *initial prices* we first accumulate the regional returns factoring in the initial price deviation from the regional median, $\hat{\alpha}_{r,t,k,\ddot{p}} = \hat{\alpha}_{r,t,k} + \frac{1}{k}\hat{b}_{t,k}\ddot{p}$, where $\hat{b}_{t,k}$ is given by Equation A.3 using $\hat{\beta}_t$, and then average the regional estimates, $\hat{\alpha}_{t,k,\ddot{p}} = \sum_r w_r\hat{\alpha}_{r,t,k,\ddot{p}}$.

To obtain model-based risk estimates we use our baseline estimates of Equation 2. For risk estimates across *time* we first obtain the k-holding period regional estimates for a median-priced house, $\hat{\sigma}_{\varepsilon,t,k,r} = \sqrt{\exp\left(\hat{\phi}_r + \hat{\rho}\left(k-1\right)\right)}\hat{\sigma}_t$, and then average the regional estimates, $\hat{\sigma}_{\varepsilon,t,k} = \sum_r w_r\hat{\sigma}_{\varepsilon,t,k,r}$. For risk estimates across *locations* we average the regional estimates over the years, $\hat{\sigma}_{\varepsilon,k,r} = \frac{1}{T}\sum_{t=1}^{T}\hat{\sigma}_{\varepsilon,t,k,r}$. For risk estimates across *holding periods* we average the yearly average returns by holding period, $\hat{\sigma}_{\varepsilon,k} = \frac{1}{T-k}\sum_{t=1}^{T-k}\hat{\sigma}_{\varepsilon,t,k}$. For risk estimates across *initial prices* we first obtain the k-holding period regional estimates for a house with initial price $\ddot{p}$, $\hat{\sigma}_{\varepsilon,t,k,r,\ddot{p}} = \sqrt{\exp\left(\hat{\phi}_r + \hat{\delta}\ddot{p} + \hat{\rho}\left(k-1\right)\right)}\hat{\sigma}_t$, and then average the regional estimates, $\hat{\sigma}_{\varepsilon,t,k,\ddot{p}} = \sum_r w_r\hat{\sigma}_{\varepsilon,t,k,r,\ddot{p}}$.

A.3.2 Raw repeat sales

Raw repeat sales estimates of return and risk are obtained by averaging and taking standard deviations, respectively, directly from the repeat sales data. Let $y_{i,t,k} = \left(p_{i,t+k} - p_{i,t}\right)/k$ stand for the annualized (log) return of a house i that was bought in year t and re-sold in year $t+k$. We then average and take standard deviations of these returns according to the dimension of interest. Across *time* we take the average and standard deviation of the annualized k-years (log) returns, $\hat{a}_{t,k} = \frac{1}{N_{t,k}}\sum_i^{N_{t,k}} y_{i,t,k}$ and $\hat{\sigma}_{a,t,k} = \text{Std}\left(\left\{y_{i,t,k}\right\}_i^{N_{t,k}}\right)$, respectively, where $N_{t,k}$ indicates the set of repeat sales included in each figure. Similar expressions are used to calculate raw repeat sales return and risk estimates across *locations* and *holding periods*, only changing the set of repeat sales to $N_{r,k}$ and N_k, respectively.

A.3.3 Index-based

Index-based return and risk estimates are generally calculated departing from annualized k-years (log) returns, $z_{(r),t,k} = \frac{1}{k} \ln \left(P_{(r),t+k} / P_{(r),t} \right)$, based on national or regional averages of house prices, $P_{(r),t} = \frac{1}{N_{(r)}} \sum_i^{N_{(r)}} P_{i,t}$. The average indices are calculated using the same sample of house sales we used to obtain our baseline model estimates. For estimates across *time* index-based returns are given by $\hat{A}_{t,k} = z_{t,k}$, while risk estimates are calculated as the standard deviation of year-on-year (log) returns over k-years moving windows, $\hat{\sigma}_{A,t,k} = \text{Std}\left(\left\{ z_{t+h,1} \right\}_{h=0}^{k-1} \right)$. For estimates across *locations* returns are the regional averages of k-years (log) returns, $\hat{A}_{r,k} = \frac{1}{T-k} \sum_{t=1}^{T-k} z_{r,t,k}$, while risk estimates are the standard deviation of the same series of index-based returns, $\hat{\sigma}_{A,r,k} = \text{Std}\left(\left\{ z_{r,t,k} \right\}_{t=1}^{T-k} \right)$. For estimates across *holding periods* we average and take the standard deviation of the national annualized (log) returns by holding period, $\hat{A}_k = \frac{1}{T-k} \sum_{t=1}^{T-k} z_{t,k}$ and $\hat{\sigma}_{A,k} = \text{Std}\left(\left\{ z_{t,k} \right\}_{t=1}^{T-k} \right)$, respectively for the return and risk estimates.

B Supplementary Sample Statistics

Table B.1: Number of observations by region.

Region	Repeat sales	Houses	Re-sold 1x	Re-sold 2x	Re-sold >2x
Auckland	325,677	183,684	96,143	51,266	36,275
Bay of Plenty	70,002	38,975	19,979	10,981	8,015
Canterbury	140,602	79,479	41,710	22,195	15,574
Gisborne	8,897	4,614	2,148	1,332	1,134
Hawke's Bay	36,375	19,932	9,972	5,766	4,194
Manawatu-Wanganui	60,777	32,407	15,489	9,381	7,537
Marlborough	13,655	7,105	3,352	2,006	1,747
Nelson	14,361	7,823	3,932	2,195	1,696
Northland	25,746	14,767	8,005	3,968	2,794
Otago	59,688	32,025	15,798	8,948	7,279
Southland	33,020	15,535	6,392	4,315	4,828
Taranaki	29,525	15,572	7,479	4,385	3,708
Tasman	8,802	4,903	2,520	1,375	1,008
Waikato	101,635	54,822	27,039	15,494	12,289
Wellington	124,574	68,836	34,941	19,353	14,542
West Coast	5,055	3,082	1,780	833	469
Total	1,058,391	583,561	296,679	163,793	123,089

Table B.2: Summary statistics of property characteristics.

	Mean	Std. dev.	Min.	Max.	N
Bedrooms	3.064	0.819	1	9	1,054,103
Bathrooms	1.432	0.693	0	6	541,315
Garages	1.464	0.871	0	7	1,026,847
Floor area (m^2)	143.106	65.163	1	885	1,029,021

Table B.3: Descriptive statistics of explanatory variables.

Variable	Mean	Std. dev.	Min.	Max.	N
Interest rate$^{(a)}$	0.050	0.023	0.004	0.094	7,088,089
	[0.048]	[0.025]	[0.004]	[0.094]	[29]
Loan-to-value ratio$^{(a)}$	0.951	0.099	0.700	1.000	7,088,089
	[0.940]	[0.109]	[0.700]	[1.000]	[29]
Internet access, 2006, region$^{(b)}$	0.607	0.048	0.466	0.656	7,088,089
	[0.568]	[0.051]	[0.466]	[0.656]	[16]
Internet access, 2006, SA2$^{(b)}$	0.605	0.131	0.183	1.000	6,785,587
	[0.605]	[0.138]	[0.183]	[1.000]	[2,125]
Total population, national (million)$^{(b)}$	4.198	0.338	3.538	5.094	7,088,089
	[4.205]	[0.435]	[3.538]	[5.094]	[29]
Total population, region (million)$^{(b)}$	0.669	0.523	0.032	1.436	7,088,089
	[0.274]	[0.340]	[0.032]	[1.436]	[16]
New build consent area, national (million)$^{(b)}$	4.255	0.962	2.618	6.158	7,088,089
	[4.291]	[1.040]	[2.618]	[6.158]	[29]
New build consent area, region (million)$^{(b)}$	0.659	0.576	0.010	2.473	7,088,089
	[0.268]	[0.387]	[0.010]	[2.473]	[464]
Average house atypicality$^{(c)}$	0.758	0.037	0.637	0.919	6,834,835
N. Sales, national (thousand)$^{(c)}$	82.375	17.089	55.505	119.701	7,088,089
N. Sales, region (thousand)$^{(c)}$	12.785	10.682	0.230	39.798	7,088,089
N. Sales, SA2$^{(c)}$	54.255	28.121	1.000	342.000	6,778,603

Sources: ^{a}Reserve Bank of New Zealand; ^{b}Stats New Zealand; ^{c}Author's construction based on data from the Real Estate Institute of New Zealand. All variables are matched to the repeat sales data extended to cover every period between purchase and sale. Original statistics for the variables not derived directly from the repeat sales data are presented between brackets.

C Supplementary Results

Table C.1: Model estimates of β_t.

t	(1)	(2)	(3)	(4)
1992	-2.868	-2.896	-2.161	-1.979
	(0.427)	(0.431)	(0.738)	(0.608)
1993	-1.220	-1.182	-0.560	-0.896
	(0.339)	(0.344)	(0.610)	(0.519)
1994	-3.772	-3.713	-3.593	-3.938
	(0.310)	(0.314)	(0.552)	(0.460)
1995	-4.628	-4.619	-4.468	-4.090
	(0.318)	(0.321)	(0.569)	(0.456)
1996	-5.118	-5.045	-4.241	-5.085
	(0.315)	(0.318)	(0.557)	(0.433)
1997	0.631	0.730	0.805	0.705
	(0.358)	(0.363)	(0.619)	(0.462)
1998	1.083	1.083	0.998	1.053
	(0.375)	(0.377)	(0.647)	(0.475)
1999	0.944	0.893	1.281	0.946
	(0.371)	(0.371)	(0.654)	(0.464)
2000	1.363	1.370	0.949	1.216
	(0.346)	(0.346)	(0.607)	(0.445)
2001	0.198	0.101	1.119	-0.047
	(0.285)	(0.285)	(0.491)	(0.382)
2002	-1.969	-1.945	-1.450	-1.449
	(0.222)	(0.223)	(0.383)	(0.333)
2003	-6.793	-6.794	-7.331	-6.929
	(0.190)	(0.191)	(0.306)	(0.303)
2004	-9.002	-9.058	-9.223	-9.304
	(0.188)	(0.190)	(0.295)	(0.296)
2005	-6.984	-7.230	-7.450	-6.375
	(0.206)	(0.210)	(0.313)	(0.316)
2006	-5.366	-5.560	-6.136	-5.043
	(0.235)	(0.241)	(0.333)	(0.331)
2007	0.400	0.461	0.678	0.176
	(0.345)	(0.355)	(0.461)	(0.423)
2008	2.462	2.378	2.878	2.405
	(0.384)	(0.396)	(0.506)	(0.456)
2009	1.493	1.565	1.785	1.420
	(0.377)	(0.390)	(0.518)	(0.440)
2010	0.155	0.132	-0.004	0.047
	(0.382)	(0.395)	(0.507)	(0.450)
2011	-0.525	-0.472	-0.332	-0.566
	(0.350)	(0.364)	(0.438)	(0.423)
2012	-1.007	-0.960	-1.545	-0.947
	(0.322)	(0.338)	(0.392)	(0.397)
2013	-0.796	-0.819	-0.883	-0.694
	(0.308)	(0.326)	(0.369)	(0.397)
2014	-2.612	-2.462	-3.208	-2.630
	(0.302)	(0.318)	(0.359)	(0.395)
2015	-2.969	-2.838	-2.875	-2.844
	(0.283)	(0.294)	(0.344)	(0.372)
2016	-4.034	-4.206	-5.133	-4.064
	(0.308)	(0.311)	(0.374)	(0.394)
2017	-5.459	-5.420	-6.217	-5.536
	(0.336)	(0.336)	(0.387)	(0.424)
2018	-6.822	-6.832	-7.693	-6.675
	(0.352)	(0.352)	(0.393)	(0.442)
2019	-3.261	-3.223	-4.178	-3.227
	(0.395)	(0.395)	(0.422)	(0.509)
2020	-2.288	-2.244	-3.866	-1.761
	(0.453)	(0.453)	(0.477)	(0.591)
Sample	All	Excluding new dwelling	Excluding missing	Excluding hold < 2 years
R-squared	0.760	0.761	0.776	0.773
N. Obs.	1,058,391	1,026,613	522,616	947,089

Notes: These are estimates of β_t from equation 1. Robust standard errors are presented between brackets.

Table C.2: Estimates of variances of idiosyncratic shocks.

t	(1)	(2)	(3)	(4)
1992	1.788	1.731	1.631	1.695
	(0.062)	(0.062)	(0.096)	(0.234)
1993	1.663	1.634	1.576	1.534
	(0.048)	(0.048)	(0.074)	(0.191)
1994	1.593	1.540	1.488	1.413
	(0.043)	(0.042)	(0.063)	(0.142)
1995	1.534	1.483	1.384	1.584
	(0.041)	(0.040)	(0.063)	(0.119)
1996	1.432	1.397	1.323	1.203
	(0.037)	(0.036)	(0.060)	(0.099)
1997	1.490	1.462	1.397	1.419
	(0.042)	(0.042)	(0.064)	(0.100)
1998	1.318	1.287	1.198	1.077
	(0.040)	(0.039)	(0.059)	(0.100)
1999	1.599	1.550	1.593	1.508
	(0.046)	(0.046)	(0.080)	(0.103)
2000	1.466	1.450	1.422	1.156
	(0.043)	(0.043)	(0.069)	(0.105)
2001	1.780	1.753	1.699	1.762
	(0.047)	(0.047)	(0.073)	(0.106)
2002	2.075	2.051	1.992	2.018
	(0.044)	(0.044)	(0.068)	(0.097)
2003	1.688	1.675	1.613	1.563
	(0.033)	(0.033)	(0.047)	(0.086)
2004	1.498	1.469	1.397	1.435
	(0.030)	(0.030)	(0.043)	(0.073)
2005	1.363	1.336	1.309	1.304
	(0.029)	(0.029)	(0.043)	(0.071)
2006	1.173	1.146	1.087	1.010
	(0.025)	(0.025)	(0.034)	(0.065)
2007	1.480	1.438	1.441	1.190
	(0.033)	(0.033)	(0.045)	(0.068)
2008	1.285	1.258	1.220	1.127
	(0.033)	(0.033)	(0.041)	(0.079)
2009	1.226	1.189	1.121	1.045
	(0.037)	(0.037)	(0.045)	(0.087)
2010	1.282	1.271	1.272	1.229
	(0.042)	(0.043)	(0.051)	(0.095)
2011	1.206	1.198	1.202	0.953
	(0.038)	(0.039)	(0.046)	(0.096)
2012	1.369	1.352	1.360	1.275
	(0.041)	(0.042)	(0.049)	(0.091)
2013	1.391	1.368	1.430	1.188
	(0.041)	(0.042)	(0.050)	(0.091)
2014	1.425	1.396	1.453	1.298
	(0.037)	(0.038)	(0.047)	(0.083)
2015	1.337	1.318	1.371	1.097
	(0.032)	(0.032)	(0.037)	(0.072)
2016	1.246	1.221	1.204	1.094
	(0.034)	(0.033)	(0.040)	(0.072)
2017	1.199	1.180	1.163	1.120
	(0.035)	(0.035)	(0.038)	(0.072)
2018	1.164	1.144	1.148	0.948
	(0.034)	(0.033)	(0.034)	(0.067)
2019	1.447	1.429	1.357	1.357
	(0.041)	(0.040)	(0.039)	(0.069)
2020	1.639	1.611	1.474	1.563
	(0.061)	(0.061)	(0.051)	(0.097)
$\hat{\delta}$	-0.278	-0.286	-0.332	-0.275
	(0.008)	(0.008)	(0.011)	(0.008)
$\hat{\rho}$	-0.156	-0.152	-0.192	-0.117
	(0.003)	(0.003)	(0.005)	(0.003)
Sample	All	Excluding new dwelling	Excluding missing	Excluding hold < 2 years
Log likelihood (avg.)	0.189	0.193	0.257	0.142
N. Obs.	1,058,391	1,026,613	522,616	947,089

Notes: These are estimates of $\exp(\phi_{Akl})\sigma_t^2$ (scaled by 100x), δ and ρ from equation 2. Note the estimates are scaled to Auckland region; see Table C.3 for the other regions' estimates of ϕ_r. Robust standard errors are presented between brackets.

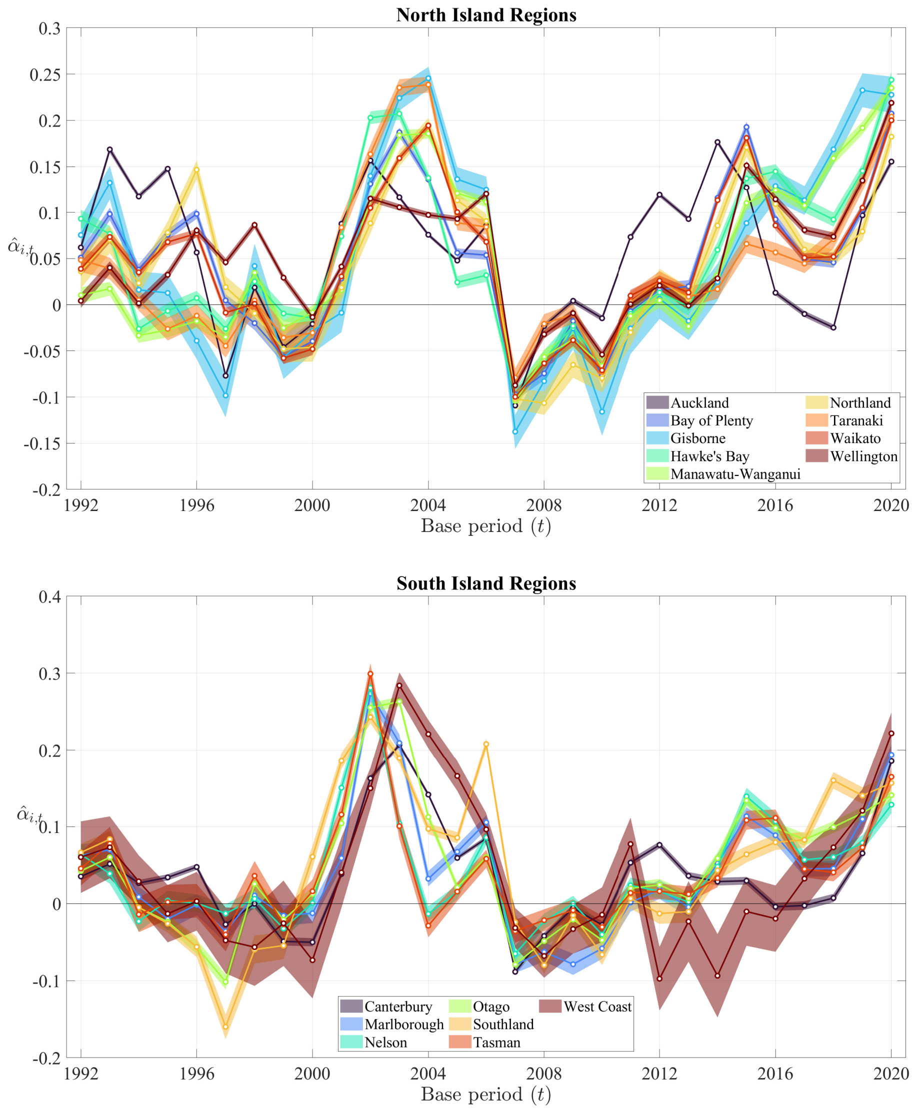
Table C.3: Estimates of regional shifts on variances of idiosyncratic shocks.

Region	(1)	(2)	(3)	(4)
Auckland	0	0	0	0
	–	–	–	–
Bay of Plenty	-0.121	-0.114	-0.096	-0.135
	(0.014)	(0.014)	(0.019)	(0.015)
Canterbury	-0.166	-0.166	-0.078	-0.179
	(0.011)	(0.011)	(0.016)	(0.012)
Gisborne	-0.046	-0.036	-0.064	-0.076
	(0.031)	(0.031)	(0.034)	(0.035)
Hawke's Bay	-0.267	-0.260	-0.220	-0.284
	(0.018)	(0.019)	(0.026)	(0.021)
Manawatu-Wanganui	-0.048	-0.045	-0.028	-0.046
	(0.015)	(0.015)	(0.020)	(0.017)
Marlborough	-0.462	-0.460	-0.462	-0.481
	(0.034)	(0.035)	(0.050)	(0.039)
Nelson	-0.607	-0.599	-0.600	-0.647
	(0.031)	(0.031)	(0.043)	(0.035)
Northland	-0.009	-0.006	-0.174	-0.019
	(0.021)	(0.021)	(0.027)	(0.023)
Otago	0.115	0.120	0.100	0.112
	(0.014)	(0.014)	(0.020)	(0.015)
Southland	0.217	0.217	0.242	0.227
	(0.018)	(0.018)	(0.027)	(0.020)
Taranaki	0.107	0.101	0.023	0.118
	(0.018)	(0.018)	(0.024)	(0.020)
Tasman	-0.540	-0.543	-0.585	-0.544
	(0.041)	(0.041)	(0.052)	(0.045)
Waikato	-0.163	-0.156	-0.208	-0.166
	(0.012)	(0.012)	(0.017)	(0.013)
Wellington	-0.271	-0.266	-0.297	-0.295
	(0.013)	(0.013)	(0.018)	(0.014)
West Coast	0.443	0.445	0.417	0.469
	(0.049)	(0.050)	(0.041)	(0.051)
Sample	All	Excluding new dwelling	Excluding missing	Excluding hold < 2 years
N. Obs.	1,058,391	1,026,613	522,616	947,089

Notes: These are estimates of ϕ_r from equation 2 and taking Auckland as the base of reference. Robust standard errors are presented between brackets.

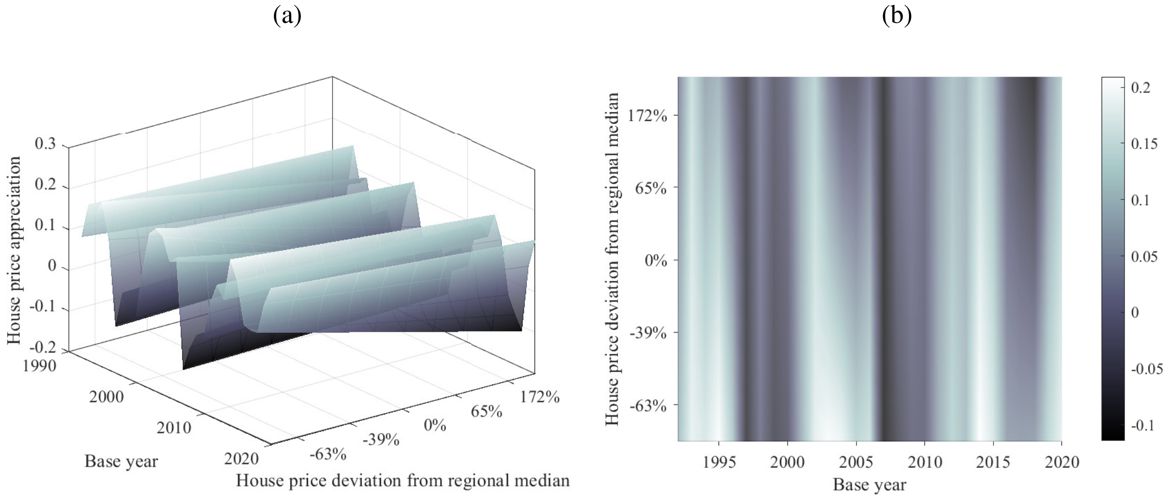
D Additional Figures

Figure D.1: Regional capital gains estimates.



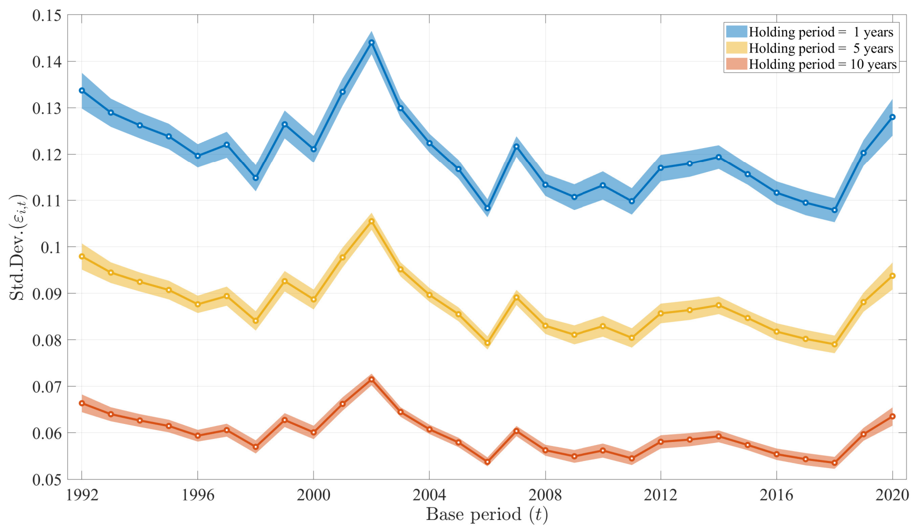
Notes: The lines depict estimates of the average (log) annual capital gains for a median house in the corresponding region. The shaded areas indicate the corresponding 90% confidence intervals.

Figure D.2: Predicted house price appreciation for Auckland.



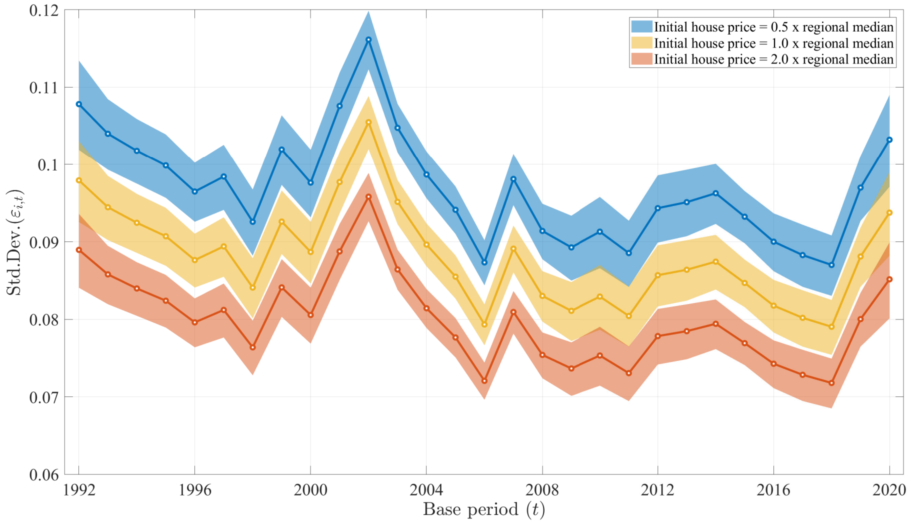
Notes: Predictions based on estimates of $\alpha_{Akl,t}$ and β_t from Equation 1 assuming no changes in house features. Panel (b) is the top view of (a).

Figure D.3: Idiosyncratic risk estimates by holding period.



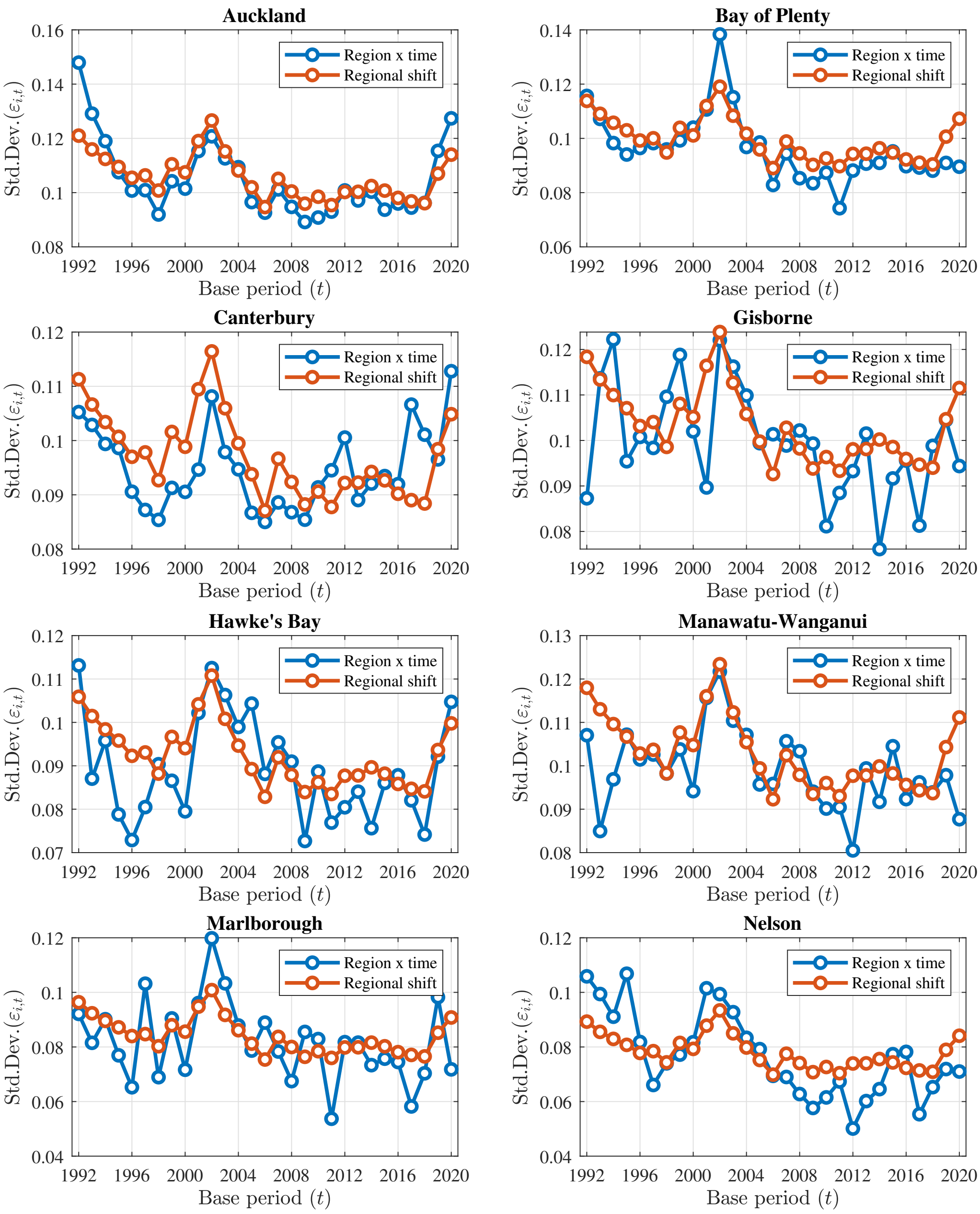
Notes: These are obtained by maximum likelihood from the residuals of the repeat sales model, assuming the idiosyncratic shocks are distributed as a Normal with variance given by equation 2. The estimates are for a median-priced house in Auckland Region. The confidence bands are based on robust standard errors and calculated at a 90% confidence level.

Figure D.4: Idiosyncratic risk estimates by initial house price.



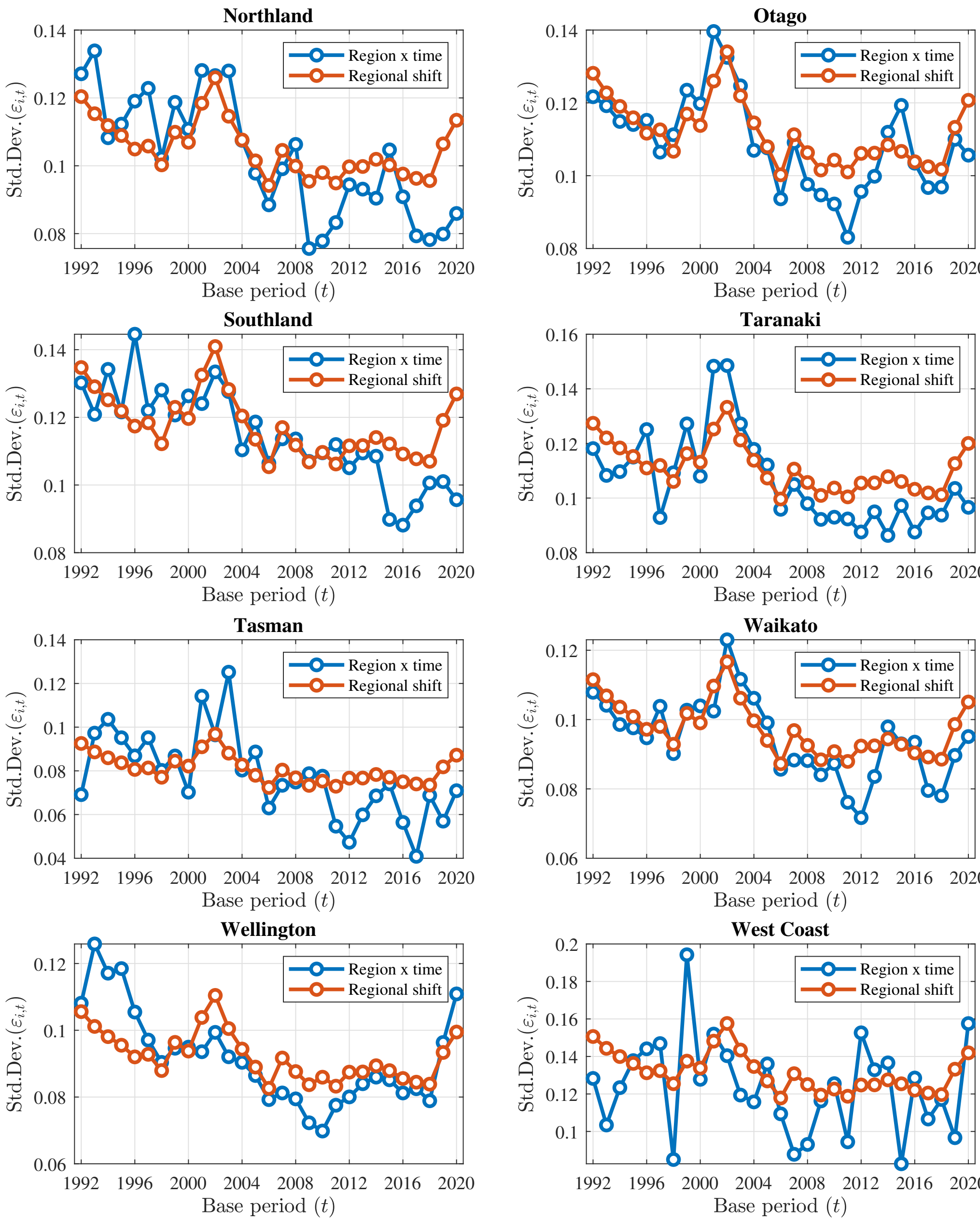
Notes: These are obtained by maximum likelihood from the residuals of the repeat sales model, assuming the idiosyncratic shocks are distributed as a Normal with variance given by equation 2. The estimates account for the effect of the initial house price, relative to the regional median. The estimates are for Auckland Region and holding period of five years between repeat sales. The confidence bands are based on robust standard errors and calculated at a 90% confidence level.

Figure D.5: Comparative of regional risk estimates (1/2).



Notes: The plots compare implied risk estimates for each region allowing for these to vary by region and time (blue lines) versus only by a regional shift (orange).

Figure D.6: Comparative of regional risk estimates (2/2).



Notes: The plots compare implied risk estimates for each region allowing for these to vary by region and time (blue lines) versus only by a regional shift (orange).