

Dwelling Price Ranking versus Socioeconomic Clustering: Possibility of Imputation

Larisa Fleishman¹, Yury Gubman¹, and Aviad Tur-Sinai¹

In order to characterize the socioeconomic profile of various geographic units, it is common practice to use aggregated indices. However, the process of calculating such indices requires a wide variety of variables from various data sources available concurrently. Using a number of administrative databases for 2001 and 2003, this study examines the question of whether dwelling prices in a given locality can serve as a proxy for its socioeconomic level. Based on statistical and geographic criteria, we developed a Dwelling Price Ranking (DPR) methodology. Our findings show that the DPR can serve as a good approximation for the socioeconomic cluster (SEC) calculated by the Israel Central Bureau of Statistics for years when the required data was available. As opposed to the SEC, the suggested DPR indicator can easily be calculated, thus ensuring a continuum of socioeconomic index series. Both parametric and nonparametric statistical analyses have been carried out in order to examine the additional social, demographic, location, crime and security effects that are exogenous to SEC. Complementary analysis on recently published SEC series for 2006 and 2008 show that our conclusions remain valid. The proposed methodology and the obtained findings may be applicable for different statistical purposes in other countries which possess dwelling transactions data.

Key words: Housing market; urban locality; index construction.

1. Introduction

The socioeconomic profile of a residential area can be identified and characterized in two different ways: by using specific demographic and socioeconomic factors, or by estimating aggregated indices based on a range of these factors. In Israel, a socioeconomic index (SEI) for local authorities was developed at the Central Bureau of Statistics (CBS) in the mid-1990s. The main and most important use of the data on SEI is its contribution to the design and implementation of different policies in various ministries and other governmental agencies relating to local authorities, including various resource allocation procedures. However, the main limitation in using SEI is that the SEI data is not available for all localities every year, since the process of SEI calculation requires a wide variety of variables from various data sources available concurrently for various localities.

In previous studies, there is well-documented evidence of the relationship between the various demographic and socioeconomic characteristics of a locality and its dwelling prices. In Israel, the main source of information about dwelling prices is the record of real estate transactions kept by the Israel Tax Authority (ITA). The Israel CBS has been receiving this database on a monthly basis since 1990. Using this data source, our study

¹ Israeli Central Bureau of Statistics, 66 Kanfei Nesharin St. Jerusalem 91342, Israel. Emails: larisaf@cbs.gov.il, yuryg@cbs.gov.il, and aviadts@cbs.gov.il

proposes a method for assessing the socioeconomic level of those localities and for those time points for which these data are missing.

The current study has two main goals: 1) to examine if dwelling prices can serve as a proxy for the socioeconomic level of various urban localities; 2) to examine the extent of correlation between social and demographic characteristics not included in the SEI calculations and the level of dwelling prices in a given locality.

In order to achieve the first goal, relevant urban localities in Israel were graded according to their dwelling price level, and this was compared to their socioeconomic level. A similar analysis was conducted for the years 2006 and 2008. To achieve the second goal, parametric models were estimated. In this way conclusions were drawn regarding the possibilities for imputation of missing data on SEI values.

This study augments the relevant research literature in several ways. First, previous studies have focused primarily on the effect of demographic and socioeconomic factors on the level of dwelling prices in specific cities or metropolitan areas. We extended this investigation to a national level. Furthermore, it should be noted that Israel is a highly urbanized country with more than ninety percent of Israelis living in urban areas. Second, our investigation focused on the correlation between socioeconomic indices and dwelling price indicators. Only limited research has been carried out on the degree of correspondence between these two indicators. Third, this study proposes, examines and discusses a reasonable alternative method for assessing the socioeconomic level of urban localities. This method does not require a range of multivariate procedures and various statistical data available concurrently, although it allows for a sound approximation to a socioeconomic level. Research thus far has not addressed this issue.

The rest of this article is organized as follows: Section 2 presents some theoretical background and a survey of relevant literature, and explains the construction of the socioeconomic index in Israel. Section 3 describes the main source of data, the construction of the key variable reflecting dwelling price level, and defines geographic units for which this variable was created. Section 4 presents and discusses the degree of correspondence between the dwelling price level and socioeconomic level, describes additional sources of information and defines the variables used in the analysis. Section 5 presents the statistical models that we used for the empirical analysis, gives the results and discusses them. Section 6 concludes the study.

2. Literature Review

2.1. Dwelling Price in Light of Socioeconomic Level Characteristics

The price of a residential property on the free market reflects the willingness of the purchaser to pay not only for the property itself, but also for a specific residential environment – in other words, for the quality of “social space” (Reed 2001).

Research findings from around the world testify to the range of factors which reflect the socioeconomic essence of a residential area, the most important of which are income, education, employment, and the demographic characteristics of its population. Earlier studies show a significant positive correlation between the three major factors characterizing the residential area’s socioeconomic level – income, education and

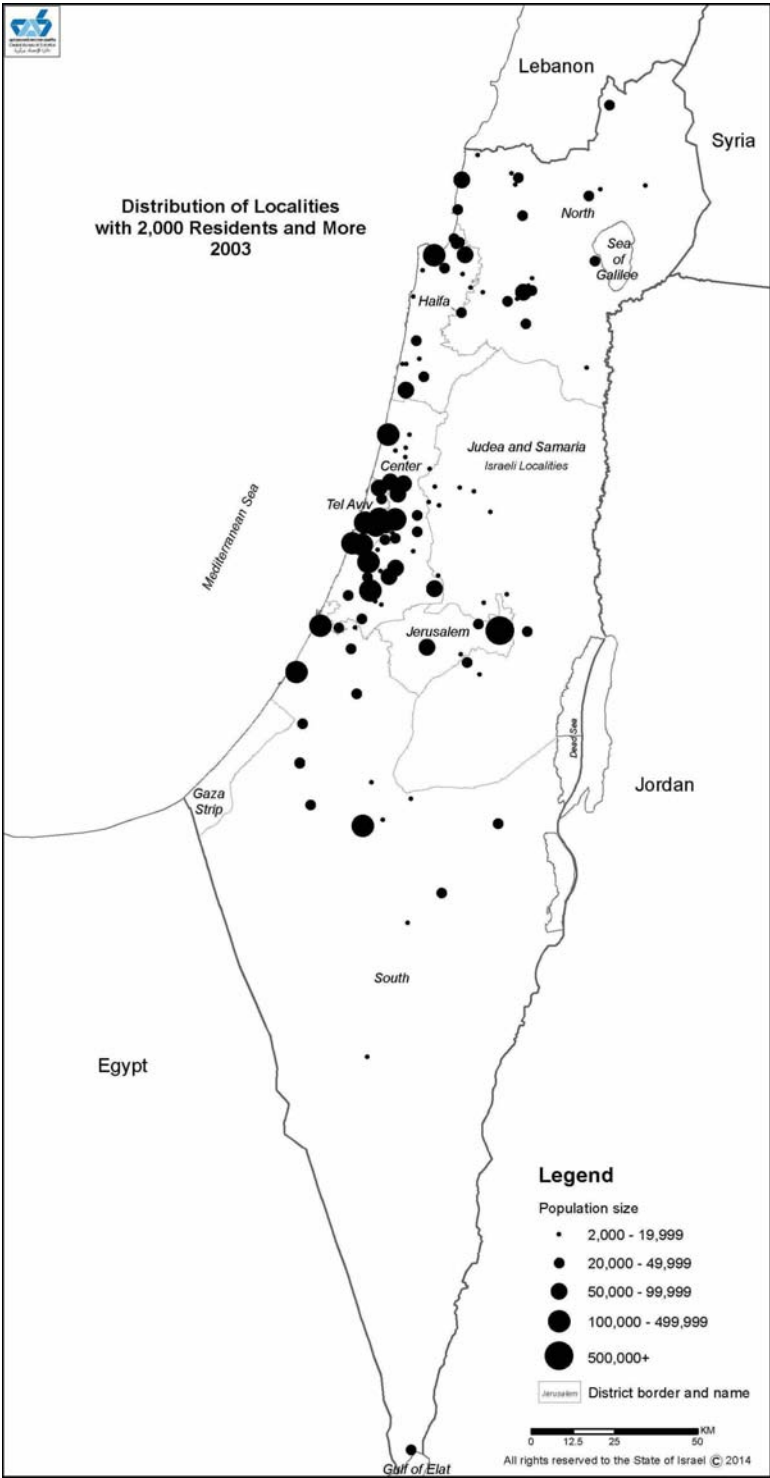
employment – and the price of dwellings (Heikkila 1992; Potepan 1996; Goodman and Thibodeau 1998; Greenberg 1999; Des Rosiers et al. 2002; Yates 2002). Income is considered the primary factor (Ozanne and Thibodeau 1983; Malpezzi et al. 1998). Those with relatively high incomes choose their residential area in an attempt to avoid neighbors with a low socioeconomic status. The popular viewpoint considers social problems, such as crime, drug use, and the neighborhood's economic decline resulting in neglected buildings, as all directly linked to neighborhoods characterized by a high proportion of unemployed and low levels of education and income (Harris 1999; Jackson et al. 2007). The study of Cummings et al. (2002) examined education as one of the dimensions of the socioeconomic level of residents of various neighborhoods in the city of Philadelphia and its influence on the price of residential dwellings in those neighborhoods. The study findings show a 21 percent increase in dwelling prices with every ten percent rise in the proportion of adults with post-high school education.

Aside from the aforementioned factors that characterize the socioeconomic space and impact dwelling price, the relevant literature has examined the relationship between the demographic characteristics of residents, such as age and marital status, and dwelling prices in that neighborhood (Myers 1990; Heikkila 1992). There are studies indicating that ethnic composition and personal security in a residential environment may also contribute to the socioeconomic space, and as a result affect the price of dwellings. In particular, previous studies provide sound evidence of a strong positive correlation between the level of personal security in a residential area and its dwelling prices (Thaler 1978; Dubin and Goodman 1982; Buck et al. 1991; Hazam and Felsenstein 2007).

Earlier studies have also addressed the effect of immigrant groups and the racial-ethnic context of residential areas on the local housing prices (Kiel and Zabel 1996). However, there is no common agreement on either the existence or the magnitude of the effect of immigration shocks on the housing market. The magnitude of the effect immigrants have on housing prices depends heavily on the reaction of natives to the presence of immigrants in the area. For example, some studies in the USA have found that blacks and Hispanics own homes of lesser value than the white population. This held true even when the researchers controlled for the characteristics of dwellings (Horton and Thomas 1998; Krivo 1995; Lewin-Epstein et al. 1997). Harris (1999) found that dwelling prices in the USA decline by an average of 16 percent when the Afro-American population exceeds ten percent in a neighborhood. Furthermore, a far more dramatic drop in prices occurs when the Afro-American population exceeds 60 percent of the neighborhood residents. According to this study, the explanation is not necessarily ethnic preference, but may be related to social problems stemming from the socioeconomic status of the Afro-American population, which is usually lower. However, a study conducted in Darwin (Australia) by Jackson et al. (2007) revealed high positive correlation between housing prices and ethnicity for people born overseas and for those who speak other languages.

Along with the abovementioned demographic and socioeconomic characteristics of a population, religiosity contributes substantially to the residential profile of a locality (Blanchard 2007). Among several religiosity patterns displayed by the Israeli Jewish population, both ultra-Orthodox and Orthodox streams play a significant role with regard to the socioeconomic, ethnic and spatial divide in Israeli society, thus notably effecting local dwelling market price level (Cahaner 2012).

Map 1. Distribution of Urban Localities in Israel



In addition, the importance of the geographical location of dwellings in the context of their price level has been emphasized in several studies (McCluskey et al. 2000; Bourassa et al. 2003). In Israel this issue is of special importance. According to the official administrative division, there are six main administrative districts in Israel. Map 1 shows the distribution of urban localities as well as administrative districts. Tel Aviv district is composed of the central city of Tel Aviv and other cities enclose it on three sides. This district is the geographical center of Israel, and is characterized by a very highly concentrated urban population. It is the financial, economic, social, and cultural center of the country.

Tel Aviv district is dominant and influential in the domains of employment and communication, as well as the domain of land and dwellings prices (Soffer and Bystrov 2006). Therefore, the geographical proximity of a locality to the center, or to the periphery, affects many aspects of life in a locality, including the socioeconomic level of the population and dwelling price level. Summarizing this short literature review, it can be concluded that a sizeable body of literature provides evidence of a strong association between various demographic, social and economic characteristics of a locality and the price of its dwellings. This evidence serves as a theoretical foundation for the premise behind this study: the price of dwellings in a specific locality can serve as an alternate measure of its socioeconomic level.

2.2. *The Socioeconomic Index and the Socioeconomic Cluster*

In order to characterize and document the socioeconomic profile of various localities, it is common practice in the official statistics of various countries to use aggregated indices (Burck and Kababia 1996, 1999; Australian Board of Statistics 2006). These indices are based on different theoretical assumptions and estimation methods, and may be classified in accordance with two main approaches. Using a “deterministic” approach, a socio-economic cluster is specified by applying predefined classification criteria based on an underlying conceptual model. For instance, Rose and Prevalin (2001) suggest the set of employment status and occupation variables for socioeconomic classification in the UK, following the social class classification methodology earlier suggested by Olausson and Vagero (1991) for Swedish register data.

The “stochastic” approach assumes the existence of a latent continuous variable (Y) for the socioeconomic level of a given locality. It is also assumed that Y may be assessed using multivariate analysis methodology on a set of observed variables. Finally, localities are clustered by the estimated values of Y . For example, Jackson et al. (2007) suggest estimating socioeconomic level using principal component methodology applied to a wide set of socioeconomic characteristics which includes income, age, family status, dwelling data, and so on. Generally, Principal Component Analysis (PCA) is a technique that is useful for the compression and classification of data. The main idea of PCA is to reduce the dimensionality of a data set which contains a large number of interrelated variables. This reduction is achieved by finding a new set of uncorrelated variables (the principal components) smaller than the original set of variables that nonetheless retains most of the variation present in the original data set (Jolliffe 2002).

The stochastic approach is widely used for socioeconomic index calculation in the official statistics of different countries. The Office for National Statistics in the UK devises the

socioeconomic index for areas within local authorities by means of principal component analysis based on population censuses. The Australian Bureau of Statistics produces five socioeconomic indices that measure various socioeconomic aspects of residential areas based on the population census. Surveys are used for updating the index in the periods between population censuses. A similar methodology is used in New Zealand.

In Israel, a socioeconomic index (SEI) was developed at the Central Bureau of Statistics (CBS) in the mid-1990s on the basis of the 1995 Population and Housing Census data (Burck and Kababia 1996). It is based on five groups of variables that include 14 variables. The variables used to construct the index reflect all of the aspects related to the socioeconomic makeup of the population of different localities, subject to the availability of the data (for more details on the selection of the variables, see CBS 2000). These five groups contain the following variables: (1) demographic characteristics (dependency ratio, median age, percentage of families with four or more children); (2) education and schooling (percentage of the students studying for a bachelor's or higher degree, percentage eligible for a matriculation certificate); (3) standard of living (level of motorization, percentage of new motor vehicles, average income per capita); (4) labor force statistics (percentage of job seekers, percentage of salaried workers and self-employed persons earning up to minimum wage, percentage of salaried workers earning more than twice the average salary); (5) support/pension (percentage receiving unemployment benefits, percentage receiving income supplements; percentage receiving old age pensions with income supplements). SEI is based on the stochastic approach and calculated using principal component analysis.

Principal components (factors) are essentially new variables, calculated as a linear combination (weighted average) of the original, standardized variables (i.e., each variable has a mean of 0 and variance of 1). The weights of the original standardized variables are determined mathematically so as to maximize the differences in the scores between the geographical units, subject to some normalization restrictions. The factors are determined sequentially, so that the first factor is the linear combination that accounts for the maximum amount of the variance of the variables. Hence, the first factor has the greatest ability to discern between the localities. The second factor accounts for a maximum variance not accounted for by the first factor, and so on. The optimal number of factors that should be used to explain the maximum amount of the variance of the variables is determined by statistical testing. It is noteworthy that since the variables are standardized, the total variance of the original variables is equal to the number of variables. These factors define an orthogonal set of axes in the multidimensional variable space where each factor is a linear combination of the original variables. This type of factor analysis can be defined as PCA (CBS 2000).

We can represent the socioeconomic index as follows:

$$SEI = a_1X_1 + a_2X_2 + \dots + a_{14}X_{14} \quad (1)$$

with SEI indicating the socioeconomic index (continuous), a_1, \dots, a_{14} the coefficients calculated using principal component analysis, and X_1, \dots, X_{14} the variables constituting the index which were specified above.

The set X_1, \dots, X_{14} has been defined based on the methodological background identified in the relevant literature, while considering local conditions and data

availability. It should be noted that, as opposed to [Jackson et al. \(2007\)](#), the dwelling prices are not included in the SEI calculations.

The SEI estimates cover most local authorities for which all the variables listed above are available at the relevant time point. For the sake of consistency and comparability of the SEI series, the set of variables and the calculation methodology have not been changed over the years.

Using cluster analysis, the local authorities, for which the SEI is calculated, are then divided into ten socioeconomic clusters (SEC), with Cluster 1 including authorities with the lowest socioeconomic level, and Cluster 10 including authorities with the highest socioeconomic level.

After 1995, the SEI was updated for the years 1999, 2001, 2003, and 2006 when the required data was available. The SEI calculated in 2008 on the basis of the 2008 Population Census is the most recent. As dwelling price data are available annually, we suggest a univariate deterministic approach for approximation of the socioeconomic level of a given locality at a given time.

In this context, it is worth noting that the relevant literature is mostly dedicated to the correlation between the different demographic and socioeconomic characteristics of a locality and the dwelling prices in it, and the examination of the degree of correspondence between aggregated socioeconomic indices and dwelling prices remains beyond the scope of research. Thus, to reach a conclusion as to the ability of dwelling prices to serve as a proxy for the socioeconomic level, we first need to check the degree of correspondence between them. Afterwards, we examine certain additional social and demographic characteristics that are not currently included in the SEI calculations, but are known to affect dwelling prices.

3. Data and Definitions

The study is based on files of dwelling transactions in the housing market in 2001 and 2003. Transaction data are provided annually by the Israel Tax Authority. In total, the basic file from 2001 included 60,851 transactions, and the file from 2003 included 57,223 transactions.

In order to compare the SEC of a given locality with its aggregate dwelling price level, the same coding scheme had to be used for both indicators. Dwelling Price Ranking (DPR) was constructed based on the following steps. First, using the transactions data, dwelling price level (DPL) for each relevant locality k was calculated:

$$DPL_k = \log(\text{median}(Y_k)), \quad (2)$$

where Y denotes price per square meter. Using price per square meter as an underlying variable for DPR, we neutralize the effect of apartment size, one of the main variables which explains differences in dwelling prices ([Lozano-Gracia and Anselin 2012](#)), and represents as far as possible the market value of a dwelling at the aggregate level for a given locality. The median is used for reasons of robustness, and the log-transformation stabilizes the variance and generally makes the data normally distributed.

Second, localities for which the DPR was created were selected using the following criteria: (1) total population of 2,000 or more in locality, which corresponds to the

definition of “urban” in Israel; (2) the number of transactions in a locality should be sufficient enough to represent the price level in the housing market (at least 15 in the current study). Localities that did not match the above criteria were excluded from the analysis. The final data set includes 104 localities in 2001 and 112 localities in 2003, covering about 90 percent of the Israeli population.

Third, the selected localities were divided into ten clusters, alongside the SEC. Localities with the lowest DPL were ranked as Level 1 ($DPR = 1$), while the localities with the highest dwelling prices were ranked as Level 10 ($DPR = 10$). Each of the resulting DPR clusters contains approximately the same number of localities.

4. Dwelling Price Ranking vs. Socioeconomic Cluster

In order to examine the degree of correspondence between the SEC and the DPR, a correlation analysis was carried out. The obtained Spearman correlation coefficients are equal to 0.69 and 0.67 for 2001 and 2003, respectively.

Table 1 presents the detailed results for 2003; the analysis for 2001 revealed similar results. In Table 1, the digit in each cell indicates the number of localities with DPR and the SEC as they appear in the rows and columns, respectively. The cases in which both rankings are identical appear in bold print; there is an exact correspondence between the SEC and the DPR for 21 localities (out of 112).

Based on these results, it can be concluded that a lack of correspondence between the SEC and the DPR is more typical for localities where the SEC is low or low-medium. The minimal gap between the SEC and the DPR (± 1 range) is observed for 28 percent of the localities. For localities where the gap between the SEC and the DPR is greater than 2, it was found that localities with a DPR higher than their SEC are mainly situated close to the Tel Aviv district, that is, close to the center of the country. Those are localities where the SEC is medium-high to high (6–9). For localities with a low-medium SEC or medium SEC (3–5), in most cases the DPR was lower than the SEC. Those localities are located in the more peripheral areas.

Table 1. The socioeconomic cluster vs. the dwelling price ranking 2003

DPR	The socioeconomic cluster										Total
	1	2	3	4	5	6	7	8	9	10	
1	–	1	2	6	1	1	–	–	–	–	11
2	–	–	–	3	7	–	1	–	1	–	12
3	–	1	–	4	4	2	–	–	–	–	11
4	–	–	2	4	2	1	–	1	1	–	11
5	1	1	–	–	5	4	–	1	–	–	12
6	–	1	–	1	1	3	4	1	–	–	11
7	–	–	–	–	1	2	5	1	–	–	9
8	–	–	–	–	1	2	3	3	2	1	12
9	–	1	–	1	–	1	5	4	–	–	12
10	–	–	–	–	–	–	–	9	1	1	11
Total	1	5	4	19	22	16	18	20	5	2	112

These findings indicate the spatial aspects contained in the correlation between the SEC and the DPR. In particular, it can be concluded that the dependence of dwelling prices on the distance from the Tel Aviv district is stronger than the spatial dependence for the SEC. That is, it is rare to find very expensive dwellings in the peripheral regions, while there are some peripheral localities with a comparatively high socio-economic profile.

It can also be seen that the degree of correspondence between the two indicators increases with the rise in the SEC scale of the localities. Maps 2 and 3 illustrate spatial distribution of the localities according to the SEC and the DPR for 2003.

We conclude that the suggested dwelling price ranking appears to be a sufficiently good approximation for the socioeconomic cluster. However, a gap is revealed between two indicators, and it is therefore reasonable to assume that there are other factors influencing dwelling prices. Using these factors, we attempted to correct the developed dwelling price ranking by reducing the gap between the DPR and the SEC.

In order to examine additional factors that are not included in the SEC calculation but are assumed to influence dwelling prices, the following administrative databases are used. First, the Population Registry from which information on population characteristics by locality was obtained (e.g., percentage of immigrants from the former USSR and Ethiopia in 2001 and 2003). Second, the “Level of Religiosity” administrative file that was developed at the CBS serves as a basis for such variables as the percentage of ultra-Orthodox and Orthodox population by locality. Third, a crime database was provided by the Israeli Police. Finally, terror incidents data for relevant years was created by using information from different sources available from the International Institute for Counter-Terrorism (ICT) at the Interdisciplinary Center (IDC) Herzliya, Ministry of Foreign Affairs and the Prime Minister’s Office. Additionally, spatial information regarding the location of the localities relative to the center of the country was provided by the Geographic Information System (GIS).

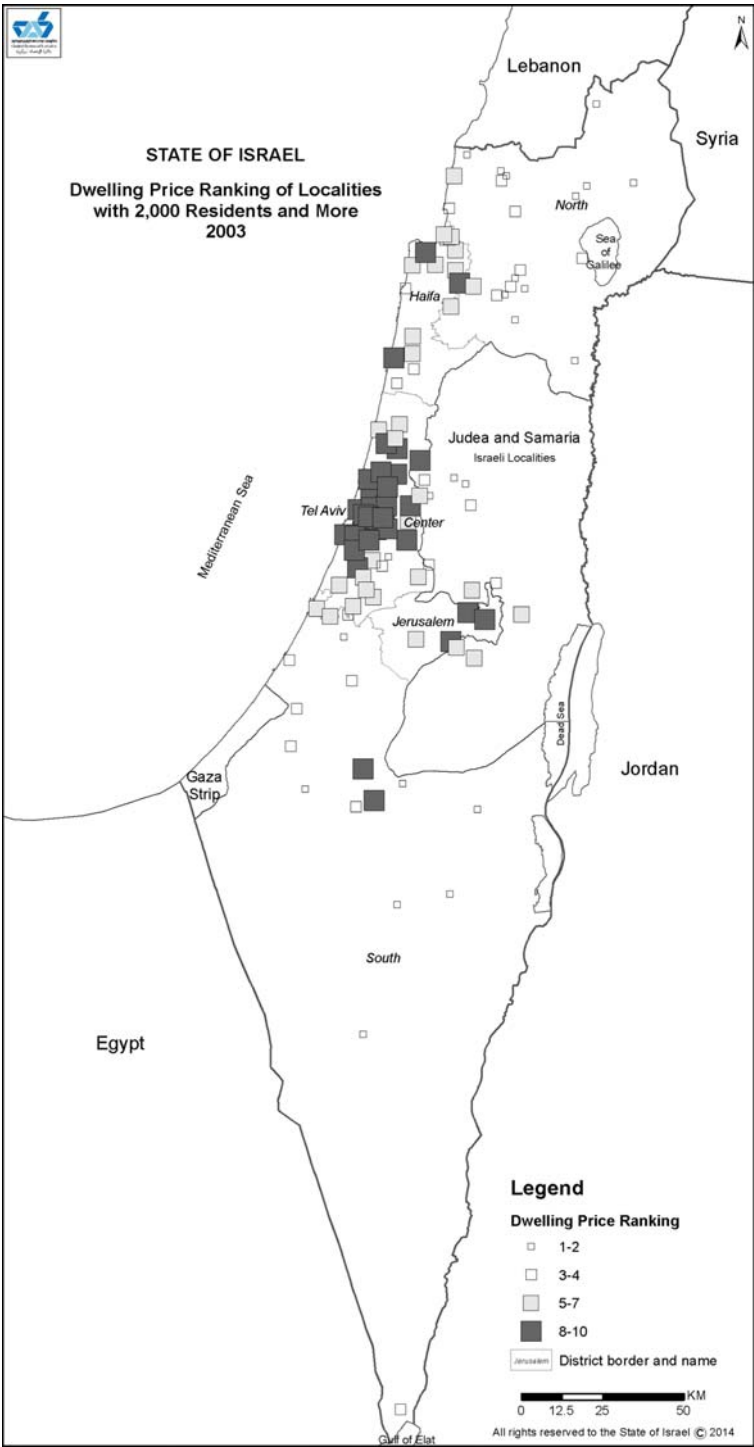
On the basis of these databases, a set of explanatory variables were selected based on the existing literature in this field partly reviewed in Subsection 2.1. Appendix 1 presents and defines the variables that we used in the study, their means, standard deviations and medians.

In order to examine the degree of correlation between the DPL and the selected variables, a correlation analysis was carried out. The Spearman correlation coefficients are presented in [Table 2](#). Note that the Spearman coefficient is used since the distribution of most explanatory variables is skewed.

Of all the variables having a significant correlation with the DPL, there are six variables characterized by a positive correlation with the DPR: the total population in a locality, the rate of cases of property crimes, the number of terror incidents and the three variables indicating the geographic district of a locality – Jerusalem District, Center District and Tel Aviv District.

The degree of correlation between the SEC and the selected variables was also examined. The results of this test are presented in Appendix 2. The correlation analysis on both the DPL ([Table 2](#)) and the SEC variables (Appendix 2) revealed similar results.

Map 2. Dwelling Price Ranking of Localities



Map 3. Socioeconomic Cluster of Localities

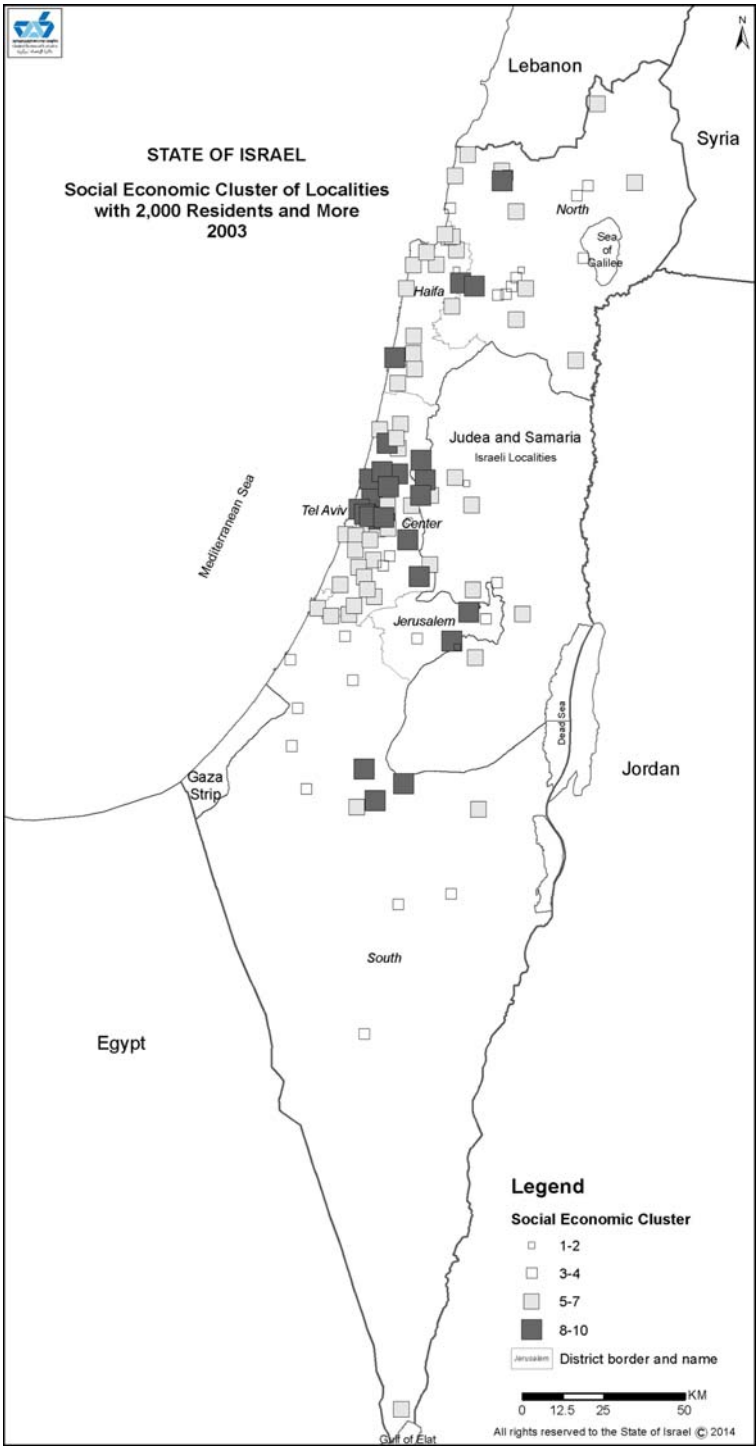


Table 2. Correlations between the DPL and the explanatory variables

Variables	2001		2003	
	Correlation coefficient	Level of significance	Correlation coefficient	Level of significance
Total population	0.169	0.086	0.264	0.005
Percentage of Arab population	-0.493	<0.001	-0.418	<0.001
Percentage of Orthodox population	-0.307	0.002	-0.340	<0.001
Percentage of ultra-Orthodox population	-0.307	0.002	-0.265	0.005
Percentage of immigrants from the former USSR since 1990	-0.423	<0.001	-0.379	<0.001
Percentage of Ethiopian immigrants	-0.205	0.040	-0.068	0.481
Rate of cases of bodily injury crimes	-0.453	<0.001	-0.459	<0.001
Rate of cases of property crimes	0.294	0.002	0.292	0.002
Number of terror incidents	0.204	0.038	0.132	0.166
Districts: Jerusalem	0.124	0.211	0.174	0.066
North	-0.446	<0.001	-0.473	<0.001
Haifa	-0.014	0.885	0.018	0.849
Center	0.346	<0.001	0.410	<0.001
Tel Aviv	0.449	<0.001	0.440	<0.001
South	-0.353	0.001	-0.315	<0.001
Distance from the Tel Aviv district	-0.695	<0.001	-0.721	<0.001

5. Parametric Models and Findings

Two parametric models were estimated: multinomial logistic regression for the SEC variable and the OLS model for the DPL. The models were estimated only for the Jewish sector for the following reasons. The housing market in the Arab sector operates under different conditions than that in the Jewish sector, and some of the explanatory variables are irrelevant to the Arab sector (such as the percentage of new immigrants from the former USSR and Ethiopia). Furthermore, the number of localities in the Arab sector with sufficient number of transactions was inadequate for performing statistical analyses for the Arab sector solely (four localities).

In order to avoid possible multicollinearity, those explanatory variables that were found to be highly correlated with other explanatory variables (Pearson correlation coefficient is more than 0.5) were excluded from the parametric models. These variables were included in the nonparametric analysis presented in Subsection 5.3.

5.1. A Multinomial Logistic Model for the SEC Variable

To estimate the marginal contribution of each of the above factors to the SEC, a regression analysis was carried out. Since the SEC is categorical, a multinomial logistic regression model was estimated, with the dependent variable being the probability of being in cluster i :

$$P(SEC = i) = \log it(\alpha_i + \beta DPR + \gamma X) \quad (3)$$

Table 3. Multinomial models for the socioeconomic cluster

Variable	2001			2003		
	Estimate	p-value	Odds ratio	Estimate	p-value	Odds ratio
Intercept for ranking = 10	-6.16	0.000	-	-5.22	0.000	-
Intercept for ranking = 9	-4.05	0.004	-	-3.58	0.005	-
Intercept for ranking = 8	-1.38	0.310	-	-0.44	0.707	-
Intercept for ranking = 7	0.75	0.589	-	2.35	0.059	-
Intercept for ranking = 6	2.42	0.083	-	4.67	0.000	-
Intercept for ranking = 5	5.52	0.000	-	8.32	<0.001	-
Intercept for ranking = 4	9.31	<0.001	-	18.85	<0.001	-
Intercept for ranking = 3	11.22	<0.001	-	21.64	<0.001	-
Intercept for ranking = 2	12.79	<0.001	-	33.82	<0.001	-
DPR	0.61	<0.001	1.85	0.50	<0.001	1.65
Percentage of Orthodox population	-0.04	0.042	0.96	-0.03	0.051	0.97
Percentage of ultra-Orthodox population	-0.09	<0.001	0.91	-0.29	<0.001	0.75
Percentage of Ethiopian immigrants	-0.15	0.318	0.86	-0.59	0.000	0.55
Percentage of immigrants from the former USSR since 1990	-0.12	<0.001	0.89	-0.12	<0.001	0.89
Rate of cases of bodily injury crimes	-0.27	<0.001	0.76	-0.30	<0.001	0.74
Rate of cases of property crimes	-0.31	0.068	0.74	-0.0003	0.43	1.00
North district	1.46	0.015	4.33	0.25	0.66	1.29
South district	1.64	0.032	5.12	0.97	0.209	2.65
Number of terror events	-0.18	0.081	0.84	-0.02	0.839	0.98
Number of observations			98			107
Percent concordant			94.2			94.6

with $\alpha_i, i = 2, \dots, 10$ being the intercepts of the model for cluster values $2, \dots, 10$ respectively, where cluster 1 was chosen to be the reference category. In (3), γ denotes a vector of the regression coefficients to be estimated and X the set of explanatory variables.

Table 3 presents the final estimated models for 2001 and 2003, with variables that are significant for at least one of the years (significance level 0.10).

It can be seen that there is a positive correlation between the DPR and the probability of appearing in a higher SEC, given all the other controlled variables.

However, it was found that this influence is partially offset as a result of the effect of minorities, for example the percentage of religious population (both ultra-Orthodox and Orthodox), the percentage of immigrants from Ethiopia, and the percentage of immigrants from the former USSR.

Given all other controlled variables, including DPR, it appears that the location in peripheral districts increases the odds for being in a higher SEC.

The effects of both bodily injuries and property crimes as well as the number of terrorism incidents were found to be significant and negative.

5.2. OLS Model on DPL

A regression model was estimated for the continuous DPL variable which served for constructing the DPR.

In order to validate the obtained estimators, we carried out appropriate statistical tests to identify possible multicollinearity, residual dependence and residual normality. Figures 1A and 2A (Appendix 3) demonstrate that the DPL is close to normally distributed, justifying use of the OLS model. Figures 3A and Table 1A (Appendix 3) display the residuals normal probability plots for the estimated OLS models, showing that the residuals' distribution is approximately normal. Statistical test results (such as the Durbin-Watson test and 1st Order Autocorrelation test for residual independence,

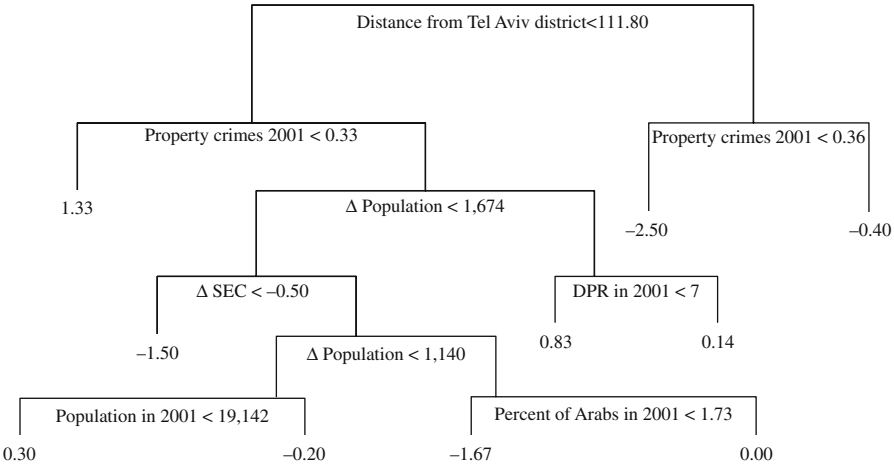


Fig. 1. Regression tree for change in the DPR between 2001 and 2003

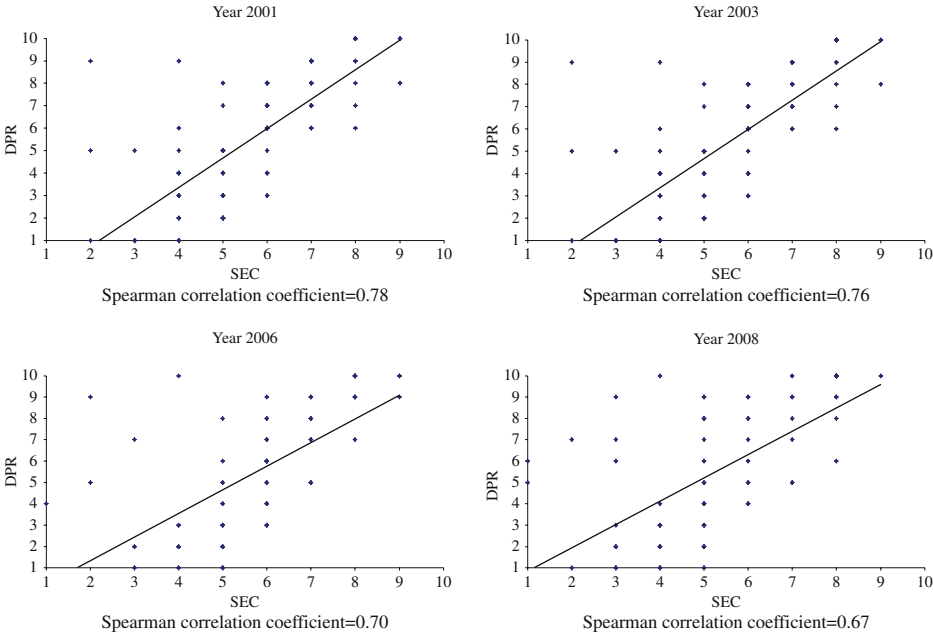


Fig. 2. Socioeconomic cluster vs. dwelling price ranking (Sources: Tax Authority, Central Bureau of Statistics)

and the Shapiro-Wilk and Kolmogorov-Smirnov tests for residual normality) show that at five percent significance level we cannot indicate residual dependence or significant deviation from normality. In addition, results from the correlation analysis do not reveal any evidence of multicollinearity.

The model can be represented as:

$$DPL_k = \alpha + \beta SEC_k + \gamma X_k + \varepsilon \quad (4)$$

In (4), β denotes the regression coefficient for SEC variable in locality k , X_k is a set of explanatory variables for this locality, γ is a vector of coefficients of X_k to be estimated, and ε denotes model residuals with zero expected value and constant variance.

Table 4 shows that most of the effects found to be significant in Model (3) on the SEC variable are also significant in (4) estimated on the DPL and follow the same directions, such as overall effect of crime and the effect of minorities (percentage of Orthodox, percentage of the immigrants from Ethiopia and from the former USSR). The positive and significant estimate of the property crimes variable in 2003 might be explained by “special” correspondence between property crime and dwelling prices, where more expensive residential areas “invite” property crimes. In addition, the negative estimate of the distance from the Tel Aviv district reflects a strong peripherality effect in Israel. However, this influence is nonlinear. A positive sign of the squared term means that the peripherality effect weakens as distance from the center of national economic activity

Table 4. OLS models on “logarithm of the median dwelling price in a locality”

Variable	2001		2003	
	Estimate	p-value	Estimate	p-value
Intercept	8.819	< 0.001	8.647	< 0.001
SEC	0.062	< 0.001	0.067	< 0.001
Percentage of Orthodox population	− 0.005	0.003	− 0.001	0.443
Percentage of immigrants from the former USSR since 1990	− 0.006	0.001	− 0.006	0.010
Percentage of Ethiopian immigrants	− 0.048	0.000	0.036	0.017
Rate of cases of bodily injury crimes	0.0001	0.975	− 0.010	0.096
Rate of cases of property crimes	− 0.0006	0.432	0.002	0.052
Distance from Tel Aviv district	− 0.005	< 0.001	− 0.005	< 0.001
Distance from Tel Aviv district –squared function	0.00001	0.000	0.00001	0.002
Small locality*	− 0.141	0.003	− 0.208	< 0.001
Total population in a locality (tens of thousands)	0.005	0.029	0.006	0.011
Number of observations	98		107	
Adjusted R ²	0.77		0.73	

*Localities with 10,000 residents or less (dummy variable)

(Tel Aviv) increases. This occurs due to the existence of additional employment centers in various peripheral towns and the decreasing effect of distance from Tel Aviv in regions that are very far from it.

Estimated Equation (4) for locality k is given by: $DPL_k = \hat{\alpha} + \hat{\beta}SEC_k + \hat{\gamma}X_k$. It follows, that the SEC variable can be expressed as: $SEC_k \approx (DPL_k - \hat{\alpha} - \hat{\gamma}X_k)/\hat{\beta}$. It should be noted that the above approximation for SEC may not be an integer due to continuous characteristics of $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$.

Therefore, this approximation for SEC in a locality k is given by:

$$\overline{SEC}_k = DPR_{corrected} = Round \left[\frac{(DPL_k - \hat{\alpha} - \hat{\gamma}X_k)}{\hat{\beta}} \right] \tag{5}$$

Using (4) and (5), we can examine whether the gap between the SEC and the DPR defined earlier can be at least partly bridged. Table 5 presents the distribution of absolute values of differences between the SEC and the DPR, before and after the correction.

Table 5. Distribution of absolute differences between the SEC and the DPR

	Mean		Min		Max		Percentiles					
							10th		50th		90th	
	2001	2003	2001	2003	2001	2003	2001	2003	2001	2003	2001	2003
Original	1.673	1.776	0	0	7	7	0	0	1	2	3	3
Corrected	0.235	0.234	0	0	1	1	0	0	0	0	1	1

The results presented in [Table 5](#) allow us to conclude that the examined socioeconomic factors that are not currently included in the calculation of the SEC may contribute to a better approximation of the suggested indicator. Furthermore, the earlier obtained Spearman correlation coefficients are also improved to some extent, now being equal to 0.70 and 0.71 for 2001 and 2003, respectively.

5.3. Regression Tree Analysis

In order to draw relevant conclusions on factors influencing the dynamics of the DPR between 2001 and 2003, a nonparametric regression tree was built.

This method of analysis was chosen for the following reasons. First, nonparametric methodology allows for the inclusion of localities in the Arab sector, despite a very small number of available observations and significant differences between the housing markets in Jewish and Arab sectors as mentioned in Section 5. Second, the variables which were removed from the regression models due to multicollinearity can be included in the nonparametric analysis (such as percentage of the Arab population). Given these explanations, the nonparametric regression method is designed to complete and enrich the results obtained from the analysis presented in Subsections 4, 5.1 and 5.2. In this analysis, the dependent variable was defined as the difference between the DPR in 2003 and the DPR in 2001.

To the explanatory variables used in Model (4) we added the differences between the values of these variables in 2003 and 2001. The regression tree method divides observations into homogeneous groups of a dependent variable, given a set of explanatory variables. A detailed description of the regression tree methodology is presented in [Breiman et al. \(1984\)](#).

The algorithm is iterative and works as follows. Initially, from the explanatory variables and their values, the algorithm finds a variable and its values which divides all the observations into two distinct groups, so that the variance of the dependent variable within each group (“leaf”) is minimal and the variance between these two groups is maximal (among all possible combinations). This value is fixed as the “split point”. The same process is repeated until a specified stopping criterion is fulfilled. At each stage, analysis is performed on the full set of the input variables; therefore, the same explanatory variable can be used several times. This method reveals nonlinear relationships between the dependent and explanatory variables.

The R^2 index was used to test goodness of fit. Let SSW_k denote the estimated variance within the final group k . Since the groups are independent, the total variance within all the groups is calculated by: $SSW = \sum_k SSW_k$. Using the definition of the index R^2 , it is given by:

$$R^2 = 1 - \frac{SSW}{SST} \quad (6)$$

where SST denotes the total variance of the dependent variable. The higher the value of the R^2 , the better the classification achieved (in terms of the homogeneity of the final “leaves”) relative to a previous iteration. In our case, the value of the R^2 index is equal to 0.65.

In [Figure 1](#), the height of the lines between the split points shows the reduction in variance within the group as a result of the division described above, while the numerical value in the final group shows the average increase/decrease in the dependent variable for those leaves. The left branch of each bifurcation corresponds to the “yes” alternative, that is, the condition being fulfilled.

It appears that the dominant factors for change in the DPR are a locality’s geographical location and its crime level. For example, in the localities that are situated rather close to the Tel Aviv district (less than 111.8 km) and where the property crime rate was less than 0.33 in 2001, the dwelling price ranking rose by an average of 1.33 (the “leaf” furthest to the left). A decrease in the SEC in a locality caused a consequent decrease in its DPR, given changes in its population and property crime rates in 2001. It also appears that given other controlled variables, an increase in total population is correlated with an increase in the DPR. Additionally, the low percentage of the Arab population in 2001 is correlated with a decrease in the DPR.

6. Conclusions

The current study examined the question of whether dwelling prices in a given locality can serve as an approximation to its socioeconomic level.

The study is based on a number of administrative databases available at a national level. It was found that during the research period (2001 and 2003) there was a strong association between the locality’s socioeconomic cluster and the value of its dwellings, with the Spearman correlation coefficients almost identical for these two years.

An analysis of recently obtained SEC data for the years 2006 and 2008 shows that the results obtained for 2001 and 2003 remained consistent, as did the results for 2006 and 2008 ([Figure 2](#)). However, a gap has been found between the SEC and the DPR. Our results show that this gap may be explained by location, other social and demographic factors, crime and security characteristics that are exogenous to SEC. In particular, a significant correlation was found between dwelling prices in a specific locality and the percentage of those belonging to defined population groups. It was also found that the size of a locality has a positive correlation with the level of the dwelling prices there. It appears that the effect of the distance from the center of Israel’s economic activity is negative, as expected.

It was found that these effects, which reflect other social and demographic characteristics that are not currently included in the SEC calculation, may bridge, at least partly, the gap between the SEC and the DPR.

Overall, we conclude that the ranking based on dwelling prices can serve as a rather good approximation to the socioeconomic level of most urban localities in Israel.

Obviously, this approximation may not always be accurate for some of the localities. Nevertheless, the proposed methodology can provide the required information on socioeconomic profile. This finding is extremely important since the process of SEI and SEC calculation requires a wide variety of variables from various data sources available concurrently for various localities, while the dwelling price ranking allows a rather simple approximation of SEC for different localities for every given year.

For the sake of the methodological consistency and comparability of the SEC series, at the current stage we do not suggest any changes in the set of variables used for the SEC calculations. Rather, we propose the DPR index as an approximation to SEC values for those localities and for years when SEC variables are not available (“intermittent points”). In such cases, using the DPR index, particularly after corrections are made according to Equation (5), can serve as an important working tool for the users of the SEC, such as the Ministry of Finance, Ministry of the Interior, planning authorities and others ensuring a continuum of the index series.

The proposed methodology and the obtained findings are likely to be valid and applicable for different statistical purposes in other countries which possess administrative data on dwelling transactions. For countries that compute socioeconomic indices, the proposed methodology may be used for assessing SEC values for time points and localities for which this index is missing. For countries that do not compute such indices, dwelling price ranking may be used to characterize the socioeconomic profile of a given locality. The suggested approximation may also be used for studying trends in SEC compared with DPR over the years.

Further development and application of an adjustment methodology for SEC imputation is a subject of future research.

Appendix 1. Descriptive Statistics

Name of variable	2001			2003		
	Mean	Std Dev	Median	Mean	Std Dev	Median
Log (DPL)	8.72	0.36	8.64	8.70	0.38	8.68
Total population in a locality	49,287	85,706	21,441	47,299	86,918	21,074
Percentage of Arab population (from the total population in a locality)	11.10	20.56	4.58	10.68	19.98	4.22
Percentage of Orthodox population (from the Jewish population in a locality)	13.56	11.26	11.41	14.90	15.16	11.79
Percentage of ultra-Orthodox population (from the Jewish population in a locality)	13.66	32.20	4.78	12.43	23.09	4.38
Percentage of immigrants from the former USSR since 1990 (from the Jewish population in a locality)	11.73	11.41	7.25	10.86	10.99	6.55
Percentage of Ethiopian immigrants (from the Jewish population in a locality)	0.87	1.47	0.18	0.81	1.43	0.14
Rate of cases of bodily injury crimes (per 1,000 residents)	0.9	0.64	0.78	0.82	0.56	0.69
Rate of cases of property crimes (per 1,000 residents)	40.36	27.40	35.59	38.14	25.81	37.76
Number of terror incidents in a locality	0.43	2.16	0	0.70	3.00	0
Districts: Jerusalem	0.03	0.17	0	0.04	0.18	0
North	0.22	0.42	0	0.20	0.40	0
Haifa	0.15	0.36	0	0.14	0.35	0
Center	0.27	0.45	0	0.28	0.45	0
Tel Aviv	0.11	0.31	0	0.10	0.30	0
South	0.14	0.35	0	0.14	0.35	0
(Dummy variables: 1 if defined district, 0 – otherwise)						
Distance from the Tel Aviv district (km)	55.31	50.76	46.24	54.60	51.12	44.31

Appendix 2. Correlations Between the SEC and the Explanatory Variables

Variables	2001		2003	
	Correlation coefficient	Level of significance	Correlation coefficient	Level of significance
Total population	−0.038	0.702	−0.037	0.702
Percentage of Arab population	−0.445	<0.001	−0.351	0.000
Percentage of Orthodox population	−0.307	0.002	−0.202	0.035
Percentage of ultra-Orthodox population	−0.363	0.000	−0.635	<0.001
Percentage of immigrants from the former USSR since 1990	−0.431	<0.001	−0.359	0.000
Percentage of Ethiopian immigrants	−0.223	0.026	−0.229	0.016
Rate of cases of bodily injury crimes	−0.487	<0.001	−0.434	<0.001
Rate of cases of property crimes	−0.081	0.413	0.055	0.562
Number of terror incidents in a locality	−0.062	0.536	−0.059	0.538
District: Jerusalem	−0.041	0.684	0.014	0.881
North	−0.263	0.007	−0.256	0.006
Haifa	0.071	0.474	0.044	0.647
Center	0.217	0.027	0.233	0.013
Tel Aviv	0.195	0.049	0.181	0.056
South	−0.138	0.163	−0.143	0.134
Distance from the Tel Aviv District	−0.301	0.002	−0.307	0.001

Appendix 3. Validation of OLS Assumptions

3.1. Distribution of the Dependent Variable

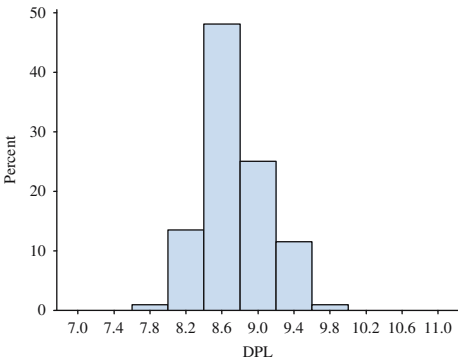


Fig. 1A. Distribution of the DPL 2001

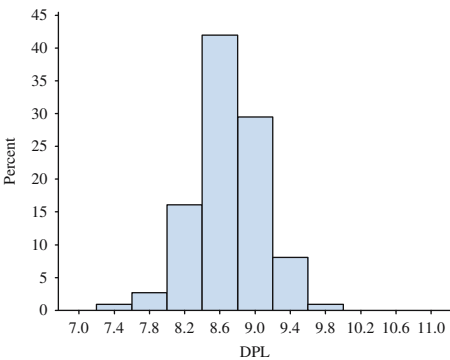


Fig. 2A. Distribution of the DPL 2003

3.2. Residual Normality

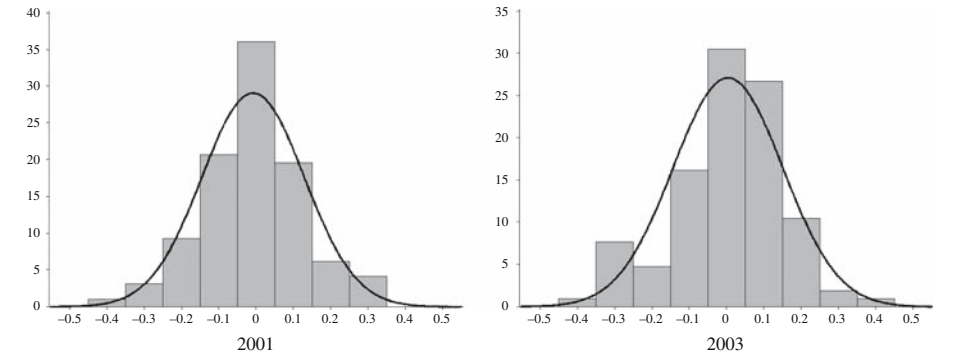


Fig. 3A. Histograms of the model residuals and normal density curves, 2001 and 2003

Table 1A. Test for normality of residuals

Year	Test (p-value)	
	Kolmogorov-Smirnov	Shapiro-Wilk
2001	0.66	0.07
2003	0.15	0.10

7. References

Australian Bureau of Statistics. 2006. “An Introduction to Socio-Economic Indexes for Areas (SEIFA).” Information Paper No. 2039.0. Available at: <http://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/2039.0Main%20Features22006?opendocument&tabname=Summary&prodno=2039.0&issue=2006&num=&view> (accessed April 2015).

Blanchard, T.C. 2007. “Conservative Protestant Congregation and Racial Residential Segregation: Evaluating the Closed Community Thesis in Metropolitan and Nonmetropolitan Counties.” *American Sociological Review* 72: 416–433. Doi: <http://dx.doi.org/10.1177/000312240707200305>.

Bourassa, S. C., M. Hoesli, and V.S. Peng. 2003. “Do Housing Submarkets Really Matter.” *Journal of Housing Economics* 12: 12–28. Available at: <http://www.sciencedirect.com/science/article/pii/S1051137703000032> (accessed April 2015).

Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone. 1984. *Classification and Regression Trees*. Wadsworth: Belmont.

Buck, A.J., J. Deutsch, J. Hakim, U. Spiegel, and J. Weinblatt. 1991. “A Von Thunen Model of Crime, Casinos and Property Values in New Jersey.” *Urban Studies* 28: 673–686. Doi: <http://dx.doi.org/10.1080/00420989120080861>.

- Burck, L. and Y. Kababia. 1996. *Characterization and Ranking of Local Authorities according to the Socio-Economic Level of the Population in 1995*. Publication No. 1039, Central Bureau of Statistics, Jerusalem, (Hebrew).
- Burck, L. and Y. Kababia. 1999. *Characterization and Ranking of Local Authorities according to the Socio-Economic Level of the Population in 1999, Based on the 1995 Census of Population and Housing*. Publication No. 1118, Central Bureau of Statistics, Jerusalem. (Hebrew).
- Cahaner, L. 2012. "Expansion Processes of the Jewish ultra-Orthodox Population in Haifa." In *Themes in Israel Geography, Special Issue of Horizons in Geography*, edited by J.O. Maos and I. Charney, 70–87, University of Haifa.
- Central Bureau of Statistics (CBS). 2000. *Characterization and Classification of Geographical Units by the Socio-Economic Level of the Population. 1995 Census of Population and Housing Publications*, No. 13, Jerusalem.
- Cummings, J. L., D. DiPasquale, and M. E. Kahn. 2002. "Measuring the Consequences of Promoting Inner City Homeownership." *Journal of Housing Economics* 11: 330–359. Available at: <http://www.cityresearch.com/pubs/dipasquale.pdf> (accessed April 2015).
- Des Rosiers, F., M. Theriault, Y. Kestens, and P. Villeneuve. 2002. "Landscaping and House Values: An Empirical Investigation." *Journal of Real Estate Research* 23: 139–161.
- Dubin, R.A. and A.C. Goodman. 1982. "Valuation of Education and Crime Neighborhood Characteristics Through Hedonic Housing Price." *Population and Environment* 5: 166–181. Doi: <http://dx.doi.org/10.1007/BF01257055>.
- Goodman, A.C. and T.G. Thibodeau. 1998. "Housing Market Segmentation." *Journal of Housing Economics* 7: 121–143. Doi: <http://dx.doi.org/10.1006/jhec.1998.0229>.
- Greenberg, M.R. 1999. "Improving Neighborhood Quality: A Hierarchy of Needs." *Housing Policy Debate* 20: 601–624. Doi: <http://dx.doi.org/10.1080/10511482.1999.9521345>.
- Harris, D.R. 1999. "Property Values Drop When Blacks Move in, Because. . .: Racial and Socioeconomic Determinants of Neighborhood Desirability." *American Sociological Review* 64: 461–479.
- Hazam, S. and D. Felsenstein. 2007. "Terror, Fear and Behaviour in the Jerusalem Housing Market." *Urban Studies* 44: 2529–2546. Doi: <http://dx.doi.org/10.1080/00420980701558392>.
- Heikkila, E. 1992. "Describing Urban Structure: A Factor Analysis of Los Angeles." *Review of Urban and Regional Development Studies* 4: 84–101. Doi: <http://dx.doi.org/10.1111/j.1467-940X.1992.tb00035.x>.
- Horton, H.D. and M.E. Thomas. 1998. "Race, Class, and Family Structure: Differences in Housing Values for Black and White Homeowners." *Sociological Inquiry* 68: 114–136. Doi: <http://dx.doi.org/10.1111/j.1475-682X.1998.tb00456.x>.
- Jackson, E., V. Kupke, and P. Rossini. 2007. The Relationship between socio-economic indicators and residential property values in Darwin. Paper presented at the 13th Annual Pacific-Rim Real Estate Society Conference. Fremantle, Western Australia. Available at: http://scholar.google.com.au/citations?view_op=view_citation&hl=en&user=KbV4jccAAAJ&citation_for_view=KbV4jccAAAJ:3fE2CS-JIrl8C (accessed April 2015).

- Jolliffe, I.T. 2002. *Principal Component Analysis*, 2nd ed. Springer Series in Statistics. New York: Springer.
- Kiel, K.A. and J.E. Zabel. 1996. "House Price Differentials in U.S. Cities: Households and Neighborhood Racial Effects." *Journal of Housing Economics* 5: 143–165. Doi: <http://dx.doi.org/10.1006/jhec.1996.0008>.
- Krivo, L.J. 1995. "Immigrant Characteristics and Hispanic-Anglo Housing Inequality." *Demography* 32: 599–615. Doi: <http://dx.doi.org/10.2307/2061677>.
- Lewin-Epstein, N., Y. Elmelech, and M. Semyonov. 1997. "Ethnic Inequality in Home Ownership and the Value of Housing: The Case of Immigrants in Israel." *Social Forces* 75: 1439–1462. Doi: <http://dx.doi.org/10.1093/sf/75.4.1439>.
- Lozano-Gracia, N. and L. Anselin. 2012. "Is the Price Right? Assessing Estimates of Cadastral Values for Bogota, Colombia." *Regional Science Policy & Practice* 4: 495–508. Doi: <http://dx.doi.org/10.1111/j.1757-7802.2012.01062.x>.
- Malpezzi, S., G.H. Chun, and R.K. Green. 1998. "New Place-to Place Housing Price Indexes for U.S. Metropolitan Areas, and Their Determinants." *Real Estate Economics* 26: 235–274. Doi: <http://dx.doi.org/10.1111/1540-6229.00745>.
- McCluskey, W. J., W. G. Deddis, I. G. Lamont, and R. A. Borst. 2000. "The Application of Surface Generated Interpolation Models for the Prediction of Residential Property values." *Journal of Property Investment & Finance* 18: 162–176. Available at: <http://eprints.ulster.ac.uk/10588/> (accessed April 2015).
- Myers, D. 1990. *Housing Demography: Linking Demographics Structure and Housing Markets*. Madison: University of Wisconsin Press.
- Olausson, P.O. and D. Vagero. 1991. "Miscellanea, A Swedish Classification Into Social Classes Based on Census Information and Comparable to The British Classification—A Proposal." *Journal of Official Statistics* 7: 93–103.
- Ozanne, L. and T. Thibodeau. 1983. "Explaining Metropolitan Housing Price Differences." *Journal of Urban Economics* 13: 51–66. Doi: [http://dx.doi.org/10.1016/0094-1190\(83\)90045-1](http://dx.doi.org/10.1016/0094-1190(83)90045-1).
- Potepan, M.J. 1996. "Explaining Intermetropolitan Variation in Housing Prices, Rents and Land Prices." *Real Estate Economics* 24: 219–245. Doi: <http://dx.doi.org/10.1111/1540-6229.00688>.
- Reed, R. 2001. "The Significance of Social Influences and Established Housing Values." *Appraisal Journal*, October 1.
- Rose, D. and D. Pevalin. 2001. *The National Statistics Socio-Economic Classification: Unifying Official and Sociological Approaches to the Conceptualization and Measurement of Social Classes*. Institute of Social and Economic Research Working Papers, November 2001–4. Available at: https://www.iser.essex.ac.uk/files/iser_working_papers/2001-04.pdf (accessed April 2015).
- Soffer, A. and E. Bystrov. 2006. *Tel Aviv State: A Threat to Israel*. Haifa, Reuven Chaikin Chair in Geostrategy, University of Haifa. Available at: http://geo.haifa.ac.il/~ch-strategy/publications/english/Tel_Aviv_State.pdf (accessed April 2015).
- Thaler, R. 1978. "A Note on the Value of Crime Control: Evidence from the Property Market." *Journal of Urban Economics* 5: 137–145. Doi: [http://dx.doi.org/10.1016/0094-1190\(78\)90042-6](http://dx.doi.org/10.1016/0094-1190(78)90042-6).

Yates, J. 2002. *A Spatial Analysis of Trends in Housing Markets and Changing Patterns of Household Structure and Income*. Positioning Paper 30 of the Australian Housing and Urban Research Institute. Sydney Research Centre. Available at: http://www.ahuri.edu.au/publications/download/ahuri_60064_fr (accessed April 2015).

Received July 2013

Revised August 2014

Accepted October 2015