

Spatial error correction and cointegration in nonstationary panel data: regional house prices in Israel

Michael Beenstock · Daniel Felsenstein

Received: 8 February 2009 / Accepted: 8 April 2010 / Published online: 13 May 2010
© Springer-Verlag 2010

Abstract We “spatialize” residual-based panel cointegration tests for nonstationary spatial panel data in terms of a spatial error correction model (SpECM). Local panel cointegration arises when the data are cointegrated within spatial units but not between them. Spatial panel cointegration arises when the data are cointegrated through spatial lags between spatial units but not within them. Global panel cointegration arises when the data are cointegrated both within and between spatial units. Spatial error correction arises when error correction occurs within and between spatial units. We use nonstationary spatial panel data on the housing market in Israel to illustrate the methodology. We show that regional house prices in Israel are globally cointegrated in the long run and there is evidence of spatial error correction in the short run.

Keywords Panel cointegration · Spatial panel data · Error correction · Regional house prices

JEL Classification C21 · C33 · R21

1 Introduction

The “spatialization” of panel data econometrics in which spatial and temporal dynamics are integrated is still in its infancy (Elhorst 2003; 2004; Giacomini and Granger 2004; Beenstock and Felsenstein 2007). In this paper, we seek to spatialize recent developments in panel cointegration (Kao 1999; Pedroni 1999) that are

M. Beenstock
Department of Economics, Hebrew University of Jerusalem, Jerusalem, Israel

D. Felsenstein (✉)
Department of Geography, Hebrew University of Jerusalem, Jerusalem, Israel
e-mail: msdfels@mscc.huji.ac.il

designed to test hypotheses in which panel data happen to be nonstationary.¹ Since economic panel data are typically nonstationary either because their means and/or their variances vary over time, the need to develop spatial panel cointegration methods requires no justification. Indeed, spatial econometricians using panel data have either tended to ignore the issue of nonstationarity,² or they have dealt with it inappropriately.³ Or they have not ignored the issues of nonstationarity and error correction, but they have ignored spatial econometrics⁴ by treating the data as if they were nonspatial. In this paper, we seek to integrate spatial econometrics and panel cointegration for spatial data that are temporally nonstationary. Specifically, we estimate spatial error correction models (SpECM) in which error correction has temporal as well as spatial dimensions.

As originally pointed out by Engle and Granger (1987), cointegration and error correction are mirror images of each other. Error correction models describe the dynamic process through which cointegrated variables are related in the long run. We refer to spatial SpECMs as the dynamic process in which spatially cointegrated variables are related in the long run. Whereas conventional ECMs only contain temporal dynamics, SpECMs incorporate spatial as well as temporal dynamics. Therefore, SpECMs generalize ECMs to the case in which the panel units are not spatially independent.

SpECMs encompass three different types of cointegration. First, we refer by “local cointegration” to the case in which nonstationary panel data are cointegrated within spatial units but not between them. Local cointegration and panel cointegration are essentially identical concepts because the cross-section or spatial units are asymptotically independent. Secondly, “spatial cointegration” refers to the case in which nonstationary variables are cointegrated between spatial units but not within them. In this case, the long-term trends in spatial units are mutually determined and do not depend upon developments within spatial units. Thirdly, “global cointegration” refers to the case in which nonstationary spatial panel data are cointegrated both locally and spatially, i.e. nonstationary variables are cointegrated within and between spatial units.

SpECMs also encompass three different types of error correction. If error correction occurs within spatial units but not between them, we refer to this as “local error correction”. “Spatial error correction” refers to the case where error correction occurs between regions but not within them. Finally, “global error correction” refers to the case where error correction occurs within and between regions.

The taxonomy of SpECMs therefore includes combinations of the three different types of cointegration (local, spatial and global) and three different types of error correction (local, spatial and global) making nine different possibilities in all. By contrast, in the case of nonspatial panel data there is only one combination (local–local).

¹ Nonstationarity in the temporal sense rather than in the spatial sense as in Fingleton (1999).

² For example, Baltagi and Li (2004, 2006).

³ For example, Badinger et al. (2004) use deterministic time trends to stationarize their data.

⁴ For example, MacDonald and Taylor (1993), Malpezza (1999) and Meen (2002).

We define and clarify these concepts below. We illustrate SpECM using spatial panel data for house prices in Israel. We use these data to test the hypothesis dating back to Smith (1969), which predicts that house prices vary directly with the demand for housing as determined by population and income, and vary inversely with the supply of housing as measured by the stock of housing. This hypothesis has been investigated extensively at the national level. We regionalize the hypothesis by assuming that households base their location decisions on relative regional house prices and by assuming that building contractors decide where to build on the basis of regional house prices. We show that economic theory predicts that regional house prices should contain the spatial lag of regional house prices in the cointegrating vector.

Our contribution is twofold. The first is concerned with spatial econometrics; we spatialize error correction models estimated from nonstationary panel data. The second contribution is concerned with regional economics as applied to housing markets. We apply these spatial econometric methods to regional panel data on house prices in Israel. Specifically, we test whether the spatial lag of regional house prices in Israel belongs in the cointegrating vector for regional house prices as predicted by our theory. Our main empirical result is that regional house prices are globally cointegrated with house prices in neighboring regions as well as other variables within regions.

The econometric analysis of regional house prices has attracted recent attention. Most authors ignore spatial econometric issues.⁵ Holly et al. (2010) focus upon spatial econometric methodology, but there are no spatial dynamics in the hypothesis that they test.⁶ Cameron et al. (2006), rightly point out, “Regional house price models are not just national house price models with regional data substituted for national data.” In their model, households take relative house prices between regions into consideration in choosing where to live, thereby inducing spatial dependence in regional house prices. They use regional panel data on UK house prices to estimate error correction models in which lagged regional house prices in contiguous regions spillover temporarily onto house prices in neighboring regions. According to the regional housing model that we propose, these regional spillovers should be permanent and not merely temporary. Indeed, this is one of the key results that we obtain for spatial panel data for house prices in Israel.

2 Econometrics

2.1 Spatial vector error correction

Let Y_{it} and X_{ikt} denote spatial panel data where $i = 1, 2, \dots, N$ labels spatial units, $t = 1, 2, \dots, T$ labels time periods and $k = 1, 2, \dots, K$ labels covariates in the model. We assume that Y and X contain spatial panel unit roots and are therefore

⁵ Such as Malpezzi (1999), Capozza et al. (2004), Gallin (2003) and Fernandez-Kranz and Hon (2006).

⁶ Their central specification is that regional house prices vary directly with national house prices, regional income and national income.

nonstationary, hence $Y \sim I(d)$ and $X \sim I(d)$ $d \geq 1$. Phillips and Moon (1999) have pointed out that nonsense⁷ and spurious regression phenomena apply to panel data models if the data happen to be nonstationary. As pointed out by Kao (1999) and Pedroni (1999), parameter estimates are not spurious or nonsense if the residuals that they generate happen to be stationary.

Consider the following homogeneous model with specific effects in which $K = 1$:

$$Y_{it} = \alpha_i + \psi Z_t + \beta X_{it} + \theta Y_{it}^* + \delta X_{it}^* + u_{it} \tag{1}$$

Asterisked variables refer to spatial lags defined as:

$$Y_{it}^* = \sum_{j \neq i}^N w_{ij} Y_{jt} \quad X_{it}^* = \sum_{j \neq i}^N w_{ij} X_{jt}$$

where w_{ij} are row-summed spatial weights with $\sum_i w_{ij} = 1$. In Eq. 1, u_{it} denotes the residual and α_i denotes the spatial specific effect. Z denotes a vector of observed common factors that are hypothesized to affect all spatial units. Spatial dependence may be present in u . However, because of the specification of spatial lags in Eq. 1 spatial dependence in u is unlikely.

In nonspatial panels, $\theta = \delta = 0$ and panel cointegration implies $u \sim (0)$ when $d = 1$. Estimates of β are not spurious or nonsense when $u \sim I(0)$. In spatial panels, $Y^* \sim I(d)$ and $X^* \sim I(d)$, i.e. spatially lagged variables must have the same order of integration as the data from which they are derived because spatially lagged variables are linear combinations of the underlying data. Therefore, if Y is difference stationary, so must Y^* be difference stationary. Or if Y is trend stationary so must Y^* be trend stationary.

Since Y^* and X^* are nonstationary and have the same order of integration as Y and X , the cointegration space is enlarged when spatial panel data are nonstationary. We define spatial panel cointegration (SPC) as follows. SPC occurs when u is nonstationary in the absence of spatial lags in Y and X , i.e. when $\theta = \delta = 0$, but is stationary otherwise.

As established originally by Stock (1987) for time series data, OLS estimates of β are “super consistent” when the model is cointegrated since $\hat{\beta}$ is T —consistent or more instead of root T —consistent. This means that $\hat{\beta}$ converges rapidly upon β even in the case when X and u are not independent. Therefore, in nonstationary time series $\hat{\beta}$ is generally consistent, so that asymptotically it is not necessary to find instrumental variables for X . Phillips and Moon (1999) have established that super consistency⁸ also arises in the case of nonstationary panel data, so that estimates of β , θ and δ are super consistent. Fortunately, therefore, Eq. 1 may be estimated without recourse to instrumental variables for Y^* and X^* . In finite samples, matters might be different, but even in finite samples the bias might be negligible (Banerjee

⁷ “Nonsense regression”, discovered by Yule (1926), occurs when two independent driftless random walks are correlated. “Spurious regression”, discovered by Yule (1897), occurs when two independent random walks with drift are correlated through common stochastic time trends.

⁸ Root N - consistency is induced by the cross-section dimension in the data, while T consistency applies to the time-series dimension.

et al. 1986).⁹ Therefore, the debate whether GMM or IV is more appropriate for estimating spatial lag coefficients (Lee 2007) does not arise asymptotically in the case of nonstationary spatial panel data.

The SpECM associated with Eq. 1 in its first-order form is:

$$\Delta Y_{it} = \gamma_{0i} + \gamma_1 \Delta Y_{it-1} + \gamma_2 \Delta X_{it-1} + \gamma_3 \Delta Y_{it-1}^* + \gamma_4 \Delta X_{it-1}^* + \gamma_5 u_{it-1} + \gamma_6 u_{it-1}^* + \gamma_7 \Delta Z_t + v_{it} \quad (2)$$

where v are residuals that are assumed to be temporally uncorrelated, but they might be spatial correlated such that $\text{cov}(v_{it}v_{jt}) = \sigma_{ij}$ is nonzero. The local error correction coefficient γ_5 is expected to be negative, since u_{it-1} is positive when Y_{it-1} is greater than its equilibrium defined in Eq. 1. Therefore, Y_{it} is expected to decrease as it corrects itself toward its equilibrium value. In the short run, X may affect Y differently to how it affects Y in the long run, hence γ_2 might differ from β in Eq. 1. Also, potential short-term inertia in Y is captured by γ_1 . If there are spatial spillovers in error correction, the dynamics of Y will be affected by u^* among neighbors. Therefore, γ_6 is the spatial error correction coefficient and is expected to have the same sign as γ_1 . The short-term SAR coefficient γ_3 might differ from its long-run counterpart θ in Eq. 1. The same applies to the short-term spatial lag for X ; γ_4 might differ from δ in Eq. 1. As in Eq. 1 where α_i is a long-run specific spatial effect, γ_{0i} in Eq. 2 is a short-term specific spatial effect. Finally, the short-term effect of Z γ_7 may be different from its long-run counterpart ψ in Eq. 1.

Note that when the panel data are difference stationary all the variables in the SpECM are stationary since u and u^* are stationary when Eq. 1 is cointegrated. If $\gamma_5 = \gamma_6 = 0$ Eq. 2 becomes a spatial autoregression since it incorporates temporal lags and spatial lags of ΔY and ΔX as well as ΔZ . When γ_5 is negative, there is local error correction. When γ_6 is nonzero, there is spatial error correction. When both types of error correction occur, we refer to this as “global error correction”.

2.2 Unit root and panel cointegration tests

The available statistical tests for panel unit roots and panel cointegration assume that there is no spatial dependence between panel units. Panel unit root and panel cointegration tests have yet to be developed for spatially dependent panels. Baltagi et al. (2007) have investigated the effects of spatial dependence on the size of several panel unit root tests, including the heterogeneous panel unit root tests proposed by Im et al. (2003). They show that the IPS test is over—sized when the spatial autocorrelation coefficient of the residuals is large (0.8), so that there is an excess tendency to reject the null hypothesis of a unit root when it is true. If, however, the spatial autocorrelation coefficient is 0.4, the size of the IPS test is close to its nominal value. Therefore, provided that spatial dependence is not very strong, the IPS test (and most other tests) is useful for detecting unit roots even in spatially dependent panel data. The IPS test has been extended by Pesaran (2007) to include

⁹ There is no literature on the finite sample properties of panel cointegration, although Pedroni (2004) investigates the power and size of alternative cointegration tests in finite samples.

cross-section dependence through a common factor (CIPS). We use this test too but stress that it is not a substitute for a genuinely spatial extension of the IPS test.

Kao (1999) has suggested residual-based panel cointegration tests under the restrictive assumption that the only form of heterogeneity takes the form of fixed effects.¹⁰ By contrast, Pedroni (1999) proposed residual-based panel cointegration tests in which apart from heterogeneity induced by fixed effects, there may be heterogeneity in the cointegrating vector and in the error correction coefficients. Since Pedroni’s tests are more general, we adopt his semiparametric group- t test statistic, which is based on the average of the Dickey–Fuller statistics estimated from the different cross-section units, and which has more power than its rivals when T is small.

Dickey–Fuller regressions¹¹ are estimated for each cross-section unit using the estimated residuals from Eq. 1:

$$\Delta \hat{u}_{it} = \rho_i \hat{u}_{it-1} + v_{it} \tag{3}$$

$$v_{it} = \sum_{j=0}^J \delta_{ij} \varepsilon_{it-j} \tag{4}$$

$$\varepsilon_{it} \approx iid(0, \sigma_i^2) \tag{5}$$

where J is the bandwidth¹² for calculating the “long-term” variance of v_i equal to $\tilde{\sigma}_i^2 = \sigma_i^2 \sum_{j=0}^J \delta_{ij}^2$ with $\delta_{i0} = 1$ and $\delta_{ij} = 0$ for $j > 0$ if the v ’s are not autocorrelated. If the v ’s are serially independent, the short- and long-term variances of v_i are identical. In the presence of autocorrelation, the long-term variance exceeds its short-term counterpart since the difference between them is:

$$d_i = \tilde{\sigma}_i^2 - \sigma_i^2 = \sigma_i^2 \sum_{j=1}^J \delta_{ij}^2 > 0 \tag{6}$$

Next, we calculate:

$$P = \sqrt{N} \bar{t} - \frac{1}{2\sqrt{NT}} \sum_{i=1}^N \frac{d_i}{\tilde{\sigma}_i \sigma_{\hat{u}_i}} \tag{7}$$

where t -bar (\bar{t}) is the average of the t statistics for the ρ ’s estimated from Eq. 5. The last term in Eq. 5 is a nuisance parameter induced by serial correlation in Eq. 5. If it is zero, P is simply equal to root N times t -bar. Finally, the cointegration test statistic is:

$$z = \frac{P - E\sqrt{N}}{\sigma_E} \Rightarrow N(0, 1) \tag{8}$$

¹⁰ Panel cointegration tests based on Johansen (1995) have been suggested by Groen and Kleibergen (2003). We restrict ourselves here to residual-based tests.

¹¹ The augmented version (ADF) adds lagged dependent variables to Eq. 5.

¹² Newey and West (1994) have suggested that J equal the nearest integer to $4(T/100)^{0.222}$. When $T = 18$, $J = 3$.

where E and σ_E are derived from Monte Carlo simulation and are provided by Pedroni (1999) for various values of K (number of covariates). For example, if $K = 4$ the critical value for t -bar is -2.47 at $P = 0.05$ (two tailed) in the absence of nuisance parameters.

These panel cointegration and unit root test statistics assume that the cross-section units are independent. Dependence may be induced in at least two ways. First, there may be unobserved common factors¹³ that affect all the cross-section units. Secondly, there may be spatial relationships between the cross-section units as in Eq. 1. Spatially lagged variables induce dependence between cross-section units. However, spatial lags such as Y^* and X^* in Eq. 1 may be treated as any other variable that is hypothesized to belong to the cointegrating vector. If, instead, the residuals in Eq. 1 happen to be spatially autocorrelated matters might be different. If they are strongly spatially autocorrelated the results of Baltagi et al. (2007) for unit roots would suggest that the cointegration tests are over sized. However, if the spatial autocorrelation is not too strong the size of the cointegration tests is most probably close to their nominal values.

Since regional housing markets are likely to be affected by common factors such as interest rates and building costs, which do not vary by region, these and related variables should be explicitly specified in the model as the Z variables in Eq. 1 instead of treating them as unobserved common factors in the cointegrating vector for regional house prices.¹⁴

3 Economics of regional housing markets

We spatialize the stock-flow model of the housing market that is quite standard in the literature.¹⁵ This is a dynamic capital asset pricing model, in which the return to housing as an asset varies directly with returns on competing capital assets, but it also varies directly with the housing stock. The model also predicts that, given everything else, house prices should vary directly with income because housing is a nontraded good.¹⁶

Suppose there are two regions A and B in which the population (Q) is fixed so that $Q_{At} + Q_{Bt} = Q$ where Q_A and Q_B are naturally positive. The population choosing to live in A is determined through the following migration model:

$$Q_{At} = \varphi_0 - \varphi_1 P_{At} + \varphi_2 P_{Bt} \quad (9)$$

¹³ Common factor models have been discussed by Bai and Ng (2004) and Pesaran (2007). Gengenbach et al. (2005) discuss the implications of common factors for Pedroni's panel cointegration tests.

¹⁴ Holly et al. (2010) specify unobserved common factors in the housing market.

¹⁵ This model dates back to Witte (1963) and has been applied in many countries including by Smith (1969) for Canada, Kearl (1979) for the United States and Bar-Nathan et al. (1998) for Israel. It also features in numerous macroeconomic texts such as Dornbusch and Fischer (1990), Sachs and Larrain (1993) and Mankiw (2003).

¹⁶ Holly et al. (2010) and Meen (2002) take an extreme position by assuming the regional house prices are exclusively determined by income.

where P_A denotes house prices in region A. The coefficients φ_1 and φ_2 reflect regional residential preferences and imply that regions are imperfect substitutes for each other. If they are perfect substitutes $\varphi_1 = \varphi_2 = \infty$. At the other extreme, if there is no substitution at all $\varphi_1 = \varphi_2 = 0$.

We assume that housing construction costs are the same in A and B, and contractors choose to build where it is more profitable. However, there is in general imperfect substitution between building in A and B. Contractors therefore build more in A if housing is more expensive in A and less expensive in B. Housing construction (h) in A and B is determined as follows:

$$h_{At} = \eta_{A0} + \eta_{A1}P_{At} - \eta_{A2}P_{Bt} \tag{10}$$

$$h_{Bt} = \eta_{B0} + \eta_{B1}P_{Bt} - \eta_{B2}P_{At} \tag{11}$$

where η_{A0} and η_{B0} express productivity in construction in regions A and B, respectively. The housing stocks at the beginning of period t in the two regions are defined as:

$$H_{jt} = H_{jt-1} + h_{jt-1} - d_{jt-1} \quad j = A, B \tag{12}$$

where d denotes demolitions. The regional housing market is in equilibrium when $Q_{jt} = H_{jt}$.

We solve the model for house prices under the simplifying assumption that $d = \delta H_{t-1}$ where δ is a common demolition rate. House prices are dynamically and spatially correlated according to the model, so that current house prices in region A are related to lagged house prices in regions A and B, as well as current house prices in region B:

$$P_{At} = \frac{1}{\varphi_1}[\varphi_0 - \eta_{A0} - \eta_{A1}P_{At-1} + \varphi_2P_{Bt} + \eta_{A2}P_{Bt-1} - (1 - \delta)H_{At-1}] \tag{13}$$

Current house prices in region A vary inversely with the local housing stock and construction productivity (η_{A0}) and vary directly with the autonomous demand to live in A (φ_0). The solution for house prices in region B has the same form as Eq. 13:

$$P_{Bt} = \frac{1}{\varphi_2}[Q - \varphi_0 - \eta_{B0} - \eta_{B1}P_{Bt-1} + \varphi_1P_{At} + \eta_{B2}P_{At-1} - (1 - \delta)H_{Bt-1}] \tag{14}$$

Equation 13 may be used to generate the following long-term solution¹⁷ for regional house prices:

$$P_A = \pi_0 + \pi_1P_B - \pi_2H_A \tag{15}$$

$$\pi_0 = \frac{\varphi_0 - \eta_{A0}}{\varphi_1 + \eta_{A1}} \quad \pi_1 = \frac{\varphi_2 + \eta_{A2}}{\varphi_1 + \eta_{A1}} \quad \pi_2 = \frac{1 - \delta}{\varphi_1 + \eta_{A1}}$$

Equation 15 establishes that the long-term spatial lag coefficient on house prices is π_1 . A similar result applies to region B.

¹⁷ Obtained by setting variables at time t to equal their value at $t-1$.

4 Data

4.1 Data sources and definitions

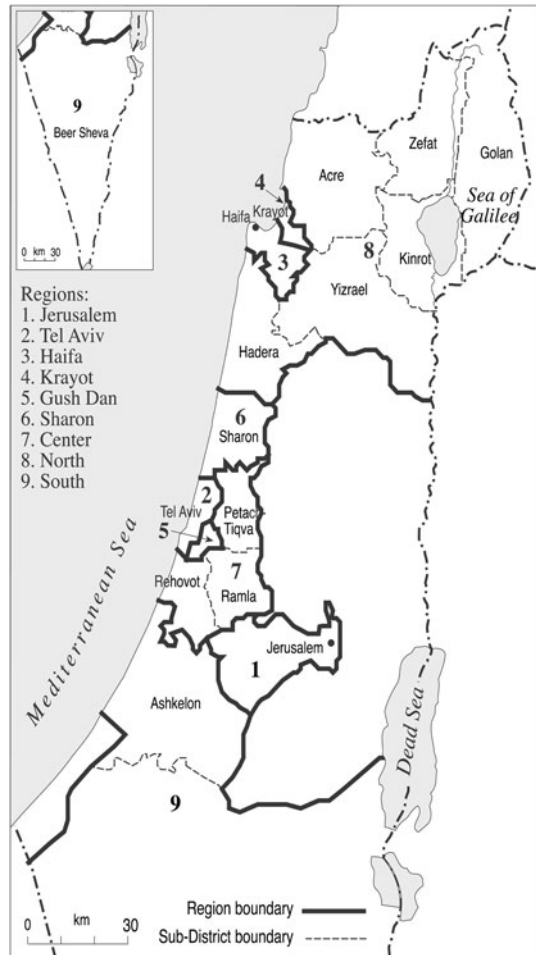
For our empirical application of SpVECM, we use annual panel data for nine regions in Israel (see Fig. 1) for the period 1987–2004. Israel is a small country and regional population size is comparable to the yardstick used for defining NUTS3 regions, i.e. 150,000 to 8000,000 population per region. Table 1 gives averages for key regional variables for 2000. The vector comprises four variables: real earnings, population, real house prices and the stock of housing (measured in 1,000's of square meters). Descriptive statistics for these variables can be found in Table 2. Hence, $T = 18$, $N = 9$ and $K = 4$.

Since these observations are too few to estimate individual models for each region, we pool the time series and cross-section data for purposes of estimation. We note that the panel unit root tests proposed by Im et al. (2003) report critical values for $T \geq 10$ and $N > 5$, in which case we feel that it is meaningful to use 18 years of data for nine regions.¹⁸ Calculations by IPS show that when $T = 18$ and $N = 9$, the size of the unit root test is about 0.05 and its power is about 0.2. This means that the probability of incorrectly rejecting the null hypothesis when it is true is about 5%, and the probability of correctly rejecting it when it is false is only about 20%. The latter would have been 26% with $T = 25$ and 75% with $T = 50$. In our opinion, what matters is the length of the observation period and not merely the number of data points. Eighteen monthly or even quarterly data points would not have been adequate because the observation period would have been only a year and half in the former case and four and half years in the latter. These periods would have been too short for observing convergence phenomena, whereas 18 years is in our opinion a sufficiently long period for these purposes.

It is natural to ask whether N is too small. Obviously, as N increases we learn more about spatial dependence. In short panels, T is small and N increases. In our case, we are short spatially because N is naturally fixed and small, while T increases. If T is sufficiently large, it does not matter that N is small. Indeed, if T were sufficiently large, there would be no need to use panel data econometrics in the first place. In our opinion, $T = 18$ is sufficiently large for meaningful statistical inference when $N = 9$. Surprisingly, this issue has not been addressed in the finite sample literature on panel data econometrics.

Real earnings in region n at time t (W_{nt}) have been constructed by us from the Household Income Surveys (HIS) of the Central Bureau of Statistics (CBS) and are deflated by the national consumer price index (CPI). The population in region n at the beginning of time t (POP_{nt}) is published by CBS. CBS also publishes indices of house prices for the nine regions, which are based on transactions data and which we deflate by the CPI. Finally, we have constructed the stock of housing in region n at the beginning of time t (H_{nt}), which is measured in (gross) square meters. We use data on housing completions in the nine regions measured in square meters,

¹⁸ For stationary panel data matters are different and N needs to be larger. Taylor (1980) shows that in short panels N needs to be about 50. See also footnote 20.

Fig. 1 Regional map of Israel

published by CBS. The change in the stock of housing is defined as completions minus our estimates of demolitions. The level of the housing stock is inferred from data in the 1995 census Fig. 1.

4.2 Panel unit root tests

The data are plotted in Fig. 2. Not surprisingly, all four variables have grown over time, hence they cannot be stationary. It should be noted that the 1990s witnessed mass immigration from the former USSR, which had major macroeconomic implications, especially for labor and housing markets. The population grew in all regions, but particularly in the South where housing was cheaper. In Table 3, we report panel unit root tests (t -bar) due to Im et al. (2003) as well as the common factor counterpart CIPS due to Pesaran (2007), which is the average of the first-order augmented Dickey–Fuller statistics for variable j in the nine regions. When

Table 1 Descriptive statistics—regional averages, 2000

	Educ (years)	Age	Pop (Th)	Total Emp.	Monthly earnings (Sh)
Center	13.67	41.19	847.70	400,107	3,658.25
Dan	13.34	40.62	676.30	257,760	3,045.94
Haifa	14.25	41.30	351.70	137,161	3,355.99
Jerusalem	14.59	42.06	823.50	180,841	3,134.52
Krayot	13.00	42.29	162.20	63,459	2,587.46
North	12.26	39.90	1147.10	324,990	2,915.78
Sharon	12.89	41.94	563.40	87,767	2,639.25
South	13.17	41.36	761.10	258,961	2,743.49
Tel-Aviv	13.98	40.53	356.20	216,234	4,088.24

$d = 0$, the absolute value of t -bar is below its critical value in the case of earnings and the housing stock, so these variables are clearly nonstationary. Surprisingly, however, Table 3 suggests that population and house prices are stationary in levels. However, the CIPS test clearly indicates that all four variables are nonstationary. When $d = 1$ absolute t -bar is greater than its critical value for all variables, hence all four variables are difference stationary. Although Table 1 suggests that earnings and the housing stock are $I(1)$ while population and house prices are $I(0)$, we assume all the variables to be $I(1)$. The data plotted in Fig. 2 are clearly trending, so that the conclusion that $d = 1$ is not controversial despite potential size distortions in Table 1.

5 Results

The discussion in this section naturally falls into two parts. We begin by testing for spatial panel cointegration in house prices. Thereafter, we estimate the error correction model for house prices that is derived from the cointegrating vector estimated in the first part. The stock-flow model described in Sect. 3 predicts that real house prices should be cointegrated with population, income and the housing stock as well as house prices in neighboring regions. Indeed, regional house prices should vary directly with the population and income in the region because these variables drive the demand for housing services, they should vary inversely with the housing stock because this variable determines the supply of housing services, and they vary directly with house prices in neighboring regions due to spatial substitution in house building and choice of location.

We investigate whether regional house prices are locally cointegrated or spatially cointegrated. In the latter case, long-run spatial lags belong in the cointegrating vector. The spatial connectivity matrix is defined in terms of relative populations so that w_{ij} is equal to the population in region j divided by the combined populations in regions i and j . If the populations are unequal, w_{ij} is different from w_{ji} , i.e. the W matrix is asymmetric. Note that W is fixed and does not vary over time and is row normalized.

Table 2 Descriptive statistics of variables in error correction model

Center	Dan	Haifa	Jerusalem	Krayot	North	Sharon	South	Tel-Aviv
Housing stock—1000 sq. meters								
Mean	20,206,530	11,109,314	17,385,091	4,828,307	21,747,100	14,379,336	15,917,012	14,529,798
SD	4,391,760	1,226,864	2,385,931	620,266.2	3,524,067	2,871,752	4,058,486	1,187,445
House prices (1991 prices)								
Mean	250.23	237.54	307.41	211.19	167.87	291.24	168.57	358.11
SD	32.15	51.31	50.99	48.84	34.34	44.68	23.77	64.38
Population (thousands)								
Mean	732.96	326.81	716.79	151.70	962.23	507.43	626.01	349.57
SD	150.78	27.80	118.75	11.74	184.27	61.59	143.13	15.22
Earnings (1991 prices)								
Mean	3,086.96	2,985.24	2,706.21	2,802.24	2,333.06	2,568.57	2,505.26	3,232.43
SD	386.75	239.30	330.98	230.08	224.42	381.67	238.17	498.53

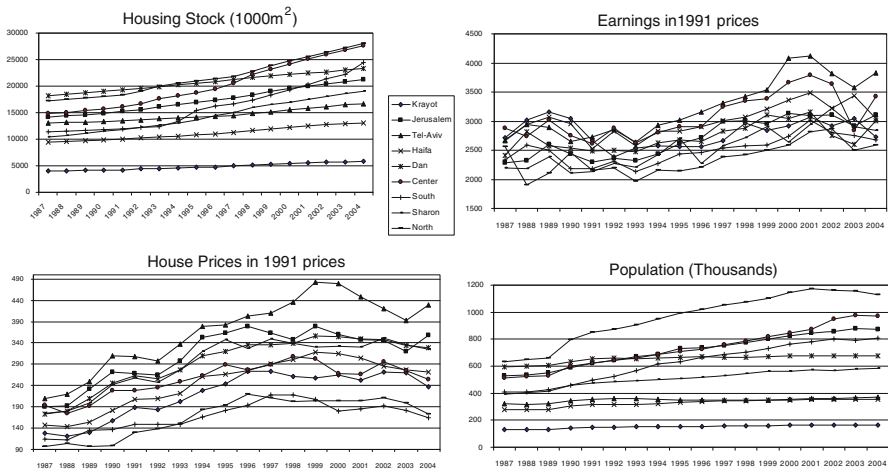


Fig. 2 Regional panel data

Table 3 Panel unit root tests (*t*-bar)

$\ln(Y_j)$	$d = 0$	$d = 1$	$d = 2$	CIPS $d = 0$
Earnings	-1.205	-3.503	-5.079	-1.67
Population	-2.707	-2.531	-6.603	-0.31
House prices	-3.030	-2.537	-5.321	-1.05
Housing stock	-0.092	-2.227	-3.410	-0.99

Auxiliary regression: $\Delta^d \ln Y_t = \alpha + \lambda \Delta^{d-1} \ln Y_{t-1} + \delta \Delta^d \ln Y_{t-1} + \varepsilon_t$. The critical values of *t*-bar with $N = 9$ and $T = 18$ are -2.28 at $P = 0.01$ and -2.17 at $P = 0.05$. CIPS denotes the common factor IPS test statistic due to Pesaran (2007)

We use the residuals generated by the cointegrating vector to estimate the error correction model for house prices. The key spatial issues here are twofold. First, is there a spatial lag in house prices in the ECM so that temporal and spatial dynamics are influential in the determination of house prices? Secondly, is there evidence of spatial error correction according to which spatially lagged residuals from the cointegrating vector directly affect house price dynamics in the short run?

5.1 Cointegration tests for regional house prices

Model A in Table 4 tests for local cointegration.¹⁹ The critical value for the group-*t* cointegration test statistic according to Pedroni (1999) is -2.02 . The calculated value is just equal to the critical value so that house prices, population, income and the housing stock are marginally panel cointegrated. Note that Pedroni’s test

¹⁹ We report parameter standard errors. Since parameter estimates from nonstationary data typically have nonstandard distributions, statistical significance cannot be measured by *t*-tests, which is measured instead by cointegration tests.

Table 4 Panel cointegration tests

Model	Population	Income	Housing stock	House prices*	Population*	Cointegration test
A	0.9140 (0.0926)	0.1163 (0.0485)	0.1325 (0.0758)			-2.02 -2.02
B	1.4661 (0.0793)	0.0181 (0.0386)	-0.5220 (0.0793)	0.1204 (0.0027)		-2.47 -2.47
C	1.1370 (0.0712)	0.0363 (0.0326)	-0.3165 (0.0661)	0.1841 (0.0061)	-0.0727 (0.0074)	-3.11 -2.82

Dependent variable is the ln real house prices in region i in year t . All variables are expressed in logarithms. Annual data for nine regions during 1987–2004 ($NT = 162$). Estimated with fixed regional effects and SUR. Standard errors in parentheses. Asterisked variables are spatial lags. Critical value of panel cointegration test statistic (group- t) in italics

statistic is asymptotic²⁰ with respect to T but refers to fixed $N = 9$. The finite sample properties with respect to T of the group- t cointegration test statistic and other test statistics are discussed in Pedroni (2004). In our case, $T = 18$ years. Therefore, model A is most probably not cointegrated. Indeed, the residuals of the regional DF cointegration test statistics are correlated.

The coefficient on the housing stock in model A, which should be negative turns out to be positive. In model B, we add a spatial lag in house prices. The group- t statistic improves (becomes more negative), but the critical value becomes more negative too so that the cointegration test statistic continues to be asymptotically marginal. However, the coefficient on the housing stock is negative instead of positive. The estimated long-run spatial lag coefficient on house prices is positive. In model C, we add a spatial lag in population. In this case, the cointegration test statistic is no longer marginal; group- t is smaller than its critical value, suggesting that model C is panel cointegrated.

The fixed regional effects are quite diverse. In the case of model C, for example, the log difference between the largest fixed effect (Tel Aviv) and the smallest (Krayot) is 0.803, which implies that controlling for covariates housing in Tel Aviv is more than twice as expensive (120%) as in Krayot. The estimated fixed effects polarize into expensive regions (Tel Aviv, Dan, Jerusalem, Center and Sharon) and cheap regions (Krayot, South and North) with Haifa in the middle.

5.2 Error correction in house prices

We use models reported in Table 4 to estimate stationary residuals for regional house prices, which are lagged and specified in the ECM for the change in log house prices. The ECM includes the lagged first difference in the variables listed in Table 4. These include lags of the spatially lagged variables as well as spatial lags of the estimated residuals. We use the residuals from model C in Table 4, in an example of a spatial ECM.

Table 5 reports an estimated spatial error correction model (SpECM) for regional house prices. We have excluded statistically insignificant terms such as income (W)

²⁰ Pedroni (2004) shows that it is oversized when $T < 120$ and underpowered when $T < 50$ for $H_0: \rho = 1$. When $T = N = 20$ and $\rho = 0.9$, its power is about 65% at $P = 0.05$. When ρ is smaller the power is naturally greater. Fortunately in our case ρ turns out to be small.

Table 5 Spatial error correction model for regional house prices (Dependent variable: $\Delta \ln p_{it}$)

	Coefficient	<i>t</i> -statistic
Intercept	-0.0005	-0.05
$\Delta \ln p_{it-1}$	0.1732	3.1
$\Delta \ln p^*_{it-1}$	0.1006	4.567
$\Delta \ln H_{it-1}$	0.6759	4.521
$\Delta \ln \text{POP}_{it-1}$	-0.3926	-3.591
$\Delta \ln W^*_{it-1}$	0.0762	1.882
EC_{it-1}	-0.7047	-8.622
EC^*_{it-1}	-0.6348	-4.63
R^2 adj	0.511	
Standard error	1.035	
DW	2.021	

EC and EC* based on Model C in Table 2. Method of estimation: Panel SUR with common effects

and the spatially lagged housing stock (H^*). EC denotes the error correcting term, which is equal to the lagged residual from model C in Table 4. EC* denotes its spatially lagged counterpart. In Table 5, both of these terms are negative and statistically significant, indicating that house prices are both spatially and locally cointegrated. Indeed, the sizes of the coefficients on EC and EC* indicate that about 70% of the local error is corrected within a year and 63% of the neighboring error spillover onto the local region. The latter also means that if house prices were too high in neighboring regions, this exerts downward pressure on local house prices, i.e. there is spatial spillover in error correction, just as there might be with any other variable.

Table 5 also incorporates spatial lags on house prices in the autoregressive component of the model. This means that the current rate of change in local house prices depends on the lagged rate of change in house prices in neighboring regions as well as the rate of change in lagged house prices in the locality. The spatial lagged coefficient is 0.1 in Table 5, whereas the coefficient on the lagged dependent variable is 0.1732. Substituting model C in Table 2 for EC and EC* in Table 5 produces the following second-order difference equation²¹ for perturbed house prices p defined as the deviation of the logarithm of house prices from their base run solution:

$$p_{it} = 0.488p_{it-1} - 0.173p_{it-2} - 0.418p^*_{it-1} - 0.1p^*_{it-2} + A_{it} \quad (16)$$

where A_{it} captures all the other variables in model C and Table 5 apart from $\log p$ and $\log p^*$. The roots of Eq. 16 are conjugate complex equal to $0.244 \pm 0.337i$ and the modulus is 0.416, which as expected lies inside the unit circle. Therefore, convergence to equilibrium is oscillatory but damped. The negative coefficients on

²¹ We have assumed for convenience that $w_{i^*} = 1/8$ since there are nine regions. Strictly speaking, we should have used the spatial weights used to estimate the model. However, the assumption of symmetric weights enables us to express Eq. 16 as a difference equation. This would not have been feasible, otherwise since the coefficients of the spatially lagged variables in Eq. 16 would inconveniently depend upon i .

p^* in equation (16) create the misleading impression that p varies inversely with p^* in the long run. This impression is misleading because p^* is not exogenous in Eq. 16 since it depends on regional prices as a whole. The long-run elasticity of P with respect to P^* is, of course, 0.1841 according to model C. Therefore, the long-run spatial lag coefficient for the logarithm of house prices is 0.1841 and the short run spatial lag coefficient is -0.418 from Eq. 16. The latter stems from the fact that EC^* is statistically significant in Table 5 so that error correction in neighboring regions spillover onto the local region. This means that when neighboring house prices increase, local house prices decrease in the next period as the error correction effect spillover. This makes local house prices too low, so that subsequently local error correction makes them increase.

Spatial lags for other variables such as income (W^*) also feature in Table 5. Indeed, whereas there is no local income effect in Table 5 there is a small but statistically significant spatially lagged effect of 0.07. The short-run effects of the housing stock and population on house prices in Table 5 are contrary to expectations with opposite signs to their long-run counterparts in model C in Table 4. The long-run effect of housing stock on house prices is of course negative according to model C. This means that shocks to the housing stock initially increase house prices, but eventually lower them, and because the roots are complex, house prices overshoot their long-run value before settling down. The same applies to the dynamic effect of population shocks on house prices, except in the opposite direction.

6 Conclusion

Our purpose has been twofold. The first has been to “spatialize” error correction models that are used in the econometric analysis of time series. In doing so, we are following a methodological research program that integrates the econometric analysis of spatial data and time series data. The SpECM extends our previous efforts to spatialize vector autoregressions into SpVARs. The next step would be to extend the single equation SpECM into a spatial vector error correction model (SpVECM) in which, for example, error correction models are estimated for regional housing construction as well as regional house prices.

The second purpose concerns the economics of regional housing markets. We suggest that regional housing models are not simply national models in which the data happen to be regional. The specification of regional housing models therefore needs to take into consideration that building contractors may choose to build in regions where house prices are higher and people may prefer to live in regions where housing is cheaper. Therefore, the choices of contractors and residents will be influenced, among other things, by the relative price of housing, especially between neighboring regions, where substitution is likely to be strongest. Our central theoretical conclusion is that when the stock-flow theory of housing markets is regionalized, a spatial lag is induced in the determination of regional house prices.

Since spatial panel data are typically nonstationary, spurious and nonsense regression phenomena may arise. We apply cointegration to test the hypothesis that

regional house prices in Israel vary directly with demand as determined by population and income and vary inversely with supply as measured by the stock of housing. We find, as predicted by our theory of regional housing markets, that adding spatially lagged house prices to the model significantly improves the degree of cointegration. This suggests that in the long-run local house prices are affected by spillovers from house prices in neighboring regions.

We also estimate a spatial error correction model for regional house prices in which error correction takes place within regions and between regions. We find clear evidence of spatial lags in error correction, suggesting that disequilibria in regional housing markets spillover onto neighboring regions in the short run. We also find that spatially lagged house prices are statistically significant in the spatial error correction model for house prices, as well as other spatially lagged variables.

Finally, we contrast our findings with other recent spatial econometric analyses of regional house prices in the United Kingdom (Cameron et al. 2006) and the United States (Holly et al. 2010). There are three mechanisms through which regional house prices may be affected by spatial lags. In the first, there is a long-run spatial lag, which implies that regional house prices depend on neighboring house prices in the long term and there is a spatial lag on house prices in the cointegrating vector. In the second, there is spatial error correction in which there is short-term spillover from disequilibrium in neighboring housing markets. In the third, there is a regular spatial lag in the error correction model. Holly et al. omit all three mechanisms. Cameron et al. omit the first two mechanisms but include the third. We find empirical evidence for all three mechanisms. This means that there are spillovers from neighboring housing markets in the short-, medium- and long terms.

References

- Badinger H, Müller WG, Tondl G (2004) Regional convergence in the European Union 1985–1999: a spatial dynamic panel analysis. *Reg Stud* 38(3):241–253
- Bai J, Ng S (2004) A PANIC attack on unit roots and cointegration. *Econometrica* 72:1127–1177
- Baltagi BH, Li D (2004) Prediction in the panel data model with spatial correlation, pp 283–295. In: Anselin L, Florax R, Rey S (eds) *Advances in spatial econometrics: methodology, tools and applications*. Springer, Berlin
- Baltagi BH, Li D (2006) Prediction in the panel data model with spatial correlation: the case of liquor. *Sp Econom Anal* 1(2):175–185
- Baltagi BH, Bresson G, Pirotte A (2007) Panel unit root tests and spatial dependence. *J Appl Econom* 22(2):339–360
- Banerjee A, Hendry DF, Dolado J (1986) Exploring equilibrium relationships in econometrics through static models: some Monte Carlo evidence. *Oxford Bull Econ Stat* 48(3):253–277
- Bar-Nathan M, Beenstock M, Haitovsky Y (1998) The market for housing in Israel. *Reg Sci Urban Econ* 28(1):21–50
- Beenstock M, Felsenstein D (2007) Spatial vector autoregressions. *Sp Econom Anal* 2(2):167–196
- Cameron G, Muellbauer J, Murphy A (2006) Was there a British house price bubble? Evidence from regional panel data, mimeo. University of Oxford, Oxford
- Capozza DR, Hendershott PH, Mack C, Mayer CJ (2004) Determinants of real house price dynamics. *Real Estate Econom* 32(1):1–32
- Dornbusch R, Fischer S (1990) *Macroeconomics*, 5th edn. McGraw Hill, New York
- Elhorst JP (2003) Specification and estimation of spatial panel data models. *Int Reg Sci Rev* 26:244–268

- Elhorst JP (2004) Serial and spatial error dependence in space-time models. In: Getis A, Mur J, Holler HG (eds) *Spatial econometrics and spatial statistics*. Palgrave Macmillan, London
- Engle R, Granger CWJ (1987) Co-integration and error correction: representation, estimation and testing. *Econometrica* 55(2):251–276
- Fernandez-Kranz D, Hon MT (2006) A cross-section analysis of the income elasticity of housing demand in Spain: is there a real estate bubble? *J Real Estate Finance Econ* 32(4):449–470
- Fingleton B (1999) Spurious spatial regression: some Monte Carlo results with a spatial unit root and spatial cointegration. *J Reg Sci* 39:1–19
- Gallin J (2003) The long-run relationship between house prices and income: evidence from local housing markets. Finance and Economics Discussion Series, Federal Reserve Board, Washington DC
- Gengenbach C, Palm FC, Urbain J-P (2005) Panel cointegration testing in the presence of common factors, mimeo, Maastricht University
- Giacomini R, Granger CWJ (2004) Aggregation of space-time processes. *J Econom* 118(1–2):7–26
- Groen JJJ, Kleibergen F (2003) Likelihood-based cointegration analysis in panels of vector error-correction models. *J Bus Econ Stat* 21:295–314
- Holly S, Pesaran MH, Yamagata T (2010) A spatio-temporal model of house prices in the US. *J Econom*. forthcoming
- Im K, Pesaran MH, Shin Y (2003) Testing for unit roots in heterogeneous panels. *J Econom* 115(1):53–74
- Johansen S (1995) *Likelihood-based Inference in cointegrated vector autoregressive models*. Oxford University Press, Oxford
- Kao C (1999) Spurious regression and residual based tests for cointegration in panel data. *J Econom* 90(1):1–44
- Kearl JR (1979) Inflation, mortgages and housing. *J Polit Econ* 87(1):1115–1138
- Lee L (2007) GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *J Econom* 137:489–514
- MacDonald R, Taylor MP (1993) Regional house prices in Britain: long run relationships and short term dynamics. *Scott J Polit Econ* 40(1):43–55
- Malpezzi S (1999) A simple error correction model of house prices. *J Hous Econ* 8(1):27–62
- Mankiw NG (2003) *Macroeconomics*. Worth, New York
- Meen G (2002) Time-series behavior of house prices: a transatlantic divide? *J Hous Econ* 11(1):1–23
- Newey W, West K (1994) Autocovariance lag selection in covariance matrix estimation. *Rev Econ Stud* 61(4):631–653
- Pedroni P (1999) Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bull Econ Stat* 61(4):653–670
- Pedroni P (2004) Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis. *Econom Theory* 20:597–625
- Pesaran MH (2007) A simple panel unit root test in the presence of cross section dependence, mimeo. *J Appl Econom* 22(2):265–310
- Phillips PCB, Moon H (1999) Linear regression limit theory for nonstationary panel data. *Econometrica* 67(5):1011–1057
- Sachs JD, Larrain FB (1993) *Macroeconomics in the global economy*. Prentice Hall, New Jersey
- Smith LB (1969) A model of the Canadian housing and mortgage market. *J Polit Econ* 77(5):795–816
- Stock J (1987) Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica* 55(5):1035–1056
- Taylor WE (1980) Small sample considerations in estimation of panel data. *J Econom* 13(2):203–233
- Witte J (1963) Microfoundations of the social investment schedule. *J Polit Econ* 74(5):441–456
- Yule GU (1897) On the theory of correlation. *J Roy Stat Soc* 60(4):812–854
- Yule GU (1926) Why do we sometimes get nonsense-correlations between time series? A study in sampling and the nature of time series. *J Roy Stat Soc* 89(1):1–64