Defining Local Housing Markets using Repeat Sales Data

Michael Beenstock

Department of Economics

Hebrew University of Jerusalem

Michael.beenstock@mail.huji.ac.il

Dan Feldman

Economic Consultant

danmsf@gmail.com

Daniel Felsenstein

Department of Geography

Hebrew University of Jerusalem

Daniel.felsenstein@mail.huji.ac.il

January 25, 2019

A new empirical approach to identify local housing markets (LHM's) is proposed, which focuses on the spatial correlation between local house price indices constructed from repeat sales data. It extends the work of Pryce (2013) who claimed that if housing in different locations are perfect substitutes, their house price indices should be perfectly correlated over time. Repeat sales data for house prices in Tel Aviv during 1998 – 2014 are used to construct house price indices for almost 100 census tracts. These price indices are used to define LHMs, the number of which varies inversely with the pairwise correlation cut-off, and with the degree of spatial contiguity. Results point to considerable spatial heterogeneity in house price movement. This belies the popular impression that the Tel Aviv housing market is relatively homogeneous, characterized by expensive housing and uniform house price movements.

1. Introduction

Defining local housing markets (LHM) is conceptually simple, but practically difficult. Conceptually, areas A and B belong to the same LHM if residing in A and B are perfect substitutes. In the nature of things, A and B are proximate and even contiguous. They are perfect substitutes when amenities in A and B are identical, and when their housing qualities are identical. Under these circumstances unit house prices in A and B should be the same. The problem is that housing is not homogeneous for physical, environmental and social reasons. Moreover, housing characteristics and amenities are imperfectly observed in practice. The empirical problem of defining LHMs has much in common with defining capital submarkets. According to capital asset pricing theory, assets belong to the same submarket if their market risk is the same. The measurement of market risk has been the focus of research as well as much controversy for more than half a century (Cuthbertson and Nitzsche, 2008).

Historically, interest in defining LHM's has been concerned with 'housing submarkets' and 'housing market segmentation'. These were invariably local in nature and geographically defined. Early work dates back to Fisher and Fisher (1954) and Grigsby (1963) who considered the housing market as a set of quasi-independent submarkets with dynamics captured by local household movement and price changes. Central to this approach was the assumption of heterogeneity of structural (physical) and socio-economic attributes of housing units. This approach was eventually superseded by the AMM (Alonso-Mills-Muth) model, which conceptualized housing markets in terms of a trade-off between accessibility to the central business district and housing space. The AMM model was rooted in the assumption of homogenous housing units yielding housing services over time and tending to long run equilibrium. Eventually this neo-classical framework was extended by hedonic theory which accommodated heterogeneity of dwellings attributes within a unitary market (Galster 1996).

Given these antecedents, the identification of LHM's has invariably been concerned with estimating hedonic house prices for these exogenously defined areas. Areas constitute an LHM if their hedonic prices differ from standard reference prices (Goodman 1981). In this respect, the willingness to pay more for physical and socio-economic attributes (such as better schools or access to amenities) within given geographical boundaries, constitutes membership of an LHM in a 'weak', de facto sense. It also implies that the attributes of dwellings or their environments are crucial in determining the LHMs to which they belong. In this vein, studies have focused on dwelling sizes (Bajic 1985) and dwelling types (Allen et al 1995). Several such studies (Basu and Thibodeau 1998, Gillen, Thibodeau and Wachter 2001, Tu, Sun and Yu 2007) allow for spatial autocorrelation in their house price regressions.

Membership of LHM's can alternatively be defined in a “strong” sense. In this case, consumer preferences for housing express themselves directly though prices rather than indirectly through locational and spatial attributes (Pryce 2013), and houses belong to the same LHMs when their prices are perfectly positively correlated, induced by perfects substitution. According to Pryce, houses that are perfect substitutes may have heterogeneous attributes.

In practice, weak LHM membership arises when hedonic prices in different locations are the same. Because hedonic house price models typically explain between 50 and 70 percent of the variance of house prices, weak LHM membership is generally inconsistent with strong LHM membership, and strong LHM membership does not imply weak membership (Appendix 1). House prices may be perfectly correlated within LHMs despite the fact that their hedonic prices differ, and if their hedonic prices happen to be the same, house prices within LHMs may even be poorly correlated.

A further dichotomy concerns the derivation of the boundaries of LHMs. In many studies LHM boundaries are imposed exogenously. For example, Palm (1979), Schnare and Stryk (1976) and Watkins (2001) use exogenously defined spatial units such as postcodes and census tracts. Israel hedonic house price studies invariably use exogenously derived administrative boundaries ('statistical areas') to capture neighborhood or environmental effects (Borukhov et al1978, Gabriel 1984, Gat 1996, Sayag 2012). Hwang and Thill (2009) use 'fuzzy' clustering where buildings and spatial units may belong to more than one LHM. A variation on this theme involves identifying LHM attributes rather LHM geometries. This allows for observing the dynamics of urban processes. Grinberger and Felsenstein (2017) for example, ascribe synthetic socio-economic attributes to residential buildings. They compare the prices of dwellings sold in these buildings with their expected prices generated from their attributes. They cluster these differences to identify ‘hot spots’ of urban change. Positive differences are used to indicate processes of urban changes such as gentrification while negative differences indicate downward mobility.

Alternatively, LHM boundaries may be left to emerge through modeling and statistical techniques. In this case, the data determine the geography of LHMs. Dale-Johnson (1982) and Watkins (1999) have used factor analysis and principal components methods to stratify LHMs. Using transactions data for Santa Clara County and Glasgow, they identify 10 and 8 homogenous market segments respectively. Goodman and Thibodeau (1998) and Bourassa et al (1999) compare hedonic house prices after controlling for amenities and housing characteristics. In these studies geographical boundaries are modeled rather than imposed. Geographical boundaries are considered the starting units of the analysis and are then aggregated up to the level of submarkets based on the significance of hedonic price differences across the units. Goodman and Thibodeau (1998) use hierarchical methods to analyze the correspondence between LHMs and school districts in Dallas, and decompose differences between LHMs in terms of price effects and hedonic effects.

Table 1 Taxonomy of LHM Studies

	Imposed	**Emergent**
Strong		Goodman and Thibodeau (2007) Pryce (2013)
Weak	Fisher and Fisher (1954) Bajic (1985) Allen et al (1995) Adair (1996) Maclennan and Tu (1996)	Bhattacharjee (2016) Bourassa (1999, 2003) Goodman and Thibodeau (1998), Helbich et al (2013)

Cross–classifying LHM membership ('weak' and 'strong') with LHM geometry ('imposed' or 'emergent') suggests a taxonomy of studies (Table 1). For example in Goodman and Thibodeau (1998) willingness to pay for better schooling constitutes a (weak) LHM whose geometry is emergent rather than imposed. Extending this approach, Goodman and Thibodeau (2007) compare a-spatial (price-determined) and non-contiguous housing submarkets with spatially contiguous submarkets, using transaction data for Dallas, Texas. The spatial submarkets emerge by combing adjacent block groups and school districts. A-spatial markets are created by combining price per sq foot with property size. They find that the two approaches yield similar results raising questions relating to the need for capturing spatial variation in house prices. Helbich et al use a 4-step data driven methodology to derive 'weak' data-driven LHM's. They generate area-wide surfaces of marginal prices using regression and interpolation in the first stage. The variability in the resultant surfaces (maps) is reduced in the next stage by principle component analysis. The principle components are subsequently clustered using a dedicated algorithm (see also Royuela and Duque 2013). Finally, the number and spatial coherence of the emergent submarkets is tested using the out-of-sample prediction error of a hedonic pricing model. In similar vein, Keskin and Watkins (2017) compare different housing submarket partitioning methods particularly highlighting the role of subjective, qualitative identification by real estate experts. Their results show that subjective identification tends to perform as well as the alternatives (imposing or clustering boundaries) and that in the absence of appropriate micro data, the agent based methods for housing submarkets definition can be considered.

In contrast, Bourrassa et al (1999) use a three-step approach to generate LHM's. (Table 1). Initially they calculate factor scores for 18 housing characteristics. They then apply cluster analysis to these factor scores (accounting for 80 percent of the variance) to form hypothesized LHMs. Finally, they estimate cross-section hedonic house price regressions for each hypothesized LHM in which (self-assessed) house prices are regressed on their factor scores. The hypothesized LHMs are expected to have different hedonic house prices.

The Bourassa-type approach however is not without criticism. First, the list of potential attributes comprising the LHM is incomplete. It excludes noise, pollution, crime, school quality and many other characteristics. Second, 20 percent of the variance of these characteristics is excluded by design. Third, since the characteristics include house prices, there is an obvious endogeneity problem in the hedonic regressions. Fourth, the residuals of the hedonic regressions are assumed to be spatially uncorrelated. Fifth, there is no spatial spillover between submarkets. In a related paper, Bourassa et al (2003) show that LHMs matter for house prices in Auckland. They test whether LHMs can improve out-of-sample hedonic predictions, and compare the effect of alternative definitions of LHMs on the accuracy of hedonic predictions. Specifically, realtor generated LHMs are contrasted with statistically generated LHMs consisting of dwellings that are similar in characteristics but not necessarily spatially contiguous. Their results underscore that the definition of LHMs depends on their purpose. Subsequently, Bourassa, Cantoni and Hoesli (2007) showed that the specification of submarket variables in hedonic price models outperforms geostatistical and lattice models.

The spatial dependence ignored by Bourassa et al (1999) is the key focus of Bhattacharjee et al (2016). Here too the authors struggle with housing heterogeneity in a Portuguese city. They introduce spatial dependence in the estimation of factor scores, and they compare spatial econometric with geographically weighted regression for estimating hedonic house prices. However, their conceptual framework is essentially similar to that in Bourassa et al (1999, 2003).

As mentioned, Pryce (2013) attaches importance to strong LHM membership. He also determines LHM boundaries emergently in terms of the spatial correlation in house prices. Whereas Bourassa et al focus on levels of hedonic house prices, Pryce focuses on changes in house prices. A practical problem with Pryce's approach is that it requires panel data on house prices. In the absence of such data, Pryce used hedonic methods to construct synthetic panel data for Glasgow. It is obvious that synthetic panel data are subject to error and are likely to generate results different to genuine panel data.

This paper extends and generalizes the Pryce approach in which LHMs are defined in the strong sense, and their boundaries are emergent rather than imposed. We make several contributions. First, we suggest some generalizations to Pryce's methodology. Second, we show that if locations A and B belong to the same LHM, house prices in A and B should be perfectly positively correlated over time. Third, in the absence of panel data for house prices, we use repeat sales data for Israel to calculate the spatial correlation matrix between intra-city house prices. Because repeat sales do not occur in each period, we use a variant of the repeat sales methodology proposed by Bailey, Muth and Nourse (1963) to construct heterogeneous house price indices by location (Peng 2012). The advantage of repeat sales data is that they obviate the need for hedonic pricing. We show that if repeat sales house price indices between locations A and B are perfectly positively correlated, A and B belong to the same submarket. We provide an empirical application using repeat sales data for house prices in Tel Aviv.

2. Pryce Theory

According to Pryce (2013), if the cross-price elasticity of demand (CPED = ε) between goods i and j is infinite, their cross-price elasticity of price (CPEP = η) should be 1. We show that this claim is true under restrictive conditions. However, if CPED is infinite, we show that the spatial correlation over time between pairs of house price indices is expected to approach 1.

We extend Pryce's 2-good model for goods i and j to include a third good (k). The equilibrium condition for good i is:

$$D_i(P_i, P_j, P_k; Z) = S_i(P_i, P_j, P_k; Y) \qquad (1)$$

where D denotes demand, S denotes supply, P denotes prices, and Z and Y are common shift variables for demand and supply respectively. It may be shown that Pryce's cross-price elasticity of price (η_{ij}) is:

$$\eta_{ij} = \frac{D_{ij} - S_{ij}}{S_{ii} - D_{ii}} \frac{P_j}{P_i} + \Omega \qquad (2)$$

$$\Omega = \frac{S_{ik} - D_{ik}}{D_{ii} - S_{ii}} \frac{P_j dP_k}{P_i dP_j} + \frac{P_j}{P_i(D_{ii} - S_{ii})} \left(S_{iY} \frac{dY}{dP_j} - D_{iZ} \frac{dZ}{dP_j} \right) \qquad (3)$$

where e.g. $D_{ij} > 0$ denotes the effect of P_j on the demand for i. Pryce omitted $S_{ij} < 0$ in the numerator of the first term in equation (2). He also assumed that Ω = 0 because good k does not feature in his model, and because he implicitly assumed that supply and demand shifters Y and Z don't change. Since $S_{ii} > 0$ and $D_{ii} < 0$, η_{ij} is positive when Ω = 0. The first component of Ω is positive if house prices increase in j and k. The second component is positive if the term in large brackets is negative. Hence, Pryce over-estimates CPEP if Ω < 0. In general, however, it is difficult to quantify the bias in Pryce's CPEP.

The first term in equation (2) may be rewritten in terms of price elasticities (ε) of supply and demand:

$$\frac{D_{ij} - S_{ij}}{S_{ii} - D_{ii}} \frac{P_j}{P_i} = \frac{\varepsilon_{ij}^D - \varepsilon_{ij}^S}{\varepsilon_{ii}^S - \varepsilon_{ii}^D} \frac{dP_j}{dP_i} \qquad (4)$$

Equation (4), which is expected to be positive, states that CPEP varies directly with CPED (ε_{ij}^{D}), and tends to 1 as CPED (ε^{D}) tends to infinity. However, since Ω is not zero Pryce's contention is incomplete. In what follows we show instead that as CPED tends to infinity prices become perfectly correlated over time.

CPED and Correlations

Continuing with a 3-good set-up we loglinearize supply and demand as follows:

$$\ln D_{it} = \alpha_i - \beta_i \ln P_{it} + \gamma_i \ln P_{jt} + \delta_i \ln P_{kt} + \phi_i \ln Z_t + u_{it} \qquad (5)$$

$$\ln S_{it} = \kappa_i + \lambda_i \ln P_{it} - \pi_i \ln P_{jt} - \theta_i \ln P_{kt} + \psi_i \ln Y_t + v_{it} \qquad (6)$$

where time periods are labelled by t, and u and v are spatially uncorrelated iid random variables. In equations (5) and (6) P refers to price indices in locations i, j and k. The CPED parameters are represented by β, δ and γ, where β exceeds γ and δ because own price elasticities exceed cross price elasticities. Also, λ exceeds π and θ. The partial equilibrium solution for house prices in i is:

$$\ln P_{it} = \frac{\alpha_i - \kappa_i + (\gamma_i + \pi_i)\ln P_{jt} + (\delta_i + \theta_i)\ln P_{kt} + \phi_i \ln Z_t - \psi_i \ln Y_t}{\lambda_i + \beta_i} + w_{it} \qquad (7a)$$

$$w_{it} = \frac{u_{it} - v_{it}}{\lambda_i + \beta_i} \qquad (7b)$$

$CPEP_{ij}$ and $CPEP_{ik}$ are represented by the coefficients of $\ln P_j$ and $\ln P_k$, which vary directly, as expected, with CPED via γ_i and δ_i. The CPEPs are less than one because $\gamma_i + \pi_i$ and $\delta_i + \theta_i$ are less than $\lambda_i + \beta_i$. However, CPEP tends to one as CPED tends to infinity, as claimed by Pryce. For example, if $CPED_{ij} = \infty$, $\gamma_i = \beta_i = \infty$ in which case $\frac{\gamma_i + \pi_i}{\lambda_i + \beta_i} = 1$, and equation (7a) simplifies to $\ln P_{it} = \ln P_{jt}$ because the coefficients of $\ln P_k$, lnZ, lnY and u - v are zero. Hence, the relative price of i and j is independent of Z, Y u and v because these goods (houses) are perfect substitutes in consumption.

The partial equilibrium solution for P_k is:

$$lnP_{kt} = \frac{\alpha_k - \kappa_k + (\beta_k + \lambda_k)lnP_{it} + (\gamma_k + \pi_k)lnP_{jt} + \phi_k lnZ_t - \psi_k lnY_t + u_{kt} - v_{kt}}{\delta_k + \theta_k} \qquad (7b)$$

The CPEPs are less than one because $\beta_k + \lambda_k$ and $\gamma_k + \pi_k$ are less than $\delta_k + \theta_k$.

The general equilibrium (reduced form) solutions for house prices are:

$$lnP_{it} = \Gamma_{i0} + \Gamma_{i1} lnZ_t + \Gamma_{i2} lnY_t + \omega_{it} \qquad (8a)$$

$$\omega_{it} = \Psi_{i0} w_{it} + \Psi_{i1} w_{jt} + \Psi_{i2} w_{kt} \qquad (8b)$$

where the Γs and Ψs are defined in Appendix 2. These reduced form parameters have a spatial structure in which each parameter depends on the structural parameters for locations i, j and k. Notice also that the reduced form residuals (ω) have a spatial structure too. Despite the fact that w is spatially uncorrelated, the reduced form residuals (ω_i) are spatially autocorrelated.

Equations (8) imply that the covariance between log prices for i and j is:

$$\text{cov}(\ln P_i \ln P_j) = \Gamma_{i1}\Gamma_{ji} \text{ var}(\ln Z) + \Gamma_{i2}\Gamma_{j2} \text{ var}(\ln Y) + (\Gamma_{i1}\Gamma_{j2} + \Gamma_{i2}\Gamma_{j1}) \text{cov}(\ln Z \ln Y) + \text{cov}(\omega_i \omega_j) \qquad (9a)$$
$$\text{cov}(\omega_i \omega_j) = \Psi_{i0}\Psi_{j0} \text{ var}(w_i) + \Psi_{i1}\Psi_{j1} \text{ var}(w_j) + \Psi_{i2}\Psi_{j2} \text{ var}(w_k)$$

and the variances of log house prices for i and j are:

$$\text{var}(\ln P_i) = \Gamma_{i1}^2 \text{ var}(\ln Z) + \Gamma_{i2}^2 \text{ var}(\ln Y) + 2\Gamma_{i1}\Gamma_{i2} \text{ cov}(\ln Z \ln Y) + \text{var}(\omega_i) \qquad (9b)$$
$$\text{var}(\omega_i) = \Psi_{i0}^2 \text{ var}(w_i) + \Psi_{i1}^2 \text{ var}(w_2) + \Psi_{i2}^2 \text{ var}(w_k)$$
$$\text{var}(\ln P_j) = \Gamma_{j1}^2 \text{ var}(\ln Z) + \Gamma_{j2}^2 \text{ var}(\ln Y) + 2\Gamma_{j1}\Gamma_{j2} \text{ cov}(\ln Z \ln Y) + \text{var}(\omega_j) \qquad (9c)$$
$$\text{var}(\omega_j) = \Psi_{jo}^2 \text{ var}(w_i) + \Psi_{j1}^2 \text{ var}(w_j) + \Psi_{j2}^2 \text{ var}(w_k)$$

The correlation between log house prices for i and j is defined as:

$$r_{ij} = \frac{\text{cov}(\ln P_i \ln P_j)}{sd(\ln P_i) sd(\ln P_j)} \qquad (10)$$

Substitution of equation (9a) into the numerator of equation (10), and equations (9b) and (9c) into the denominator, defines the relationship between r_{ij} and the structural parameters, including CPED. Since house prices in i and j are affected by common factors Z and Y, but with different loadings (Γ_{i1}, Γ_{j1} and Γ_{i2}, Γ_{j2}), the correlation is expected to be less than 1. Also, the correlation between ω_i and ω_j is less than 1. For both of these reasons house prices in i and j are expected to be imperfectly correlated.

Appendix 2 demonstrates that when goods i and j are perfect substitutes in consumption, the correlations between their prices tends to unity. This result stems from the fact that under perfect substitution equation (7a) implies $P_{it} = P_{jt}$.

Measurement Error and Differences between Correlations

If house prices are measured with error (p) such that $ln\tilde{P} = lnP + p$ and p_i and p_j are independent, the correlation between house prices is attenuated and equals:

$$\tilde{r}_{ij} = \frac{\text{cov}(\ln P_i \ln P_j)}{sd(\ln \tilde{P}_i)sd(\ln \tilde{P}_j)} \qquad (13)$$

If the correlation is calculated over T periods, Fisher's test implies that correlation is significantly less than r* when:

$$\frac{\sqrt{T-3}}{2}\ln\left[\frac{(1+r^*)(1-r)}{(1-r^*)(1+r)}\right] > 1.64 \qquad (14)$$

at p = 0.05 (one-tail). Note that smaller absolute differences between r and r* are more likely to be significantly different as r* tends to 1.

Table 2 Measurement Error and 95% Confidence Intervals for Correlations

r*	$\tilde{r}$	σ_p
0.995	0.9876	0.0037
0.99	0.975	0.0076
0.95	0.88.	0.039
0.9	0.769	0.0818

Table 2 uses equation (14) to generate (one sided) 95% confidence intervals for measured correlations when the true correlation is r*, and uses equation (13) to determine the implicit value of measurement error (p) that accounts for the difference between these correlations. For example, if the true correlation is 0.95, the correlation at the 95% confidence interval is 0.88. Measurement error of 3.9% (σ_p = 0.039) would account for the difference between these correlations. As expected, the confidence interval varies inversely with r* and tends to zero as r* tends to 1. If r* = 0.995 a correlation of 0.9876 is at the confidence interval, which is generated by measurement error of only 0.37%.

In summary, we apply Pryce's conceptual model to the spatial correlations between house prices rather to their CPEPs. These correlations are calculated using location specific house price indices constructed from repeat sales data, which finesses the need to take account of hedonics.

3. Repeat Sales

Repeat sales data obviate the need for estimating hedonics under the assumption that the attributes of housing do not change over time (Bailey, Muth and Nourse 1963). If unobserved housing attributes follow a random walk process, Case and Shiller (1989) suggested that the weight on housing should vary inversely with the repeat sales interval because attributes are more likely to change with the passage of time. We have reservations about this weighting proposal. First, the price of land that is implicit in house prices is not directly affected by attributes. Therefore, the proportionate effect of an attribute on house prices will be smaller the more expensive the housing. Second, if homeowners undertake major restructuring on purchase, the random walk model will be incorrect. In any case, Nagaraja, Brown and Wachter (2014) show that results for BMN and CS are correlated 0.999969 in Minneapolis. In what follows, we ignore complications associated with changing attributes.

Let $\pi_{\tau t}$ denote the average log change in house prices purchased in period τ and sold in period $t > \tau$ in some location. In a 3-period setup, the data reveal π_{12}, π_{23} and π_{13}. Let π^* denote log changes in the repeat sales index (RSI). Using the fact that $\pi^*_{13} = \pi^*_{12} + \pi^*_{23}$ we may solve for the log changes in RSI between periods 1 and 2, and periods 2 and 3:

$$\pi^*_{12} = \alpha_1 \pi_{12} + (1-\alpha_1)(\pi_{13} - \pi^*_{23}) \qquad (15a)$$

$$\pi^*_{23} = \alpha_2 \pi_{23} + (1-\alpha_2)(\pi_{13} - \pi^*_{12}) \qquad (15b)$$

$$\alpha_1 = \frac{n_{12}}{n_{12} + n_{13}} \qquad \alpha_2 = \frac{n_{23}}{n_{23} + n_{13}}$$

where $n_{\tau t}$ is the number of repeat sales in period t of housing purchased in period τ. For example, the second term in equation (15a) refers to information on the repeat sales index (RSI) during period 1 embodied in π_{13}. The solutions to equations (15) for the RSIs are:

$$\pi^*_{12} = \frac{\alpha_1\pi_{12} - \alpha_2(1-\alpha_1)(\pi_{23} - \pi_{13})}{\alpha_1 + \alpha_2 - \alpha_1\alpha_2} \qquad (16a)$$

$$\pi^*_{23} = \frac{\alpha_2\pi_{23} - \alpha_1(1-\alpha_2)(\pi_{12} - \pi_{13})}{\alpha_1 + \alpha_2 - \alpha_1\alpha_2} \qquad (16b)$$

If $\alpha_1 = \alpha_2 = ½$ ($n_{\tau t} = n/3$), $\pi_{12} = 3$, $\pi_{23} = 4$ and $\pi_{13} = 7$, equations (16) imply $\pi^*_{12} = 3, \pi^*_{23} = 4$ as expected. If $\pi_{13} = 8$ $\pi^*_{12} = 3.33, \pi^*_{23} = 4.33$ as expected. These numbers replicate the solutions to equations (6) in BMN, and demonstrate that BMN's regression approach is equivalent to solving simultaneous equations such as equation (15).

When there are T periods the principles are the same. The restriction $\sum_{\tau=t+1}^{T-1} \pi^*_{t\tau} = \pi^*_{\tau T}$ induces a set of T – 1 simultaneous equations, which solve for T - 1 values for π^*_{tt+1} in terms of ½T(T-1) data points for $\pi_{\tau t}$:

$$\pi^*_{tt+1} = w_{tt+1}\pi_{tt+1} + \sum_{\tau=2}^{T}\sum_{j=1}^{t} w_{j\tau}\left[\pi_{j\tau} - \left(\pi^*_{j\tau} - \pi^*_{tt+1}\right)\right] \qquad (17)$$

$$w_{j\tau} = \frac{n_{j\tau}}{N_t}$$

$$w_{tt+1} = \frac{n_{tt+1}}{N_t}$$

$$N_t = n_{tt+1} + \sum_{\tau=2}^{T}\sum_{j=1}^{t} n_{j\tau}$$

Notice that π_{t-1t} and before do not feature in (17). Notice also that with the passage of time RSIs require updating historically to take account of new data on repeat sales, especially those with large repeat sales intervals.

Equations (17) may be vectorized as:

$$A\pi^* = B\pi \qquad (18)$$

where π^* is T-1 vector with elements π^*_{tt+1}, A is a T-1xT-1 asymmetric matrix with 1 along the diagonal and positive off-diagonal elements depend on $w_{j\tau}$, π is ½T(T-1) vector with elements

$\pi_{j\tau}$, and B is a T-1x½T(T-1) matrix with elements depending on $w_{j\tau}$. In the case of equations (15):

$$A = \begin{bmatrix} 1 & 1-\alpha_1 \\ 1-\alpha_2 & 1 \end{bmatrix} \quad B = \begin{bmatrix} \alpha_1 & 1-\alpha_1 & 0 \\ 0 & 1-\alpha_2 & \alpha_2 \end{bmatrix} \quad \pi^* = \begin{bmatrix} \pi_{12}^* \\ \pi_{23}^* \end{bmatrix} \quad \pi' = [\pi_{12} \quad \pi_{13} \quad \pi_{23}]$$

If the rank of A is full, equation (17) solves for all T-1 elements of π*. Notice that identification of π* does not require that transactions took place in all the ½T(T-1) combinations of $n_{j\tau}$. If the rank is R < T-1 equation (18) solves for only R elements of π*. These solutions and their standard deviation should be the same as their BMN counterparts. Since the T-1 elements of π* are inferred from N repeat sales, the degrees of freedom equal N – (T-1). In most applications N is very large relative to T because the data are national or for cities. Matters are different, however, when space is disaggregated to the sub-city level as discussed below.

4. Data

We use transactions price data to construct repeat sales indices for house prices by statistical area in Israel. These administrative data are recorded when house buyers pay stamp duty according to the value of the transaction in the conveyance. These data are available electronically since 1998 and include about 1.3m transactions of which about a quarter involve repeat sales. The distribution of the interval between repeat sales (Figure 1) is censored from both sides. It is censored from the left because the data exclude repeat sales of properties bought before 1998, and it is censored from the right because it excludes repeat sales that occurred after 2014. Consequently, Figure 1 creates the misleading impression that turnover is rapid and that most repeat sales occur within about 7 years. Because the data commence in 1998, the number of repeat sales naturally increases over time (Figure 2). Unfortunately, uncensored data on the natural history of repeat sales are not available, and it will take many years until they are.

Ben-Tovim, Zussman and Yachin (2014) used these data to construct repeat sales indices for two locations (center and periphery), and show that BMN and CS methodologies generate almost identical results. They also show that between 1998 and 2008 housing was less expensive according to repeat sales methods than according to hedonic methods, but the opposite was true subsequently (until 2012).

Figure 1 Time Interval between Repeat Sales

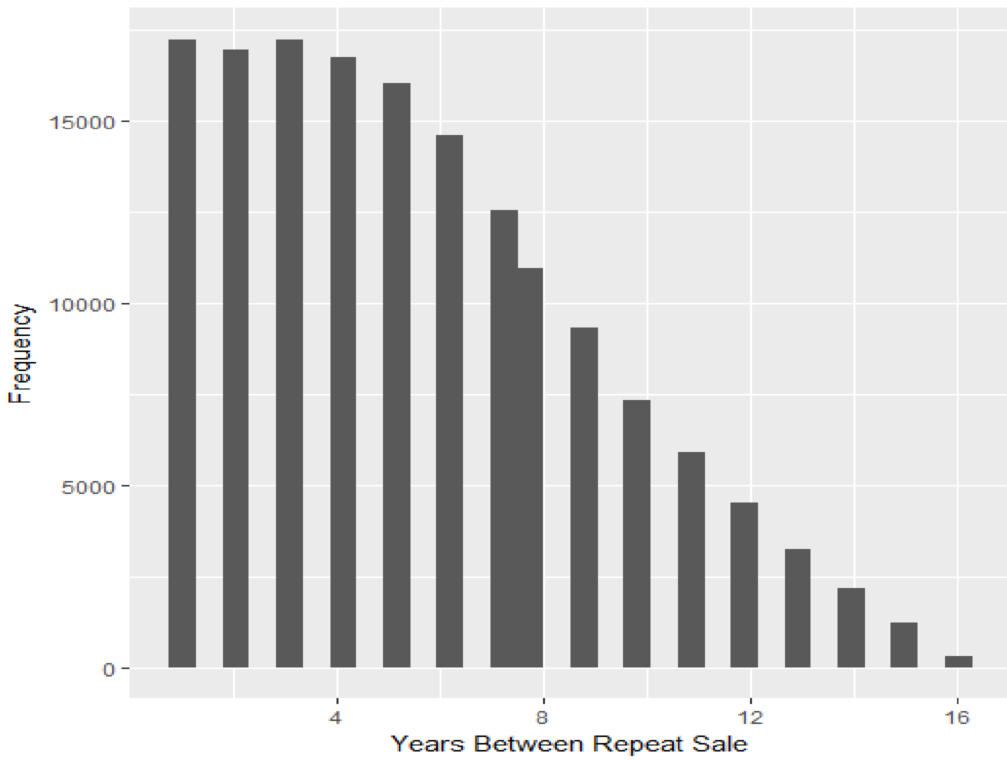
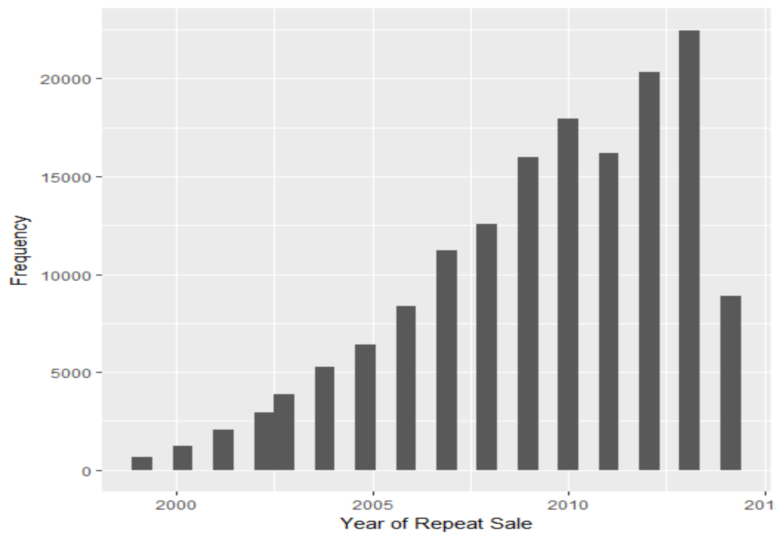


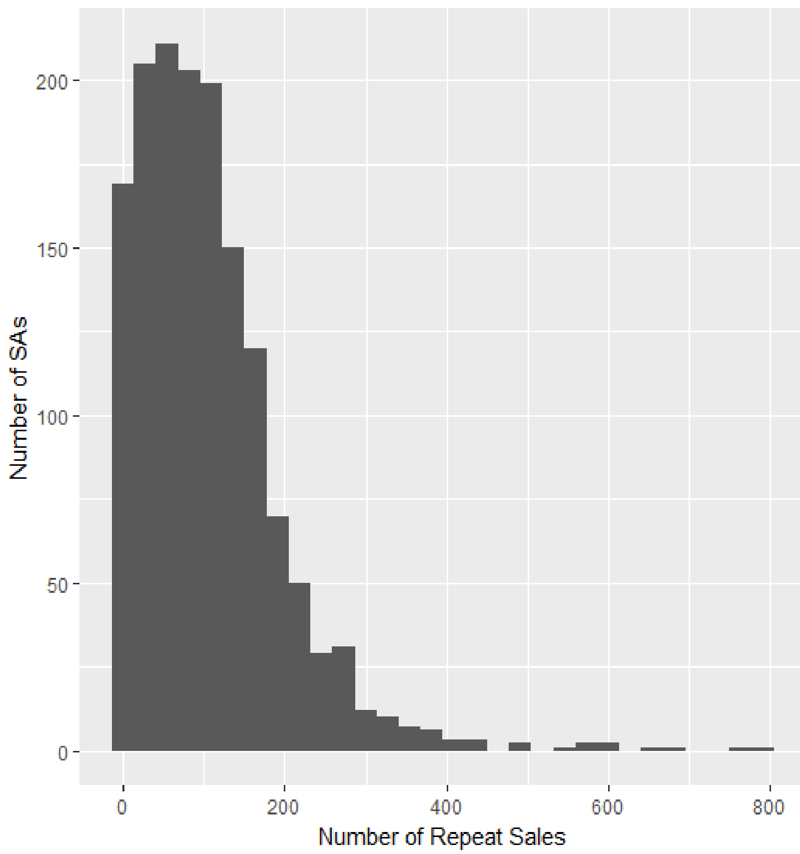
Figure 2 Repeat Sales by Year



Israel is divided by the Central Bureau of Statistics into over 3000 statistical areas (SAs), or census tracts, of which about 1800 are populated. Since these SAs are designed to be roughly equally populated, they are geographically smaller in more densely populated areas. SAs in large cities are generally small, homogenous administrative units with roughly 3000 inhabitants. In Tel Aviv city, average SA size is 0.306 sq. km. in contrast to the Tel Aviv metropolitan average of 0.819 sq. km.[1]. The frequency distribution of the number of repeat sales nationwide during 1998 - 2014 by statistical area (Figure 3) is almost exponential. The vast majority of SAs have up to 200 repeat sales, but some have as many as 600 or more.

[1] While the average SA size in other large cities is larger than in Tel Aviv, for example Jerusalem (0.627 sq.km.), Haifa (0.614 sq. km.) and Beer Sheva (1.894 sq. km.) these are still small areas.

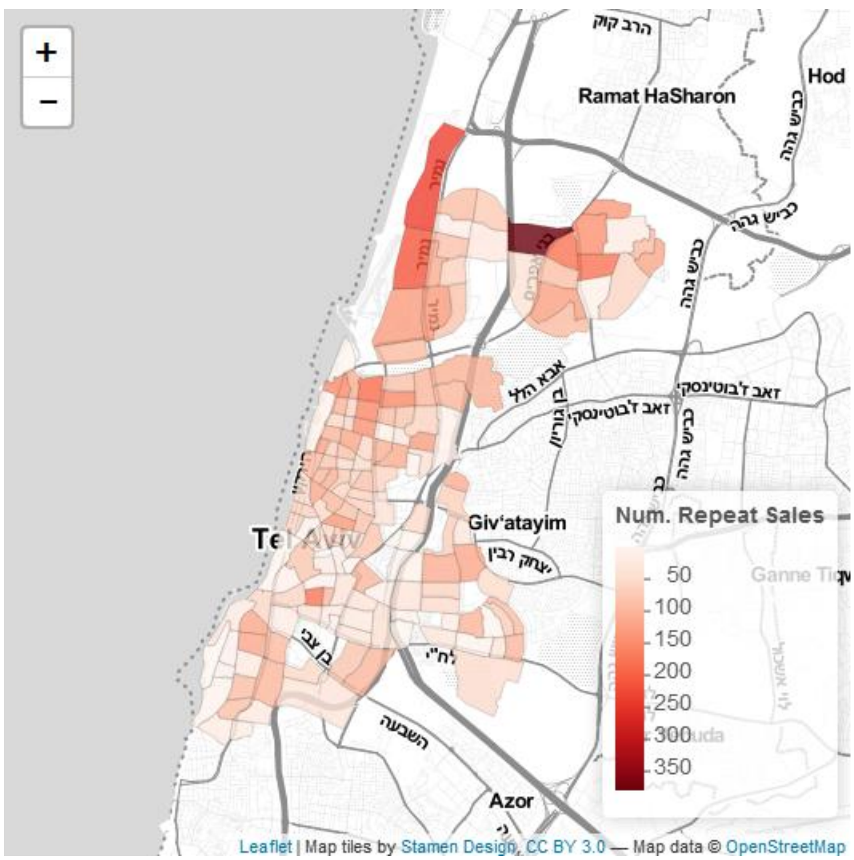
Figure 3 Number of Repeat Sales by Statistical Area



5. Identifying Housing Submarkets in Tel Aviv

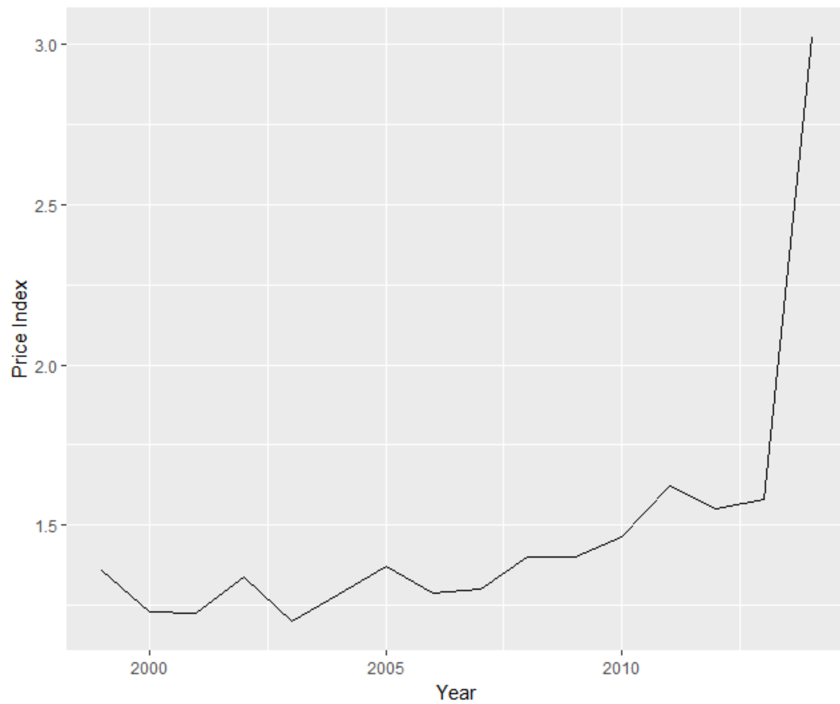
The city of Tel Aviv comprises 169 statistical areas, which are mapped in Figure 4. On average, there were about 170 repeat sales per statistical area during 1998 – 2015. Figure 5 plots the repeat sales indices (RSI's) for Tel Aviv as a whole. House prices nationwide had peaked in 1998-9 having doubled in real terms since 1989 in the wake of mass immigration from the former Soviet Union, which increased the population of Israel by 20 percent. During 1989 – 2007 house prices decreased nationally by about 25 percent. Subsequently, house prices began to increase nationally, surpassing their previous peak by 2010, and they are currently almost double what they were in 1989. These national trends are reflected in Figure 5. However, house prices in Tel Aviv have grown faster than the national average.

Figure 4 Map of Tel Aviv Statistical Areas



In Figure 5 the large increase in house prices in 2014 is most probably induced by data censoring. It will take several years before it will be possible to determine whether this development was genuine or whether it was induced by censoring.

Figure 5 Repeat Sales Price Index: Tel Aviv



We construct RSIs for almost S = 100 statistical areas in Tel Aviv (see Fig 4) excluding those for which the number of repeat sales is less than 50. Since T = 17 these data are used to estimate 16 elements of π* for each SA. In 4 cases some elements could not be calculated because matrix A in equation (18) was not full rank. Next, we calculate ½S(S - 1) pairwise correlations between the log price indices (lnP* not π*) generated by π*. These correlations are expected to be spatially clustered because SAs that are closer to each other are more likely to share common amenities such as accessibility, crime, and services. The distribution for these correlations is presented in Figure 6. It has an extended right tail and most of the correlations are below 0.5. Despite measurement error in RSIs, many correlations exceed 0.8. Since measurement error attenuates correlations away from 1, Table 1 suggests that these correlations may be regarded as being close to 1.

Figure 7 presents these correlations in a heat map. The pixels in Figure 7 refer to pairwise correlations, which are clustered according to their size on a raster generated density surface. Along its diagonal, the correlations equal 1 by definition. The dominant color in Figure 7 is yellow, indicating that the vast majority of correlations are small. However, the purple off-diagonal patches indicate clusters of high correlation. If these statistical clusters belong to the same LHM they should be spatially clustered. If highly correlated statistical areas happen to be remote from each other they cannot belong to the same LHM.

Figure 6: Distribution of Pairwise Correlations between Repeat Sales Indices

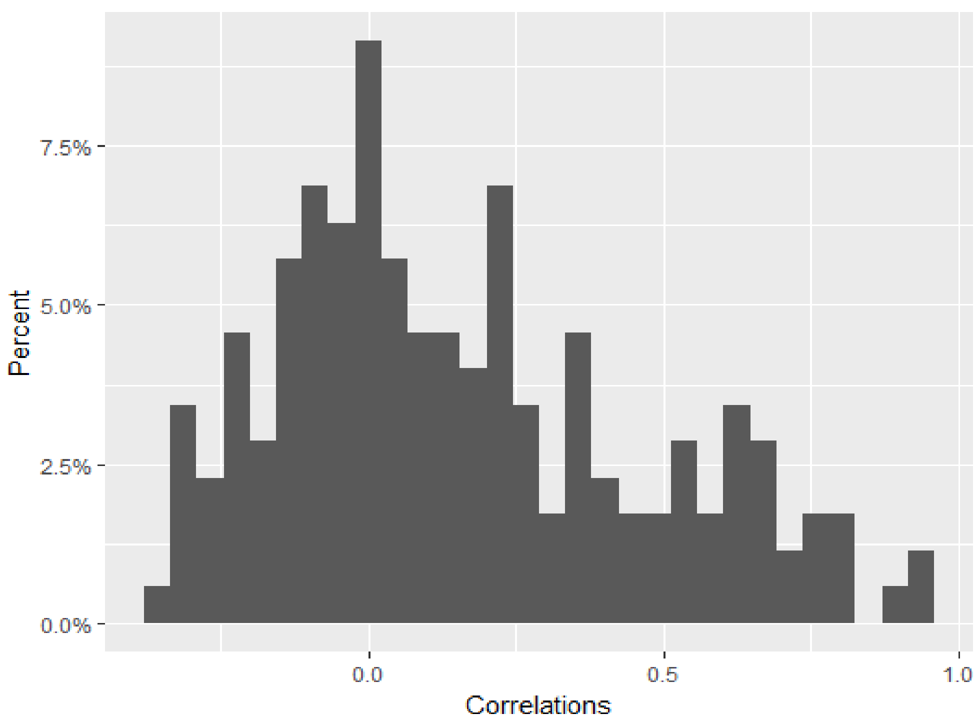


Figure 7: Heat Map of Pairwise Correlations

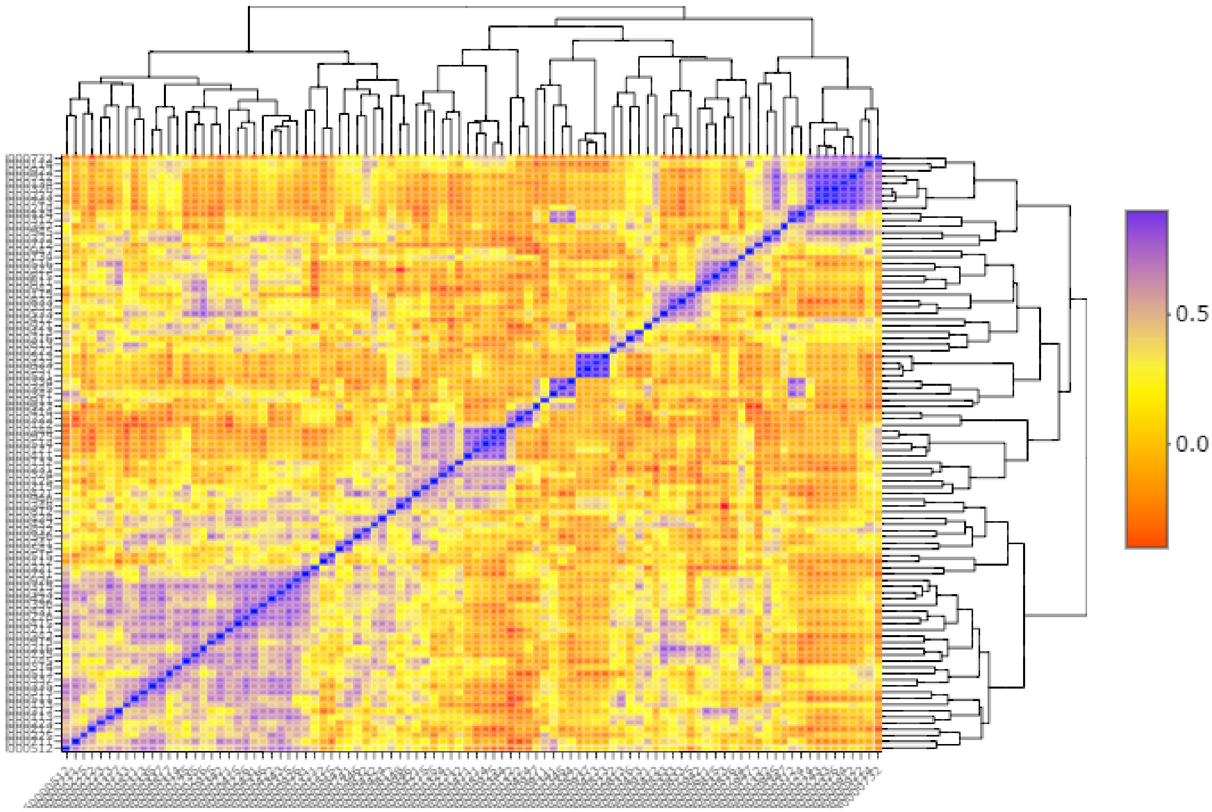
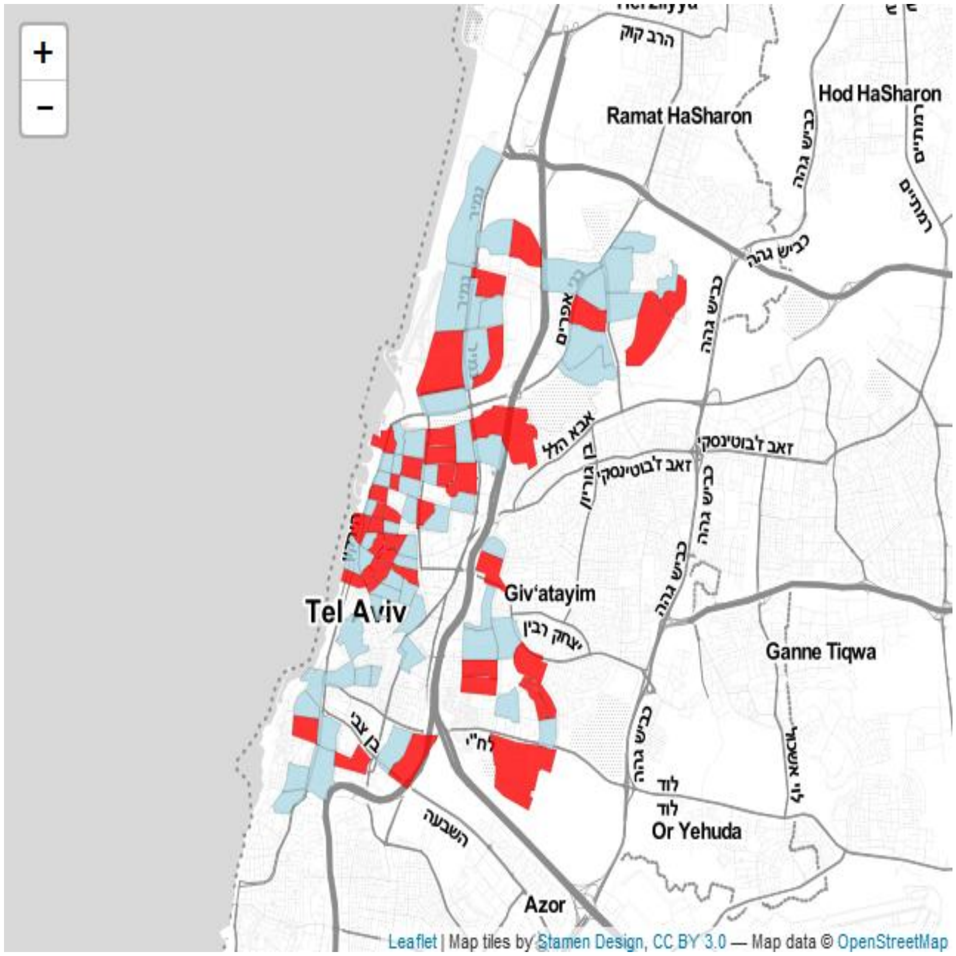


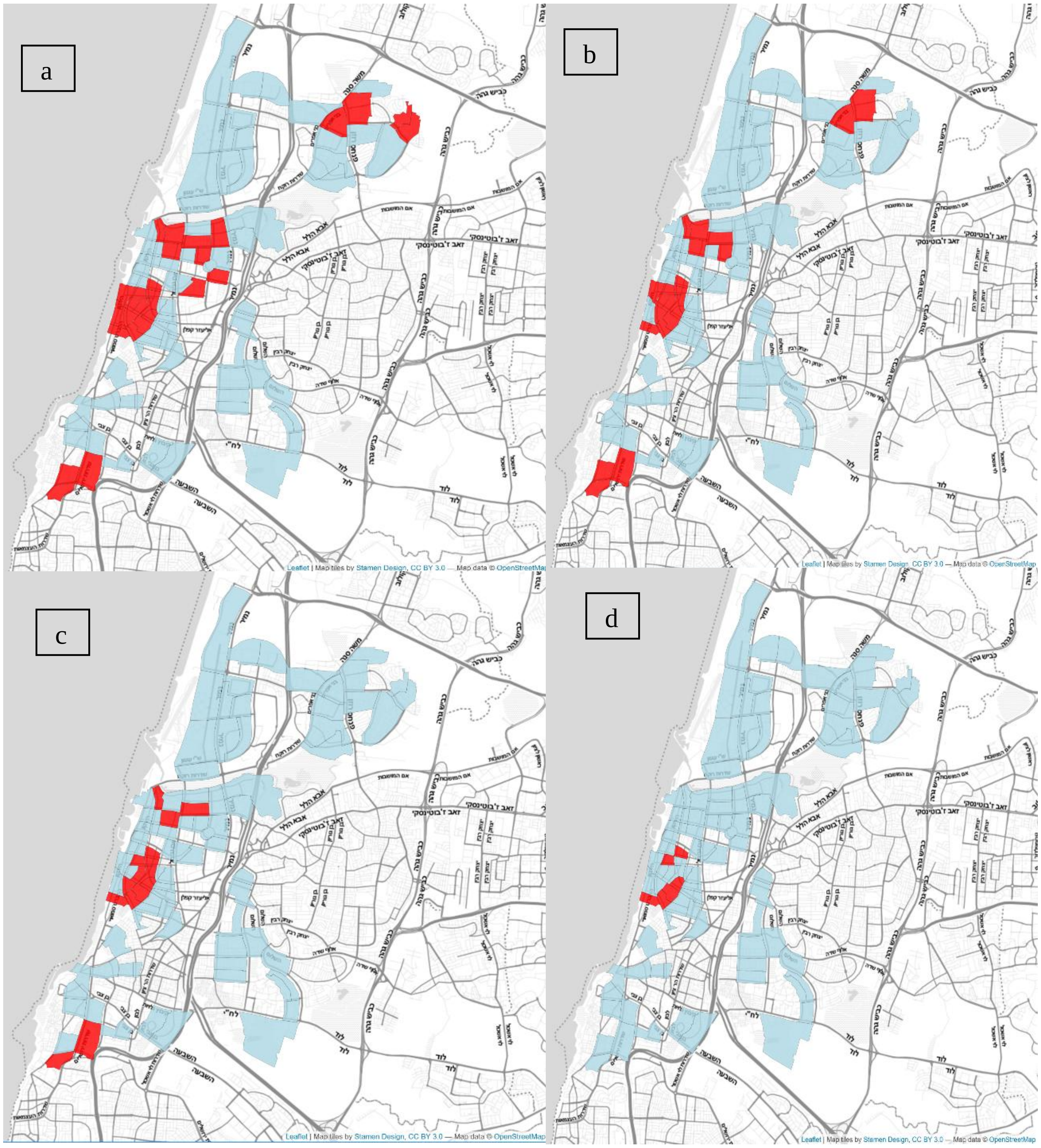
Figure 8 maps the locations where house prices are not only highly correlated, but they are also spatially clustered. We identify about 15 LHMs generated with no regard to contiguity. The red zones indicate statistical areas where the correlations with other statistical areas exceed 0.8. Recall from Table 2 that such correlations might not be significantly smaller than correlations that exceed 0.9. In Figures 9 and 10 contiguity restrictions are imposed, and correlation cut-offs varied. With contiguity, the number of LHMs naturally declines as correlation cut-offs are successively raised from 0.55 to 0.85 (Figure 9). Conversely, if we relax the contiguity requirement to include 2nd order (neighbor's neighbors) and 3rd order (neighbors of neighbor's neighbors) contiguity, the number of LHMs increases (Figure 10).

Figure 8 Local Housing Markets: r > 0.8, no contiguity restrictions*



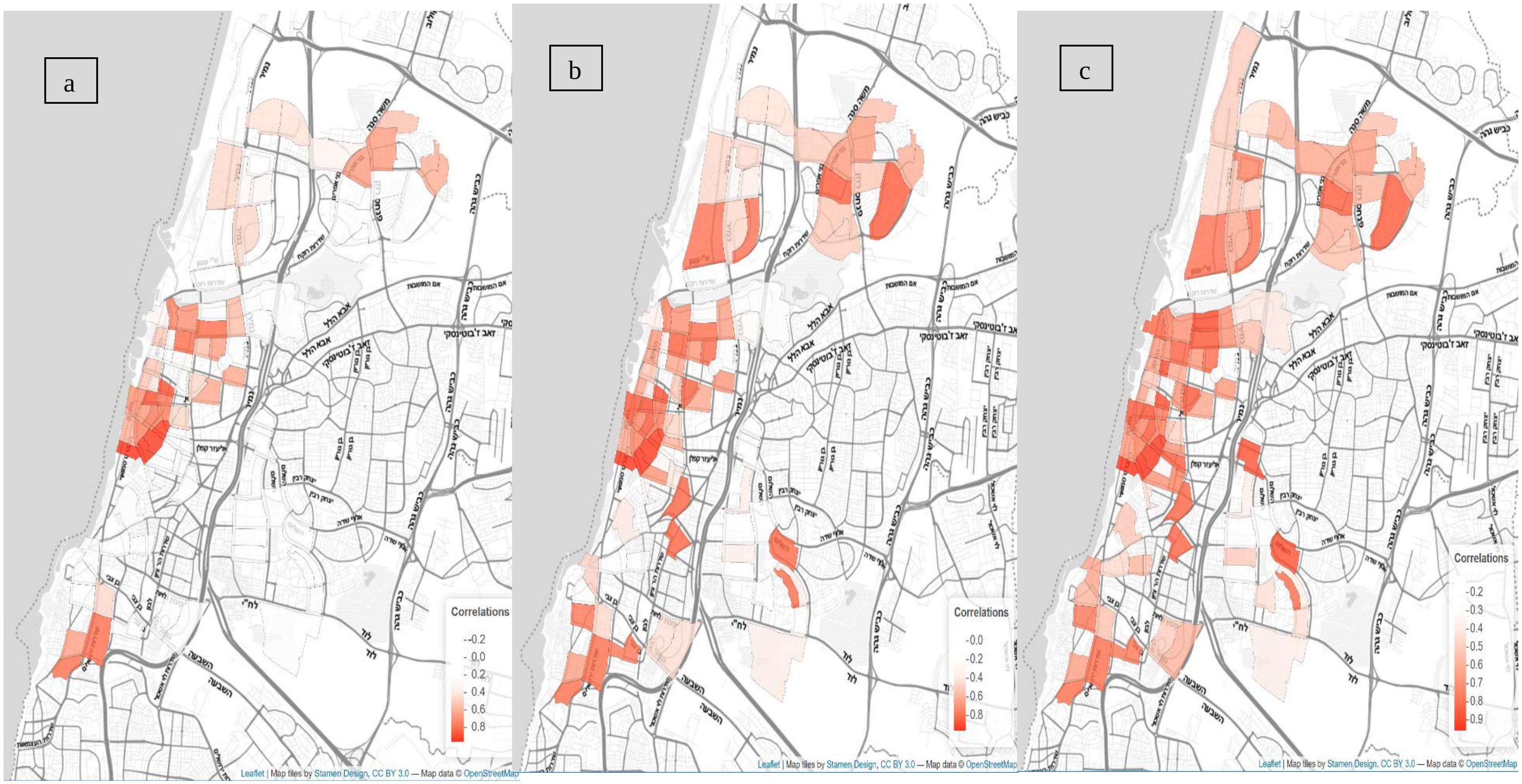
*Red zones indicate statistical areas where pairwise correlations with other statistical areas exceed 0.8.

Figure 9 Local Housing Markets: (a) r > 0.55, (b) r > 0.65, (c) r > 0.75, (d) r > 0.85 with contiguity*



*Red zones indicate statistical areas where pairwise correlations with other statistical areas meet the benchmarks outlined in the caption.

Fig 10: Correlations with contiguity relaxed: (a) 1st order neighbors, (b) 2nd order neighbors, (c) 3rd order neighbors*



*Orange shading denotes level of correlation

Conclusions

Identifying local housing markets is important for planners, realtors, property tax assessment, mortgage evaluation and price prediction. In this paper, we therefore propose a new empirical approach to identifying LHMs, which focuses on the spatial correlation between local house price indices constructed from repeat sales data.

We extend the insight of Pryce (2013) on strong membership of local housing markets, who claimed that if houses in different locations are perfect substitutes, their house price indices should be perfectly correlated over time. We therefore focus on changes in house prices indices rather than levels of hedonic house prices, as in Bourassa et al (1999). We implement Pryce's method by using repeat sales data in the absence of panel data on house prices.

Results for almost 100 statistical areas in Tel Aviv illustrate the highly heterogeneous nature of its housing market. Similar heterogeneity is reported by Goetzmann and Speigel (1997) who use a repeated sales method for San Francisco to decompose a city-wide housing returns index into its neighborhood constituents. Strictly speaking, strong LHM membership requires that the correlations should be 1 and locations should be contiguous. These strict criteria imply that there are no LHMs in Tel Aviv. In practice, correlations are expected to be less than 1 due to measurement error. For example, there are two LHMs when the correlation exceeds 0.85 and locations are required to be contiguous. The number of LHMs naturally varies inversely with the degree to which house prices in locations that belong to the same LHM are expected to be correlated, and varies directly with the order of contiguity of these locations. When these criteria are relaxed the number of LHMs increases towards 20.

Our purpose is not to declare the number of LHMs in Tel Aviv. Instead, it is to illustrate empirically the use of repeat sales data to identify LHMs in the strong sense. However, we are surprised by the heterogeneity of house prices across Tel Aviv. The popular impression that Tel Aviv constitutes a relatively homogeneous housing market in which housing is expensive and house prices move together if not in unison, is not supported by the data. Our results show the opposite; the correlations between house prices by statistical areas are, on the whole, rather low. The vast majority of correlations are less than 0.5 and some are even negative. In the absence of studies for other cities, we do not know whether Tel Aviv is unusual.

Our approach may be sensitive to the modifiable area unit problem (MAUP) (Fotheringham and Wong 1991) since the spatial units in the analysis defined by the Central Bureau of Statistics (SA's) might influence the outcome. This critique applies to the large literature using census tract data. However, because the statistical areas are geographically very small (0.306 sq km.on average), we do not think that the MAUP is a serious problem. In principle, we could carry out robustness checks by redefining statistical areas by their geographical coverage to see whether MAUP matters. The alternative is to use data on individual dwellings rather than spatial panel data. In the absence of panel data for dwellings, the methodology of Bourassa et al (1999) is the default, which has numerous problems. We think that the use of repeat sales data by census tract, which are geographically small, strikes a sensible compromise between methodological imperfections induced by MAUP and imperfections related to hedonic pricing.

Appendix 1: The Relation between Weak and Strong LHM Membership

The hedonic house price models for locations A and B are:

$$P_{Ait} = \alpha_A + \beta_A C_{1it} + \gamma_A C_{2it} + u_{it}$$

$$P_{Bit} = \alpha_B + \beta_B C_{1it} + \gamma_B C_{2it} + v_{it}$$

where C_1 and C_2 denote hedonic characteristics, and u and v are iid. R^2 for these models are r_A^2 and r_B^2. If $\beta_A = \beta_B$ and $\gamma_A = \gamma_B$ housing in A and B belong weakly to the same LHM.

Average house prices in A and B in period t are denoted by P_{At} and P_{Bt}. The correlation between them over T periods is:

$$r_{AB} = \frac{cov(P_A P_B)}{sd(P_A)sd(P_B)}$$

A and B belong strongly to the same LHM when $r_{AB} = 1$. Suppose, for simplicity, the first two moments of C_{1t} and C_{2t} are the same in A and B, so that the hedonic component of house prices in period t, denoted by H_t, is the same in A and B. Hence:

$$cov(P_A P_B) = var(H) + cov(u\ v)$$

It may be shown that the panel correlations r_A and r_B equal their time series counterparts. Since $r_A^2 = var(H)/var(P_A)$ and $r_B^2 = var(H)/var(P_B)$, the standard deviations for average prices over T periods are $sd(P_A) = sd(H)/r_A$ and $sd(P_B) = sd(H)/r_B$. Substituting these results into r_{AB} implies:

$$r_{AB} = r_A r_B \left(1 + \frac{cov(uv)}{var(H)}\right)$$

In the absence of unobserved common factors cov(uv) = 0, in which case the correlation between average prices over T periods equals the product of the panel correlations for the hedonic price models. If R^2 for the hedonic price regressions are of the order of 0.6, $r_{AB} = 0.6$. Weak membership of LHMs does not imply strong membership and vice-versa. Matters would be different if r_A and r_B equaled 1.

When there are unobserved common factors $cov(uv) = r_{uv} sd(u) sd(v) > 0$, where r_{uv} denotes the correlation between u and v. Since, for example, $var(u) = (1 - r_A^2) var(P_A)$, the general solution for r_{AB} may be rewritten as:

$$r_{AB} = r_A r_B + r_{uv}[(1 - r_A^2)(1 - r_B^2)]^{\frac{1}{2}}$$

If $r_{uv} < 1$ then r_{AB} must be less than 1. If $r_{uv} = 1$ then $r_{AB} = 1$ even when $r_a = r_b = 0$. When $r_A = r_B = r$, the solution simplifies to:

$$r_{AB} = r^2 + r_{uv}(1 - r^2)$$

If, for example, $R^2 = 0.6$ in the hedonic price regressions and $r_{uv} = 0.25$ then $r_{AB} = 0.7$.

Appendix 2: Definitions of Γ and Ψ Coefficients in Equations (8)

For good i:

$$\Gamma_{i1} = -\frac{1}{D}\left[\phi_i A_{ii} + \phi_j A_{ij} + \phi_k A_{ik}\right]$$

$$\Gamma_{i2} = \frac{1}{D}\left[\psi_i A_{ii} + \psi_j A_{ij} + \psi_k A_{ik}\right]$$

$$\omega_i = \left[\Psi_{ii}(u_i - v_i) + \Psi_{ij}\left(u_j - v_j\right) + \Psi_{ik}(u_k - v_k)\right]$$

$$\Psi_{im} = \frac{A_{im}}{D} \quad m = i, j, k$$

$$A_{ii} = \left(\gamma_j + \pi_j\right)(\delta_k + \theta_k) - \left(\delta_j + \theta_j\right)(\gamma_k + \pi_k) > 0$$

$$A_{ij} = (\gamma_i + \pi_i)(\delta_k + \theta_k) + (\delta_i + \theta_i)(\gamma_k + \pi_k)$$

$$A_{ik} = (\gamma_i + \pi_i)\left(\delta_j + \theta_j\right) + (\delta_i + \theta_i)\left(\gamma_j + \pi_j\right)$$

$$D = (\lambda_i + \beta_i)A_{ii} + \left(\lambda_j + \beta_j\right)A_{ij} + (\lambda_k + \beta_k)A_{ik}$$

For good j:

$$\Gamma_{ji} = -\frac{1}{D}\left[\phi_i A_{ji} + \phi_j A_{jj} + \phi_k A_{jk}\right]$$

$$\Gamma_{j2} = \frac{1}{D}\left[\psi_i A_{ji} + \psi_j A_{jj} + \psi_k A_{jk}\right]$$

$$\omega_j = \left[\Psi_{ji}(u_i - v_i) + \Psi_{jj}\left(u_j - v_j\right) + \Psi_{jk}(u_k - v_k)\right]$$

$$\Psi_{jm} = \frac{A_{jm}}{D} \quad m = i, j, k$$

$$A_{ji} = \left(\beta_j + \lambda_j\right)(\delta_k + \theta_k) + \left(\delta_j + \theta_j\right)(\beta_k + \lambda_k)$$

$$A_{jj} = (\beta_i + \lambda_i)(\delta_k + \theta_k) + (\delta_i + \theta_i)(\beta_k + \lambda_k)$$

$$A_{jk} = (\beta_i + \lambda_i)\left(\delta_j + \theta_j\right) + (\delta_i + \theta_i)\left(\beta_j + \lambda_j\right)$$

For good k:

$$\Gamma_{k1} = -\frac{1}{D}\left[\phi_i A_{ki} + \phi_j A_{kj} + \phi_k A_{kk}\right]$$

$$\Gamma_{k2} = \frac{1}{D}\left[\psi_i A_{ki} + \psi_j A_{kj} + \psi_k A_{kk}\right]$$

$$\omega_k = \left[\Psi_{ki}(u_i - v_i) + \Psi_{kj}\left(u_j - v_j\right) + \Psi_{kk}(u_k - v_k)\right]$$

$$\Psi_{km} = \frac{A_{km}}{D} \quad m = i, j, k$$

$$A_{ki} = (\gamma_k + \pi_k)\left(\beta_j + \lambda_j\right) + (\beta_k + \lambda_k)\left(\gamma_j + \pi_j\right)$$

$$A_{kj} = (\beta_i + \lambda_i)(\gamma_k + \pi_k) + (\gamma_i + \pi_i)(\beta_k + \lambda_k)$$

$$A_{kk} = \left(\gamma_j + \pi_j\right)(\beta_i + \lambda_i) + \left(\beta_j + \lambda_j\right)(\gamma_i + \pi_i)$$

Asymptotes of Γ and Ψ under perfect substitution in consumption:

If i and j are perfect substitutes in demand, β and γ for i and j tend to a common value (S), which tends to infinity under perfect substitution. The asymptotic order of these elasticities is denoted by O(S). Since D involves products of β and γ for i and j, it tends to:

$$D_\infty = \left[2(\delta_k + \theta_k)O(S^2) + (\delta_j + \delta_i + \theta_j + \theta_i)O(S)\right]$$

All the As involve sums of these elasticities except A_{kk}, which involves their products ($\beta_i\gamma_j$). Hence, A_{kk} tends to $O(S^2)$ while other As tend to O(S). Therefore, for goods i and j:

$$\Gamma_{i1,\infty} = \Gamma_{j1,\infty} = -\frac{\left[(\phi_i + \phi_j)(\delta_k + \theta_k) + \phi_k\left(\delta_j + \delta_i + \theta_j + \theta_i\right)\right]O(S)}{D_\infty} \sim O(S^{-1})$$

$$\Gamma_{i2,\infty} = \Gamma_{j2,\infty} = \frac{\left[(\psi_i + \psi_j)(\delta_k + \theta_k) + \psi_k\left(\delta_j + \delta_i + \theta_j + \theta_i\right)\right]O(S)}{D_\infty} \sim O(S^{-1})$$

$$B = -\frac{\left(\phi_i + \phi_j\right)(\delta_k + \theta_k) + \phi_k\left(\delta_i + \delta_j + \theta_i + \theta_j\right)}{2(\delta_k + \theta_k) + \left(\delta_i + \delta_j + \theta_i + \theta_j\right)}$$

$$C = \frac{\left(\psi_i + \psi_j\right)(\delta_k + \theta_k) + \psi_k\left(\delta_i + \delta_j + \theta_i + \theta_j\right)}{2(\delta_k + \theta_k) + \left(\delta_i + \delta_j + \theta_i + \theta_j\right)}$$

$$\Psi_{ii,\infty} = \Psi_{ij,\infty} = \frac{(\delta_k + \theta_k)}{2(\delta_k + \theta_k)O(S) + \left(\delta_i + \delta_j + \theta_i + \theta_j\right)} \sim O(S^{-1})$$

$$\Psi_{ik,\infty} = \frac{\delta_i + \delta_j + \theta_i + \theta_j}{2(\delta_k + \theta_k)O(S) + \left(\delta_i + \delta_j + \theta_i + \theta_j\right)} \sim O(S^{-1})$$

For good k:

$$\Gamma_{k1,\infty} = -\frac{\phi_k}{2(\delta_k + \theta_k) + \delta_j + \delta_i + \theta_j + \theta_i}$$

$$\Gamma_{k2,\infty} = \frac{\psi_k}{2(\delta_k + \theta_k) + \delta_j + \delta_i + \theta_j + \theta_i}$$

$$\Psi_{kk,\infty} = \frac{1}{2(\delta_k + \theta_k) + \delta_j + \delta_i + \theta_j + \theta_i}$$

$$\Psi_{ki,\infty} = \Psi_{ik,\infty} = \Psi_{kj,\infty}$$

Implications of asymptotes for price correlations:

Substituting these results into equation (9b) implies:

$$var(lnP_i) = \Gamma_{i1,\infty}^2 var(lnZ) + \Gamma_{i2,\infty}^2 varln(Y) + 2\Gamma_{i1,\infty}\Gamma_{i2,\infty} cov(lnZ\ lnY) + \omega_i^2 \sim O(S^{-2})$$

$$\omega_i^2 = \Psi_{ii,\infty}^2\left[var(u_i - v_i) + var\left(u_j - v_j\right)\right] + \Psi_{ik,\infty}^2 var(u_k - v_k) \sim O(S^{-2})$$

The same applies to equation (9c), in which case var(lnP_i) = var(lnP_j) asymptotically. The asymptotic covariance in equation (9a) becomes:

$$cov\left(lnP_i lnP_j\right) = \Gamma_{i1,\infty}\Gamma_{j1,\infty} var(lnZ) + \Gamma_{i2,\infty}\Gamma_{j2,\infty} varln(Y) + \left(\Gamma_{i1,\infty}\Gamma_{j2\infty} + \Gamma_{i2,\infty}\Gamma_{j1,\infty}\right) cov(lnZ\ lnY) + \omega_j^2 = var\left(lnP_j\right) \sim O(S^{-2})$$

Finally, substituting these results into equation (10) proves that r_{ij} tends to 1 as goods i and k tend to perfect substitutes in consumption, because the covariance equals the variance.

Matters are different for r_{ik} and r_{jk}. The variance of lnP_k does not depend on S:

$$var(lnP_k) = \Gamma_{k1,\infty}^2 var(lnZ) + \Gamma_{k2,\infty}^2 var(lnY) + 2\Gamma_{k1,\infty}\Gamma_{k2,\infty} cov(lnZlnY) + \Psi_{kk,\infty}^2 var(u_k - v_k)$$

Its covariance with with lnP_i or lnP_j is $O(S^{-1})$ instead of $O(S^{-2})$:

$$cov(lnP_i lnP_k) = \Gamma_{i1,\infty}\Gamma_{k1,\infty} var(lnZ) + \Gamma_{i2,\infty}\Gamma_{k2,\infty} var(lnY) + \left(\Gamma_{i1,\infty}\Gamma_{k2,\infty} + \Gamma_{i2,\infty}\Gamma_{k1,\infty}\right) cov(lnZlnY) + cov(\omega_i\omega_k)$$

$$\mathrm{cov}(\omega_i\omega_k) = \Psi_{ii,\infty}\Psi_{ki,\infty} var(u_i - v_i) + \Psi_{ij,\infty}\Psi_{kj,\infty} var\left(u_j - v_j\right) + \Psi_{ik,\infty}\Psi_{kk,\infty} var(u_k - v_k)$$

$$= \left[var(u_i - v_i) + var\left(u_j - v_j\right)\right] O(S^{-2}) + var(u_k - v_k) O(S^{-1}) \sim O(S^{-2})$$

Since $cov(lnP_i lnP_k) \sim O(S^{-1})$ and $sd(lnP_i) \sim O(S^{-1})$, r_{ik} does not depend on S. When goods i and j are perfect substitutes in consumption the correlation tends to:

$$r_{ik} = \frac{\Gamma_{i1,\infty}\Gamma_{k1,\infty} var(lnZ) + \Gamma_{i2,\infty}\Gamma_{k2,\infty} var(lnY) + \left(\Gamma_{i1,\infty}\Gamma_{k2,\infty} + \Gamma_{i2,\infty}\Gamma_{k1,\infty}\right) cov(lnZlnY) + \Psi_{ik,\infty}\Psi_{kk,\infty} var(u_k - v_k)}{\sqrt{\Gamma_{i1,\infty}^2 var(lnZ) + \Gamma_{i2,\infty}^2 varln(Y) + 2\Gamma_{i1,\infty}\Gamma_{i2,\infty} cov(lnZ\ lnY) + \omega_i^2}\sqrt{\Gamma_{k1,\infty}^2 var(lnZ) + \Gamma_{k2,\infty}^2 var(lnY) + 2\Gamma_{k1,\infty}\Gamma_{k2,\infty} cov(lnZlnY) + \Psi_{kk,\infty}^2 var}}$$

Whereas the correlations between the prices of i and j tend to 1 due to perfect substitution, the correlations between i and k (and j and k) are less than 1, because they are imperfect substitutes.

References

Adair A.S., Berry J, and McGreal W.S. (1996) Hedonic modelling, housing submarkets, and residential valuation. *Journal of Property Research*, 13(1), 67-84.

Allen MT, Springer T and Waller NG (1995) Implicit pricing across residential submarkets. *Journal of Real Estate Finance and Economics*, 11(2), 137-151

Bajic V. (1985) Housing-market segmentation and demand for housing attributes: some empirical findings. *AREUEA Journal*, 13, 58-75.

Bailey M.J, Muth R.F. and Nourse H.O. (1963) A regression method for real estate price index construction. *American Statistical Association Journal*, 58: 933-942.

Basu S. and Thibodeau T.G. (1998) Analysis of spatial autocorrelation in house prices. *Journal of Real Estate Finance and Economics*, 17, 61 – 85.

Ben-Tovim N., Yakhin Y. and Zussman N. (2014) Repeat sales and the measurement of house price indices. DP 2014.1 Research Department, Bank of Israel (Hebrew).

Bhattacharjee A., Castro E., Maiti T. and Marques J. (2016) Endogenous spatial regression and delineation of submarkets: a new framework with application to housing markets. *Journal of Applied Econometrics*, 31: 32-57.

Borukhov E, Ginsburg Y and Werczberger E. (1978), Housing Prices and Housing Preferences in Israel. *Urban Studies*, 15, 187-200

Bourassa S., Hamelink F., Hoesli M. and MacGregor B.D. (1999) Defining housing submarkets. *Journal of Housing Economics*, 8: 160-183.

Bourassa S., Hoesli M. and Peng V.S. (2003) Do housing submarkets really matter? *Journal of Housing Economics*, 12: 12-28.

Bourassa S., Cantoni E. and Hoesli M. (2007) Spatial dependence, housing submarkets and house price predictions. *Journal of Real Estate Finance and Economics*, 35, 143-160.

Case K.E. and Shiller R.J. (1987) Prices of single-family homes since 1970: new indices for four cities. *New England Economic Review*, pp 45 – 56.

Case K.E. and Shiller R.J. (1989) The efficiency of the market for single-family homes. *American Economic Review*, 79: 125-137.

Cuthbertson K. and Nitzsche D. (2008) *Quantitative Financial Economics*, 2nd edition, Wiley, Chichester.

Dale-Johnson D (1982) An alternative approach to housing market segmentation using hedonic price data. *Journal of Urban Economics*, 11, 311-332.

Fisher EM and Fisher RM, (1954) *Urban Real Estate* , Henry Holt, NY

Fotheringham AS and Wong DWS (1991) The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A* 23, 1025–1044.

Gabriel S (1984) A note on housing market segmentation in an Israeli development town. *Urban Studies*, 21, 189-194

Galster G (1996) William Grigsby and the analysis of submarkets and filtering. *Urban Studies*, 33(10), 1797-1806.

Gat D. (1996) A Compact Hedonic Model of the Greater Tel Aviv Housing Market. *Journal of Real Estate Literature* 4 (2), 162-172

Gillen K., Thibodeau T.G. and Wachter S. (2001) Anisotropic autocorrelation in house prices. *Journal of Real Estate Finance and Economics*, 23, 5 – 30.

Goetzmann W.N. and Speigel M. (1997) A spatial model of housing returns and neighborhood substitutability. *Journal of Real Estate Finance and Economics*, 14, 11–31.

Goodman, A. C. (1981). Housing submarkets within urban areas: definitions and evidence. *Journal of Regional Science,* 21, 175–185.

Goodman A and Thibodeau T.G (1998) Housing market segmentation, *Journal of Housing Economics,* 7, 121-143.

Goodman A.C. and Thibodeau T.G (2007) The spatial proximity of metropolitan area housing submarkets, *Real Estate Economics*, 35(2), 209-232

Grigsby W.G. (1963) *Housing Markets and Public Policy*, University of Pennsylvania Press, Philadelphia PA.

Grinberger, A. Y., and Felsenstein, D. (2017) Using big (synthetic) data for identifying local housing market attributes, pp. 109-120 in Schintler L and Chen Z (eds) *Big Data for Regional Science*, Routledge, Abingdon, Oxford.

Helbich M., Brunauer W., Hagenauer J and Leitner M (2013) Data-driven regionalization of housing markets. *Annals of the Association of American Geographers*, 103 (4), 871-889.

Hwang, S. and Thill, J.-C. (2009) Delineating urban housing submarkets with fuzzy clustering. *Environment and Planning B*, 36, pp. 865–882.

Keskin B. and Watkins C. (2017), Defining spatial housing submarkets: exploring the case for expert delineated boundaries. *Urban Studies*, 54 (6), 1446-1462.

Maclennan D and Tu Y(1996) Economic perspectives on the structure of local housing markets. *Housing Studies*, 11, 387-406

Nagaraja C, Brown L.and Wachter S (2014) Repeated sales house price index methodology. *Journal of Real Estate Literature*, 22 (1), 23-46.

Palm R, (1978) Spatial segmentation of the urban housing market. *Economic Geography,* 54, 210-221.

Peng L. (2012) Repeat sales regression on heterogeneous properties. *Journal of Real Estate Finance and Economics*, 45, 804 – 827.

Pryce G. (2013) Housing markets and the lattice of substitution. *Urban Studies*, 50: 2682-2699.

Rosen S (1974) Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy*, 82, 34-55

Royuela V.and Duque J.C. (2013) HouSI: Heuristic for delimitation of housing submarkets and price homogeneous areas. *Computers, Environment and Urban Systems*, 37, 59-69.

Sayag D (2012) Measuring local house price movement in Israel and estimating price elasticity. *Israel Economic Review,* 10 (1), 39-94

Schnare A and Struyk R, (1976) Segmentation in urban housing markets. *Journal of Urban Economics,* 3, 146-166.

Tu Y., Sun H. and Yu S-M. (2007) Spatial autocorrelations and urban housing market segmentation. *Journal of Real Estate Finance and Economics*, 34, 385 – 406.

Watkins C (1999) Property valuation and the structure of urban housing markets, *Journal of Property Investment and Finance*, 17, 157-175.

Watkins C. (2001) The Definition and identification of housing submarkets, *Environment and Planning A*, 33, 2235-2253.