ARC Centre of Excellence in Population Ageing Research

Working Paper 2015/19

The Income elasticity of House Prices in New South Wales: A Panel Study

Xiangling Liu[1]

[1]Research Fellow, ARC Centre of Excellence in Population Ageing Research (CEPAR), Business School, UNSW Australia email: xiang.liu@unsw.edu.au

The Income Elasticity of House Prices in New South Wales: A Panel Study

Xiangling LIU
xiang.liu@unsw.edu.au
ARC Centre of Excellence in Population Ageing Research (CEPAR)
UNSW, Sydney 2052, Australia

June 2015

Abstract

This paper estimates the income elasticity of house prices over a long-term time period of 1991 to 2012 for 144 LGAs in New South Wales of Australia. The income elasticity of house prices is estimated to be 0.69 by multi-factor panel data models accounting for cross-section dependence and serial correlation. The estimate confirms a co-integrated long-run relationship between real house prices and real income. Alternatively, the income elasticity is estimated to be 0.46 using traditional spatial autoregressive models, where the spatial matrix is specified as a distance weighted matrix. The spatial effect of house price in one location is estimated significantly to be 0.84.
JEL classification: R12,R31,C33
Keywords: Income elasticity; cross-section dependence; house price dynamics; co-integration

1 Introduction

Housing prices in Australia have been increasing over more than two decades and this is obviously seen in New South Wales. Since housing wealth dominates household wealth for a typical household, volatility in housing prices influences household wealth and also the macroeconomic stability. In consequence, the dramatic increases in house prices pose some concern if housing prices have been overvalued. This paper investigates if prices are moving away from their market fundamental drivers, which are normally monitored by the affordability indicator of real disposable income per capital, the mortgage condition indicator of real interest rates and population growth (Case and Shiller 2003 and McCarthy and Peach 2004; Andr, C., & Girouard, N. 2010).

A panel dataset for house prices and income for 144 Local Government Areas (LGAs) in New South Wales (NSW) for the period 1991 to 2012 is used in the study. The statistical analysis has a particular focus on testing if there has been a long run equilibrium in price to income ratios by examining the existence of a co-integration relationship of real housing price to real income. The models estimated in this Chapter allow for the fact that the behaviour of real house prices is likely to be correlated across LGAs. The spatial patterns in house price distributions are considered to arise when cross section units are subject to observed or unobserved common effects (Holly, Pesaran and Yamagata, 2010) and (or) if there are diffusion effects due to spatial or other forms of local dependencies (Alexander and Barrow 1994; Ashworth and Parker 1997 and Cook 2003). Therefore, a key element of the econometric analysis is an attempt to account for possible cross-section (spatial) dependence in the error terms of the panel data models, which is modelled by a multifactor error structure, and also, by a spatial autoregressive process.

Abelson,Joyeux, Milunovich and Chung (2005) found a co-integrated relationship between real income per capita and real house prices and a number of other variables in Australia from 1970 to 2003. Bodman and Crosby (2003) found no significant relationships between house prices and income per capita to model changes in real quarterly house prices in five capital cities in Australia from 1980 to 2003. Instead, they find evidence that housing prices in Sydney and Brisbane were overvalued. These studies suggest that local housing markets have variations in the desirability of location and culture and the value of obtainable land (which naturally lead to the heterogeneity in house prices in regions). However, it is very difficult to control these effects when the focuses are on aggregate time series analyses and the use of fixed effects and time dummies may not be sufficient to capture these idiosyncratic factors. Housing markets are also likely to be correlated due to the existence of some spatial or non-spatial diffusion effects in the development of a housing market. The impacts of these factors will also lead to bias in estimates if not dealt with appropriately.

In the existing literature, there are two forms of spatial patterns in housing price distributions. One is caused by common factors, mainly described as macroeconomic conditions, including changes in interest rates and oil prices, or unobserved, such as technological change. Meen (1999) suggests that the impacts of common factors result from the heterogeneity in the responses of sub-markets in a given geographical area to the overall state of the macro-economy. The second is the diffusion effect due to spatial or other forms of local dependencies (Alexander and Barrow 1994; Ashworth and Parker 1997) in house prices that cannot be captured by spatial effects alone. For example, interactions between sub-housing markets and migration patterns may result in some pattern of diffusion effects to the development of housing markets (Cook 2003 and Alexander and Barrow 1994). Market participants would be crowded out from one particular preferred location to the next when prices in one geographic location change.

I take account of the spatial influences due to common factors by making use of the common correlated effects (CCE) estimator (Pesaran 2006; Pesaran and Tosetti 2011; Chudik, Pesaran, and Tosetti 2011) which is consistent under heterogeneity and cross-sectional dependence. The CCE estimator treats spatial temporal dynamics as general dynamics, which are captured by common factors as well as exogenous individual-specific time series dynamics. This approach may have some disadvantage that it does not explicitly model spatial patterns in house prices. Yet, as shown by Pesaran and Tosetti, this approach continues to provide consistent estimates of the slope coefficient, even in the presence of a spatial error process. Holly, Hashem Pesaran, and Yamagata (2011) apply the approach to study the co-integrated relationship between real house prices and real income per capita for U.S. states using annual data from 1975 to 2003. They find that the observed common factors (such as real interest rate) or unobserved common factors (proxied by the average of observed individual-specific factors) significantly impact the long-run equilibrium of house prices. With controls for common shocks, house prices and income are found to be cointegrated with a unity coefficient. As an alternative, we also apply the spatial SAR panel data models to account for the spatial effects in house prices and a model's error term. This traditional spatial econometric approach explicitly models cross section dependence by using an N x N spatial matrix that specifies a relationship between a cross-section unit and its 'nearest' neighbors (Anselin 1988; Baltagi, Song, and Koh 2003; Kapoor, Kelejian, and Prucha 2007; Baltagi, Egger, and Pfaffermayr 2007; Anselin, Le Gallo, and Jayet 2008; LeSage and Pace 2010 and Lee and Yu 2010). However, a potential problem with the approach is that the spatial weight matrix may be mis-specified because the spatial effects are multidimensional or because there are unobserved common factors that are not adequately captured by the spatial contiguity matrix and end-up in the model's error term.

The majority component of this paper focus on providing consistent estimates for the income elasticity to house prices with consideration to the cross-section dependence and serial autocorrelation in the panel data models. Multi-factor panel data models and spatial SAR models are estimated respectively. The income elasticity of house prices is estimated to be 0.69 by multi-factor panel data models which are proved to be successfully filtered from cross-section dependence and serial correlation. This estimate confirms a co-integrated long-run relationship between the real house prices and real income however only when there is a control for the linear time trend. In addition, house price in one location is proved to be connected with prices in other spatial units through a contiguous distance weighted matrix with a coefficient of 0.84. After controlling for such an impact, somewhat lower estimate of the long-run income elasticity is obtained to be 0.46. The marked upturn in home prices is largely attributable to strong market fundamentals. Home prices have essentially moved in line with increases in family income and declines in nominal mortgage interest rates.

The reminder of this paper has the following structure. In Section 2, the main econometric methods used in this paper are described. A description of the panel dataset and a preliminary data analysis is presented in Section 3. The empirical results, including income elasticities and factor loadings are reported in Section 4. Section 6 concludes.

2 Econometric Framework

2.1 Model and Estimators

In estimating the co-integrating relationship of house prices in NSW the following model due to Holly, Pesaran, and Yamagata (2010) is employed:

$$y_{it} = a_i + \boldsymbol{\beta}_i' \boldsymbol{X}_{it} + \boldsymbol{\gamma}_i' \boldsymbol{f}_t + e_{it} \tag{1}$$

where $i = 1, 2, .., 144$; t=$1, 2, .., 22$. The dependent variable y_{it} denotes the log of real house prices in the ith LGA during time t. The vector $\boldsymbol{X}_{it}$ is a $k \times 1$ vector of observed individual-specific regressors, namely the exogenous economic fundamentals, including log real income and log population growth on the ith cross-section unit at time t. Another exogenous variable included is the nation-wide observation of the real interest rate which represent the mortgage market condition. Structural changes in housing and mortgage markets in some economy are likely to be shifting equilibrium levels. For example, a reduction of mortgage rates resulting from lower inflation expectations or increased efficiency and competition in financial markets may lower borrowing costs in a durable way and thereby increase equilibrium house prices. But a fall in mortgage rates driven by the underestimation of risks is likely to be reversed sooner or later as risk premiums are reassessed, leading to a correction in house prices (Andr, C., & Girouard, N. 2010). The vector $\boldsymbol{f}_t$ is assumed to be an m-dimensional vector of unobservable common factors[1], which might influence house prices across all regions of NSW in a similar manner, such as policy effects, expectations, or common technologies influencing all groups in a similar way. $\boldsymbol{\gamma}_i'$ is a $1 \times m$ vector of heterogeneous factor loadings. The idiosyncratic error term, e_{it} is assumed to be independent of $\boldsymbol{X}_{it}$ and $\boldsymbol{f}_t$, but can be weakly dependent (spatially and temporally correlated)[2]

The model given by (1) nests a number of simpler panel data specifications. If there are no unobserved common factors, $\gamma_i = 0$ and the slope coefficients are homogeneous, $\boldsymbol{\beta}_i' = \boldsymbol{\beta}'$, then the (1) collapses to a standard fixed-effects model (possibly with spatially correlated errors). In the empirical analysis a fixed-effects model is estimated as a benchmark for comparison with the more general specifications.

Maintaining the assumption of no unobserved common factors, but allowing for heterogeneity in the slope coefficients yields the dynamic heterogeneous panel model due to Pesaran and Smith (1995). Pesaran and Smith show that consistent estimates of $\boldsymbol{\beta}_i'$ coefficients can be obtained by estimating separate regressions for each cross-section unit and then averaging the coefficients over the units. They call this a mean group (MG) estimator. Such an estimator is likely to work best for large N and large T. Finally the most general version of model (1) can allow for cross-section dependence in house prices by use of the common correlated effects (CCE) estimator (Pesaran 2006). A pooled or mean group version of the CCE estimator can be employed, depending upon what is assumed about heterogeneity in the slope coefficients.

More details for the various estimators used in this paper are described in Appendix C.

2.2 Cross-section Dependence

Various tests for cross-section dependence in panel data have been proposed for the case of (small) T and (large) N (Frees 1995; Pesaran 2004; Sarafidis, Yamagata, and Robertson 2009). In this study, the cross-section dependence (CD) test due to Pesaran (2004) is used. One advantage of the CD test is that it can be applied to a wide variety of models, including heterogeneous dynamic models with multiple breaks and non-stationary dynamic models with small/large N and T.

[1] In this study, m is assumed to be no greater than the number of included exogenous regressors.

[2] In Pesaran (2006) and Pesaran and Tosetti (2011), e_{it} has this form: $\boldsymbol{e}_{.t} = \boldsymbol{R}_t\boldsymbol{e}_{.t} + \boldsymbol{\varepsilon}_{.t}$, where $\boldsymbol{R}_t$ is a given $N \times N$ weighted spatial matrix for catching up the remaining weak spatial dependence in $\boldsymbol{e}_{.t}$, and $\boldsymbol{\varepsilon}_{.t}$ allows for serial autocorrelation to capture the individual specific short-run dynamics. In this study, the cross-section dependences are successfully filtered out by $\boldsymbol{f}_t$ in panel regressions as proved by the CD test in Table 7-8, The weak spatial effects in $\boldsymbol{e}_{.t}$ are not estimated. Instead, I use robust non-parametric estimators shown in (28-29) to estimate the standard error of $\boldsymbol{e}_{.t}$. The spatial autocorrelation in real house prices and panel regression residuals are estimated by Model (18).

Suppose y_{it} is the panel data variable to be tested, denoting $\hat{\rho}_{ij}$ as the sample estimate of the pair-wise correlation of the ith and jth cross-section units of y_{it},

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t=1}^{T} y_{it} y_{jt}}{(\sum_{t=1}^{T} y_{it}^2)^{1/2} (\sum_{t=1}^{T} y_{jt}^2)^{1/2}} \tag{2}$$

The value of cross-section dependence is then computed as,

$$\bar{\rho}_{ij} = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij}. \tag{3}$$

The null hypothesis is that y_{it} is independently and identically distributed across cross-section units, written as,
$H_0 : \rho_{ij} = \rho_{ji} = 0$ for all $i \neq j$,
$H_1 : \rho_{ij} = \rho_{ji} \neq 0$ for some $i \neq j$,
The CD statistic proposed by Pesaran (2004) has the following form,

$$CD = \sqrt{\frac{2T}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij} \rightarrow N(0,1), \tag{4}$$

for $N \rightarrow \infty$ and T sufficiently large.

2.3 Panel Unit Root Tests

Most of the variables in (1) enter in levels or log-levels. This implies that some variables may be non-stationary due to the presence of a unit root. Furthermore if some variables contain unit roots, a valid model specification and hence consistent estimates, will require the existence of one (or more) co-integrating relationships. A standard framework used in testing for unit roots in panel data is the following:

$$\Delta y_{it} = \mu_{it} + \phi_i y_{it} + u_{it} \tag{5}$$

where additional deterministic trends and lags of the dependent variable can be added as necessary.

Many standard tests for panel unit root tests are based on the assumption of cross-section independence (Im and Pesaran 1997; Harris and Tzavalis 1999; Maddala and Wu 1999; Hadri 2000; Choi 2001 and Levin, Lin, and James Chu 2002). Under this assumption, Choi (2001) finds in a Monte Carlo analysis that a test based on the inverse normal distribution has the best trade-off in terms of size and power. This Z statistic is calculated as

$$Z = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Phi^{-1}(pv_i) \tag{6}$$

where pv_i is the p-value on the ith panel associated with an Augmented Dickey Fuller (ADF) test for a unit root test applied to each of the N cross-section units and $\Phi^{-1}(.)$ is the inverse of the standard normal cumulative distribution function. Under the null hypothesis that all N times series have a unit root, $Z \sim N(0,1)$. The null hypothesis is rejected for values of $Z < C_{za}$ where C_{za} is from the lower tail of the normal distribution.

A number of recently developed tests have relaxed this assumption of cross-section independence (Cerrato and Sarantis 2007; Pesaran 2007, Breitung and Das 2008 and Pesaran, Vanessa Smith, and Yamagata 2013). In this analysis, the CIPS test by Pesaran (2007) and the CIPSM test by Pesaran, Vanessa Smith, and Yamagata (2013) are employed. These approaches control for cross-section dependence by using factor error specification models and also allow for serial correlations in the idiosyncratic components. One advantage of CIPS (CIPSM) is that unlike the principle component methods by Bai and Ng (2004) and Moon and Perron (2004), CIPS does not require estimating the number of unobserved common factors for obtaining valid individual CADF statistics. In addition, CIPS statistics are based on the simple averages of the individual CADF statistics and are asymptotically consistent as long as the number of individual specific variables of x_{it}, k, is greater than the true number of common factors.

The CIPS test procedure is outlined below. The data generating process for y_{it} for CIPS is assumed to be:

$$y_{it} = (1 - \Phi_i)\mu_i + \Phi_i y_{i,t-1} + u_{it},\ i = 1, 2, 3, ..., N;\ t = 1, 2, ..., T \tag{7}$$

with the residual, u_{it}, accounting for serial correlation,

$$u_{it} = \rho_i u_{i,t-1} + \eta_{it}, |\rho_i| < 1 \text{ for } i = 1, 2, ..., N, \tag{8}$$

and η_{it} accounting for the unobserved factors by a one-factor error specification model,

$$\eta_{it} = \gamma_i f_t + \varepsilon_{it},\ \varepsilon_{it} \sim i.i.d(0, \sigma_i^2) \tag{9}$$

Combing the equations (7)-(9), y_{it} could be defined by

$$\Delta y_{it} = (1 - \Phi_i)\mu_i(1 - \rho_i) - (1 - \Phi_i)y_{it-1} + \rho_i(1 - \Phi_i)\Delta y_{i,t-1} + \gamma_i f_t + \varepsilon_{it} \tag{10}$$

The null hypothesis is that all series in the panel have a unit root and is given by $H_0 : 1 - \Phi_i = 0,\ i = 1, 2, ..., N_1$. The alternative allows for some series to be stationary and some to have unit roots and is given by $H_1 : 1 - \Phi_i < 0$, $i = 1, 2, ...N_1$, and $1 - \Phi_i = 0$,

$i = N_1 + 1, N_1 + 2, ..., N$, where N_1/N represents the fraction of the individual processes that are stationary and tends to the fixed value Θ such that $0 < \Theta \leqslant \infty$ as $N \to \infty$.

The test is performed by running a cross-section augmented Dickey Fuller test (CADF) on each of the N time series in the panel. For the ith cross-section unit the CADF test regression takes the following form;

$$\Delta y_t^i = r_0^i + r_1^i y_{t-1}^i + r_2^i \bar{y}_{t-1} + r_3^i \Delta y_{t-1}^i + r_4^i \Delta \bar{y}_{t-1} + \omega_t^i, \;\; \omega_{it} \sim i.i.d(0, \sigma^2) \tag{11}$$

where $\bar{y}_{t-1}$ and $\Delta \bar{y}_{t-1}$ are cross-section averages for the N units at time t. In line with reasoning of Pesaran (2006), the above regression filters out any cross-section dependence caused by the unobserved common factor f_t, by augmenting the standard ADF regression with the cross-section averages $\bar{y}_{t-1}$ and $\Delta \bar{y}_{t-1}$. The associated CADF statistic is the t-statistic for r_1^i, denoted as $\widetilde{t}_i$. The CIPS test statistic is obtained as the cross-section average of the N values of $\widetilde{t}_i$, denoted as

$$\overline{CADF} = N^{-1} \sum_{i=1}^{N} \widetilde{t}_i \tag{12}$$

Critical values for the CIPS test are given in Pesaran (2006). The CADF test regression can be generalized to include trends and additional lags of the dependent variable.

In the CIPSM test, due to Pesaran, Vanessa Smith, and Yamagata (2013), the CIPS procedure is extended to allow for a vector of unobserved factors $\boldsymbol{f}_t$. The additional unobserved factors are proxied by including additional stationary regressors $\Delta \boldsymbol{X}_t^i$ and their cross-section average $\Delta \bar{\boldsymbol{X}}_t^i$ in the CADF test regressions. Critical values for the CIPS test is given in Pesaran (2006).

3 Data and Preliminary Data Analysis

A panel dataset is constructed for 144 LGAs in NSW over the period 1991 to 2012. Median sales prices for non-strata dwellings are used to measure house prices [3]. For each

[3] There might be arguments that median sales prices such as those used in this thesis are likely to overstate income elasticities because they embody improvements to dwellings, and "house prices" might better be described as "house values" in recognition of the fact that "house values" embody both quantity and quality components as well as a price component. However, there are several reasons that I am not able to use "house values" to proxy "house prices". Firstly, the measurement of house values is at the individual household level and has the micro nature. If I need to carry out the empirical analysis using the household level data, the current analyses might have to be changed substantially due to a set of new data, which is switched from the current aggregated LGA level data to the household individual data. Secondly, the study of a microl level analysis was not the starting point of this thesis. This thesis has the feature of doing empirical analysis in relevance to macro housing economy, it can be reasonable to use the median housing sales prices to reflect the development trend of a local housing market and use the aggregated real income per tax payer as the proxy for real income per capita for each location.

LGA, the median price is converted to real (or constant) prices using the consumer price index (CPI) for Sydney. Real income for each LGA is measured using real income per taxpayer. The other variables used in the empirical analysis are the resident population for an LGA and a non-LGA specific measure of the real government bond interest rate. The symbols used for each variable are given in Table 1. LGAs in NSW are frequently grouped or clustered into twelve geographical regions. These regions and their abbreviations are listed in Table 2. The LGAs that comprise each region are listed Tables 11-13 of the Appendix. More details for data including data sources are included in Appendix ??.

Table 1: Variables Names

P_{it}	Logarithm of real house prices
Y_{it}	Logarithm of real income per taxpayer
Pop_{it}	Logarithm of population
R_t	Real interest rate

Table 2: Regions in the State of New South Wales

CW:	Central West	FW:	Far West	HT:	Hunter
IW:	Ilawarra	MN:	Mid North Coast	MB:	Murrumbidgee
MR:	Murry	NW:	North Western	NT:	Northern
ST:	Richmond Tweed	SE:	South Eastern	SY:	Sydney

As a preliminary analysis of the panel data, time series plots of house prices and real income at the regional-level are presented. The CD statistics is used to test for cross-section dependence in the level and changes in house prices, real income and population across regions. Finally, the pair-wise contemporaneous cross-correlations between the regions for house prices and income, respectively, are computed and reported.

3.1 Regional Profiles for Real House Prices and Real Income

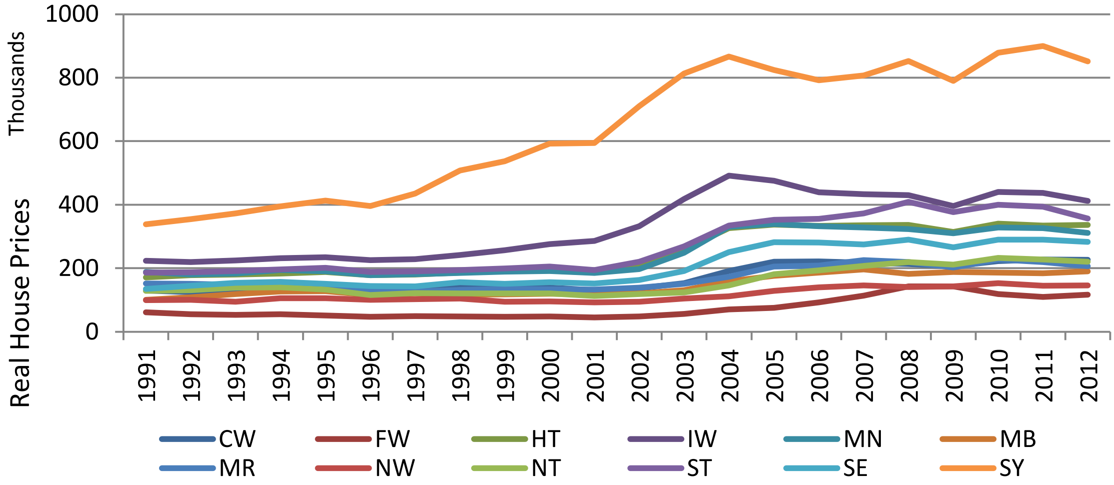
Figure 1 plots the regional profiles which are derived by averaging over the regional LGAs to illustrate the course of real house prices during 1991-2012 for the 12 regions. I can see that even though there is heterogeneity in the initial values across LGAs, the growth patterns across regions are fairly similar. Prices for almost all of regions have shown persistent increases through 1991 to 2004 and then mild fluctuations from 2004 on. The period of 2002-2004 seems to display specific time effects when real house prices experienced the most dramatical increase, on average by around 36.8 per cent[4].

[4]LGAs where real house prices have grown over 100 per cent over the years 2002 to 2004 are Ballina, Bega Valley, Boorowa, Byron, Carrathool, Cessnock, Clarence Valley, Coffs Harbour, Eurobodalla, Great

The mean profiles for the covariate of real income are plotted in Figure 2. Similarly, for the level of real income, there is a consistent growth over time and growth patterns across the regions are quite similar.

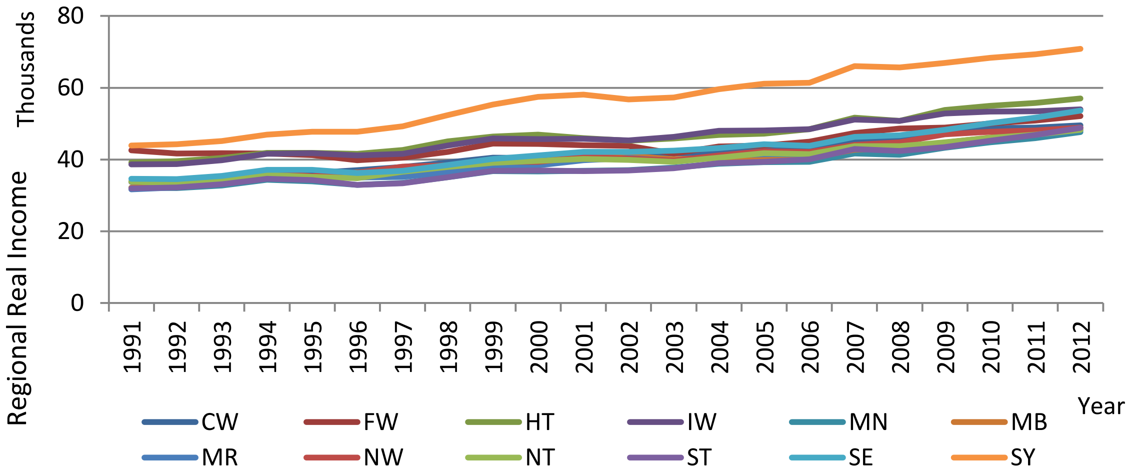
I also plot the LGA-level time profile in Figure 8 and Figure 9 to illustrate the course of real housing prices and real income during 19912012 for the 144 individual LGAs.

Figure 1: Mean Time Profiles for Regional Real House Prices



Notes. Data for each region is the average of real house prices over the regional LGAs. Curves of regions are distinguished by color.

Figure 2: Mean Time Profiles for Regional Real Income



Notes. Data for each region is the average over the regional LGAs. Curves of regions are distinguished by color.

Lakes, Harden, Lake Macquarie, Maitland, Nambucca, Newcastle, Shoalhaven, Tweed.

3.2 Testing for Cross-section Dependence

The variable cross-section dependence within all the cross-section units and the corresponding CD statistics for P_{it}, Y_{it}, Pop_{it}, ΔP_{it}, ΔY_{it} and ΔPop_{it} are computed using (2) and (4) respectively . Results are tabulated at Table 3.

It is evident that the magnitudes of cross-section dependence for all the relevant variables are estimated to be very significant for P_{it} and Y_{it}, and ΔP_{it} and ΔY_{it}. However, this significance is not found in the change of population, ΔPOP_{it}, despite the cross-section dependence is estimated to be significant for the level of population, POP_{it}. The presence of cross-section dependence in both P_{it} and Y_{it}, and ΔP_{it} and ΔY_{it} demonstrate the importance of accounting for the cross-section dependence in all the panel data model specifications. Since only ΔPOP_{it} may be used as a covariate in the remaining panel data models, consideration of cross-section dependence for this variable is excluded. In the remaining panel data model estimations, the possible cross-section (spatial) dependence in panel data models is modeled by a multifactor error structure, for which CCEMG (common correlated effects mean group) and CCEP (common correlated effects pooled) estimators are used.

Table 3: Variable Cross-section Dependence Testing Results

Variable	CD-Statistics	$\bar{\hat{\rho}}_{ij}$	Variable	CD-Statistics	$\bar{\hat{\rho}}_{ij}$
P_{it}	370.81	0.779	ΔP_{it}	139.6	0.300
Y_{it}	453.33	0.953	ΔY_{it}	213.19	0.458
Pop_{it}	69.75	0.147	ΔPop_{it}	27.78	0.060

Notes. The cross-section dependence $\bar{\hat{\rho}}_{ij}$ is defined as the average of the pair-wise correlation, $\hat{\rho}_{ij}$, of y_{it} and y_{jt}. Test statistics for cross-section dependence is $CD = \sqrt{\frac{2T}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij} \xrightarrow{d} N(0,1)$.

3.3 Correlations across Regions

In Table 4[5], the diagonal presents the first lag serial correlations of the first difference of log real house prices, ΔP_{it}, in each region, and the off-diagonal numbers are the pairwise between-region correlations. Table 5 shows the estimates for the first difference of log real income, ΔY_{it}.

[5]The aggregation into regions as listed in this table is for presentation purposes only, which demonstrate the presence of spatial correlation in changes of log real house prices between regions. This presentation is only to show the importance of accounting for the cross-section dependence in model estimations. The estimates are not intended to provide implications for the spatial correlation pattern between regions in the variables estimated.

In Table 4, we could see that there are apparent cross-section connections among different regions and regions close to each other seem to be more likely to have similar response rates. For example, the changes of the log real house prices in metropolitan Sydney demonstrate negative relationship with regions further from Sydney, such as Far West, Murrumbidgee, and North-Western, but is positively connected with the regions Illawarra, Hunter and Mid North Coast. Besides, the significant positive diagonal estimates show that there are significant positive serial correlations for changes in log real house prices within a region. In contrast, despite the regional correlations of changes in the log real income are strong, there seem to be a spatial homogeneity in the pairwise correlations, since the estimates for the pairwise correlations are similar. But changes in the log real income do not show to be serially auto correlated, which either present insignificant or negative estimates (with an exception of 0.153 for Far West).

Table 4: Pairwise Cross-section Dependence between Regions for the Growth of Log Real House Prices, ΔP_{it}

Region	CW	FW	HT	IW	MN	MB	MR	NW	NT	ST	SE	SY
CW	0.386											
FW	0.308	0.556										
HT	0.783	0.347	0.448									
IW	0.588	0.097	0.851	0.459								
MN	0.830	0.408	0.939	0.830	0.489							
MB	0.646	0.370	0.568	0.260	0.607	0.341						
MR	0.741	0.312	0.640	0.512	0.645	0.496	0.302					
NW	0.593	0.373	0.430	0.192	0.447	0.525	0.636	0.129				
NT	0.738	0.450	0.531	0.273	0.533	0.658	0.808	0.603	0.348			
ST	0.731	0.501	0.889	0.800	0.870	0.461	0.675	0.407	0.647	0.342		
SE	0.864	0.351	0.865	0.688	0.846	0.568	0.739	0.463	0.696	0.834	0.359	
SY	0.364	-0.040	0.561	0.802	0.494	-0.014	0.288	0.039	0.174	0.628	0.473	0.092

Notes. Data series for calculating the estimates are the first order difference of log real house prices at the region-level. Data for each region is calculated by the average across regional LGAs. Labels in the first row and the first column are explained in Table 2. The diagonal numbers are the estimates of the first-lag serial correlations for each region. The non-diagonal numbers are the between-area pairwise correlations.

Table 5: Pairwise Cross-section Dependence between Regions for the Growth of Log Real Income, ΔY_{it}

Region	CW	FW	HT	IW	MN	MB	MR	NW	NT	ST	SE	SY
CW	-0.271											
FW	0.289	0.153										
HT	0.558	0.487	-0.024									
IW	0.647	0.524	0.852	-0.061								
MN	0.647	0.584	0.814	0.868	-0.194							
MB	0.792	0.340	0.349	0.433	0.531	-0.033						
MR	0.718	0.561	0.469	0.671	0.743	0.815	-0.181					
NW	0.841	0.331	0.400	0.483	0.518	0.770	0.630	-0.126				
NT	0.686	0.664	0.650	0.652	0.781	0.647	0.702	0.760	-0.353			
ST	0.422	0.659	0.797	0.818	0.941	0.396	0.676	0.318	0.709	0.020		
SE	0.627	0.586	0.638	0.717	0.885	0.624	0.757	0.543	0.838	0.818	0.097	
SY	0.476	0.599	0.748	0.795	0.695	0.428	0.549	0.435	0.762	0.707	0.745	0.045

Notes. Data series for calculating the estimates are the first order difference of log real income at the region-level. Also see notes in Table 4

4 Empirical Results

4.1 Unit Root Tests

Panel unit root tests are performed to examine the stationarity/non-stationarity of real house prices, real income and population. The Z-statistic recommended by Choi (2001) is calculated and is valid under the assumption of no cross-section dependence. To account for possible cross-section dependence results for the CIPS and CIPSM tests are also reported. The tests are applied to the (log) levels of the series and also to the series in first-differences. Results are reported in Table 6.

Table 6 reports results for a number of different specifications for ADF test regressions. In the upper half of the table the ADF regression includes a constant and allows for either one or two lags of the dependent variables. In the bottom half the ADF regression includes both a constant and a time trend.

The Z-statistics for the variables in levels imply that the null hypothesis of a unit root in all series cannot be rejected, except in one case for real income. In contrast when the variables are first-differenced, the presence of unit root in all series is strongly rejected.

Use the CIPS and CIPSM tests produce somewhat mixed results. At the 5 % level of significance, the model that includes a time trend generally provides little evidence against the null of a unit root in the series. In the case where the time trend is omitted, the CIPSM test with one lag leads to the rejection of the null for all three variables. In general, the other tests do not reject the null. When the variables are first-differenced, there is strong evidence against a unit root.

On balance the results from the panel unit root tests in Table 6 suggest that real house prices, real income and population for NSW LGAs could reasonably be treated as being non-stationary due to a unit root, and the series of these variables in first-differences as being stationary.

Finally, since the economy-wide real interest rate is a potential explanatory variable for house prices, the series is also tested for a unit root [6]. In this case the results are sensitive to whether or not a time trend is included in the model. In the absence of a time trend the ADF test (with one or two lags) does not reject the null hypothesis of a unit root. When a time trend is included in the ADF regressions, the t-statistics are -4.49 for one lag and -3.27 for two lags, so these results point to a real interest rate that is stationary around a deterministic trend.

[6] In response to policy induced disinflation, real interest rates appear to have trended down from the early 1990s to sometime in the 2000s but to have been more or less stable since then. The current unit root test for real interest rates did not take into account the possible structural break in the trend approximately half way through the time period considered. However, macro influences to real house price dynamics, including the impacts of real interest rates, are controlled by common factors, the possible impact from the structural break of real interest will also be captured by the common factors.

Table 6: Panel Unit Root Testing Results

	Z-test (1)	Z-test (2)	CIPS(1)	CIPS(2)	CIPSM(1)	CIPSM(2)
With an intercept						
P_{it}	4.82	6.91	-1.82	-1.51	-2.32	-1.56
Y_{it}	14.64	17.49	-2.06	-1.48	-2.32	-1.85
Pop_{it}	4.95	5.62	-1.69	-1.07	-2.24	-1.91
ΔP_{it}	-22.15	-12.74	-3.35	-2.31	-3.45	-2.20
ΔY_{it}	-30.22	-20.65	-3.14	-2.25	-3.10	-2.06
ΔPop_{it}	-15.07	-10.37	-2.39	-1.83	-3.11	-2.61
With an intercept and a linear trend						
P_{it}	3.76	6.39	-2.28	-1.63	-2.39	-1.77
Y_{it}	-3.79	1.37	-2.04	-1.53	-1.93	-1.54
Pop_{it}	4.62	4.30	-1.69	-1.64	-2.82	-2.17
ΔP_{it}	-14.50	-5.33	-3.63	-2.62	-3.88	-2.80
ΔY_{it}	-23.87	-15.88	-3.57	-2.74	-3.61	-2.80
ΔPop_{it}	-11.90	-8.94	-3.00	-2.31	-3.54	-2.57

Notes. We test for panel unit root for the first lag and the second lag of the panel-specific ADF regressions using Z-test, CIPS and CIPSM tests. The reported values of Z-test are the inverse normal Z-statistics computed by (6). Values for CIPS are computed as the simple average of the individual-specific CADF statistics ($\overline{CADF}$) using (11). Extra factors for computing the CIPSM statistics are $\overline{\Delta P_{it}}$ for Pop_{it}, $\overline{\Delta P_{it}}$ for Y_{it}, $\overline{\Delta P_{it}}$ for Y_{it}. Critical values for CIPS tests or CPISM tests are -2.14, -2.04 and -1.99 for 1 %, 5% and 10% significance respectively in the case of an intercept only, and -2.65, -2.55 and -2.49 for 1%, 5% and 10% significance respectively in the case of an intercept and a linear trend.

4.2 Income Elasticity

Using the econometric framework given in equation (1), the elasticity of real house prices to real income is estimated using a number of different model specifications. In terms of explanatory variables, the most general model estimated is given by:

$$P_{it} = a_i + \beta_{inc,i} Y_{it} + \beta_{pop,i} \Delta Pop_{i,t-1} + \beta_{inter,i} R_t + \beta_{iD} D_{0204} + e_{it} \tag{13}$$

where real house prices are related to real income, the real interest rate, population growth and a dummy variable that takes the value of 1 in years 2002, 2003, and 2004, but zero elsewhere.

As an initial benchmark, the potential effects of cross-section dependence are ignored and estimates are reported using both fixed-effects(FE) and mean group (MG) estimators. The FE estimator assumes there is no heterogeneity in the slope coefficients and is based on pooled data. The MG estimator allows for coefficient heterogeneity and is derived from averaging the coefficient estimates from separate time-series regressions for each LGA.

To account for the possible effects of cross-section dependence, common correlated effects (CCE) estimators are used. With CCE estimators the above model is augmented with the time t cross-section averages of the dependent variable and the (relevant regressors) as follows:

$$P_{it} = a_i + \beta_{inc,i} Y_{it} + \beta_{pop,i} \Delta Pop_{i,t-1} + \beta_{inter,i} R_t + \beta_{iD} D_{0204} + + \gamma_p \bar{P}_t + \boldsymbol{\gamma}_i' \bar{\boldsymbol{X}}_t + e_{it}, \quad (14)$$

where $\bar{\boldsymbol{X}}_t$ includes only those regressors with a cross-section dimension. Two versions of the CCE estimator are used; a mean group estimator (CCEMG) and a pooled estimator (CCEP).

Formulas for standard errors are given in Appendix C. The FE variance-covariance estimator is given in (24), while the estimator for the CCEP is in (29). Variance-covariance formulas for the MG and CCEMG estimators are given in (23) and (28 respectively.

Estimates of different versions of the above models are reported in Tables 7-8. In addition to the coefficient estimates, a CIPS test is applied to the residuals from each regression (see the $\overline{CADF}$ statistic) as a check on whether the specified model is a valid co-integrating relationship. Results for the CD test applied to the residuals are reported as is the estimate of the average pair-wise cross-section dependence in the residuals $\bar{\rho}$.

Columns (1) and (2) of Table 7 report FE and MG estimates for the most restricted model that includes real income as the only regressor. The estimated income elasticity is broadly similar for both estimators; around 2 for the MG estimator and slightly lower at 1.9 for the FE estimator. However in both cases the CD tests strongly reject the hypothesis that the cross-section units are independent. The average correlation coefficient for the residuals is estimated to be approximately 0.4. Cross-section dependence is clearly evident in the data and is likely to affect the FE and MG estimators.

The third and fourth columns of Table 7 report the estimates obtained using the CCEMG (common correlated effects mean group) and CCEP (common correlated effects pooled) estimators. For both of these estimators the average correlation coefficient for the residuals is effectively zero and the CD test statistics are much smaller. The CD statistics for both the CCE estimators do not reject the hypothesis of no cross-section dependence. Both CCE estimators yield the same estimate for the income elasticity of 0.69, which is considerably lower than what is produced by the estimators that do not account for cross-section dependence. The significant difference (between the estimates of 2.0 from FE estimators and 0.69 from CCE estimators) explains the impacts of spatial effects to the estimates and the importance of accounting for the cross section dependence in the panel data models in order to obtain consistent estimates.

In the last four columns of Table 7 results for a model that includes the real interest rate as a regressor are reported. Since the real interest rate has only time variation, it

Table 7: Estimation Results: Income Elasticity of Real House Price: 1991-2012

P_{it}	MG	FE	CCEMG	CCEP	MG	FE	CCEMG	CCEP
$\hat{\beta}_{inc,i}$	2.02	1.86	0.69	0.69	1.36	1.17	0.69	0.69
	(0.060)	(0.180)	(0.220)	(0.170)	(0.070)	(0.223)	(0.215)	(0.179)
$\hat{\beta}_{inter,i}$					-0.08	-0.09	-0.00	0.00
					(0.008)	(0.023)	(0.005)	(0.010)
$\overline{CADF}$	-2.64	-1.79	-2.15	-2.01	-3.01	-2.08	-2.34	-2.10
CD	201.09	171.39	2.28	-0.15	199.49	178.72	2.34	-0.98
$\hat{\bar{\rho}}$	0.42	0.36	0.00	0.00	0.42	0.38	0.00	0.00
R^2	0.67	0.62	0.84	0.83	0.70	0.65	0.84	0.83

Notes. Numbers in the parenthesis are robust standard errors. Robust standard errors for $\hat{\beta}_i$ is given by (23) and (28) for MG and CCEMG respectively, and given by (24) and (29) for FE and CCEP respectively. Critical values for CIPS tests are -2.14, -2.04 and -1.99 for 1%, 5% and 10% significance respectively in the case of an intercept only.
CD is the statistics testing for the significance of cross-section dependence computed by (4). $\hat{\bar{\rho}}$ is the estimate for the average of the pair-wise correlations of the cross-section units computed by (3).

Table 8: Estimation Results with the Addition of Population Growth and the Specific Time Dummy: Income Elasticity of Real House Price: 1991-2012

P_{it}	MG	FE	CCEMG	CCEP	MG	FE	CCEMG	CCEP
$\hat{\beta}_{inc,i}$	1.35	1.12	0.44	0.49	1.41	1.17	0.76	0.69
	(0.078)	(0.460)	(0.120)	(0.210)	(0.070)	(0.215)	(0.150)	(0.179)
$\hat{\beta}_{inter,i}$	-0.09	-0.12	0.00	-0.00	-0.07	-0.09	0.00	0.00
	(0.010)	(0.054)	(0.005)	(0.030)	(0.007)	(0.020)	(0.010)	(0.010)
D_{0204}					0.02	0.01	0.00	0.00
					(0.010)	(0.020)	(0.010)	(0.010)
$\hat{\beta}_{pop,i}$	-2.16	0.21	-1.98	-0.01				
	(0.580)	(0.846)	(0.500)	(0.164)				
$\overline{CADF}$	-3.09	-2.52	-2.68	-1.93	-2.70	-2.09	-3.02	-2.01
CD	171.52	183.69	118.36	113.30	246.82	175.56	4.18	-0.15
$\hat{\bar{\rho}}$	0.38	0.405	0.26	0.24	0.52	0.39	0.01	0.00
R^2	0.86	0.86	0.92	0.92	0.74	0.65	0.89	0.83

Notes. See notes to Table 7. D_{0204} denotes the time dummy for years 2002 to 2004.

could act as a common factor in driving cross-section dependence. Comparing the results obtained from the FE and MG estimators when the real interest rate is included in the model it is evident that there is little reduction in cross-section dependence in the residuals. However the estimated coefficient on the real rate is negatively signed and statistically significant.

When the model including the real interest rate is estimated using the CCEMG and CCEP estimators - while there is no cross-section variation in the real interest rate, real interest rate no longer displays a significant effect. The impacts of real interest rate are absorbed by the augmented terms of common factors.

Table 8 presents estimates for models that include a dummy variable D_{0204} for the period 2002-04, and population growth rates in LGAs. While population growth is found to be statistically significant when the MG and CCEMG estimators are used, it is estimated to have a negative effect on real price levels. This finding is counter-intuitive. In addition use of the FE and CCEP estimators do not produce statistically significant coefficient estimates. Interestingly when population growth is included in the model, there is evidence of residual cross-section dependence, even when the model is estimated using the CCE procedures. There seems to be no robust result for population growth in affecting house prices. This finding is broadly consistent with results in Bodman and Crosby (2003), Otto (2007) and Liu and Otto (2014).

Including the dummy variable to account for the widespread rapid growth in real house prices in the period 2002 to 2004 has little effect on the results. The dummy variable is not statistically significant and the income elasticity for real house prices is estimated to be around 0.7 to 0.8.

4.3 Co-integration of Real House Prices and Real Income

For each of the models reported in Tables 7-8, the $\overline{CADF}$ statistic is computed using the model-residuals. This test statistic can be used to conduct a CIPS test for the present of a unit root in the residuals. Critical values for CIPS test (based on a test regression with intercept, but no time trend) are -2.14, -2.04 and -1.99 for the 1 percent, 5 percent and 10 percent significance levels (respectively). For most of the models estimated the $\overline{CADF}$ statistics are sufficiently negative to reject the null hypothesis of a unit root (no co-integration) in favour of a co-integrating relationship.

As a further test for evidence of co-integration the following set of residuals are constructed,

$$u_{it} = P_{it} - \hat{\beta}_{inc} Y_{it} + \hat{a}_{it} \tag{15}$$

where $\hat{\beta}_{inc} = 0.69$ and is the CCEMG estimate for income elasticity and $\hat{a}_i$ is the LGA-specific intercept term from the individual time series regressions. A CIPS panel unit root test is applied to $\hat{u}_{it}$ (using a test regression with both an intercept and a time trend), and this produces a $\overline{CADF}$ statistic of -2.49. This figure indicates that the null hypothesis of a unit root in $(P_{it} - \hat{\beta}_{inc} Y_{it})$ can be rejected at the 5 percent level of significance [7]. Overall there seems to be support for co-integration between real house prices and real income across NSW.

[7] The inclusion of a time trend is required because a common elasticity is being imposed across all LGAs.

4.4 Panel Error Correction Models

Based on the results in the previous section, it is assumed that there is a long-run or co-integrating relationship between the log of real house prices,P_{it} and the log of real income, Y_{it} with the co-integrating vector (1, -0.69). Since co-integration implies an error-correction model, this model can be used to examine the short-run dynamics of real house prices and provide an estimate of the speed at which real house prices return to long-run equilibrium following a change in real income. The baseline panel error correction model is assumed to be:

$$\Delta P_{it} = a_i + \phi(P_{i,t-1} - \hat{\beta} Y_{i,t-1}) + \varphi \Delta P_{i,t-1} + \omega \Delta Y_{i,t-1} + \mu_{it} \tag{16}$$

where $\hat{\phi}_i$ provides a measure of the speed of adjustment of log real house prices to a shock. $\hat{\varphi}$ describes the long-run impact of a change in Y_{it} on ΔP_{it}. μ_{it} reflects random shocks to the housing markets (e.g. shocks of consumer confidence that affect consumption). ΔP_{it} and ΔY_{it} are first differenced terms of P_{it} and Y_{it} and both of them are proved to be I(0). In a variety of experiments for (16), we also include the population growth rate, and the combination of real interest rate and a linear time trend in the specifications of ECMs[8].

Estimates of the error correction models are reported in Table 9. As an initial benchmark, results for FE and MG estimators are presented. For both estimators the CD statistic provides strong evidence of cross-section dependence. Use of the CCEMG and CCEP estimators lead to a substantial reduction in evidence of residual cross-section dependence and it seems likely that more reliable estimates are obtained from these two procedures. For the CCEP estimator $\hat{\phi}_i$ is -0.14 and this figure does not varying when additional regressors (real interest rate, population growth and time trend) are included in the model. The estimate of $\hat{\phi}_i$ obtained from the CCEMG estimator is somewhat sensitive to the inclusion of additional regressors. It is -0.35 in the most restricted model, but is -0.66 (-0.67) in the more general specifications. In all cases the estimated error-correction coefficient is negative and statistically significant. The only other robust finding is that the estimated coefficient on the current growth rate of real income is positive and significant. Overall these estimates tend to confirm an important role for real income as a fundamental driver of house prices across NSW.

4.5 Factor Loadings

Use of the multi-factor model given by (1) appears to have been quite successful in controlling for cross-section dependence in the panel regressions. It is interesting to estimate the individual factor loadings for each LGA in response to the unobservable

[8]the ADF test in Section 4.1 provides evidence that a linear combination of real interest rate and a time trend forms a stationary co-integrated relationship.

Table 9: Estimation Results: Panel Error Correction Models

P_{it}	MG	FE	CCEMG	CCEP	MG	FE	CCEMG	CCEP	MG	FE	CCEMG	CCEP
$\hat{\phi}$	-0.152	-0.068	-0.349	-0.141	-0.371	-0.086	-0.663	-0.141	-0.375	-0.086	-0.671	-0.141
	(0.012)	(0.007)	(0.034)	(0.013)	(0.019)	(0.008)	(0.040)	(0.013)	(0.021)	(0.047)	(0.041)	(-0.253)
$\hat{\varphi}$	0.154	-0.082	-0.023	-0.253	0.239	-0.075	0.043	-0.252	0.239	-0.075	0.044	-0.253
	(0.026)	(0.019)	(0.025)	(0.019)	(0.026)	(0.018)	(0.024)	(0.019)	0.026	0.019	(0.025)	(0.019)
$\hat{\omega}$	0.373	0.134	0.484	0.137	0.364	0.150	0.420	0.137	0.371	0.150	0.397	0.137
	(0.066)	(0.076)	(0.099)	(0.087)	(0.077)	(0.077)	(0.108)	(0.087)	(0.082)	(0.077)	(0.114)	(0.087)
$Rinter$				0.001	-0.002	0.004	0.001	0.003	0.002	0.003	0.001	0.001
					(0.005)	(0.006)	(0.004)	(0.006)	(0.005)	(0.006)	(0.004)	(0.001)
$Trend$					0.011	-0.002	-0.003	0.001	0.012	0.002	0.002	0.000
					(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)
Pop_{it}									0.431	-0.012	-0.057	0.009
									(0.283)	(0.139)	(0.280)	(0.137)
CD	126.53	142.66	4.518	3.100	133.374	138.568	4.354	2.826	128.045	138.520	4.271	2.824
$\hat{\bar{\rho}}$	0.279	0.314	0.009	0.007	0.294	0.305	0.009	0.003	0.282	0.107	0.009	0.006
R^2	0.107	0.107	0.433	0.329	0.240	0.204	0.529	0.329	0.224	0.224	0.548	0.329

Notes. Estimation model is (16). CD is the statistics for testing for the significance of cross-section dependence computed by (4). $\hat{\bar{\rho}}$ is the estimate for the average of the pair-wise correlations of the cross-section units computed by (3).

common factors that are captured by State-wide average prices and income. The factor loading γ_i for a particular LGA can be estimated using the following model:

$$(P_{i,t-i} - 0.69 * Y_{i,t-1}) = a_i + \gamma_i(\bar{P}_{i,t-1} - 0.69 * \bar{Y}_{i,t-1}) + \tau_{it} \qquad (17)$$

where the dependent variable is the cointegrating relation for real price and real income for the ith LGA and this is regressed on the average of the cointegrating relationships across all LGAs at time t. Based on the previous results these variables should be stationary and by construction the cross-section average of γ_i is unity and of a_i is zero.

Equation (17) has a similar form to the Capital Asset Pricing Model (CAPM) in finance, which relates the equilibrium expected return on a single asset to an aggregate market risk factor (Fama and French 2004). The CAPM implies the expected return on an asset depends on its correlation with the overall market portfolio; its market Beta. The larger is Beta the more risky the asset and the higher expected equilibrium return required to compensate for the risk. In a recent paper Hanewald and Sherris (2013) apply the CAPM to suburb-level house prices in Sydney, however in their analysis Hanewald and Sherris do not initially control for the effect of real income on house prices. They find some evidence that suburbs with higher Betas experience higher ex-post capital gains.

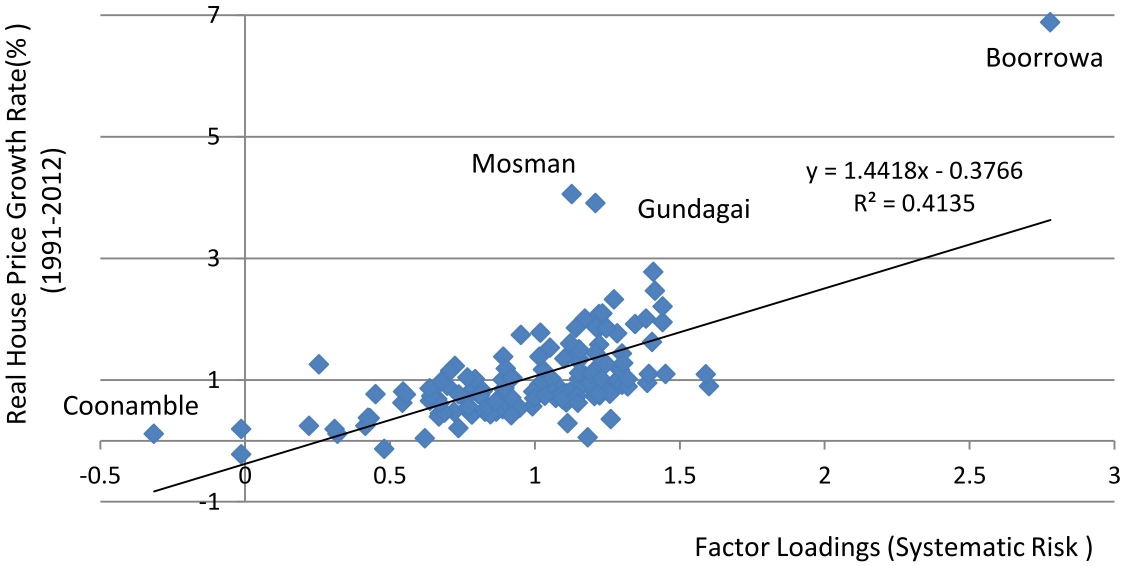
Analogous to the CAPM, the estimate of γ_i indicates how closely house prices in an LGA after adjusting for real income are related to average State-wide house prices again after adjusting for average real income. Thus γ_i is a measure of the market risk of houses in an LGA. A relatively high γ_i implies a higher market risk for the LGA and the CAPM predicts that higher market or systematic risk should be compensated by higher expected returns. This prediction can be tested by considering the relationship between average real house price growth in an LGA and its estimated value of γ_i.

Figure 3 presents a scatter-plot of average real growth rates of house prices (1991-2012) against the estimates of γ_i obtained from (17). The data point to a reasonably strong positive relationship between growth rates of real house prices and the use of $\hat{\gamma}_i$ to measure systematic risk. There are some particularly large outliers in the LGAs of Coonamble ($\hat{\gamma}_i = -0.31$), Boorrowa, Gundagai and Mosman, possbily due to the prexisting relative high land values in these areas. However, overall the results are generally consistent with the CAPM and with Hanewald and Sherriss results for Sydney. One limitation of this analysis and that by Hanewald and Sherris is that real returns to housing only considers capital gains and not rental returns.

4.6 The Ratio of Log Real House Prices to Log Real Income

Primary data analyses in the previous section have shown that despite the heterogeneous initial values in 1991, real house prices for almost all of LGAs have experienced

Figure 3: Factor Loadings vs. Real Capital Gain

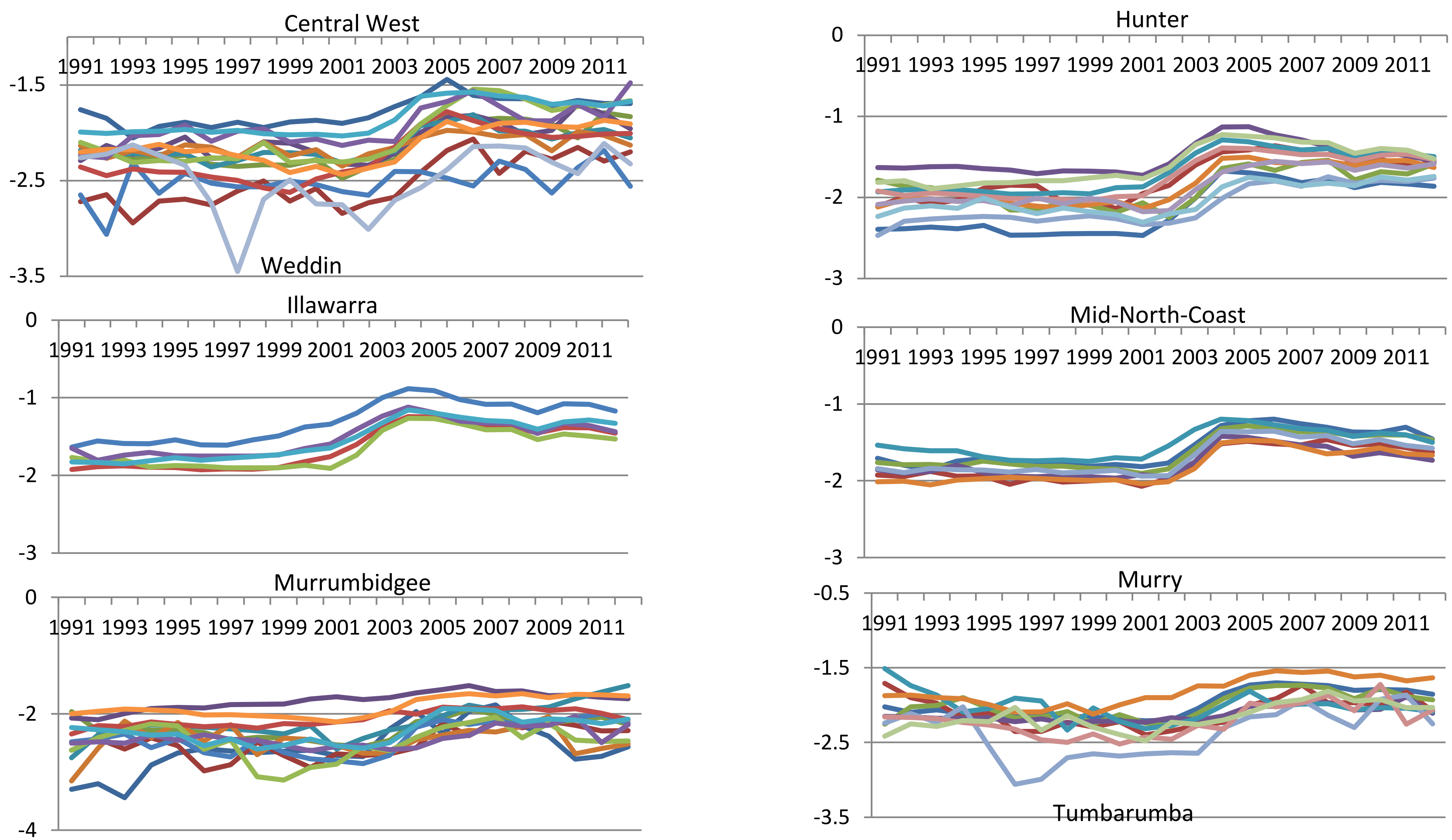


persistent increases with quite parallel growing patterns across the individuals or regions through 1991 to 2004. After 2004, the growth rates across the LGAs are relatively more volatile with some negative reverts but in general, real house prices have maintained their growing trends. There is a very similar case with the development of log real income, but the response rates over time appear to be relatively more stable and with mild volatility.

We have confirmed the long run cointegrating relationship between log real house prices and log real income. It is interesting to look at the LGA specific time profile for the short run dynamics of this relationship, computed as $(P_{i,t-1} - 0.69 * y_{i,t-1})$ or $\Delta(P_{i,t-1} - 0.69 * y_{i,t-1})$. The profiles are plotted in Figures 4-7.

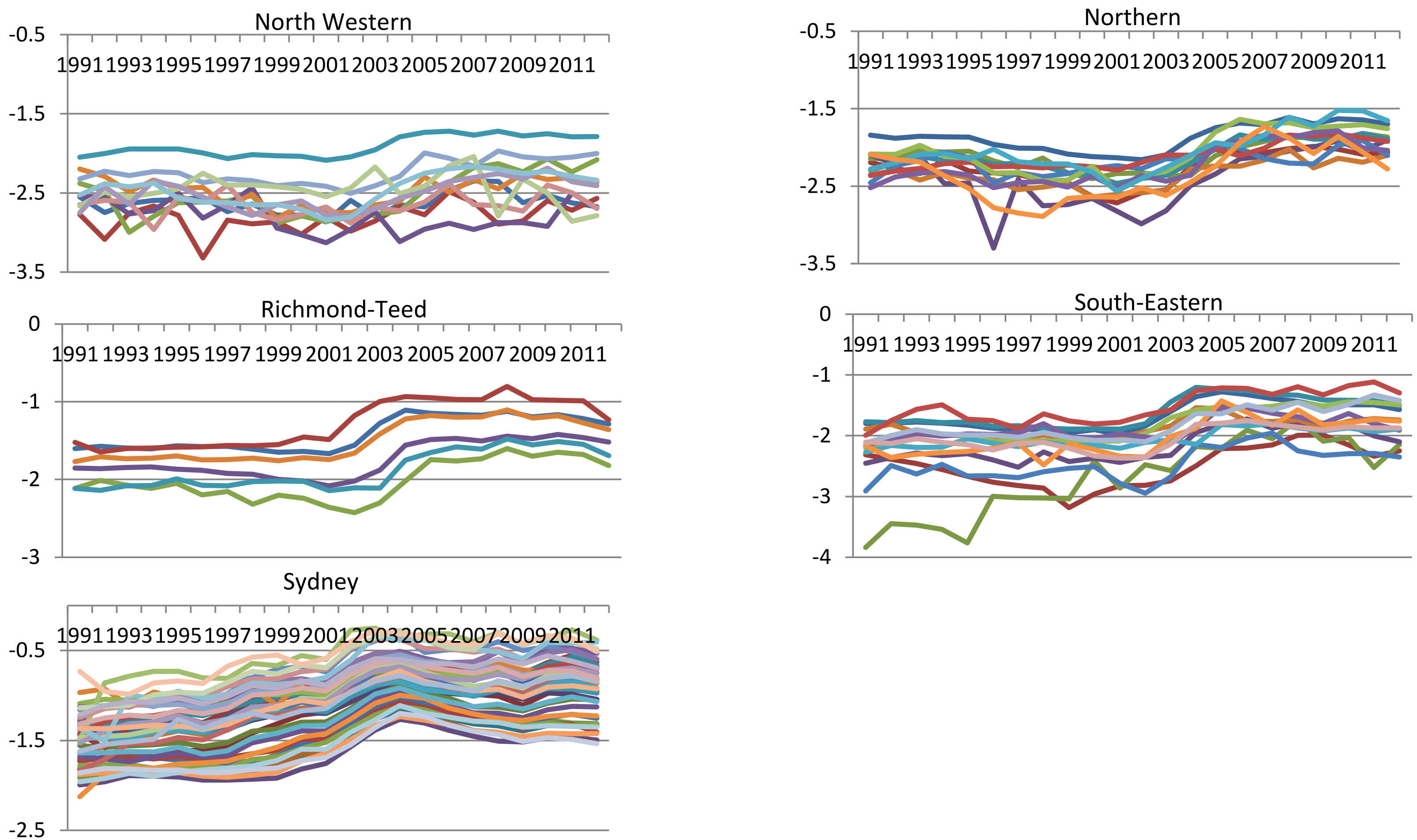
Figures 4-5 shows that there is a linear upward time trend in the development of price to income ratio for the most of LGAs and the patterns are parallel across 80% of LGAs. The excess rise in real house prices appears to be associated with the increasing ratio of price to income. However, the first difference of the time profiles in Figures 6-7 show that the positive changes are followed by the negative reverts to maintain the general equilibrium within LGAs.

Figure 4: LGA Time Profiles for the Weighted Log Ratio of Real House Price to Real Income



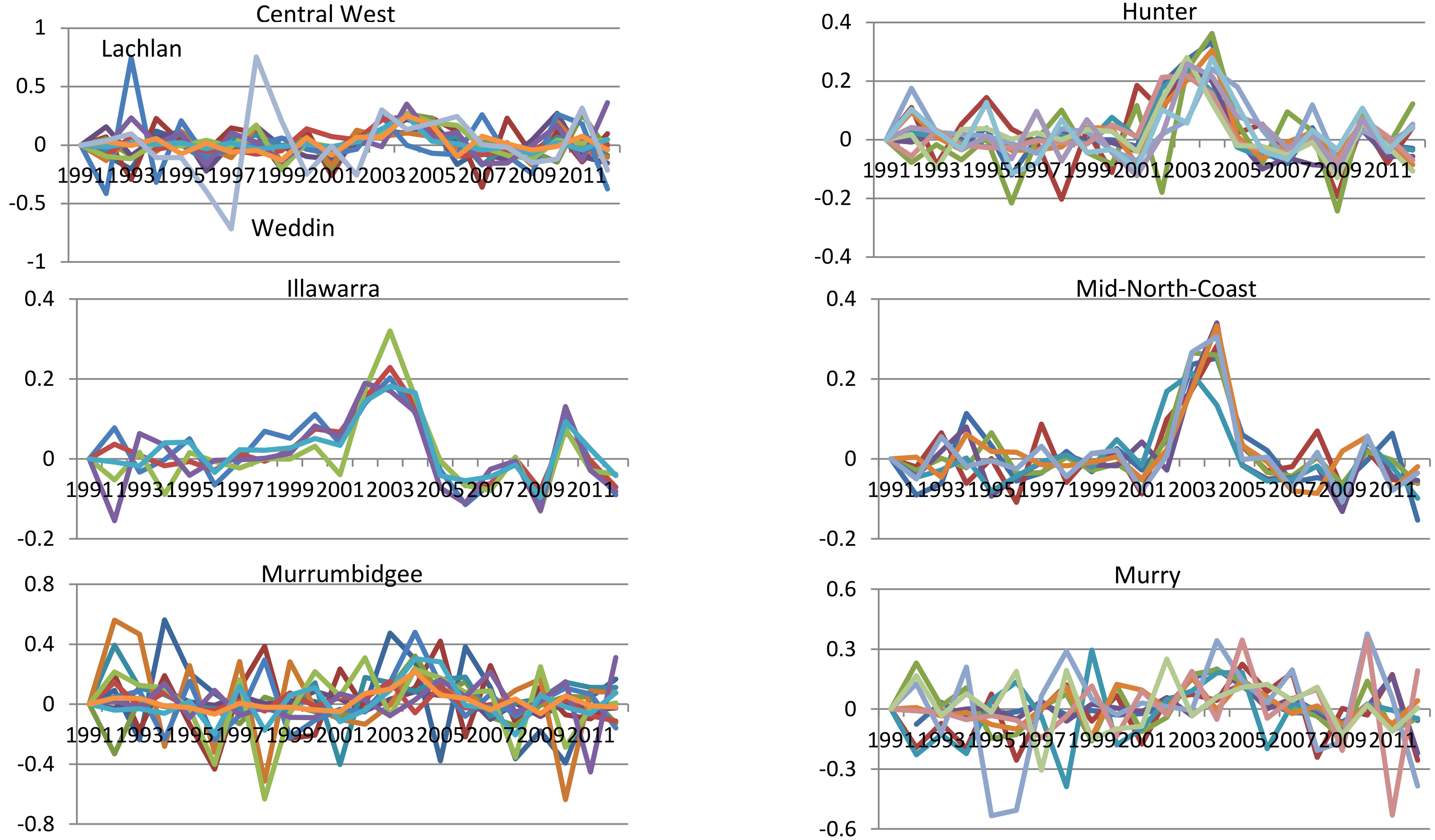
Notes: Each Figure plots LGA-specific values of $P_{it-1} - 0.69 * y_{it-1}$, where P_{it} and y_{it} are the natural log forms of real house prices and real income.

Figure 5: LGA Time Profiles for the Weighted Log Ratio of Real House Price to Real Income- continued



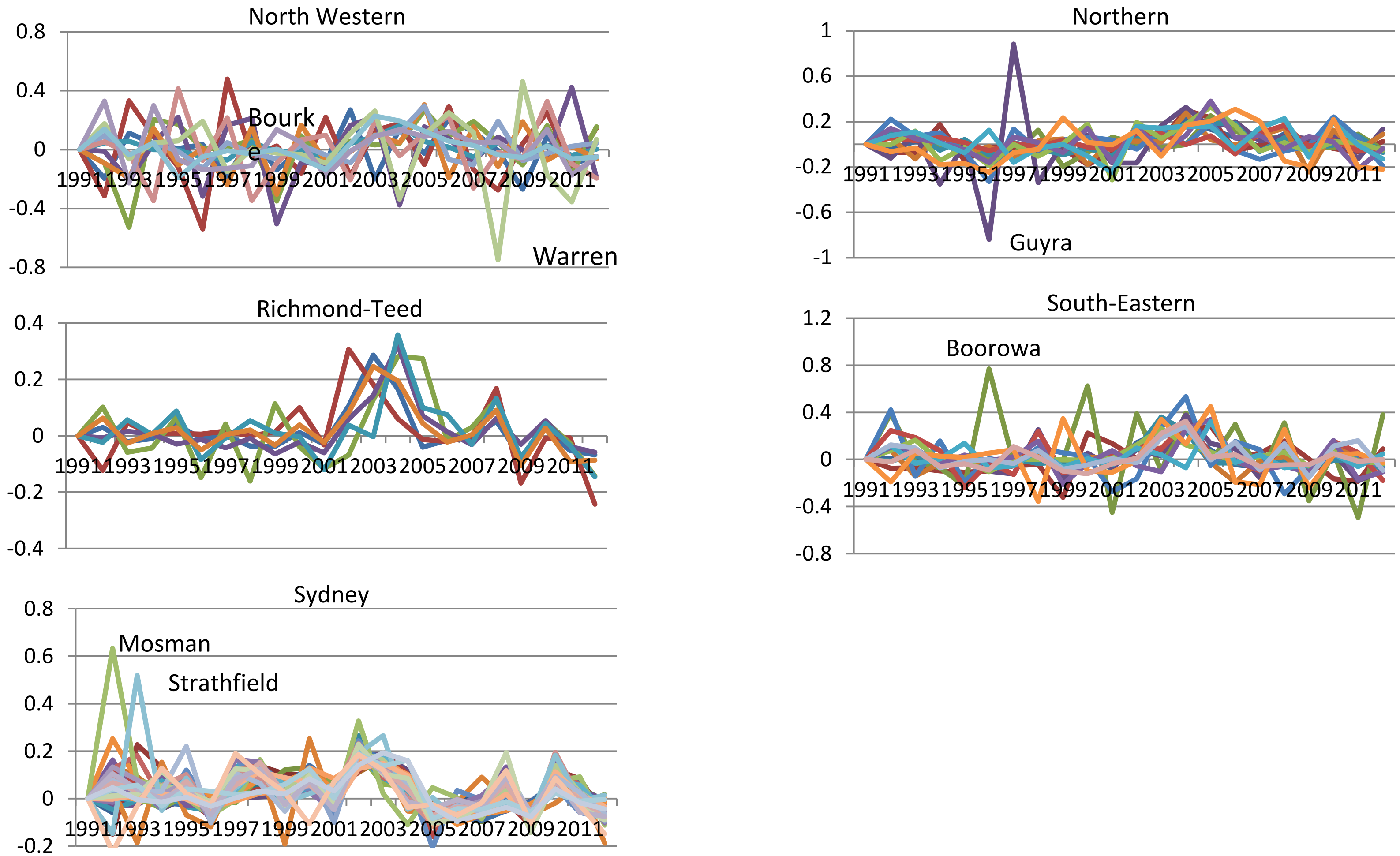
Note: See notes to Figure 4.

Figure 6: LGA Time Profiles for Changes in the Weighted Log Ratio of Real House Price to Real Income - continued



Notes: Each Figure plots the LGA-specific values of $\Delta(P_{it-1} - 0.69 * y_{it-1})$ in the region, where P_{it} and y_{it} are the natural log forms of real house prices and real income.

Figure 7: LGA Time Profiles for Changes in the Weighted Log Ratio of Real House Price to Real Income - continued



Notes: See notes to Figure 6.

5 Estimating Spatial Autocorrelation

Estimates for Model (14) from the previous section have shown that by augmenting the regressors with the cross-section average terms of P_{it} and y_{it}, CCE estimators have successfully reduced the extent of general cross-section dependence in the cross-section units of residuals. But CCE estimators so far do not provide solutions for measuring the extent of such cross-section dependence. In this section, we estimate SAR panel data models with fixed effects and SAR disturbances, which will provide estimates of the response rate of house prices at one location to changes of house prices at its surrounding sites. The model is written as,

$$P_{it} = a_i + \tilde{\rho} \sum_{j=1}^{N} w_{ij} P_{ij} + \boldsymbol{X}_{it} \tilde{\boldsymbol{\beta}} + \tilde{e}_{ij}, \text{where } \tilde{e}_{ij} = \tilde{\lambda} \sum_{j=1}^{N} w_{ij} \tilde{e}_{ij} + \tilde{\varepsilon}_{it} \tag{18}$$

where spatial variation is captured by an additional explanatory variable, $\sum_{j=1}^{N} w_{ij} P_{ij}$, measured as a distance-weighted function of neighboring response values of P_{it} and is usually known as the first order spatial lag of P_{it} in literature. $\hat{\tilde{\rho}}$ is then viewed as the spatial covariate and is intended to capture spatial autocorrelation originating from endogenous processes such as locational attraction, transport congestion, contagious population growth, and movement of censused individuals between the different locales. $\tilde{e}_{ij}$ is the SAR disturbance term and $\tilde{\varepsilon}_{it}$ is the idiosyncratic term assumed as $\tilde{\varepsilon}_{it} \sim i.i.dN(0, \tilde{\sigma}^2)$. w_{ij} is the weight given to i_{th} influence over the j_{th} LGA. Spatial matrix of w_{ij} is designed as follows: 1) compute a pairwise Euclidean distance matrix among the geographic centroid of the LGAs, 2) choose a threshold value D and 3) define $w_{ij} = 0$ if $i = j$ or $d_{ij} > D$, and $w_{ij} = \frac{1}{d_{ij}}$ if $i \neq j$. The spatial matrix using this approach works for the common cases of either binary or distance matrices and may coincident with the spatial dispersion characteristics for the localized house prices. Estimation of this spatial model will provide an alternative set of estimates for the impacts of log real income, population growth, and the real interest rate.

Define $S_i(\lambda)$, and $R_i(\rho) = I_i - \rho W_i$, the log likelihood function of this model is shown as,

$$LnL = -\frac{NT}{2} ln(2\pi\tilde{\sigma}^2) + T[ln|S_i(\lambda)| + ln|R_i(\rho)|] - \frac{1}{2\tilde{\sigma}^2} \sum_{t=1}^{T} \tilde{V}_{it}(\Re) \tilde{V}_{it}(\Re)' \tag{19}$$

where $\tilde{V}_{it}(\Re) = R_i(\rho)[S_i(\lambda)P_{it} - \boldsymbol{X_{it}}\tilde{\boldsymbol{\beta}}]$. Derivatives for this equation is shown in Lee and Yu (2010).

The estimation results are presented in Table 10. Estimates for spatial autocorrelations of either $\hat{\tilde{\rho}}$ or $\hat{\tilde{\lambda}}$ across the cross-section units of log real house prices is estimated to be very strong, 0.84 and 0.87 respectively, and consistent to the addition of different explanatory variables. The income elasticity is robust to model specifications at 0.42

Table 10: Estimation Results: SAR Models with Spatial Disturbances for P_{it}

Spatial Panel Data Models for P_{it}				
$\hat{\hat{\rho}}$	0.833	0.857	0.834	0.857
	(0.032)	(0.024)	(0.032)	(0.024)
$\hat{\hat{\lambda}}$	0.876	0.866	0.874	0.864
	(0.025)	(0.021)	(0.025)	(0.021)
Y_{it}	0.409	0.433	0.403	0.426
	(0.089)	(0.088)	(0.090)	(0.088)
$\Delta Pop_{i,t-1}$			0.511	0.498
			(0.193)	(0.194)
Rinter		0.048		0.044
		(0.044)		(0.043)

Notes. The spatial panel data model for P_{it} is (18).

and statistically significant, which is broadly in line with the CCE estimate of 0.69. The population growth has a positive significant effect of 0.50, while the effect of the real interest rate is not significant.

6 Conclusions

This paper estimates the income elasticity of house prices over a relatively long-term time period of 1991 to 2012 for 144 LGAs in New South Wales of Australia. I find that real house prices across all the LGAs in the State of New South Wales have experienced persistent upward increases over 1991 to 2012. The period of 2002-2004 is unique as the real house prices increased 36.76 % on average, and in 17 regional LGAs the value of houses doubled. The confirmation of a positive co-integrating relationship between real income and real house prices implies that real income across most of the LGAs have also witnessed persistent increases (possibly) to certain extent to maintain the equilibrium level of the housing price affordability.

The income elasticity of house prices is estimated to be 0.69 using panel data methods that can account for non-stationarity and cross-section dependence. After controlling for a linear time trend, the estimate confirms a co-integrated relationship exists between real house prices and real income. Real house prices and real income moves together co-integratedly in the long run. Positive changes in the relationship are followed by the negative changes around a linear time trend to maintain the equilibrium relationship. House prices in one location are proved to be connected with prices in other spatial units through a distance weighted matrix, where the constant coefficient is 0.84. The impacts for prices in a location from the closer locations are much stronger than the further locations. After controlling for such an impact, an alternative estimate of the income elasticity(0.46) is obtained.

Appendices

Appendix A Data for Economic Factors

A.1 How are LGA boundary changes over time taken into account?

The Australian Standard Geographic Classification (ASGC) provided a common framework of statistical geography which enabled the production of statistics that were comparable and could be spatial. For our focus research period, there are ASGC 1991, ASGC 1996, ASGC 2001, ASGC 2006 and ASGC 2011 [9]. Data during years of 1991 to 1996 are recorded under the code of ASGC 1991; data from 1997 to 2001 are recorded under the code of 1996, and hence so forth. In the meantime, ABS provides concordances in ASGC if there are changes in statistical geography boundary, such as LGA amalgamations. In thesis, I use ASGC 2006 as the standard for geographical boundary to concord all of variables involved. In consequence, for population, income, and regulatory variables, all data have been concorded to ASGC 2006 at the LGA level using concordances for ASGC 1991, ASGC 2001, ASGC 2006, and ASGC 2011. In addition, when I process the satellite based data to obtain the geographical variables, I also use 2006 ASGC LGA level digital boundary. Therefore, all the data used for analysis in thesis are consistently observed at LGA level with geographical boundary based on ASGC 2006.

A.2 Property Prices

Median prices for LGAs are sourced from the quarterly Rent and Sales Reports published by Housing NSW.

http://www.housing.nsw.gov.au/About+Us/Reports+Plans+and+Papers/Rent+and+Sales+Repo

A.3 Population

The following sources provide data for Estimated Resident Population for LGAs on end of financial year basis (end-June) for the period 1995-96 to 2011-12. Source: Regional

[9] In 2011 census of Population and Housing, the Australian Bureau of Statistics (ABS) introduced the Australian Statistical Geography Standard (ASGS), which replaces the Australian Standard Geographical Classification (ASGC). However, the historical ASGC standard is also available to classify the 2011 Census data at the level of Statistical Local Areas (SLAs).

Population Growth, Australia, 1996-2006, 2008-09, 2011 (cat. no. 3218.0).

For data for the period 1990-91 to 1994-95 we use data from tables produced by the UNSW Local Grants Commission. This reports (preliminary) population estimates over the period 1990-91 to 2010-11. No data are reported for 1991-92, so we simply use the average of 1990-91 and 1992-93 for this year. Source: ABS Estimated Resident Population of Statistical Local Areas New South Wales at 30 June 1990 Preliminary (Cat. No. 3210.1)

A.4 Income

Estimates of income for LGAs are based on data from the Australian Taxation Offices Taxation Statistics. The Bureau of Infrastructure, Transport and Regional Economics (BITRE) has a database derived from the ATO data that report by LGA real income per taxpayer (in 2007-08 prices) for the period 1990-91 to 2005-06. The original figures for nominal taxable income are deflated using the CPI for Australia.

For the period 2006-07 to 2009-10 the ABS reports data by LGA for nominal income per taxpayer in their National Regional Profile 2007-2011. We convert these figures to constant 2007-08 prices using the CPI for Australia.

As we have no income data by LGA for the financial years 2010-11 and 2011-12, we simply assume that the growth rate of real income per taxpayer in both 2010-11 and 2011-12 is equal to the growth rate for 2009-10.

A.5 Real Interest Rate

Data on the real interest rate is obtained from the RBA spreadsheet Capital Market Yields (F2). It is the yield on the Australian Government inflation-indexed bond with the longest maturity.

A.6 Consumer Price Index

The consumer price index for Sydney is obtained from ABS release 6401.0 - Consumer Price Index, Australia.

Appendix B Names of Regions and Regional LGAs

Table 11: Regions and Regional Local Government Areas Names

Region	LGAs	Region	LGAs
Central West	Bathurst Regional	Illawarra	Shellharbour
Central West	Bland	Illawarra	Shoalhaven
Central West	Blayney	Illawarra	Wingecarribee
Central West	Cabonne	Illawarra	Wollongong
Central West	Cowra	Mid North Coast	Bellingen
Central West	Forbes	Mid North Coast	Clarence Valley
Central West	Lachlan	Mid North Coast	Coffs Harbour
Central West	Lithgow	Mid North Coast	Greater Taree
Central West	Mid-Western Regional	Mid North Coast	Hasting
Central West	Oberon	Mid North Coast	Kempsey
Central West	Orange	Mid North Coast	Nambucca
Central West	Parkes	Murrumbidgee	Carrathool
Central West	Weddin	Murrumbidgee	Coolamon
Far West	Broken Hill	Murrumbidgee	Cootamundra
Hunter	Cessnock	Murrumbidgee	Griffith
Hunter	Dungog	Murrumbidgee	Gundagai
Hunter	Gloucester	Murrumbidgee	Hay
Hunter	Great Lakes	Murrumbidgee	Junee
Hunter	Lake Macquarie	Murrumbidgee	Leeton
Hunter	Maitland	Murrumbidgee	Lockhart
Hunter	Muswellbrook	Murrumbidgee	Narrandera
Hunter	Newcastle	Murrumbidgee	Temora
Hunter	Port Stephens	Murrumbidgee	Wagga Wagga
Hunter	Singleton	Murry	Albury
Hunter	Upper Hunter Shire	Murry	Berrigan
Illawarra	Kiama	Murry	Corowa Shire

continued on next page

Table 12: Regions and Regional Local Government Areas Names- continued

Region	LGAs	Region	LGAs
Murry	Deniliquin	Northern	Tamworth Regional
Murry	Greater Hume Shire	Northern	Tenterfield
Murry	Murray	Northern	Uralla
Murry	Tumbarumba	Northern	Walcha
Murry	Wakool	Richmond-Tweed	Ballina
Murry	Wentworth	Richmond-Tweed	Byron
North Western	Bogan	Richmond-Tweed	Kyogle
North Western	Bourke	Richmond-Tweed	Lismore
North Western	Cobar	Richmond-Tweed	Richmond Valley
North Western	Coonamble	Richmond-Tweed	Tweed
North Western	Dubbo	South-Eastern	Bega Valley
North Western	Gilgandra	South-Eastern	Bombala
North Western	Narromine	South-Eastern	Boorowa
North Western	Walgett	South-Eastern	Cooma-Monaro
North Western	Warren	South-Eastern	Eurobodalla
North Western	Warrumbungle Shire	South-Eastern	Goulburn Mulwaree
North Western	Wellington	South-Eastern	Harden
Northern	Armidale Dumaresq	South-Eastern	Palerang
Northern	Glen Innes Severn	South-Eastern	Queanbeyan
Northern	Gunnedah	South-Eastern	Snowy River
Northern	Guyra	South-Eastern	Tumut Shire
Northern	Inverell	South-Eastern	Upper Lachlan Shire
Northern	Liverpool Plains	South-Eastern	Yass Valley
Northern	Moree Plains	South-Eastern	Young
Northern	Narrabri		

Notes: Local government areas in Sydney are listed in Table 13.

Table 13: Sydney's Local Government Areas

Sydney		
Ashfield	Holroyd	Penrith
Auburn	Hornsby	Pittwater
Bankstown	Hunter's Hill	Randwick
Baulkham Hills	Hurstville	Rockdale
Blacktown	Kogarah	Ryde
Blue Mountains	Ku-ring-gai	Strathfield
Botany Bay	Lane Cove	Sutherland Shire
Burwood	Leichhardt	Sydney
Camden	Liverpool	Warringah
Campbelltown	Manly	Waverley
Canada Bay	Marrickville	Willoughby
Canterbury	Mosman	Wollondilly
Fairfield	North Sydney	Woollahra
Gosford	Parramatta	Wyong
Hawkesbury		

Appendix C Estimators for (1)

Case 1 FE estimator - no Common Factors but with Spatial error. For the panel data model (1)-(2), with the assumptions (1-3,4a and 5a) in Appendix hold, MG and FE estimators are summarized as followed. Derivation details for MG estimators could be found in (Pesaran and Smith 1995).
let $M_D = \boldsymbol{I_T} - \mathbf{1(1'1)1'}$, assuming, $\boldsymbol{\gamma_i'} = 0$, for $i = 1, 2, 3..., N$,

$$\hat{\beta}_i = (X_i' M_D X_i)^{-1} X_i' M_D y_i \tag{20}$$

$$\hat{\beta}_{MG} = N^{-1} \sum_{i=1}^{N} (X_i' M_D X_i)^{-1} X_i' M_D y_i \tag{21}$$

$$\hat{\beta}_P = (\sum_{i=1}^{N} (X_i' M_D X_i)^{-1} \sum_{i=1}^{N} X_i' M_D y_i \tag{22}$$

Consistent non-parametric estimators of the asymptotic variance of $\widehat{\beta}_{MG}$ and $\widehat{\beta}_P$ provided by Pesaran (2006) are,

$$\widehat{Asy.var} \hat{\beta}_{MG} = \frac{1}{N(N-1)} \sum_{i=1}^{N} (\hat{\beta}_i - \hat{\beta}_{MG})(\hat{\beta}_i - \hat{\beta}_{MG})' \tag{23}$$

$$\widehat{Asy.var} \hat{\beta}_P = \frac{1}{N} Q_{NT}^{-1} \Lambda_{NT} Q_{NT}^{-1} \tag{24}$$

where

$$Q_{NT}^{-1} = \frac{1}{N-1} Q_{NT}^{-1} \sum_{i=1}^{N} T^{-1} X_i' M_D X_i$$

$$\Lambda_{NT} = \frac{1}{N-1} Q_{NT}^{-1} \sum_{i=1}^{N} T^{-1} X_i' M_D X_i (\hat{\beta}_i - \hat{\beta}_{MG})(\hat{\beta}_i - \hat{\beta}_{MG})' T^{-1} X_i' M_D X_i$$

Case 2 CCE Estimator individual fixed effects, unobserved Common Factors, with Spatial error. For the panel data model (1)-(2), with the assumptions (1-3,5,6 and 7) in Appendix hold, CCEMG and CCEP estimators are summarized as followed. Derivation details for MG estimators can be viewed in Appendix.
let $\bar{M} = I_T - \bar{H}(\bar{H}'\bar{H})^{-1})\bar{H}'$, where $\bar{H} = (\mathbf{1}, \bar{Z}), and \bar{Z}_{.t} = (\bar{Y}_{.t}, \bar{X}_{.t}')'$,

$$\hat{\beta}_{CCE,i} = (X_i'\bar{M}X_i)^{-1} X_i'\bar{M}y_i \tag{25}$$

$$\hat{\beta}_{CCEMG} = N^{-1} \sum_{i=1}^{N} \hat{\beta}_{CCE,i} \tag{26}$$

$$\hat{\beta}_P = (\sum_{i=1}^{N} (X_i'\bar{M}X_i)^{-1} \sum_{i=1}^{N} X_i'\bar{M}y_i \tag{27}$$

A consistent non-parametric estimator of the asymptotic variance of $\hat{\beta}_{CCEMG}$ and $\hat{\beta}_{CCEP}$ provided by Pesaran (2006) are,

$$\widehat{Asy.var}\hat{\beta}_{CCEMG} = \frac{1}{N(N-1)} \sum_{i=1}^{N} (\hat{\beta}_{CCE,i} - \hat{\beta}_{CCEMG})(\hat{\beta}_{CCE,i} - \hat{\beta}_{CCEMG})' \tag{28}$$

$$\widehat{Asy.var}\hat{\beta}_{CCEP} = \frac{1}{N} Q_{NT}^{*-1} \Lambda_{NT}^{*} Q_{NT}^{*-1} \tag{29}$$

where

$$Q_{NT}^{*-1} = \frac{1}{N-1} \sum_{i=1}^{N} T^{-1} X_i' \bar{M} X_i$$

$$\Lambda_{NT}^{*} = \frac{1}{N-1} \sum_{i=1}^{N} T^{-1} X_i' \bar{M} X_i (\hat{\beta}_{CCE,i} - \hat{\beta}_{CCEMG})(\hat{\beta}_{CCE,i} - \hat{\beta}_{CCEMG})' T^{-1} X_i' M_D X_i$$

Remark: The non-parametric variance-covariance estimators given by (23) or (24) are no longer consistent when the idiosyncratic term is spatially correlated (Pesaran and Tosetti 2011). However, (28)(29) continue to yield consistent estimates and are robust to both spatial and serial error correlations (Pesaran 2006, and Pesaran and Tosetti 2011). Besides, in the case of homogeneous slopes, an alternative variance matrix estimator allowing for the spatial temporal effects could be a non-parametric variance matrix estimator that adapts the Newey and West (1987)s heteroskedasticity autocorrelation consistent (HAC) procedure called spatial, heteroskedasticity, autocorrelation (SHAC) estimator (Pesaran and Tosetti 2011).

Appendix D Individual Time Profiles

Figure 8: Individual Time Profiles for Real House Prices NSW

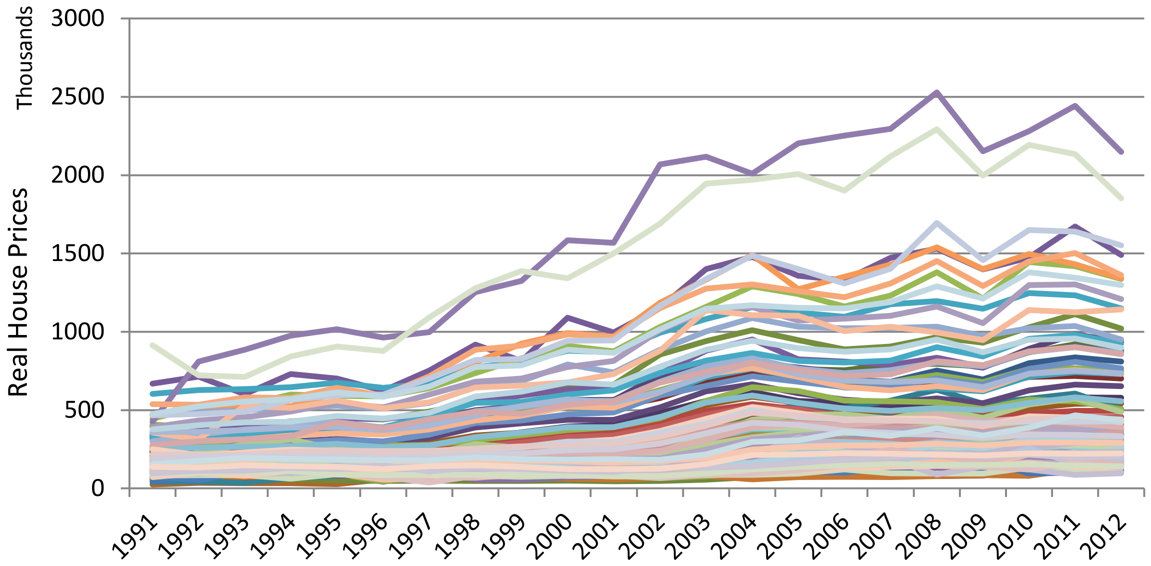
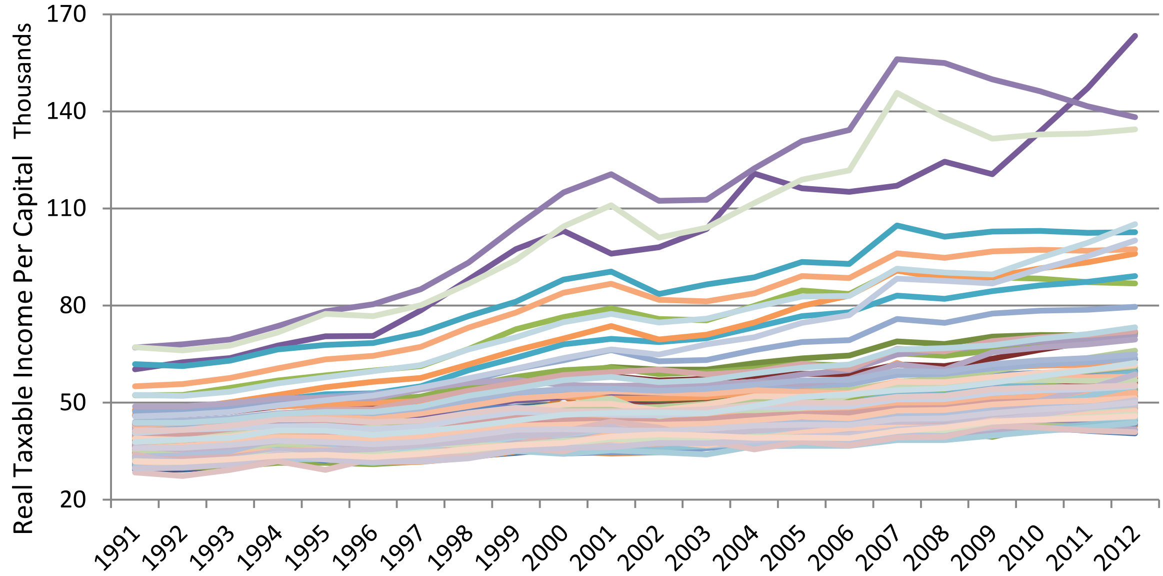


Figure 9: Individual Time Profiles for Real Income



Appendix E CCE Estimator Procedure

CCE estimator starts from assuming that the observed individual-specific regressors $\boldsymbol{X}_{it}$ could be approximated by the common factors, d_t and $\boldsymbol{f}_t$, shown as,

$$\boldsymbol{X}_{it} = \boldsymbol{A'd}_t + \boldsymbol{\Gamma}'\boldsymbol{f}_t + \boldsymbol{v}_{it} \tag{30}$$

where $\boldsymbol{A}_i$ and $\boldsymbol{\Gamma}_i$ are $1 \times k$ and $m \times k$ factor loading matrices, where m is the number of unobserved common effects factors, and k is the number of observed variables. $\boldsymbol{v}_{it}$ is the random components of $\boldsymbol{X}_{it}$, which is assumed to be distributed independently of the common effects, $\boldsymbol{f}_t$, across i.

The primary characteristic of CCE estimator is the application of using the cross-section averages of Z_{it} as proxies for the unobserved common factor with $\bar{Z}_{.t} = (\bar{Y}_{.t}, \bar{X}'_{.t})$, where $\bar{Y}_{.t} = N^{-1}\sum_{i=1}^{N} y_{it}$, $\bar{X} = N^{-1}\sum_{i=1}^{N} x_{it}$. Assuming rank $(\Gamma'_i)= m < k+1$, the asymptotic property of f_t using $\bar{h}_t = (\bar{d}_t, \bar{z}'_t)'$ as observable proxies is shown as,
$f_t = (\bar{C}\bar{C}')^{-1}\bar{C}(\bar{z}_t - \bar{B}'\bar{d} - \bar{\mu}_t)$
$f_t - (\bar{C}\bar{C}')^{-1}\bar{C}(\bar{z}_t - \bar{B}'\bar{d}_t) \xrightarrow{p} 0$, if $C \xrightarrow{p} \bar{C}$ and $\bar{\mu}_t \xrightarrow{p} 0$, as $N \xrightarrow{p} 0$, as $N \xrightarrow{p} \infty$

This suggests using $\hat{\bar{h}}_t = (\bar{d}, \bar{z}_t)$ as observable proxies for f_t . This procedure shows that β_i and their means β can be consistently estimated by augmenting the OLS or pooled regressions of Y_{it} on X_{it} with the cross-section averages $\bar{h}_t$ (Pesaran 2006, Pesaran and Tosetti 2011).

Assumption 1. R has bounded row and finite fourth-order cumulants
Assumption 2. Individual-Specific Errors
Assumption 2a. ε_{it} is assumed to be independently distributed of the independent variables, X_{it}, cross-sectionally uncorrelated but follow a linear stationary process with absolute summable autocovariances, that is,
$\varepsilon_{it} = \sum_{j=0}^{p} a_{ij}\xi_{i,t-j}$, where $\xi_{i,t-j} \sim i.i.d(0,\epsilon^2)$ and $var(\varepsilon_{it}) < \infty$ Assumption 2b. v_{it} is the random components of X_{it}, which is assumed to be distributed independently of the common effects, f_t, and across i, and follow a linear stationary process with absolute summable autocovariances, that is,
$v_{it} = \sum_{j=0}^{p} b_{ij}\vartheta_{i,t-j}$, where $\vartheta_{i,t-j} \sim i.i.d(0, v_i^2)$ and $var(v_{it}) < \infty$
Assumption 3. Random slope coefficients
Individual slope coefficients β_i are allowed to be heterogeneous across LGAs and follow a random coefficient model denoted as,
$\beta + i = \beta + \delta$, $\delta_{i.} \sim i.i.d(0, \Omega_\beta)$ where $\|\beta\| < k$, $\|\Omega_v < k$, Ω_v is a $ktimesk$ symmetric nonnegative definite matrix, and the random deviations δ_i are distributed independently of the factor loadings of γ_i, Γ_i, and the individual specific effects, ε_{it}, and v_{it}.
Assumption 4. Model Identification Condition
Defining,

$\hat{\Psi}_{iT} = T^{-1}(X_i'\bar{M}_D X_i)$
$\hat{\Psi}_{NT} = \dfrac{1}{N}\sum_{i=1}^{N} T^{-1}(X_i'\bar{M}_D X_i)$
4a) To identify β_i, the $ktimesk$ individual observation matrix $\hat{\Psi}_{iT}$ is bounded and non-singular.
4b) To identify β, the $k \times k$ pooled observation matrix $\hat{\Psi}_{NT}$ is bounded and non-singular. $\hat{\Psi}_{iT}^{-1}$ and $\hat{\Psi}_{NT}^{-1}$ have finite second order moments for all i.
Assumption 5. Common effects
5a)$(d_t', X_{it}',)'$ and ε_{js} are independently distributed for all I, t, j and s.
5a) The $(n+m) \times 1$ vector $g_t = (d_t', f_t')$ is a covariance stationary process, with absolute summable auto-covariance, distributed independently of the individual specific errors of ε_{it}, and v_{it}, but allowed to be presented with unit roots and deterministic trends, which in turn would introduce units roots in the individual-specific regressors X_{it}(Kapetanios, Pesaran, and Yamagata,2006).
Remark: It is worth noting that the common feature dynamics across i are captured through the serial correlation structure of the common effects, and individual specific dynamics are allowed through serial correlation in ε_{it}.
Assumption 6. Ranking condition Let $\tilde{\Gamma} = E(\Gamma_i, \gamma_i) = (\Gamma, \gamma)$ and Rank($\tilde{\Gamma}$)= m
Remark: The rank condition Rank($\tilde{\Gamma}$)= m ensures that under Assumptions 1-7, $T^{-1}(\bar{H}'\bar{H}')$ converges to a positive definite matrix for a fixed T as $N \to \infty$. $T^{-1}(X_i'\bar{M}X_i)$ and its limit exist even if the rank condition is not satisfied.
Assumption 7. Model Identification Condition for Common factors. Defining
$\bar{M} = I_T - G(G'G)_{-1}G'$, with $G = (D, \bar{Z})$, D is a $T \times 1$ vector of unit, and $\bar{Z}$ is a $T \times (k+1)$ matrix of observations on $\bar{z}_{.t} = (\bar{y}_{.t}, \bar{x}_{.t})$,
$\hat{\Psi}_{iT}^{*} = T^{-1}(X_i'\bar{M}X_i)$
$\hat{\Psi}_{NT}^{*} = \dfrac{1}{N}\sum_{i=1}^{N} T^{-1}(X_i'\bar{M}X_i)$
7a) To identify β_i, the $k \times k$ individual observation matrix $\hat{\Psi}_{iT}$ is bounded and non-singular.
7b) To identify β, the $k \times k$ pooled observation matrix $\hat{\Psi}_{NT}$ is bounded and non-singular. $\hat{\Psi}_{iT}^{*-1}$ and $\hat{\Psi}_{NT}^{*-1}$ have finite second order moments for all i.

E.1 Asymptotic Property of Individual Slope Coefficients Estimation by CCE Estimators

Defining the individual slope coefficient as,

$$\hat{\beta}_i = (X_i'\bar{M}X_i)_{-1}X_i'\bar{M}y_i \tag{31}$$

Using (1) and (**??**),we have,

$$\hat{\beta}_i - \beta = (\frac{X_i'\bar{M}X_i}{T})_{-1}(\frac{X_i'\bar{M}F}{T})_{-1}\gamma_i + (\frac{X_i'\bar{M}X_i}{T})_{-1}(\frac{X_i'\bar{M}\varepsilon_i}{T})_{-1} \tag{32}$$

This equation shows the direct dependence of $\hat{\beta}_i$ on the unobserved factors through $\frac{X_i'\bar{M}F}{T}$. With Assumption 7, $\frac{X_i'\bar{M}X_i}{T}$ is bounded irrespective of the rank condition. In addition, where $F \subset G$, then $\bar{M}F = 0$ and,

$$\frac{X_i'\bar{M}F}{T}\gamma_i = O_P(\frac{1}{\sqrt{N}}) + O_P(\frac{1}{\sqrt{T}}) \tag{33}$$

Therefore, with conditions being satisfied in Assumption 4 , $(\hat{\beta}_i - \beta_i)$ could be written as,

$$(\hat{\beta}_i - \beta_i) = (\frac{X_i'\bar{M}X_i}{T})_{-1}(\frac{X_i'\bar{M}\varepsilon}{T})_{-1} + O_P(\frac{1}{\sqrt{N}}) + O_P(\frac{1}{\sqrt{T}}) \tag{34}$$

This equation tells that the finite T-distribution of $(\hat{\beta}_i - \beta_i)$ will not depend on the factor loadings as $N \to \infty$, but depend on the probability density of ε_i. Recalling ε_i is a stationary process independently distributed of X_i and G, the consistency of $\hat{\beta}^i$ could be obtained with $(\frac{X_i'\bar{M}X_i}{T})_{-1} \xrightarrow{p} \sum_i$. Therefore, in the case when T is fixed or N is relatively larger than T, namely, $T/N \to c$, as $N \to \infty$, the asymptotical distribution of $\sqrt{T}(\hat{\beta}_i - \beta_i)$ will be shown as,

$$\sqrt{T}(\hat{\beta}_i - \beta_i) \xrightarrow{d} N(0, \sum \beta_i) \tag{35}$$

where $\sum \beta_i$ could be consistently estimated as,

$$\sum \beta_i = \sum_i^{-1} S_{i\varepsilon} \sum_i^{-1}$$

$$\sum_i^{-1} = p \overset{T\to\infty}{lim}\ T^{-1}(X_i'\bar{M}\Omega_{\varepsilon i}\bar{M}X_i)$$

$$\Omega_{\varepsilon i} = E(\varepsilon_i \varepsilon_i')$$

E.2 Asymptotic Property of CCEMG Estimator

CCEMG estimator is designed in consideration to the case that there might be a substantial amount of unobserved heterogeneity among the cross-section units which lead to slope coefficients of the model cross-section specific. It is shown as,

$$\hat{\beta}_{MG} = N^{-1}\sum_{i=1}^{N}\hat{\beta}_i with \hat{\beta}_i defined by 43 \tag{36}$$

Using (32) and Assumption 3,

$$\sqrt{N}(\hat{\beta}_{CCEMG} - \beta) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\delta_i + \frac{1}{N}\sum_{i=1}^{N}(\frac{X_i'\bar{M}X_i}{T})(\frac{\sqrt{N}X_i'\bar{M}F}{T})\gamma_i + \tag{37}$$

$$\frac{1}{N}\sum_{i=1}^{N}(\frac{X_i'\bar{M}X_i}{T})(\frac{\sqrt{N}X_i'\bar{M}\varepsilon_i}{T})$$

Using Assumption 5,

$$\frac{1}{N}\sum_{i=1}^{N}(\frac{X_i'\bar{M}X_i}{T})(\frac{\sqrt{N}X_i'\bar{M}\varepsilon_i}{T}) = O_P(\frac{1}{\sqrt{N}} + O_P(\frac{1}{\sqrt{T}}) \tag{38}$$

Using (33) and Assumption 2,

$$\frac{1}{N}\sum_{i=1}^{N}(\frac{X_i'\bar{M}X_i}{T})(\frac{\sqrt{N}X_i'\bar{M}F}{T})\gamma_i = \Delta_{NT} + O_P(\frac{1}{\sqrt{N}} + O_P(\frac{1}{\sqrt{T}}) \tag{39}$$

where $\Delta_{NT} = \frac{1}{N}\sum_{i=1}^{N}(\frac{X_i'\bar{M}X_i}{T})(\frac{\sqrt{N}X_i'\bar{M}\varepsilon_i}{T})$, with $E(\Delta_{NT}) = 0$, and $Var(\Delta_{NT}) = O_P(\frac{1}{T})$ Combing (37-39),

$$\sqrt{N}(\hat{\beta}_{CCEMG} - \beta) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\delta_i + O_P(\frac{1}{\sqrt{N}} + O_P(\frac{1}{\sqrt{T}}) \tag{40}$$

The distribution of $\sqrt{T}(\hat{\beta}_{CCEMG} - \beta)$ will be asymptotically normal shown as,

$$\sqrt{N}(\hat{\beta}_{CCEMG} - \beta) \xrightarrow{d} N(0, \sum_{MG}^{*}) \tag{41}$$

Since the time-invariant variability of β_i dominates the other sources of randomness in the model (Pesaran, Tosetti, 2011). Robust estimators for MG can be obtained following the non-parametric approach employed in Pesaran (2006), which makes use of estimates of β computed for different cross-sectional units. Using non-parametric approach by Pesaran (2006), the estimator of $\sum_{CCEMG}^{*}$ is consistently estimated by,

$$\frac{1}{(N-1)}\sum_{i=1}^{N}(\hat{\beta}_{CCEMG} - \beta)(\hat{\beta}_{CCEMG} - \beta)' \tag{42}$$

E.3 Asymptotic property of CCEP Estimators

When the individual slope coefficients β_i are homogeneous, CCEMG estimators lose efficiency in its estimates whereas pooled estimators provide consistent and efficient estimates (Pesaran, Shin and Smith, 1999). A pooled estimator of β that accounts for common effects is given as,

$$\hat{\beta}_i = (\sum_{i=1}^{N} X_i' \bar{M} X_i)_{-1} \sum_{i=1}^{N} X_i' \bar{M} y_i \tag{43}$$

References

Alexander, Carol and Michael Barrow (1994). "Seasonality and cointegration of regional house prices in the UK". In: *Urban Studies* 31.10, pp. 1667–1689.

Anselin, Luc (1988). *Spatial econometrics: methods and models*. Vol. 4. Springer.

Anselin, Luc, Julie Le Gallo, and Hubert Jayet (2008). "Spatial panel econometrics". In: *The econometrics of panel data*. Springer, pp. 625–660.

Ashworth, John and Simon C Parker (1997). "Modelling regional house prices in the UK". In: *Scottish Journal of Political Economy* 44.3, pp. 225–246.

Bai, Jushan and Serena Ng (2004). "A PANIC attack on unit roots and cointegration". In: *Econometrica* 72.4, pp. 1127–1177.

Baltagi, Badi, Seuck Heun Song, and Won Koh (2003). "Testing panel data regression models with spatial error correlation". In: *Journal of econometrics* 117.1, pp. 123–150.

Baltagi, Badi H, Peter Egger, and Michael Pfaffermayr (2007). *A generalized spatial panel data model with random effects*. Tech. rep. Working Paper, Syracuse University.

Bodman, Philip M and Mark Crosby (2003). "How far to fall? Bubbles in major city house prices in Australia". In: *Australian Property Journal* 38.3.

Breitung, Jörg and Samarjit Das (2008). "Testing for unit roots in panels with a factor structure". In: *Econometric Theory* 24.01, pp. 88–108.

Cerrato, Mario and Nicholas Sarantis (2007). "A bootstrap panel unit root test under cross-sectional dependence, with an application to PPP". In: *Computational Statistics & Data Analysis* 51.8, pp. 4028–4037.

Choi, In (2001). "Unit root tests for panel data". In: *Journal of international money and Finance* 20.2, pp. 249–272.

Chudik, Alexander, M Hashem Pesaran, and Elisa Tosetti (2011). "Weak and strong cross-section dependence and estimation of large panels". In: *The Econometrics Journal* 14.1, pp. C45–C90.

Cook, Steven (2003). “The convergence of regional house prices in the UK”. In: *Urban Studies* 40.11, pp. 2285–2294.

Fama, Eugene F and Kenneth R French (2004). “The capital asset pricing model: theory and evidence”. In: *Journal of Economic Perspectives*, pp. 25–46.

Frees, Edward W (1995). “Assessing cross-sectional correlation in panel data”. In: *Journal of econometrics* 69.2, pp. 393–414.

Hadri, Kaddour (2000). “Testing for stationarity in heterogeneous panel data”. In: *The Econometrics Journal* 3.2, pp. 148–161.

Hanewald, Katja and Michael Sherris (2013). “Postcode-Level House Price Models for Banking and Insurance Applications”. In: *Economic Record* 89.286, pp. 411–425.

Harris, Richard DF and Elias Tzavalis (1999). “Inference for unit roots in dynamic panels where the time dimension is fixed”. In: *Journal of econometrics* 91.2, pp. 201–226.

Holly, Sean, M Hashem Pesaran, and Takashi Yamagata (2011). “The spatial and temporal diffusion of house prices in the UK”. In: *Journal of Urban Economics* 69.1, pp. 2–23.

Holly, Sean, M Hashem Pesaran, and Takashi Yamagata (2010). “A spatio-temporal model of house prices in the USA”. In: *Journal of Econometrics* 158.1, pp. 160–173.

Im, KS and M Hasham Pesaran (1997). “Y. Shin (1997):Testing for Unit Roots in Heterogeneous Panels”. In: *Cambridge: University of Cambridge*.

Kapoor, Mudit, Harry H Kelejian, and Ingmar R Prucha (2007). “Panel data models with spatially correlated error components”. In: *Journal of Econometrics* 140.1, pp. 97–130.

Lee, Lung-fei and Jihai Yu (2010). “Estimation of spatial autoregressive panel data models with fixed effects”. In: *Journal of Econometrics* 154.2, pp. 165–185.

LeSage, James P and R Kelley Pace (2010). “Spatial econometric models”. In: *Handbook of Applied Spatial Analysis*. Springer, pp. 355–376.

Levin, Andrew, Chien-Fu Lin, and Chia-Shang James Chu (2002). “Unit root tests in panel data: asymptotic and finite-sample properties”. In: *Journal of Econometrics* 108.1, pp. 1–24.

Liu, Xiangling and Glenn Otto (2014). *Housing Supply Elasticity in Sydney Local Government Areas*. SSRN: Working Papers Series. UNSW Australian School of Business Research Paper No. 2014-13.

Maddala, Gangadharrao S and Shaowen Wu (1999). “A comparative study of unit root tests with panel data and a new simple test”. In: *Oxford Bulletin of Economics and statistics* 61.S1, pp. 631–652.

Meen, Geoffrey (1999). “Regional house prices and the ripple effect: a new interpretation”. In: *Housing studies* 14.6, pp. 733–753.

Moon, Hyungsik Roger and Benoit Perron (2004). “Testing for a unit root in panels with dynamic factors”. In: *Journal of econometrics* 122.1, pp. 81–126.

Newey, Whitney K and Kenneth D West (1987). “Hypothesis testing with efficient method of moments estimation”. In: *International Economic Review* 28.3, pp. 777–787.

Otto, Glenn (2007). “The growth of house prices in Australian capital cities: What do economic fundamentals explain?” In: *Australian Economic Review* 40.3, pp. 225–238.

Pesaran, M Hashem (2004). *General diagnostic tests for cross section dependence in panels*. CESifo working paper series.

— (2006). “Estimation and inference in large heterogeneous panels with a multifactor error structure”. In: *Econometrica* 74.4, pp. 967–1012.

— (2007). “A simple panel unit root test in the presence of cross-section dependence”. In: *Journal of Applied Econometrics* 22.2, pp. 265–312.

Pesaran, M Hashem and Ron Smith (1995). “Estimating long-run relationships from dynamic heterogeneous panels”. In: *Journal of econometrics* 68.1, pp. 79–113.

Pesaran, M Hashem and Elisa Tosetti (2011). “Large panels with common factors and spatial correlation”. In: *Journal of Econometrics* 161.2, pp. 182–202.

Pesaran, M Hashem, L Vanessa Smith, and Takashi Yamagata (2013). “Panel unit root tests in the presence of a multifactor error structure”. In: *Journal of Econometrics* 175.2, pp. 94–115.

Sarafidis, Vasilis, Takashi Yamagata, and Donald Robertson (2009). “A test of cross section dependence for a linear dynamic panel model with regressors”. In: *Journal of econometrics* 148.2, pp. 149–161.