



BUSINESS SCHOOL
Te Kura Pakihi

ISSN 1178-2293 (Online)

University of Otago
Economics Discussion Papers
No. 1714

DECEMBER 2017

Geographic variation in intra-city house price appreciation over the boom-bust cycle: evidence from Auckland, NZ

Leon Stirk-Wang and Paul Thorsnes
University of Otago

Address for correspondence:

Paul Thorsnes
Department of Economics
University of Otago
PO Box 56
Dunedin
NEW ZEALAND
Email: paul.thorsnes@otago.ac.nz
Telephone: 64 3 479 8359

Geographic variation in intra-city house price appreciation over the boom-bust cycle: evidence from Auckland, NZ

Leon Stirk-Wang and Paul Thorsnes
University of Otago

20 December 2017

1. Introduction

This study contributes to the empirical literature on intra-urban house-price dynamics. A demand shock affects metropolitan-area house prices, but the effect of the shock varies across neighborhoods. For example, in a relatively recent paper, Guerrieri et al. (2013) provide evidence that the variation in house-price appreciation across the census tracts in relatively large US cities is greater than the variation across cities overall; the intra-city standard deviation in house prices is about 0.5 relative to the inter-city standard deviation of about 0.2. Of interest is why this happens. More specifically, this paper contributes to the literature that addresses the question of how and why the effects of a demand shock flow through the metro-area housing market.

The key data needed for this kind of analysis consist of price indices at relatively small geographic scales. Guerrieri et al., for example, use annual Case-Shiller repeat-sales indices at the zip code level. They also use much less frequent information from the decennial census at the smaller census tract level. In this study we use observations on estimates of median house value on 1 July in each of 111 census ‘area units’ – similar to US census tracts – in Auckland, New Zealand from year 2000 through 2016. This gives us relatively high spatial resolution and, conveniently, this time period covers a general boom, a mild bust, and then another boom in house prices.

A novel methodologic aspect of this study is that we use a cluster analysis technique to identify four clusters of census area units in each of which house values follow a similar pattern of price appreciation over the 15-year time period. That is, we let the variation in the trends in the house value data at a relatively small geographic scale reveal spatial relationships and the propagation of price shocks over time across the urban area.

Not surprisingly, area units with similar price paths tend to cluster geographically. Or, to put it another way, areas that are similar in relevant aspects are large relative to census area units. In terms of the most easily observed of these aspects, the clusters are distinguished by house price and distance from the CBD. In order of house price and distance, the four clusters are: (1) high-price, centrally-located; (2) middle-high price, close-in; (3) several distinct middle-price suburban areas, and (4) a lower-price, more distant suburban area.

Of interest is that during the boom of the early 2000s, rates of price appreciation appear to flow over time from high-priced centrally located neighborhoods progressively through the house-price/distance tiers. This seems sensible in that the boom resulted from rapid growth in ability to pay for houses: the most desirable areas would be targeted first, others would follow motivated by changes in relative prices. Appreciation rates appear least volatile in the suburban middle-priced neighborhoods, consistent with a relatively elastic supply response. They appear most volatile in the distant lower-priced cluster, perhaps reflecting modest boom-time gentrification and subsequent bust-time distress.

The remainder of the paper is organized as follows. Section 2 develops the conceptual foundation in the context of the existing literature. Section 3 describes the Auckland context. Section 4 describes the data. Section 5 cluster methodology and the resulting clusters. Section 6 reports statistical analyses of the distinguishing characteristics of each cluster and of the propagation of price shocks through the house-price/distance tiers. Section 7 concludes the paper.

2. Conceptual foundations

Well established by now is that land and house prices vary across an urbanized area. A typical empirical hedonic house-price function includes, in addition to structural characteristics of the house and lot, measures that vary over space, e.g., measures of location relative to employment centers and measures of neighborhood characteristics that influence local public goods. The theoretical foundations for systematic intra-urban variation in house prices are the Alonso-Muth-Mills monocentric-city model and the Tiebout model of spatial household sorting. Results reported in probably hundreds of cross-sectional analyses of sale prices verify that sale prices vary systematically across an urbanized area.

The question that arises is: Do *rates of change* in house prices over time also vary systematically across an urbanized area? On the face of it, a house and lot are effectively capital assets. If asset markets are efficient and asset buyers rational, well-informed, and risk-neutral, then we would expect the market to yield equal total returns to housing equity, all else constant. If rates of house-value appreciation vary spatially, then there must be some unanticipated shocks that affect parts of an urban area more than others and/or there is at least some delay in the reactions of market participants. The existing literature consists either of purely empirical estimates of spatial variation in rates of house-price appreciation or of a description of a type of shock and an empirical test usually with results consistent with the impacts of the described shock.

Archer et al. (1996) argue that the foundations of standard empirical hedonic model specifications for investigating intra-urban variation in house sale prices can also provide rationales for intra-urban variation in rates of house-price appreciation. In the basic monocentric city model, for example, an improvement to a radial transportation system or a general increase in incomes should lead to a relative increase (faster appreciation) in land values farther from the city center. A demand shock to a particular household demographic type

affects demand for house sizes and neighborhood amenities, which affect relative appreciation rates. And systematic spatial variation in the characteristics of existing houses can affect appreciation rates: e.g., the newer houses in the suburbs depreciate faster than older houses closer in, which slows observed appreciation rates in the suburbs.

Archer et al. use a standard repeat sales approach to estimate appreciation rates for each of 78 census tracts independently from 1971 to 1992 in the Miami, Florida MSA; only sales of houses within each census tract contribute to the tract's price index. However, the sample consists only of those houses that sold at least twice, which may not be representative of the population of houses and reduces the precision of the estimates. Appreciation rates for roughly half of these census tracts differ significantly from the overall urban-area rate. They conclude that "house price appreciation appears to vary spatially; that is, it varies by municipality and with distance from the CBD, with the level of house prices, as well as with local changes in population, housing units, and in ethnic mix."

Interestingly, Geotzmann and Spiegel (1997) and McMillen (2003) develop *distance weighted* repeat sales approaches that treat the imprecision in the Archer et al. estimates. McMillen finds higher appreciation rates from 1990 to 1996 in low-income census tracts close to the Chicago city center. Geotzmann and Spiegel find "widely differing rates of return" across San Francisco Bay Area zip codes over the period 1980 to 1994. Zip codes with lower household incomes and better educated households had better returns. They also find that "contiguous zip codes typically 'lump' together in the same capital appreciation quartile, due to continuities in socioeconomic variables within the region".

Some papers focus on the *propagation* of house price appreciation across an urban area. Clapp and Tirtiroglu (1994) find empirical support for the hypothesis that a demand-induced increase in sale prices in one area provides local information that then diffuses or propagates across the larger urban area; when house prices increase in one neighborhood, house prices in

neighboring areas rise with a lag of about a year. More recently, Ho et al. (2008) report evidence that a demand shock specifically to the low-quality end of an urban-area housing market flows upward over time through the quality tiers. Guerrieri et al. (2013) provide evidence from Chicago that lower-priced housing near higher-priced neighborhoods tends to appreciate faster in response to a positive demand shock, consistent with gentrification.

Bostic et al. (2007) argue that house-price appreciation rates following a demand shock should be positively correlated with the value of land as a proportion of the market value of the bundle of land and structure. The argument is that because urban land is more inelastic in supply than are structures, most of the long-run change in the market value of the house-land bundle is in the price of land. The implication is that the effect of a given percentage shock to land values will have a smaller effect on the market value of properties in which the value of land is a smaller proportion of the value of the house-land bundle. Empirical support is provided using data from Wichita Kansas. Davis et al. (2017), however,

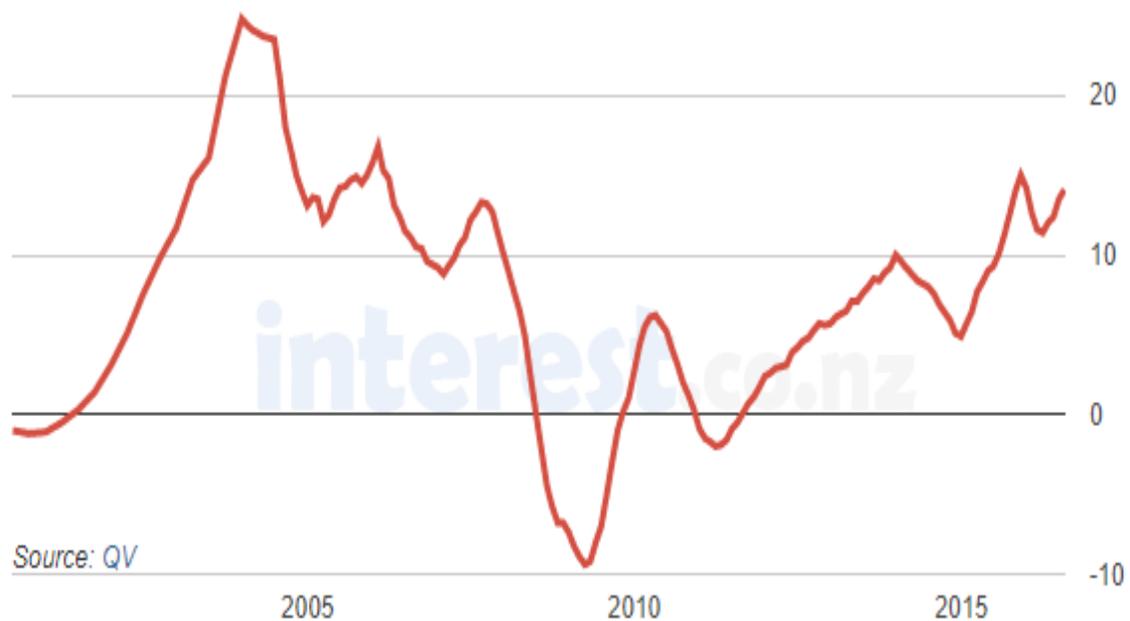
Genesove and Han (2013) provide evidence that house prices tend to appreciate more slowly in suburban than in central areas and interpret this as a result of higher housing supply elasticities in areas with easier opportunities for development. In nation-wide samples, Ferreira and Gyourka (2012) also find central areas appreciating faster, and Glaeser et al. (2012) note that central areas with low-income households appreciate faster.

Finally, the recent period of demand-shock boom and bust in housing markets around the world has attracted attention. Liu et al. (2014), for example, report that relative prices across house-size tiers in Phoenix, Arizona were stable during the mid-2000s boom, but the prices of smaller houses depreciated faster during the subsequent bust, consistent with more distressed sales. In contrast, Waihl (forthcoming) reports that appreciation rates in Sydney, Australia over the recent boom-bust cycle were highest for relatively low-priced housing in suburban areas and lowest for high-priced inner-city housing.

3. Auckland in the New Zealand context

We observe house prices over the boom-bust-boom period from year 2000 to 2016. Figure 1 shows the trend in the average annual growth rate in house prices for the nation overall. New Zealand in year 2000 was gradually recovering from a recession in the late 1990s caused by the financial crisis in NZ's Asian trading partner countries. Once the recovery started, however, it proceeded with vigor, which is reflected in house prices during the early mid-2000s. The global financial crisis affected NZ in the late 2000s. The economy recovered quickly, however, aided in part by government spending in response to the earthquake in February 2011 that destroyed the Christchurch central business district and destroyed or damaged infrastructure and a large number of houses. The economy has since performed well.

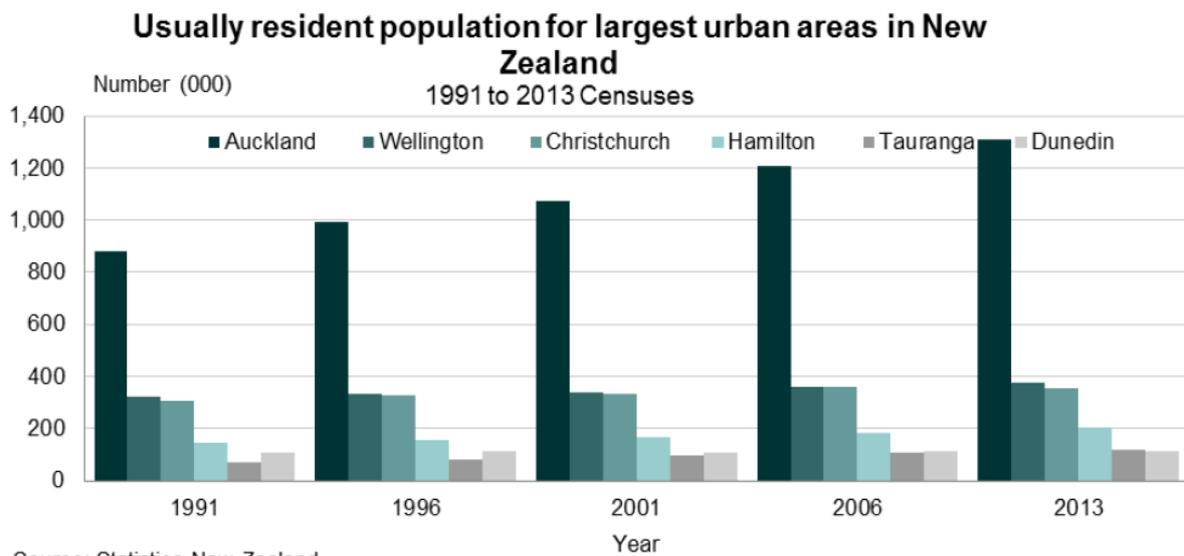
Figure 1. Average percent change in NZ house prices.



The population of NZ has grown over the sample period from about 3.9 to about 4.7 million people, or by just over 20%. The population of the urbanized area of Auckland grew correspondingly from 1.17 to nearly 1.5 million, or by about 33%. The likely reason for Auckland's relatively fast population growth is agglomeration economies. Auckland is by far

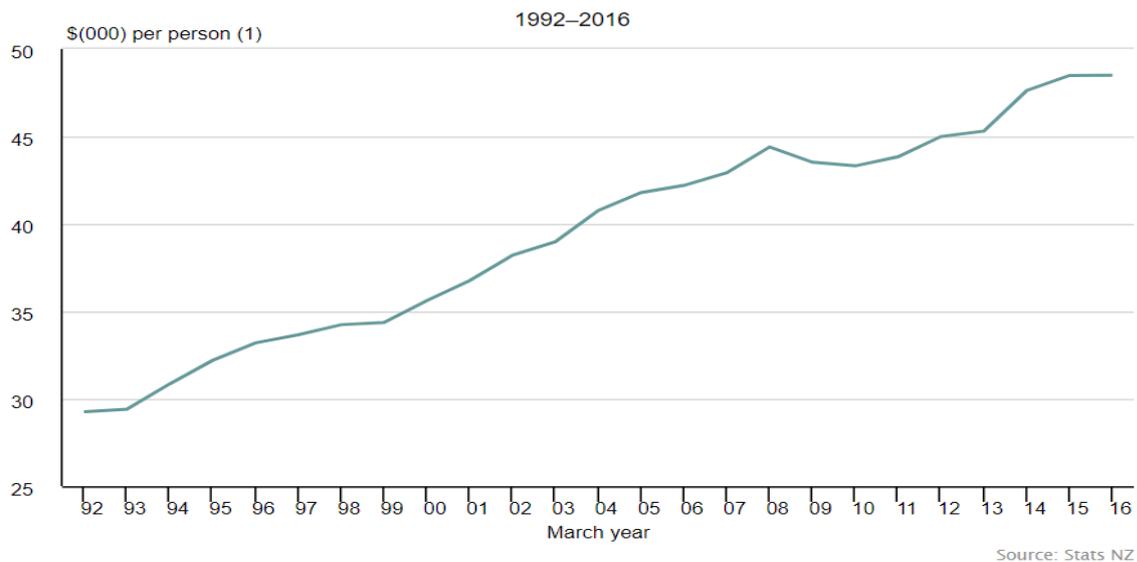
NZ's largest city, with national capital and the North Island's southern regional service center Wellington currently in second place at about 470,000 and the South Island's largest city and central regional service center, Christchurch, in third place at about 370,000. Auckland is where the employment opportunities are greatest and where the bulk of growth has occurred among NZ's larger urban areas as shown in Figure 2 below.

Figure 2. Growth in population of NZ's six largest urban areas



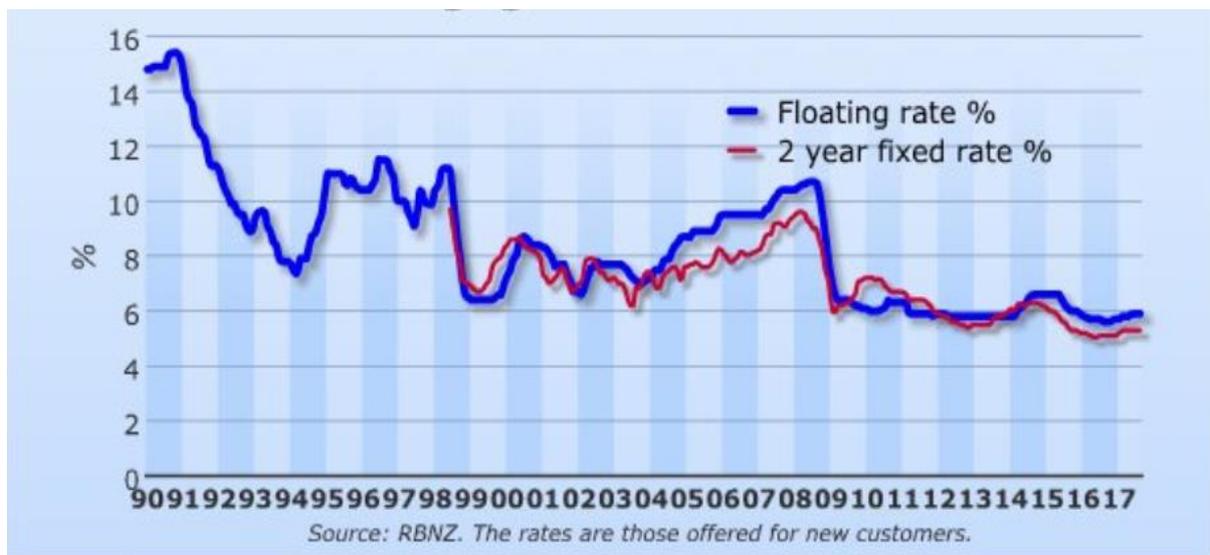
More people implies greater demand for housing. However, the purchasing power of these people has also increased markedly. Figure 3 shows that real disposable income per person rose by about 33% from 2000 to 2016. This large increase in income came in the wake of large-scale liberalization of the NZ economy in the late 1980s and corresponds with a more than 40% increase in the number of people employed.

Figure 3. Real disposable income per person



In addition, as shown in Figure 4, interest rates have recently been at historic NZ lows. Home buyers in NZ can take usually up to a 30-year mortgage, but common practice is to fix the interest rate for periods of less than five years, often two years or floating as shown in the Figure. As elsewhere, banks in NZ were generous in their lending in an environment of rising property prices in the early-mid 2000s boom. Mortgage rates, however, rose with restrictive monetary policy and then fell with easing of monetary policy in the wake of the GFC.

Figure 4. Trend in mortgage interest rates



4. Data

The unit of analysis in this study is the NZ census “area unit”, similar to a US census tract, with an average population of about 4000. Census area units have a couple of advantages as a unit of observation. The first is that detailed data from the census of households are reported at the area-unit level. The census normally is taken at 5-year intervals in New Zealand. In our time frame, censuses were taken in 2001 and 2006. Unfortunately the census scheduled for March of 2011 was not taken due to disruption caused by the damaging earthquake in Christchurch in February 2011. That census went ahead in March of 2013, with the next one scheduled for March 2018. So, we have data from three censuses.

The second advantage is that area units are geographically relatively small in areas such as Auckland that are developed at urban densities. Yet they are sufficiently large for observations of house sales that contribute to estimation of price indices. To help ensure sufficient numbers of sales, our sample consists of the 111 area units that contain at least 400 detached single family dwellings as identified in the most recent (2013) census of population and that were reasonably fully developed prior to the year 2000 (we excluded from the sample 8 distant suburban area units that developed over the course of the sample time period). This set of census tracts includes those areas that are normally thought of as single-family residential areas. Excluded are area units with large amounts of commercial and multi-family residential development.

Census tract price indices

Our area-unit price indices consist of an estimate of the median house value in each of the sample area units as at 1 July each year from year 2000 through 2016. The data come from CoreLogic New Zealand, the leading commercial supplier of property information and analysis in New Zealand. CoreLogic maintain an “automated valuation model” based on “recent, nearby comparable sales in the area” to value every house in Auckland once a month, thereby

taking advantage of the most recent sales. Thus, every house in every area unit, regardless of whether or not it sold, receives an updated market value each month. The data in our sample consist of this median estimated house value at mid-year for each of seventeen years, which yields estimates of the median annual percent increase in median area-unit house value.

There clearly can be concerns about these data. To start, the nature of the “model” that analysts at CoreLogic use to estimate house values is not described in detail on the company website, presumably due to commercial sensitivity. The US and Australia sites, however, discuss hedonic valuation models, but again not in detail, so at this stage we do not know the type or specification of the model used by CoreLogic New Zealand. Thus, it is difficult to evaluate the quality of the price indices.

To be fair, however, estimating price indices at small geographic scales is difficult. McMillen (2012), for example, nicely describes issues in estimating price indices for small geographies using both hedonic and repeat-sales approaches. In the present sample, the estimate is of the median of the population of houses in the area unit rather than of those that sold, and so is representative of the characteristics of the population of houses. The estimates cannot be regarded as a constant-quality index as the trend in prices will reflect trends in both depreciation and renovation in the area unit. And the estimates are influenced by sales nearby but outside the area unit, which will tend to blur any boundaries between neighborhoods identified by trends in prices. However, census area unit boundaries were not chosen to reflect differences in house values or appreciation rates. As a practical matter, CoreLogic claims reasonable levels of predictive accuracy and these estimated house values are widely used by local governments in New Zealand for property tax assessment. As an additional practical matter, using these estimated median values is a cost-effective way to explore the usefulness of our cluster analysis methodology; we avoid the high cost of purchasing information on a very large number of transactions.

Figure 6 shows the trend in the sample average census tract median appreciation rates over the sample period. The trend is similar to that shown for NZ overall in Figure 1. Prices rise considerably prior to the effects of the GFC, fall modestly for a couple of years, and then rise increasingly quickly in the early 2010s. The effects of the GFC were relatively brief.

Figure 7 superimposes each of the individual 111 area-unit trends in annual price-appreciation rates. Not surprisingly, the overall trend is similar to that shown in Figure 6. However, there is clearly significant variation in appreciation rates in any given year. Moreover, there appears to be considerable variation in the timing of relative peaks and troughs in the trends in appreciation rates; prices appear not to rise and fall tightly in tandem across the area units in urban area. Less clear from this superimposition is the extent to which the sizes of the peaks and troughs in appreciation rates vary across the area units.

5. Cluster analysis

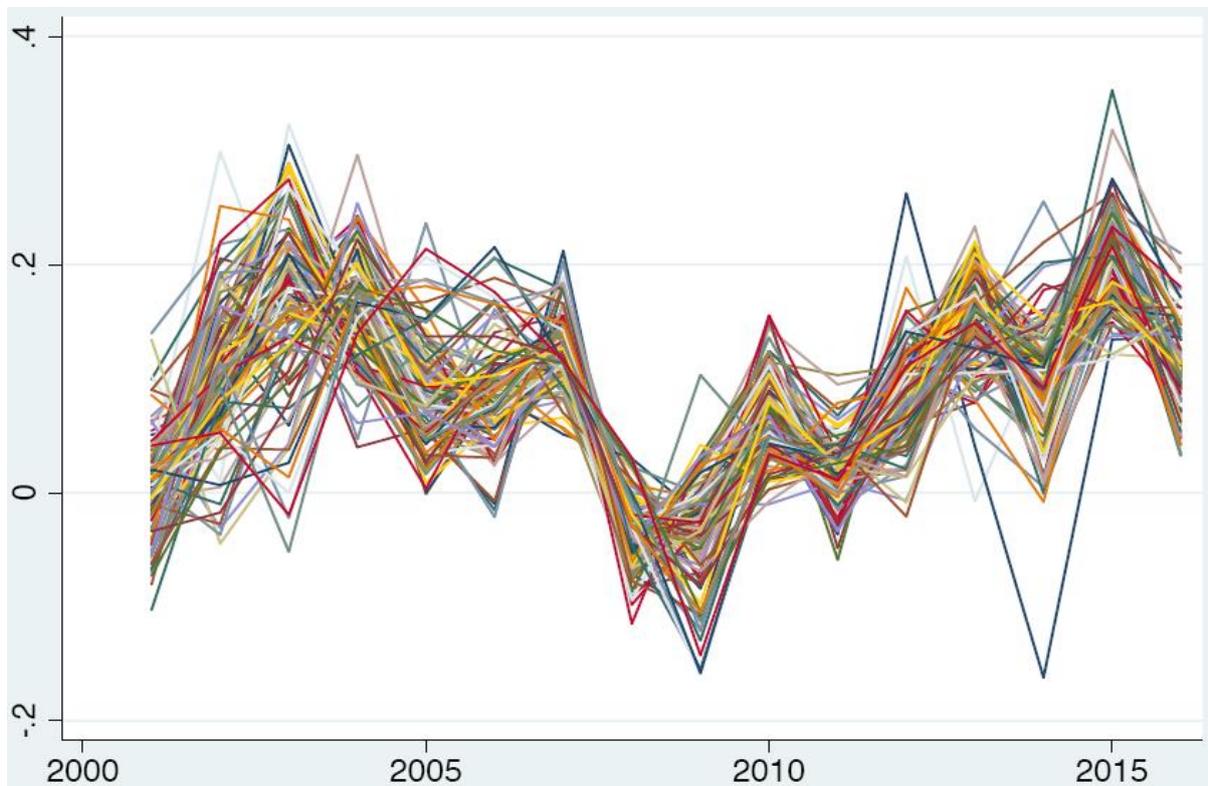
Our approach is to identify ‘clusters’ of area units with similar patterns in relative price appreciation over the entire time period. If Auckland is typical of the other metropolitan areas studied in the literature, rates of price appreciation vary across Auckland ‘submarkets’, where a submarket is defined as an area across which prices appreciate at similar rates. Our unit of analysis is an area unit, and we expect that area units are small relative to the size of submarkets as defined, though this is an empirical question. The clustering exercise, then, can be thought of as identifying submarkets as groups of area units with similar patterns of annual price appreciation rates from year 2000 through 2016. This subsection reports the results of this cluster analysis.

We use the k-means clustering routine in Stata statistical software. The ‘k’ in k-means stands for the number of clusters, which is chosen by the researcher. The clustering routine starts by choosing the 16-year price-appreciation vectors of k participants at random from the

Figure 6. Trend in the mean of the sample annual price appreciation rates



Figure 7. Variation across area units in the trend in annual price appreciation



111 area units in the sample and using these vectors as the initial centroids of the k clusters. Each of the remaining $n-k$ area units are then assigned to the cluster with the centroid closest in Euclidean distance to their own 16-year price-appreciation vector. From there the routine iterates: once all area units are assigned to one of the k clusters, the vector of cluster-average values becomes the new centroid and each area unit is again allocated to the cluster with centroid closest in Euclidean distance to their own price-appreciation vector. The routine continues iterating until there are no further movements of area units among the clusters.¹ The end result is k clusters of area units with each area unit allocated to the cluster with mean pattern of price appreciation closest in Euclidean distance to their own pattern.

An important parameter is the number of clusters, k , which is chosen by the researcher. The routine works quickly, so the researcher can easily inspect the results from various values of k . It is not clear that there is an ‘optimal’ number of clusters as one can evaluate a cluster solution in a variety of ways: the distinctiveness of the clusters, the stability of the solution over repeated trials, and how interesting or informative the results are. After experimentation we chose to study the 4-cluster solution as it is stable, with clusters that are relatively distinctive in terms of the pattern of price appreciation, and, perhaps of most interest, that consist mostly of spatially contiguous sets of area units that comport plausibly to distinct regions as defined in terms *other than* price appreciation rates, such as distance to the CBD.

Figure 8 shows each of the four clusters of price-appreciation vectors. The labels assigned to each reflect their locational characteristics as mapped in Figure 9. The group of area units shaded in red in Figure 9 are the most centrally located. The area units shaded in tan in the figure correspond to ‘close in’ suburban areas. The effectively three separate groups of area units shaded in light blue are typical suburban residential areas. And the area in the south shaded in dark blue is a residential area relatively distant from the CBD.

¹ See, for example, Fielding (2007) for details.

Figure 8. Four cluster solution

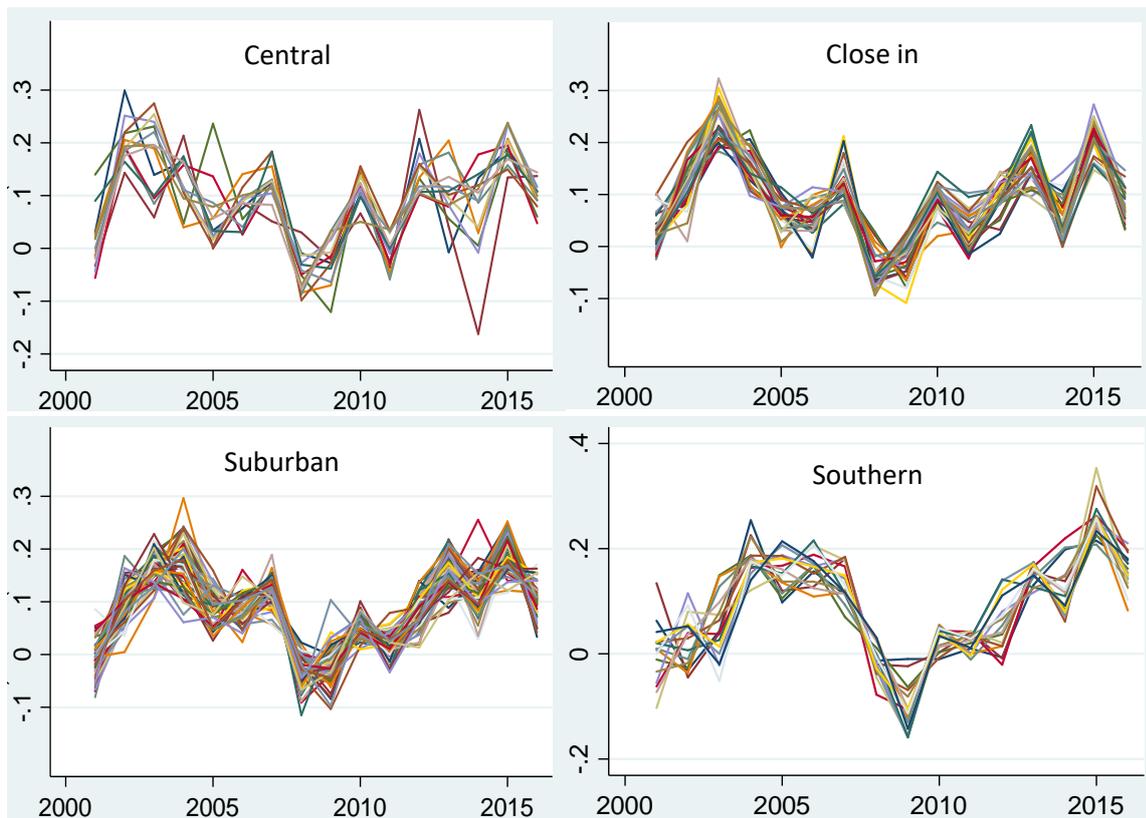


Figure 9. The spatial arrangement of the clusters.

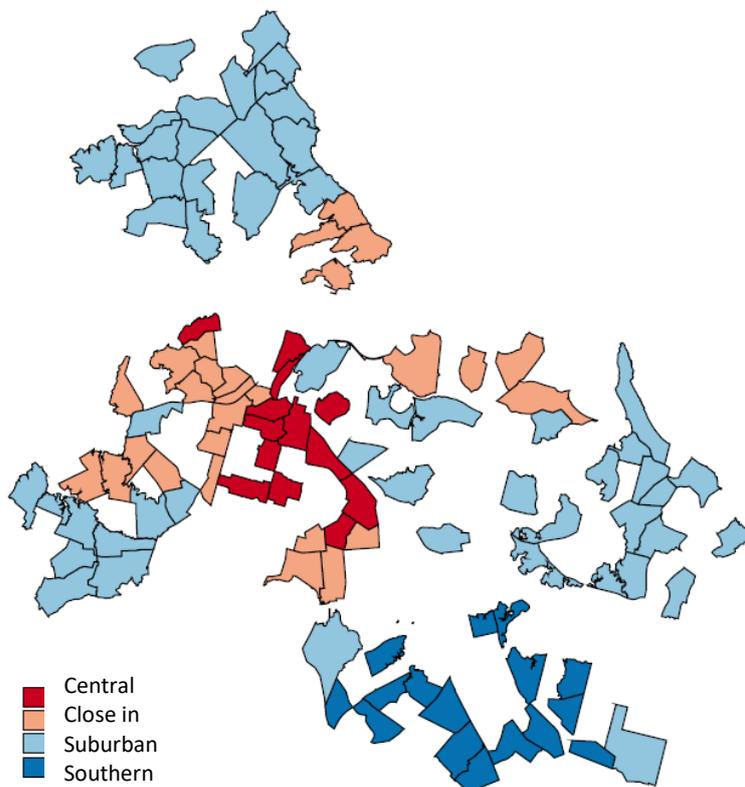


Table 1 provides summary statistics. The column labelled ‘Mean’ shows the difference of each cluster’s overall mean appreciation rate from that of the sample mean. For example, the mean appreciation rates over the sixteen years in the Southern and Suburban clusters are slightly less than the sample mean appreciation rate. The mean appreciation rates in the Close-in and Central clusters are slightly correspondingly higher than the sample mean appreciation rate. The means do not sum to zero because the numbers of area units varies across the clusters. As shown in the second row corresponding to each cluster (labelled ‘n’ in the right-hand column), the Suburban cluster is largest with 55 area units and the Central cluster is smallest with 13 area units.

Table 1. Summary of adjusted price appreciation series by cluster

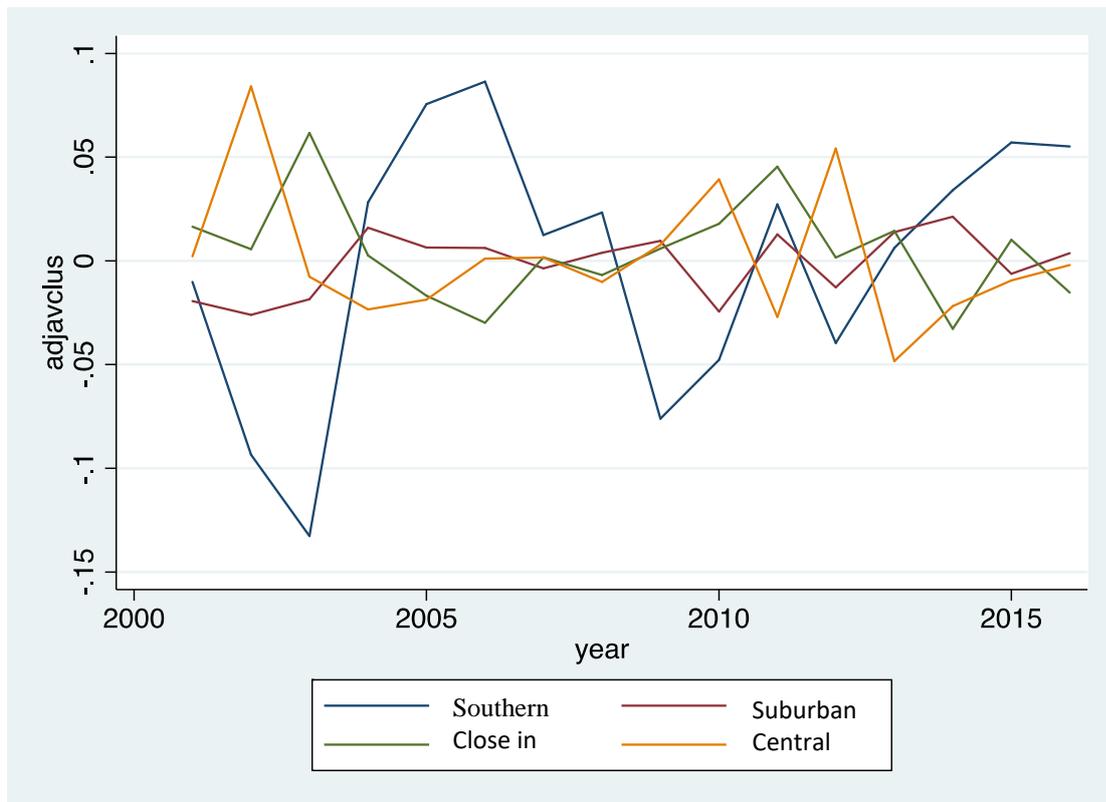
		Mean*	Std. Dev.	Obs	
Central	Overall	0.001720	0.05725	208	N
	Between		0.00940	13	n
	Within		0.05653	16	T
Close-in	Overall	0.003226	0.04000	432	N
	Between		0.00629	27	n
	Within		0.03951	16	T
Suburban	Overall	-0.001941	0.03512	880	N
	Between		0.00458	55	n
	Within		0.03482	16	T
Southern	Overall	-0.000958	0.07185	256	N
	Between		0.00478	16	n
	Within		0.07170	16	T

*Overall mean relative to the urban-area mean.

Looking across the clusters in Figure 8, the area-unit trends in the Central cluster appear perhaps to fit less well together than in the other clusters. This is borne out by the ‘Between’ standard deviations in Table 1, which indicate the similarity among the individual trends in each cluster. The Southern and Suburban clusters have the lowest ‘Between’ standard deviations, though both show some variability. The Southern cluster stands out with

the highest ‘Within’ standard deviation, consistent with the relatively high volatility over time in appreciation rates. Figure 10 superimposes the plots of the mean appreciation rates of the trends in each cluster (again relative to the overall sample mean appreciation rates). The trend in the Southern cluster, shown in blue, appears more volatile than the others throughout the time period.

Figure 10. Mean in the trend of each of the cluster relative to the sample mean



The trends in cluster means in Figure 10 also suggest, working from left to right, that spikes in price-appreciation rates may ripple through time from the central cluster away from the city center to the more distant suburbs. Early in the period at least, these spikes appear to be about a year apart. In the next section we test this and other hypotheses concerning systematic spatial variation price appreciation.

6. Tests of systematic spatial variation in patterns of price appreciation

In this section we report the results of statistical tests of several hypotheses suggested by the spatial patterns in the clusters shown in Figure 9. We begin by exploring the extent to which house prices and demographic characteristics predict assignment to each cluster. Of relevance is that Auckland differs from older, historically manufacturing cities in other developed countries in that some of the highest income and highest amenity residential areas are near the city center. We expect that house prices and demographic characteristics associated with income and wealth will predict assignment to clusters.

Given these results we estimate a distributed lag model to test the significance of a ‘ripple effect’ of price appreciation propagating from the Central through to the Southern clusters, i.e., from the highest to the lowest priced neighborhoods. We might expect that a positive income/purchasing-power demand shock beginning in the early-2000s to affect house prices in the most desirable area first, and then to ripple outwards to less desirable areas. In the case of Auckland that rippling would proceed from the center outward toward the more distant suburbs.

Predictors of assignment to clusters

In this subsection we report estimates from multinomial logit analyses of the association of first house prices and then census demographic characteristics on assignment of area units to each of the four clusters.

Table 2 provides summary statistics for median area-unit house prices. Mainly for convenience, the mean for each cluster is the natural log of the median price of each area unit averaged over the area units in the cluster and over the 17-year time period, i.e., year 2000 – 2016. So the number of observations for each cluster is the number of years (17) times the number of area units in the cluster. As a result, the standard deviations and ranges reflect both variation across area units and (more so) over time.

Table 2. Summary statistics for natural log of median area-unit house prices by cluster

Cluster	Mean	(Level)	Std. Dev	Min	Max	Obs
Central	13.65	(\$847.5k)	0.497	12.46	14.78	221
Close in	13.33	(\$615.4k)	0.470	12.18	14.46	459
Suburban	13.12	(\$499.0k)	0.496	12.02	14.56	935
Southern	12.52	(\$273.8k)	0.429	11.69	13.55	272

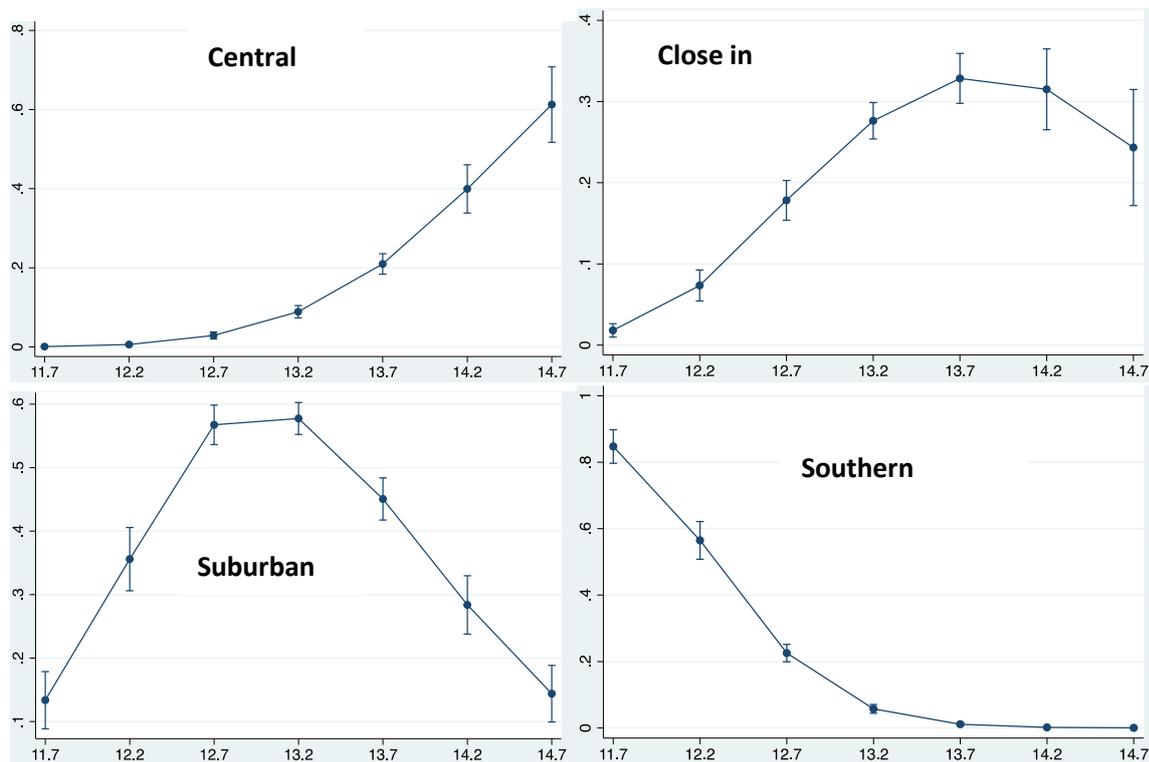
Not surprisingly, prices fall with distance from the city center, though the magnitude of the drop from the Central to Southern clusters surely reflects more than the cost of commuting. Table 3 reports the results of a multinomial logistic analysis of cluster membership on area unit median sale price in each area unit and year. We report estimates using the Close-in cluster as the base (omitted) category and using the Suburban cluster as the base so that the significance of differences can easily be seen. In this case, the standard errors are all relatively small, so median area-unit sale price does significantly predict cluster membership.

Table 3. Multinomial logistic regression of cluster membership on sale price

Predictor variable	Central	Close in	Suburban	Southern
Natural log of cluster median house value	2.213 (0.175)*	0.842 (0.120)*	-0.842 (0.120)*	-2.769 (0.189)*
	1.372 (0.180)*			-3.610 (0.212)*
* p-value < 0.05		Pseudo R ² 0.141		# obs = 1887

Figure 11 plots the estimated probabilities (with 95% confidence bands) of assignment to each cluster as a function of log median sale price. The pattern is clearly consistent with that expected given the numbers in Tables 2 and 3 above.

Figure 11. Estimated probabilities of cluster membership w/AU median price



In addition to distance from the central business district, house prices (and house and neighborhood characteristics) reflect the characteristics of householders. Table 4 reports descriptive statistics for demographic characteristics of sample area units obtained from the national censuses of 2001, 2006, and 2013. The variables include a measure of income, in this case the natural log of personal income; education level as measured by the proportion of those with a tertiary degree; two measures of household composition, and an admittedly crude, but commonly-used measure of variation in ethnicity. The number of observations varies with the number of area units in the cluster and the number of censuses in which data are available: data on education level and child-bearing are available only from the 2006 and 2013 censuses.

Table 4. Selected area unit demographic characteristics

	Mean	Std Dev	Min	Max	Obs.
Central cluster					
Ln personal income	10.235	0.377	9.240	10.960	39
Proportion w/tertiary degree	0.439	0.066	0.296	0.540	26
Proportion single	0.509	0.104	0.392	0.761	39
Proportion of females w/children	0.443	0.155	0.115	0.603	26
Proportion w/NZ European ancestry	0.617	0.146	0.319	0.906	39
Close-in cluster					
Ln personal income	10.304	0.278	9.540	10.810	81
Proportion w/tertiary degree	0.399	0.079	0.210	0.541	54
Proportion single	0.513	0.067	0.317	0.615	81
Proportion of females w/children	0.494	0.120	0.201	0.662	54
Proportion w/NZ European ancestry	0.664	0.155	0.252	0.932	81
Suburban cluster					
Ln personal income	10.158	0.279	9.409	10.795	165
Proportion w/tertiary degree	0.313	0.084	0.117	0.530	110
Proportion single	0.568	0.066	0.283	0.601	165
Proportion of females w/children	0.593	0.060	0.363	0.689	110
Proportion w/NZ European ancestry	0.598	0.174	0.156	0.939	165
Southern cluster					
Ln personal income	9.760	0.202	9.409	10.111	48
Proportion w/tertiary degree	0.107	0.052	0.041	0.239	32
Proportion single	0.475	0.038	0.454	0.600	48
Proportion of females w/children	0.561	0.027	0.519	0.633	32
Proportion w/NZ European ancestry	0.189	0.125	0.068	0.648	48

Worth noting is the relatively large variation in demographic characteristics across the area units within each individual cluster. In particular, standard deviations and ranges tend to be lowest for the Southern cluster. Some ranges seem surprisingly large: e.g., in the Central cluster, which has the highest house prices, the proportion of households with children ranges across the 13 area units in the cluster from 11.5% to 60.3%. The corresponding range in the

Close-in cluster is nearly as large. Much of the explanation for these large ranges is that the Central and Close-in area units are centrally located and therefore contain relatively large areas of higher-density housing, e.g., apartment complexes. The demographics in these higher-density areas likely differ from those in the single-family residential areas. For this reason, data from single-family residential blocks would likely be superior, though much more costly to collect (a task perhaps for the future).

Table 5 reports multinomial logistic regression estimates of area-unit level census demographic characteristics as predictors of cluster membership. Again, for convenience, the values of the demographic characteristics are the averages of those available across the 2001, 2006, and 2013 censuses. The coefficient estimates that are bolded are significant predictors, at the 5% level, of cluster membership relative to the omitted category. Again, we report results from omitting the Close-in and Suburban clusters independently to ease interpretation of significant differences in estimated coefficients.

The signs on the coefficients of personal income surprisingly appear opposite to expectations as a predictor of cluster membership: the sign is negative on the Central cluster where house prices are highest, and relatively large and positive (though insignificant) on the Southern cluster. As shown in Table 4, average incomes are in fact highest in the Central and Close-in clusters, smaller in the Suburban cluster and lowest in the Southern cluster. The most likely explanation of these unexpected results is that other predictor variables are correlated with, and picking up the effects of, income.

The prime example is the proportion of householders with a tertiary (e.g., university) qualification. As expected this proportion, which is correlated with income and house prices, is a strong and significant predictor of cluster membership. Indeed, the Central cluster appears to be distinguished by its level of higher education, which is reflected in the high quality of the primary and secondary schools located in that area.

Table 5. Multinomial logistic regression of cluster membership demographic traits

	Central	Close in	Suburban	Southern
Ln personal income	-4.39 (2.91) -6.34* (2.57)	1.95 (1.96)	-1.95 (1.96)	11.11 (6.73) 9.16 (6.96)
Proportion w/tertiary qualification	40.67* (8.39) 34.11* (8.17)	6.56 (4.35)	-6.56* (4.35)	-32.71* (15.9) -39.28* (16.46)
Proportion single	9.08 (10.2) -11.37 (9.97)	20.45* (7.32)	-20.45 (7.32)*	16.81 (25.3) -3.64 (26.1)
Proportion w/children	-7.27 (6.66) -1.75 (5.20)	-5.52 (5.67)	5.52 (5.67)	29.31 (25.0) 34.84 (25.6)
Proportion with European ancestry	-2.68 (4.42) -6.87 (4.25)	4.19* (2.67)	-4.19* (2.67)	-22.75* (9.30) -26.94* (9.66)
* p-value < 0.05	# of observations: 222		Pseudo R ² : 0.543	

The coefficients on the proportions of singles appears to distinguish the Close-in and Suburban clusters. This is likely to reflect the population of young, mostly professionals who live in conveniently located, higher-density dwellings and perhaps elderly singles who prefer to stay in their high-quality close-in neighborhoods. The negative coefficient on the Suburban cluster reflects the family nature of these suburban areas.

The signs and magnitudes of the coefficients on the proportion of households with children are consistent in that they are opposite in sign to those on the proportion of singles. The coefficients are especially large in the Southern cluster, though insignificant. Their insignificance likely reflects the negative correlation with proportion single.

The defining demographic characteristic of the Southern cluster is its relatively low average proportion of households who claim European ancestry. The mean proportion is less

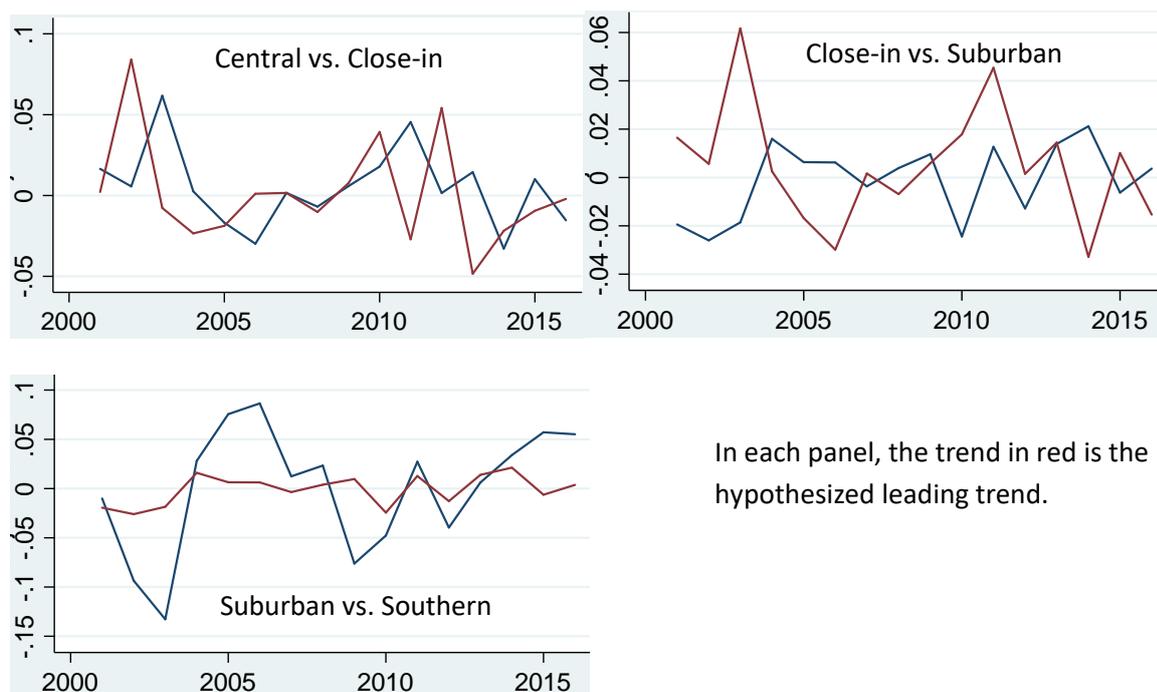
than 20% of the population, whereas the mean proportion in other clusters equals or exceeds 60% of the population (though still with considerable within-cluster variation).

These results as a whole are consistent with results from earlier studies that house-price appreciation rates, and in this case, patterns in house price appreciation rates, vary statistically in plausible ways with variation in house and household demographic characteristics.

Tests of a “ripple effect”

The trends in cluster-mean area-unit appreciation rates superimposed in Figure 10 above suggest that changes in price appreciation start in the Central cluster and then ripple consecutively through the Close-in, Suburban, and Southern clusters. To de-clutter the trends somewhat, Figure 12 plots them in pairs. The ripple effect appears most clearly in the Central vs. Close-in clusters. There may be something similarly in the other pairs but it seems less clear in the plots.

Figure 12. Pairwise plots of cluster trends



In each panel, the trend in red is the hypothesized leading trend.

We use a simple distributed-lag regression model to test the ripple hypothesis statistically. Specifically, we estimate:

$$r_{j,t} = a + \beta_0 r_{j+1,t} + \beta_1 r_{j+1,t-1} + \varepsilon_{j,t}$$

where r is the appreciation rate, j is cluster, and t is year. Suppose, as an example, that j is the Central cluster and $j+1$ is the Close-in cluster. The estimated coefficient β_0 in this case represents the fraction of Close-in's contemporaneous appreciation rate to Central's. If our hypothesis is that changes in price appreciation rate radiate outward from the center, then the estimate of β_1 is the coefficient of interest: it represents the contribution to Close-in's current-year appreciation rate of last year's Central appreciation rate, i.e., the influence of Central's appreciation rate lagged one year.

Table 6 reports the average over the entire time period of estimating the model above for each of the pairs of trends in Figure 12. The lag term is positive and significant at the 5% level for each of the pairs of clusters, which supports the hypothesis of a ripple effect on average over the time sample time period.

Table 6. OLS estimates of lag in adjacent cluster

	β_0	β_1	R^2	Durbin Watson
Central vs. Close-in	0.865* (0.122)	0.338* (0.118)	0.831	1.64
Close-in vs. Suburban	0.737* (0.090)	0.197* (0.088)	0.847	1.37
Suburban vs. Southern	0.866* (0.210)	0.540* (0.196)	0.771	2.50

p-value < 0.05*

What might explain the ripple effect from central, high-income, high-house-price areas ultimately to distant, low-income, low house-price areas? Returning to the context described in Section 3, the explanation seems to be an increase in demand driven by immigration and greater willingness and ability to pay combined with imperfect foresight. Coming out of the

recession of the late 1990s, desirable central areas become more affordable and attractive, bidding up prices there. Prices in close-in neighborhoods then look more attractive, followed by more distant suburbs, and then finally by distant, lower-income suburbs.

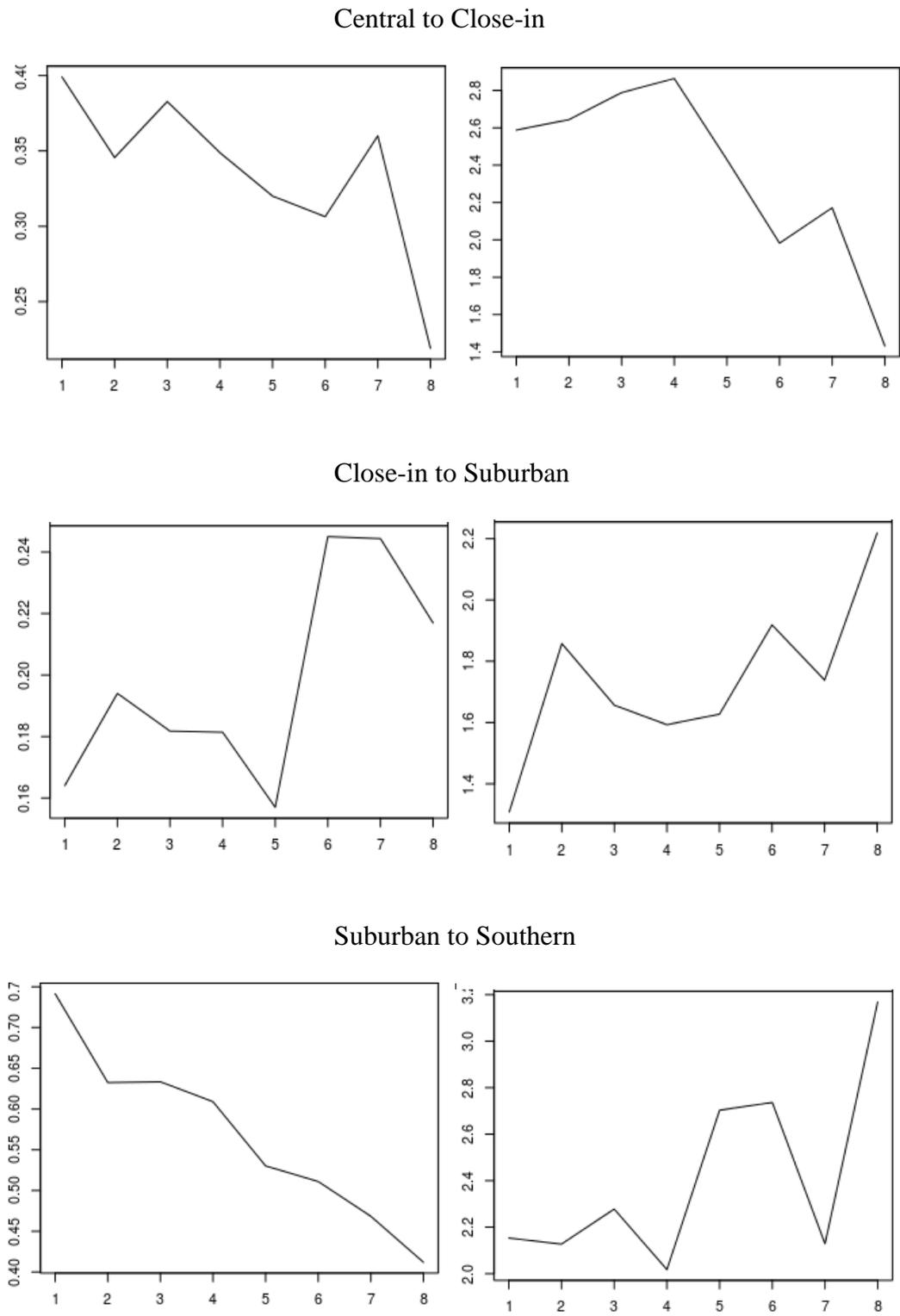
This explanation raises the question of whether the same pattern holds throughout the boom-bust-boom cycle. That is, does the bust and subsequent boom in appreciation rates start in the center and work outward? As we have only annual observations on appreciation rates, we explore the time-consistency of the trend by estimating the nine-year moving average in β_1 for each of the cluster pairs reported in Table 6.

The series of three plots on the left of Figure 13 plot the trends in the estimated moving average of β_1 for each pair of clusters moving from Central through to Suburban. Thus, the plots show the variation over time around the estimates of β_1 reported in Table 6 above. The horizontal axis indicates the sequence of eight nine-year samples; the first one centers on the beginning of 2005 and the other seven center on the beginning of each subsequent year to 2013. The right-hand side of the plots depict the trend in the t-statistic associated with each of the estimates in the corresponding left-hand diagram. Note that the scaling on the vertical axes vary as one proceeds top to bottom on both sides of the figure.

The trend in the estimates of β_1 is generally downward over time from Central to Close-in and from Suburban to Southern, consistent with a waning ripple effect. In contrast, the lagged effect from Close-in to Suburban is small and insignificant during the first boom and grows just enough to become significant in the second boom. Why this difference from the other two is unclear, though the Suburban cluster is spatially much larger and more dispersed than any of the other clusters with perhaps more variation in the timing of influence of central-area prices. Taken together, these results suggest that the ripple effect persists over the boom-bust-boom period (at least over successive nine-year averages), but diminishes

overall, which would be expected if home buyers gradually gain an understanding of the spatial dynamics.

Figure 13. Moving average estimates of β_1 and corresponding t-statistic.



7. Conclusions

The novelty of this paper is in the use of cluster analysis to identify areas in which house-price appreciation follows a similar trend over an interesting boom-bust-boom cycle. The k-means cluster analysis methodology appears to have worked reasonably well. The four cluster solution is stable and the trends in each cluster appear to fit reasonably well throughout the time period. The clusters also seem sensible geographically in that they follow a sensible pattern of house values and demographic characteristics.

We find evidence of a ripple effect from central, high-income, high-house-price areas through ultimately to distant, low-income, low-house price areas. This seems explainable as a result of a substantial increase in population and especially in purchasing power combined with limited foresight: prices in the most attractive areas were bid up first, flowing fairly quickly toward less desirable areas. There is evidence however that the ripple effect diminished over time, consistent with participants in the housing market better anticipating this effect over time.

References

- Archer, W.R., D.H. Gatzlaff, and D.C. Ling (1996). Measuring the importance of location in house price appreciation, *Journal of Urban Economics*, 40: 334-353.
- Bostic, R.W., S.D. Longhofer, and C.L. Redfeard (2007). Land leverage: Decomposing home price dynamics, *Real Estate Economics*, 35(2): 183-208.
- Clapp, J.M. and D. Tirtiroglu (1994). Positive feedback trading and diffusion of asset price changes: Evidence from housing transactions, *Journal of Economic Behavior and Organization*, 24: 337-355.
- Davis, M.A., S.D. Oliner, E.J. Pinto, and S. Bokka (2017). Residential land values in the Washington, DC metro area: New insights from big data, *Regional Science and Urban Economics*, 66: 224-246.
- F. Ferreira and J. Gyourko (2012), Heterogeneity in neighborhood-level price growth in the United States, 1993-2009, *American Economic Review: Papers and Proceedings*, 102: 134-140.
- Genosove, D. and L. Han (2013). A spatial look at housing boom and bust cycles in *Housing and The Financial Crisis*, Glaeser, E. and T. Sinai editors, University of Chicago Press, pp. 105-141.
- Goetzmann, W.N. and M. Spiegel (1997). A spatial model of housing returns and neighborhood substitutability, *Journal of Real Estate Finance and Economics*, 14: 11-31.
- Guerrieri, V., D. Hartley, and E. Hurst (2013). Endogenous gentrification and housing price dynamics, *Journal of Public Economics*, 100: 45-60.
- Ho, L.S., Y. Ma, and D. Haurin (2008). Domino effects within a housing market: The transmission of house price changes across quality tiers, *Journal of Real Estate Finance and Economics*, 37: 299-316.
- Liu, C.H., A. Nowak, and S. Rosenthal (2014). Bubbles, post-crash dynamics, and the housing market [Electronic article]. The Center for Real Estate and Finance Working Paper Series, 2014-001, 1-40.
- McMillen, D. 2003. Neighborhood house price indexes in Chicago: A Fourier repeat-sale approach. *Journal of Economic Geography*, 2: 57-73
- McMillen, D.P. (2012). Repeat sales as a matching estimator. *Real Estate Economics*, 40(4): 745-773.
- Waltl, S.R. (forthcoming). Variation across price segments and locations: A comprehensive quantile regression analysis of the Sydney housing market. *Real Estate Economics*.