Mass Appraisal at the Census Level: Israeli Case

Larisa Fleishman\(^1\) (corresponding author) Yury Gubman\(^2\)

\(^1\)Israeli Central Bureau of Statistics  
\textit{E-mail: larisaf@cbs.gov.il}  
Tel. 972-26592843  
Fax: 972-26593203

\(^2\)Israeli Central Bureau of Statistics  
\textit{E-mail: yuryg@cbs.gov.il}

Central Bureau of Statistics, 66 Kanfe Nesharim St., P.O.B. 34525, Jerusalem 91342, Israel

Abstract

Mass appraisal valuation is increasingly based on multiple regression analysis in order to fit valuation statistical model for calculating market value of all properties in a target population. In this paper we develop and assess the predictive performance of different models for the estimation of dwelling values for all types of dwellings at a nationwide level in Israel. In order to investigate the representativeness of transaction data, the regression coefficients derived from the hedonic models estimated on transaction prices and on self-reported dwelling valuations, are compared. The results of this analysis are used to build several prediction models. Regression coefficients’ stability is tested by a quantile regression method that allows properties to be divided into sufficiently homogeneous estimation cells parsed by geographic and economic criteria. Our prediction model allowed obtaining acceptably precise dwelling value estimators at a census tract level. The proposed method applied to additional time points, gains stable estimators with rather small fluctuations in accuracy indices and standard deviations. Dwelling values estimated at the nationwide level enable to produce new statistical products at high geographic resolutions on a range of topics, e.g., the behavior of the housing market, the economic profile of residential areas, well-being and inequality, etc.

Keywords: dwelling value, accuracy indices, prediction model
1. Introduction

The acceptable practice for mass appraisal of residential property is using multiple regression analysis in order to fit valuation statistical model for calculating market value of all properties in a target population [29]. Compared to full professional appraisals, such valuation modeling procedure might produce less accurate results, but is also less costly while covering a large number of properties [7,31]. As in Europe mass appraisal services are mainly proprietary businesses [55] very few papers address the methodological issues of mass appraisal modeling and the performance of statistical models [3,11,24]

This study aimed at developing an effective and efficient prediction modeling approach to mass appraisal which accounts for economic, spatial and timing perspectives, and meets quality requirements of official statistics. Following the basic principal for official statistics, we address valuation accuracy and robustness of estimates as major criteria in all statistical procedures and tests. The proposed methodology was developed according to established principles in order to produce model-based dwelling value estimates that are rigorous and can be approved for publication by the National Statistical Bureau; this makes the current study unique in the field of register-based statistics.

Our paper augments the research literature on this topic in several ways. First, we develop original valuation methodology for the estimation of dwelling values for all types of dwellings at a nationwide level. We consider and treat the issue of representativeness of the existed data sources on dwelling values by applying clustering procedure based on quantile regression analysis along with geographical stratification. Spatial aspects were taken into account in all steps of analysis: hedonic model estimation, stratification, model selection and estimation of dwelling values at the census level. Second, in the prediction model, out-of-sample performance of value estimators is used as a main criterion for model selection. This approach allows significant improvement in estimators’ accuracy and stability, while reducing variance. Using regression specification with the log price, we apply the inverse transformation to examine a predictive performance of property values. All these are consistent with the practical purpose of estimated property values. Third, in addition to acceptable and commonly used accuracy measures (mean percentage error and median percentage error), in the validation step we assess statistical properties of the obtained predictors by calculating their standard deviation. This issue is rarely addressed in relevant literature. However, ignoring the fact that mass appraisal utilizes methods with statistical uncertainty might result in selection of a model which is statistically unstable.
The chosen prediction model specification provides rather high accuracy level of predicted values, promising the effective usage of a new register-based dataset on dwellings value for different purposes, including official statistics. The robustness of our prediction model along with rather low standard deviation of the obtained value predictors allow the application of the proposed methodology to different national dwelling samples at different time points.

The rest of this paper is organized as follows: Section 2 presents theoretical background and a survey of literature on mass appraisal valuation and some related methodological issues. Section 3 describes our sources of information and the construction of the database used in the study, defines the variables used in the analysis, and presents descriptive statistics of the variables. Section 4 presents the statistical methodology that we use for the empirical analysis as well as issues related to the prediction accuracy tests. Section 5 presents the main findings resulted from the hedonic models, and displays the performance of prediction models in terms of accuracy of the estimated dwelling values. Section 6 discusses the developed methodology and the obtained results. Section 7 concludes this paper.

2. Literature review

2.1 Mass appraisal valuation: applied issues

Mass appraisal may be defined as a systematic appraisal of groups of properties using standardized procedures [31]. Mass appraisal may be based on data from a periodic housing census performed as part of a population census (U.S., UK, Canada, Spain, etc.). In countries such as Denmark, Sweden, and Finland, in contrast, dwelling valuation may be produced on the basis of administrative sources that are updated at predetermined frequency, such as Denmark’s dwelling value register. In Israel, the population census includes a housing census but does not provide, in its current form, information about housing stock at the national level. Recently, however, an annually updated Dwelling and Building Register (DBR) has been established; it furnishes the information needed for estimation of the value of all dwellings included in it. Like most countries that have dwelling and building registers, Israel’s register, constructed by the Central Bureau of Statistics (CBS), is based on data that municipal authorities gather for tax purposes.

Estimations of dwelling value by census coverage have been produced in different countries for purposes of both official statistics and research. In official statistics, nationally predicted dwelling values are currently used mainly for the development of tax models and the creation of property level value data for municipal taxation (ad valorem tax purposes), [1,50,56]. In most OECD countries the estimated value of a property is the basis for the calculation of a property tax
From this perspective, one of the most essential issues in the design of a mass valuation system is how values are communicated to taxpayers and stakeholders. Complex predicting models are difficult to explain. One approach is to strive for simple, logical models that can be presented in a “base home” format, which allows taxpayers to understand how the features of their properties affect values [37]. Another, more complicated approach is to convert multivariate models into a series of tables that display prices per unit of area for different classes of properties. For example, in Netherlands, models are made public, and taxpayers can request a valuation report that includes valuation data for several comparable properties [2].

Apart from the property based ‘ad valorem’ taxation systems, there are additional important applications of low cost appraisal of any residential property in official statistics, first of all as macroeconomic indicators or as input into other indicators. These indicators provide national governments and banks with the statistics they need for monitoring the owner-occupied sector. This data serves essential information for banking mortgage policy and regulations, including loan advances, mortgage equity withdrawals and remortgaging, mortgage underwriting, current loan-to-value ratios at both national and local levels, required for bank risk monitoring, particularly in light of the last economic and financial crisis. Specifically, the growth in use of valuation modeling for processing loan valuations has been established. For example, in 2005 the Danish Financial Supervisory Authority authorized a major mortgage bank to use valuation modeling in loan origination [16]. Main product of mass appraisal valuations is also used in real estate investment industry for long-term mean investment return assessment, investment performance measurement, etc. [21].

Assessment values data may also augment a repeat sales dataset used for house price indices construction, as approximations for past or current values of houses that have not been resold during the sample period [40]. De Vries et al. [14] investigated the reliability of the official Dutch appraisal values and found that the quality of this data is sufficient enough for computing a house price index.

In addition, mass appraisal data can be used in inequality analysis. It is common practice to estimate inequality indices based on income distribution, while ignoring the fact that some persons are owners of dwellings and others are not. Households which are traditionally defined as income poor may, in fact, own some assets and thus be in a financially better position. Therefore, using totally register-based data on income and dwelling value in official household wealth statistics, can gain a more comprehensive picture of household economic wealth [19]. From this perspective, information about the value of the entire national housing stock may serve as a basis for the development of such statistical outputs, as small area estimation using dwelling data at a census level. In particular,
the construction of precise welfare estimates for small areas can allow policy makers that make use of "poverty maps" to allocate funds and improve targeting of welfare programs [51].

Accessible information on market value of any property increases the transparency of residential property markets at any spatial level [46]. That is, property owners could anchor their expectations regarding their property’s value, thus may contribute to households’ economic decision making and behavior.

In addition, data on value of the entire housing inventory may serve as a basis for sample frames for statistical surveys [48,54] and, in the years to come, be a unique data source for longitudinal analyses, for instance in respect to studying trends in household savings over the years [19,27,55].

In different countries the existing legal acts and ordinance determine the conditions needed for performing the mass appraisal in accordance with deliberate policy choices and to practical considerations. In Canada, Denmark, Netherlands, United Kingdom, and United States, the laws merely establish standards, and then there is a considerable discretion regarding methods and the valuation models employed. In other countries (Brazil, Germany), the law governing valuation often requires that valuation models be formally adopted by the government and published in a regulation containing the necessary rates and, coefficients [2].

2.2 Valuation modeling: methodological issues

According to the IAAO definition [29], mass appraisal of properties is carried out by using uniform estimation methods, predefined set of variables and specific validation procedures for a given point in time. In most cases, it is the practice to predict dwelling value for all residential properties using hedonic price model. In such a model, the marginal contribution of the characteristics of the property and its location to property value is estimated [44]. Several groups of factors affect the value of a dwelling: (1) physical characteristics, such as area, number of floors in a building, its age, etc. [18,58]; (2) characteristics of the residential environment, e.g., air quality, proximity to the sea or to open areas, exposure to road or airport noise, etc. [18,38]; and (3) location characteristics, e.g., demographic, social, and economic characteristics of the neighborhood’s population, location of dwellings relative to city center, and so on [4,32].

In most cases, information about dwelling value is harvested from files of transactions actually consummated in the specified period [7,19, 24]. If transaction data are unavailable, several alternative sources of information are used to estimate the imputation model, e.g., estimates from assessor surveys [55] or asking prices in media advertisements [35]. In these cases, the samples are rather small and previous years’ data must be included to attain a reasonable sample size. Thus, in
Wersing [55], the imputation model was estimated on the basis of more than 9,000 observations that had to be collected from twenty-seven years of assessment estimation data (in 2000 prices). Yet, despite the conversion into constant prices, final estimates may be biased due to macroeconomic changes over such a lengthy period, including crises and bubbles in the real estate market. The use of asking prices may also skew the results due to the obvious discrepancy between asking price and transaction price [5].

In most cases, the explained variable in the valuation model is logged dwelling price [19] or logged price per square meter [35]. Benjamin et al. [8] discuss direct prediction of the dwelling price variable without log transformation, but do not accompany the discussion with a comprehensive empirical analysis. In different models, the set of explanatory variables varies corresponding with their availability, completeness, and marginal contribution to estimation accuracy. The literature proposes that a distinction be made between models for the explanation of the phenomenon and models for its prediction [47]. In the latter type of model, the main criterion for the insertion of a variable is its forecasting ability as opposed to explanatory ability or the statistical significance of the regression coefficient. In estimating dwelling value, some scholars propose that a classic hedonic model including a large number of property characteristics that may affect its value be used [24,55]. Other scholars contest this, noting that additional information that would improve the accuracy of the final estimates is not provided for some of these variables [8,19, 39].

In several studies, a division into estimation cells [35] and the estimation of a separate model for each cell is proposed. The division into cells enhances the accuracy of imputed values by allowing the heterogeneity of a country’s housing markets to be taken into account. The proposed variables which could be used for the division into cells include geographical regions, types of property (private house vs. multi-storey building), regional economic situation, etc.

To examine estimation quality, the studies noted above use several indices: Mean Absolute Percentage Error (MAPE), Median Absolute Percentage Error (MedAPE), Mean Square Error (MSE), and Forecast Error (FE), to name only the most common. These indices are based on the principle of the measurement of the difference between a predicted value and an actual value. The MAPE index is the most common; the smaller its value is, the more accurate the prediction. The values reported in the aforementioned studies range from 10 percent [39] to 35 percent [35]. It is standard practice to calculate these indices on the basis of out-of-sample observations, i.e., those not used in the estimation of the model. Furthermore, Goodman and Thibodeau [24] stress the need to estimate the variance of each value estimator as an additional indicator of results reliability.
In building the transaction price model, attention should be paid to the issue of cycles in the real estate market and the need to update estimates within a time window that will reflect market conditions in the estimation year. For example, if a model is estimated for “boom” years in the market, its estimates will be irrelevant in “bust” years. (In particular, a model estimated for a “bubble” period in the housing market will be invalid once the bubble bursts.) Hence the importance of updating and even redefining the model at predetermined intervals (e.g., annually, as is done in Denmark).

2.3 Sale transactions vs. subjective valuation

The advantages of using actual sales prices for mass appraisal purposes are that sales prices (from arm’s length, non-coerced property transactions) represent the most accurate and reliable indicator of the actual market value of any given property [29]. Yet, one of the issues discussed in the professional literature is the extent to which dwellings sold in the free market in a given period of time are representative of the housing stock at large. It has been argued that the use of sale transaction prices may lead intrinsically to a bias because these data reflect only the prices of dwellings sold and not the price level of the housing stock at large [49]. This issue, known in the literature as sample selection bias, may also skew a dwelling valuation produced by a statistical/econometric model based on transaction data [39].

The additional source of dwelling value information is self-reported owners’ dwelling valuations reported in the surveys. As for this data, researchers claim that no such bias exists since such data reflect the price level more correctly by yielding a representative random population of properties sampled in surveys [57]. Thus, allowing for subjectively driven behavioral and perceptual components may improve the conceptual soundness of the value model [13].

However, subjectively estimated dwelling valuations that are reported in surveys are susceptible to several problems relating to cognition, selectivity of participation, and survey response rate. First, people who are asked to evaluate their dwellings vary considerably in their awareness of the state of the housing market, generally, and the price level in the region where the dwelling is located, particularly. Self-valuation of properties is usually based on information about transactions made in the homeowners’ residential surroundings [22, 43]. Second, the response rate in a survey depends on the respondents’ willingness to cooperate with the surveyor; absent this willingness might result in an interview nonresponse (to the entire survey) or item nonresponse [25]. Third, subjective dwelling valuations are based on different degrees of importance that homeowners assign to a range of property characteristics and surroundings, including physical-structural, locational, neighborhood quality, and environmental amenities and hazards [7]. Overall,
like the “asking price,” subjectively estimated dwelling valuations have an average upward bias [4,23,28,33].

Taking the above mentioned aspects into consideration, the main impediment to using subjective valuations for prediction of value data is caution over inaccuracy. When developing models, there is an assumption regarding the accuracy of information, on both attribute data (physical characteristics) and market data. The quality of the values produced is directly impacted by the quality of the data which are analyzed and used to produce the value estimates [52]. Therefore, survey data on property owners’ valuation might be relevant, if no good market data exist, or if market data are considered invalid for the appraisal task, or if for some special purpose it is extremely important to ascertain the behavioral nuances involved [31]. Yet, possible application of survey data on subjective valuation to mass appraisal is barely appropriate, and transactions information does remain the most credible source of market data.

3. Research population and databases

The research population comprises privately owned residential properties in CTs in which more than 50 percent of inhabitants are Jews, such regions situated in localities that have populations in excess of 2,000 (hereinafter: “the Jewish urban sector”). The 2011 DBR file yields 1,745,635 records that belong to the Jewish urban sector. This population was chosen for reasons including relatively high quality of data in both the DBR and the Israel Tax Authority (hereinafter ITA) files. As opposed to the Jewish urban sector, for Arab sector, very small number of transactions is available. Moreover, the physical, environmental, socio-economic and socio-cultural characteristics of the Israeli Arab housing market constitute a major factor in differentiating it from the Jewish housing market. Thus, the dwelling value estimation in the Arab sector should be based on different methodology fitted for the current stage of housing market in this sector. The development and implementation of such methodology is beyond the scope of this paper and is a topic for the future research.

In Israel, the main source of information about dwelling value and its characteristics is the record of real-estate transaction prices kept by the ITA. The main and most important use of the data in the ITA files by CBS is the calculation of the regularly published housing price index. The Household Expenditure Survey (hereinafter: HES), conducted by Israeli CBS annually, provides subjective estimate of property value by its owner and characteristics of the property. This information is used to generate estimates of the consumption of housing services and facilitate research on the housing economy. In most cases, these sources do not overlap: the HES data relate to dwellings not sold
during the survey year [43], whereas the ITA data cover only dwellings sold during the reporting year.

The 2011 ITA file and the 2011 HES file were used for comparative analysis of the distribution of transaction prices (ITA) as against property owners’ valuations (HES). The 2011 ITA file was used as a basis for the prediction of dwelling prices of properties registered in the DBR. In addition, the 2011 Population Register and the Tax Authority income data were also used. On the basis of these databases and in accordance with hedonic model theory, explanatory variables that are expected to affect property values were constructed.

The 2011 ITA file contained more than 75,000 property transactions that were carried out in the Israeli housing market. For the purposes of this study, the relevant transactions were chosen under the following criteria: (1) transactions conducted in the free market (excluding records labeled “Inheritance”, “Gift”, etc.) between private individuals or between a construction company (seller) and a private individual; (2) transactions taken place in urban localities in Jewish sector excluding regional council jurisdictions; and (3) transactions geocoded at a building level. Outlying and uncompleted observations were omitted from the work file. The final sample size comprised 39,244 transactions for the year 2011, with the average transaction price NIS 1,092,162 (New Israeli Shekels, local currency).

Of the 6,000 households that were sampled in the 2011 HES, records pertaining to respondents who lived in dwellings that they owned and belong to the target population, comprise 2,570 households that response to the specific item of subjective dwelling valuation with the average assessment NIS 1,499,539. The 2011 ITA and HES variables that were used for the analysis in this study are presented in Appendix 1.

The average price in the ITA data is perceptibly different from the corresponding owners’ valuation in the HES data. One explanation for the differences is that dwellings sold are smaller, on average, than those in the HES data in both area (82.4 square meters vs. 112.2, respectively) and number of rooms (3.5 as against 4.2, as it is shown in Appendix 1). Also, dwellings in ITA are generally situated in less “prestigious” CTs (according to the property tax rate and average income data for the region). In addition, there is an upward bias in respondents’ report on the value of their dwellings, as explained in Section 2.3 above, justifying attention to the measuring error in Model (1) (Section 4.1 below).

Notably, in the other indicators shown in Appendix 1, there is no meaningful difference between dwellings sold and those unsold.
4. Methodology

The process of producing the register-based dataset on dwellings values includes two main stages. Firstly, a hedonic model was used to investigate the representativeness of the transaction file, with the economic conditions and the annual number of sales transactions treated as given. In this matter, differences in the effect of the characteristics of a property on its value were examined for dwellings sold versus dwellings not sold. Secondly, based on the results of the foregoing analysis, models were constructed and estimated for the valuation of the target population of dwellings. Then, with the help of accuracy indices, the best prediction model was chosen.

4.1 Hedonic models

Hedonic models were estimated for data from both ITA and HES sources. Following the hedonic theory, these models estimated the effect of the characteristics of the property and its location on property value, and incorporated the same set of explanatory variables that was determined in accordance with the literature partly reviewed in Section 2 above. The use of the same variables made it possible to compare the regression coefficients and make inferences about the difference between the two data sources on dwelling value.

The hedonic model that was used for the comparison may be written as follows:

(1) \[ Y_{ijkl} = \log(P_{ijkl}) = \lambda_0 + \lambda_i \text{Asset}_i + \lambda_j \text{Building}_j + \lambda_k \text{CT}_k + \lambda_l \text{Locality}_l + u_{ijkl} \]

where \( P_{ijkl} \) denotes the value of property \( i \) in building \( j \) in region \( k \) and locality \( l \). \( \text{Asset}_i \) denotes the characteristics of dwelling \( i \) (area, number of rooms, property tax rate), \( \text{Building}_j \) - the characteristics of building \( j \) (year of construction, building type, number of dwellings in building), \( \text{CT}_k \) - locational variables of a census tract \( k \) and residents’ demographic and economic characteristics (number of buildings, residents’ median age, residents’ average annual income, proximity of CT to center of locality, proximity of CT to center of Tel Aviv, and whether CT fronts the sea; \( \text{Locality}_l \) - locality characteristics (population and district of locality); and \( u_{ijkl} \) - random noise with variance \( \sigma \). After log transformation, the explained variable is approximately normally-distributed, justifying the use of the ordinary least squares (OLS) method to estimate Equation (1). This transformation also stabilizes the variance of \( Y_{ijkl} \).

Following the literature on average upward bias in subjectively estimated dwelling valuations (Section 2.3), we assume the multiple measurement error in the survey data. Under this assumption, the log transformation yields \( \hat{Y}_{ijkl} = Y_{ijkl} + e_{ijkl} \), where \( \hat{Y}_{ijkl} \) denotes the (log) value reported in the
survey by homeowner. Assume a normal distribution of the measuring error: \( e_{ijkl} \sim N(\mu, \sigma^2) \) with expectation \( \mu \) and variance \( \sigma^2 \). Assuming that the parameters of the measurement error distribution are constant (in log), we get: 
\[
\hat{Y}_{ijkl} = Y_{ijkl} + \mu + \epsilon_{ijkl},
\]
where \( \epsilon \) denotes random noise with variance \( \sigma^2 \).

This brings us back to Equation (1) where, in the case of HES, the intercept is equal to \( \lambda_0 + \mu \) and the residual equals \( u_{ijkl} + \epsilon_{ijkl} \) with distribution \( N(0, \sigma^2 + \sigma^2) \).

After Equation (1) is estimated for the HES and the ITA data, the explanatory variable coefficients are compared. Insofar as no meaningful differences among the estimated regression coefficients in terms of sign and magnitude are found (Section 5.1 below), one may infer that, given the controlling variables, the real estate transaction data in the ITA file are at least approximately representative of the value of the research population’s housing stock.

4.2 Prediction models

Based on the results of the foregoing analysis, several models are estimated for the prediction of dwelling value for the research population, and the best model, in terms of accuracy and robustness (Section 4.3 below), is chosen. The model used to predict the value of the target population’s entire dwelling stock is based on Equation (1).

Different functional forms of the dependent variables were checked. The literature offers two approaches toward defining the dependent variable in the prediction model. First, there is an advantage in fitting a direct model to the target variable (in our case: (log) dwelling price) [35]. Conversely, the creation of a dependent variable called “(log) price per square meter” makes it possible to improve the numerical stability of the estimates, reduce their variance, and, as a result, enhance prediction accuracy [19]. Moreover, log transformed dwelling price/price per square meter is approximately normally distributed. In the present study, both approaches are tested. Empirical analyses show that the model with dependent variable defined as log price per square meter better fit to the given data in terms of R² criteria.

The preliminary analysis shows that the dwelling area variable explains most variance of the dependent variable (log price or log price per square meter), consistent with the findings of the literature review in Section 2.2. Following the literature, we use log transformation for the area variable because its effect on price is not linear.

Furthermore, the literature that deals with the analysis of dwelling value offers a lengthy discussion of the spatial dependence of property price (an autoregressive spatial effect). What this means is that the value of a specific dwelling affects, and is affected by, the value of dwellings in its vicinity [30]. This phenomenon creates a strong correlation between the value of a specific
dwelling and the average value of dwellings in its neighborhood. One cannot, however, take into account the structure of spatial dependence at the individual property level in DBR that contains more than a million records. Therefore, it was decided to include an aggregate variable, “(log) mean dwelling prices in CT” in the prediction model, as a proxy for dwelling prices in the neighborhood where the specific property is located. In several studies on housing prices, it is the practice to use the median price of dwellings in the region instead of the mean because the mean is sensitive to outlier values [26,30]. Since such values were excluded in the preliminary stage, it was decided to use the mean price. Notably, to avoid a multicollinearity problem in the hedonic model and possible bias in the estimated coefficients, this variable was not used in Equation (1) (for instance, a strong correlation was found between log mean housing prices in CT and property location characteristics, such as average income). Therefore, the inclusion of the log mean housing prices in CT variable in Equation (1) will necessitate the extraction from the model of several important variables needed to be examined.

The importance of including a variable that reflects the geographic location of the property in the prediction model is emphasized in several studies, such as McCluskey et al. [36] and Bourassa et al. [6]. In Israel, Tel Aviv is the center of national economic, financial, and cultural activity; therefore, the proximity of a CT to this important metropolis reflects the extent of the region’s peripherality [53].

Stepwise model selection algorithm was applied, with cut-off probability of 0.1, i.e. the final model includes the statistically significant (at 10 percent level) variables only. A number of competing model selection procedures were applied: (a) AIC maximization, (b) BIC maximization, (c) R-square maximization, (d) forward and (e) backward selection, (f) including "important" variables, unless they are insignificant in the particular model, (g) applying the same model for every estimation cells. It appears that methods (a)-(c) provide similar models with similar indices for the prediction accuracy; however, using (d), (e) and (g) may reduce the accuracy in some cases. Applying (f) leads to a significant increase in the standard deviations of the estimated dwelling values. Stepwise selection has been chosen as it provides more accurate and stable estimates with low standard deviations, compared to other methods being tested. Moreover, stepwise selection provides rather stable models over years; this aspect is of a special importance for official statistics.

Since the log dwelling value \( Y_{ijk} \) is approximately normally distributed, one may infer that the value of the dwellings is log-normally distributed. For each value estimate, the standard deviation was calculated. The formula used for the calculation of the predicted value and its variance is given by the following:
\[ \hat{P}_{ijk} = \exp( \hat{Y}_{ijk} + 0.5\sigma_{ijk}^2 ) \]

\[ \text{Var}( \hat{P}_{ijk} ) = (\exp(\hat{\sigma}_{ijk}^2) - 1)\exp(2\hat{Y}_{ijk} + \hat{\sigma}_{ijk}^2) \]

where \( \hat{\sigma}_{ijk}^2 \) denotes the variance of residuals estimated in the regression model.

### 4.3 Accuracy assessment and robustness tests

To test for the most accurate model, the MAPE index is used. It is given by:

\[ \text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{|\hat{P}_i - P_i|}{P_i} \right) \times 100 \]

where \( \hat{P}_i \) denotes estimated value of unit \( i \) and \( P_i \) denotes its actual value, according to relevant data source. Unit \( i \) may denote a dwelling (in cross-validation) or CT (comparison with external sources). In addition, the robust MedAPE index is calculated:

\[ \text{MedAPE} = \text{Median}\left( \left( \frac{|\hat{P}_i - P_i|}{P_i} \right) \times 100 \right) \]

To test accuracy, a cross-validation (CV) method with \( m \) iterations is used. In CV, random sampling is used to divide the original data set into two sub-sets: training sample on which the model is estimated (in the current study, 80 percent of the original data) and the learning sample on which prediction values and accuracy indices are calculated (the rest 20 percent of observations). The literature indicates that 10 iterations are sufficient (\( m=10 \)) [17]. The final estimate is defined as a mean value of those obtained in each individual iteration. Below, the ten-iteration CV method is abbreviated as CV10.

A number of statistical procedures were applied in order to check a robustness of the final estimates. Notably, the extreme outliers were excluded in the preliminary step. Using cross validation methodology, the final prediction model was tested. The following tests were carried out: (1) with different percentage of observations randomly assigned to a training sample (50%, 60%, 80%); (2) non-random training sample (60%, 80%) by the lowest/highest values of the key independent variable "dwelling area"; (3) non-random training sample (60%, 80%) by the lowest/highest values of the first principal component of the matrix of independent variables. Application (1)-(3) provide similar results in terms of the MAPE and MedAPE indices. It follows that the final model and the estimated dwelling values are robust to high/low values of independent variables, as well as to the size of training sample. Based on foregoing results, the proposed methodology is applicable for additional years and for different geographical regions.
In some studies, [e.g., 19], an internal validation method is used. The use of this method, however, may introduce a downward bias in the estimates of prediction accuracy (i.e., the estimates appear more accurate than they really are). This bias is the result of double use of the same information, once to estimate the model and once to check accuracy. In the current study, external (out-of-sample) validation method is used.

5. **Estimation of dwelling values**

5.1 *Comparing the results of the hedonic models: ITA data vs. HES data*

To perform the estimation, a stepwise selection algorithm was employed with only variables that are significant at least at the 10 percent level (Table 1). Table 1 shows that, with the exception of two variables, Year of Construction and Number of Buildings in CT, the coefficients of Models (A) and (B) are identical in direction and not much different in values. The $R^2$ index receives values of 0.73 and 0.74 for Models (A) and (B), respectively, validating the findings of the model.

The findings of Models (A) and (B) show that the marginal contribution of the dwelling size (area and number of rooms) to its value is very similar, whether the dwelling is sold or not.

As for the effect of the year of construction of a residential building on property value, both models show an upward trend in the regression coefficients over the years, although the coefficients for the HES data are lower and even negative until the mid-1980s. This finding is consistent with the descriptive statistics in Appendix 1.

The positive effect of the property tax rate variable reflects the correlation between type of dwelling and classification of region by the municipal authority for the collection of municipal taxes. Notably, in Israel municipal property taxes are not directly based on property value.

As expected, the geographic location of a dwelling has a significant effect on its value. Thus, peripherality affects property value at two different levels of spatial resolution. The proximity of center of locality to center of CT, in which a dwelling is located, reflects the effect of peripherality at the municipal level, while the proximity of CT center to center of Tel Aviv reflects the effect of peripherality at the national level. It may be seen that the peripherality effect is negative at both spatial levels investigated. The proximity to center of Tel Aviv variable was inserted as a quadratic function so that the non-linear effect of peripherality on dwelling value could be taken into account. Where the quadratic term has a positive sign, it means that the peripherality effect, generally speaking, weakens as distance from the center of national economic activity (Tel Aviv) grows.

[Table 1 about here]
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 that city. Also, locality size has a positive effect on housing prices, as expected.

The confidence interval analysis (at 95 percent confidence level) revealed that for the most important variables confidence interval of the coefficients of Model (B) contains the respective regression coefficients from Model (A): intercept, dwelling area, interaction between dwelling area and number of rooms, multi-storey building, property tax rate per square meter, CT fronting Mediterranean Sea, and proximity of CT to center of Tel Aviv (including a quadratic function of this variable). For other variables with the exception of Year of Construction, the difference is significant but relatively small (0.01).

It follows that both the ITA and the HES data sources may potentially serve as a basis for mass appraisal. However, the literature on the subject supports the premise that transaction price is the best proxy for property market value [4,7,21]. Thus, the model based on ITA data may be used for mass appraisal of the entire housing stock in a given area and at a certain time, economic conditions and annual number of sale transactions treated as givens. As well, we can conclude that underrepresentation of certain types of dwellings in the ITA data is unlikely to be crucial. Yet this issue is worth considering at a prediction stage.

5.2 Estimation of dwelling values

In accordance with the two approaches mentioned in Section 4.2 toward defining the dependent variable in the prediction model, two models were estimated on the whole sample. Table 2 presents the variables included in this estimation. The table shows that in terms of accuracy indices, Model (A2) estimated on the ‘price per square meter’ dependent variable, is preferable. Therefore, the rest of this study will relate to Model (A2) specification only.

[Table 2 about here]

Following the literature reviewed in Sections 2.2, it is common practice to apply a stratification procedure prior the estimation step. We analyze the possibility of transactions’ sample division into rather homogeneous cells in terms of housing prices. For this purpose, we first examined the distribution of mean price per square meter over geographical units as it is suggested in relevant research. Figure 1 presents the officially defined administrative division of Israel: Six districts and the Judea and Samaria Area (JS Area).

[Figure 1 about here]
Figure 2 presents the distribution of mean price per square meter over the districts.

Figure 2 shows that a distribution of mean price per square meter over the districts is far from being uniform. Differences between the districts are found to be statistically significant in a pairwise comparison between every two geographical units (using t-test). These findings support stratification by these geographical units.

In addition, a quantile regression analysis was performed to test the distribution of the estimated coefficients in Model (A2) over centiles of the dependent variable. In this method, the regression coefficients were calculated for specific percentiles of distribution of the dependent variable [34]. This yielded a vector of regression coefficients for each variable and each given percentile, making graphic analysis possible. Figure 3 shows the results of this analysis for regression coefficients of the main variables in the prediction model: log mean price per square meter in CT, log mean income in CT, log area, and distance from Tel Aviv.

One may see that the coefficients of the regression model are stable enough across most of the distribution: from the 10th centile to approximately the 95th. At the extremes, the coefficients are unstable and differ from the values obtained in the middle of the distribution range. This characteristic was also observed for additional variables in Model (A2), justifying separate estimation of the least expensive and the most expensive dwellings (bottom decile and five uppermost centiles, respectively). The same analysis conducted separately for each district reveals a picture similar to that presented in Figure 2, where the level of the coefficients varying from district to district.

The results of these analyses suggest the following stratification: (1) inexpensive dwellings (lowest decile, all districts) (2) expensive dwellings (five uppermost centiles, all districts), and (3) seven cells differentiated by district for dwellings remaining after the removal of those in (1) and (2). This division for 9 cells is consistent with the literature as it takes account of not only the spatial aspect, but also the differences in the contribution of the factors that affect the prices of most expensive and inexpensive dwellings as against the rest of the housing stock [41]. In such a way we also addressed and treated the issue of the representativeness of the ITA transaction data.

Regions that have expensive dwellings and those that have inexpensive ones were identified by means of the mean price per square meter in CT.
A prediction model was estimated independently for each cell. Table 3 shows the distribution of the absolute percentage errors calculated by the CV10 method. Table 3 demonstrates that the proposed stratification into nine estimation cells reduces the MAPE index to 20.12 percent, when in 75 percent of cases this index is no more than 22.44 percent. The comparison of accuracy indices presented in Table 2 and Table 3 shows that the division into estimation cells as described above improves accuracy of prediction by more than 7 percent on average.

[Table 3 about here]

The accuracy indices presented in Table 3 are superior to those reported in most of the studies reviewed in Section 2.2. For example, in a recent project that estimated dwelling value for all records in the residential properties register in Norway [19], the median prediction error was about 20 percent as against 12 percent in the current study.

Table 4 contrasts the distribution of predicted values for 2011 with the distributions of transaction prices and subjective valuation data for the same year.

[Table 4 about here]

Table 4 shows that the mean, median, and percentiles of the distribution of predicted values are greater than those of the distribution of transaction prices, indicating that the prediction method chosen corrects for the underrepresentation of expensive dwellings in the ITA files.

As stated in Section 4.2, standard deviations of the predicted values were estimated by (3). The obtained standard deviation was NIS 16,079 on average, about 1.3 percent of the mean value presented in Table 4, showing that the proposed predictors are sufficiently stable.

5.3 **Dwelling values and socio-economic level**

Property price level in a given residential area is often analyzed in context of its socio-economic profile [15,20,24,41]. This approach stems from the well-known correlation between property prices and various effects reflecting socio-economic characteristics of population in a given area. From this perspective, to analyze the distribution of the accuracy indices given a certain socio-economic level is of a particular interest. Such an analysis is assumed to serve an additional validation of the dwelling value estimators obtained by the proposed prediction method.

In order to characterize and document the socio-economic profile of various geographical units, it is common practice to calculate aggregated indices [5,9,10,42]. In Israel, socio-economic clustering was developed at the Central Bureau of Statistics (CBS) in the mid-1990s on the basis of the 1995 Population and Housing Census data, and recently updated on the base of 2008 Population Census data [12]. Variables used for clustering different geographical units reflect all of the aspects
related to the socio-economic makeup of the population of these units (such as demographic characteristics, education, unemployment rates, income, etc.). At the CT level, 20 socio-economic clusters were defined, with Cluster 1 including CTs with the lowest socio-economic level, and Cluster 20 including CTs with the highest socio-economic level. We used this clustering for the above mentioned analysis.

Figure 4 shows the distribution of MAPE and MedAPE indices over CT’s socio-economic clusters.

[Figure 4 about here]

Figure 3 demonstrates that there is certain variance in the accuracy indices over socio-economic clusters. In particular, the attained accuracy of the predicted values (MAPE) is greater for the socio-economically strongest and weakest regions than in the remaining regions, justifying the distribution into estimation cells parsed by the value of properties in the region as described above. MedAPE curve shows that the robust estimator of accuracy in all clusters is smaller than MAPE index and no meaningful variance in its distribution is observed among the clusters.

These findings yield the following conclusions. The proposed prediction method allowed obtaining acceptably precise dwelling value estimators not only at a national level (Table 3), but also for CTs belonging to specific socio-economic clusters. Moreover, the lowest MAPE index for both extremes of the socio-economic cluster range supports applied stratification, estimating separate prediction models for the most expensive and the cheapest CTs, bearing in mind mentioned above correlation between dwelling values and socio-economic level.

5.4 Applying the proposed methodology for 2012 and 2013

In order to examine the performance of the proposed methodology at different time points, we used data from 2012 and 2013. Notably, the Israeli property market in 2011-2013 is characterized by precipitous increase in housing prices. The same stratification procedure was conducted, and the same set of independent variables in prediction model was used. The DBR contain 1,748,254 and 1,865,012 records related to the relevant population, in 2012 and 2013, respectively. Table 4 shows the main results: means and standard deviations of the predicted values, as well as their accuracy indices. Two standard deviation indices are presented in table 5: standard deviation of the predicted values distribution and the square root of the variance in Formula (3), on average.

[Table 5 about here]

The main conclusion that can be drawn from Table 5 is that the proposed method applied to additional time points gains the accuracy indices stable during the addressed period. A slight
decrease in standard deviations $Std(\hat{P}_{ijkl})$ over the years may be explained, at least partly, by increased number of transactions in years 2012 and 2013 and larger rate of geocoded records.

6. Discussion

This study developed and applied a methodology for the prediction of dwelling values at the individual-record level in census coverage in Jewish urban localities.

Initially, hedonic models were estimated on the basis of sale transactions and subjective dwelling valuations. Analysis of the estimated models reveals certain similarity in regression coefficients for most variables. It follows that the model based on transaction data as reported to the Tax Authority may be used for mass appraisal, despite the possible lack of representativeness of certain market segments. The findings of the hedonic models are consistent with those reported in the extensive literature on this topic.

In prediction modeling process we faced several challenges. First, hedonic model fitted for explaining a specific phenomenon as property value, might not be the best option for its prediction. On one hand, in hedonic model (1) there are variables with rather low predictive ability, despite the statistical significance of their regression coefficients. On the other hand, some variables with strong predictive ability might not be of research interest from the viewpoint of phenomenon explanation. In this study, we define predictor’s stability and accuracy as main criteria for the prediction model selection. Consequently, only effects which are strongly correlated with the dependent variable and stable over the years have been included in the prediction model.

Second, as national transaction data for a given time period may not precisely represent the entire property stock, some adjustment procedure should be applied in order to overcome possible bias. To tackle this problem, we adopted stratification approach considering two different issues. First of all, in such a way we could address the existing heterogeneity of national housing market, suggesting different effects of the property characteristics in different housing segments. As a result, regression coefficients of the same variables are supposed to be different in models estimated separately for various segments. According to empirical evidence presented in Section 5.2, we divided our data bases using both economic and spatial criteria. Utilizing these criteria is consistent with literature, and significantly improved the accuracy of the predicted values. In addition, each estimation cell resulted from the stratification procedure is assumed to be rather homogeneous in terms of property value, thus allowing to treat possible non-representativeness bias.
Third, application of a stepwise selection algorithm with cut-off probability of 0.1 provides accurate and robust dwelling value estimates. Moreover, the developed prediction model provides stable and accurate estimates over the years. However, applying stepwise selection algorithm requires thorough check of the obtained model; otherwise it may adversely affect estimators' quality in terms of stability and robustness. Furthermore, using this method as a "black box" may lead to meaningless and/or biased results, as the relative importance of different independent variables may change over years and regions. For example, a model estimated for “boom” years in the market, becomes irrelevant for “bust” years. Therefore, information criteria as AIC, BIC, R² index should be addressed, and a robustness test should be performed.

Finally, for the official statistics purposes it is extremely important to perform additional analysis of the accuracy indices distribution. In particular, these indices are expected to be uniformly distributed over predefined stratum used in official statistics, like socio-economic level, administrative units, etc.

7. Conclusions

Dwelling values estimated at the nationwide level allow one to produce new statistical products at high geographic resolutions on a range of topics, e.g., the behavior of the housing market, the economic profile of residential areas, and welfare and inequality, to name only a few. Value data at the individual-record level also facilitates estimations and analysis of the distribution of dwelling values by income level, demographic characteristics and geographical units.

The practical application of the current study for official statistics, decision and policy makers in housing economics, investment, urban planning etc., requires predefined sufficiently high level of the predictors’ accuracy and stability over time. The accuracy attained in this study resembles, if not exceeds, that reported in the relevant literature. Moreover, it appears that the proposed method applied to additional time points, gains stable estimators with rather small fluctuations in their accuracy indices and standard deviations. Application of the prediction modeling selection and stratification procedures allow for more accurate estimations by considering wider segmental variation of the housing inventory in Israel. To assure robustness of the obtained predictors, appropriate outliers’ treatment was applied.

In line with recent trends in official statistics, application of Geographical Information System (GIS) tools allows visualization of the obtained estimates in various geographical resolutions. Both interactive and static maps provide user-friendly data dissemination and presentation for better understanding of spatial distribution of dwelling values.
Completing this study by including the Israeli Arab sector poses an important research challenge that may contribute to a more profound understanding of the distribution of physical residential properties in both sectors.
References


document&tabname=Summary&prodno=2039.0&issue=2006&num=&view


[38] H. Nijland and B. van Wee, Noise Valuation in Ex-ante Evaluations of Major Road and Railroad Projects, EJTIR, 3 (August 2008).


Table 1: Results of the models

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model A</th>
<th></th>
<th>Model B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2011 transaction data</td>
<td>2011 HES data (weighted)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Estimate (Model)</td>
<td>S.D. (Model)</td>
<td>Estimate (weighted)</td>
<td>S.D. (weighted)</td>
</tr>
<tr>
<td>Intercept</td>
<td>12.477*</td>
<td>0.018</td>
<td>12.485*</td>
<td>0.062</td>
</tr>
<tr>
<td>No. of rooms</td>
<td>0.218*</td>
<td>0.003</td>
<td>0.253*</td>
<td>0.013</td>
</tr>
<tr>
<td>Area (sq.m.)</td>
<td>0.006*</td>
<td>0.0001</td>
<td>0.006*</td>
<td>0.0005</td>
</tr>
<tr>
<td>Interaction: area * no. of rooms</td>
<td>-0.0006*</td>
<td>0.00001</td>
<td>-0.0007*</td>
<td>0.0001</td>
</tr>
<tr>
<td>Year of construction:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1955–1964</td>
<td>0.059*</td>
<td>0.008</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1965–1974</td>
<td>0.050*</td>
<td>0.007</td>
<td>-0.047***</td>
<td>0.025</td>
</tr>
<tr>
<td>1974–1984</td>
<td>0.071*</td>
<td>0.007</td>
<td>-0.071*</td>
<td>0.020</td>
</tr>
<tr>
<td>1985–1989</td>
<td>0.092*</td>
<td>0.012</td>
<td>-0.057*</td>
<td>0.017</td>
</tr>
<tr>
<td>1990–1999</td>
<td>0.161*</td>
<td>0.007</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>2000–2011</td>
<td>0.249*</td>
<td>0.008</td>
<td>0.061*</td>
<td>0.017</td>
</tr>
<tr>
<td>Multi-storey bldg. (Y/N)</td>
<td>-0.118*</td>
<td>0.006</td>
<td>-0.094*</td>
<td>0.018</td>
</tr>
<tr>
<td>No. of dwellings in building</td>
<td>-0.001*</td>
<td>0.0001</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Property tax rate per sq.m. in bldg. (NIS)</td>
<td>0.001*</td>
<td>0.0001</td>
<td>0.001*</td>
<td>0.0004</td>
</tr>
<tr>
<td>No. of bldgs. in CT</td>
<td>-0.0001*</td>
<td>0.00002</td>
<td>0.0002*</td>
<td>0.00004</td>
</tr>
<tr>
<td>CT fronting sea (Y/N)</td>
<td>0.095*</td>
<td>0.007</td>
<td>0.045***</td>
<td>0.025</td>
</tr>
<tr>
<td>Proximity of CT to center of locality (km.)</td>
<td>-0.035*</td>
<td>0.0001</td>
<td>-0.022*</td>
<td>0.004</td>
</tr>
<tr>
<td>Proximity of CT to center of Tel Aviv (hundreds of km.)</td>
<td>-0.835*</td>
<td>0.014</td>
<td>-0.889*</td>
<td>0.045</td>
</tr>
<tr>
<td>Proximity of CT to center of Tel Aviv (quadratic term)</td>
<td>0.225*</td>
<td>0.005</td>
<td>0.239*</td>
<td>0.015</td>
</tr>
<tr>
<td>Median age of CT’s residents</td>
<td>-0.007*</td>
<td>0.0003</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Mean labor income in CT (thousands of NIS)</td>
<td>0.007*</td>
<td>0.0001</td>
<td>0.004*</td>
<td>0.0002</td>
</tr>
<tr>
<td>No. of residents in locality (thousands)</td>
<td>0.099*</td>
<td>0.001</td>
<td>0.088**</td>
<td>0.004</td>
</tr>
<tr>
<td>Southern District</td>
<td>-0.096*</td>
<td>0.008</td>
<td>-0.154*</td>
<td>0.024</td>
</tr>
<tr>
<td>Northern District</td>
<td>-0.056*</td>
<td>0.009</td>
<td>-0.068*</td>
<td>0.028</td>
</tr>
<tr>
<td>$R^2$ index</td>
<td>0.73</td>
<td></td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>Observations (N)</td>
<td>39,244</td>
<td></td>
<td>2,570</td>
<td></td>
</tr>
</tbody>
</table>

Significant at: (*) 1%, (**) 5%, (***) 10%
Table 2: Variables included in prediction model

<table>
<thead>
<tr>
<th>Explanatory variables</th>
<th>Model A1</th>
<th>Model A2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable</td>
<td>Log price</td>
<td>Log price per sq. m.</td>
</tr>
<tr>
<td>At property level</td>
<td>Log area</td>
<td></td>
</tr>
<tr>
<td>At building level</td>
<td>Multi-storey building</td>
<td></td>
</tr>
<tr>
<td>At CT level</td>
<td>Log mean price</td>
<td>Log mean price per sq. m.</td>
</tr>
<tr>
<td>At locality level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At national level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At locality level</td>
<td>Log mean labor income</td>
<td></td>
</tr>
<tr>
<td>At national level</td>
<td>No. of buildings</td>
<td></td>
</tr>
<tr>
<td>At locality level</td>
<td>Median year of construction</td>
<td></td>
</tr>
<tr>
<td>At national level</td>
<td>No. of residents</td>
<td></td>
</tr>
<tr>
<td>Proximity of CT to center of Tel Aviv</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAPE</td>
<td>29.24</td>
<td>27.73</td>
</tr>
<tr>
<td>MedAPE</td>
<td>14.85</td>
<td>13.21</td>
</tr>
</tbody>
</table>
Table 3: Distribution of absolute percentage errors

<table>
<thead>
<tr>
<th>Index</th>
<th>Value (pct.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (MAPE)</td>
<td>20.12</td>
</tr>
<tr>
<td>Median (MedAPE)</td>
<td>12.28</td>
</tr>
<tr>
<td>Percentile 10</td>
<td>2.21</td>
</tr>
<tr>
<td>Percentile 25</td>
<td>5.65</td>
</tr>
<tr>
<td>Percentile 75</td>
<td>22.44</td>
</tr>
<tr>
<td>Percentile 90</td>
<td>36.86</td>
</tr>
</tbody>
</table>
Table 4: Distribution of predicted values, actual prices, and subjective valuations

<table>
<thead>
<tr>
<th>Source</th>
<th>Index</th>
<th>Mean</th>
<th>Median</th>
<th>10(^{th}) percentile</th>
<th>25(^{th}) percentile</th>
<th>75(^{th}) percentile</th>
<th>90(^{th}) percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBR (predicted values)</td>
<td>1,262,668</td>
<td>1,089,102</td>
<td>458,974</td>
<td>735,989</td>
<td>1,559,817</td>
<td>2,178,519</td>
<td></td>
</tr>
<tr>
<td>ITA data</td>
<td>1,092,162</td>
<td>900,000</td>
<td>350,000</td>
<td>550,000</td>
<td>1,350,000</td>
<td>1,940,001</td>
<td></td>
</tr>
<tr>
<td>HES data</td>
<td>1,499,539</td>
<td>1,300,000</td>
<td>625,000</td>
<td>900,000</td>
<td>1,800,000</td>
<td>2,500,000</td>
<td></td>
</tr>
</tbody>
</table>
Table 5: Performance of the proposed method in 2011-2013

<table>
<thead>
<tr>
<th>Year</th>
<th>Distribution of predicted values*</th>
<th>Estimated $\text{Std}(\hat{P}_{ijkl})$ (Formula 3)</th>
<th>Accuracy indices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Standard deviation</td>
<td>MAPE</td>
</tr>
<tr>
<td>2011</td>
<td>1,262,688</td>
<td>869,568</td>
<td>16.079</td>
</tr>
<tr>
<td>2012</td>
<td>1,292,192</td>
<td>840,143</td>
<td>14.050</td>
</tr>
<tr>
<td>2013</td>
<td>1,426,019</td>
<td>881,445</td>
<td>13.616</td>
</tr>
</tbody>
</table>

* In current prices
### Appendix 1: Description of variables

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dwelling value, (thousands)</td>
<td>1,092</td>
<td>897</td>
<td>43,342</td>
<td>45</td>
<td>993</td>
<td>1,500</td>
<td>1,300</td>
<td>16,000</td>
<td>30</td>
<td>996</td>
</tr>
<tr>
<td>Area (sq.m.)</td>
<td>82.4</td>
<td>74.0</td>
<td>1207.0</td>
<td>15.0</td>
<td>43.0</td>
<td>112.2</td>
<td>101.0</td>
<td>445.0</td>
<td>18.0</td>
<td>48.9</td>
</tr>
<tr>
<td>No. of rooms</td>
<td>3.5</td>
<td>3.0</td>
<td>15.0</td>
<td>1.0</td>
<td>1.1</td>
<td>4.2</td>
<td>4.0</td>
<td>10.0</td>
<td>1.0</td>
<td>1.1</td>
</tr>
<tr>
<td>Year of const.:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before 1964</td>
<td>0.1</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.3</td>
<td>0.1</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.3</td>
</tr>
<tr>
<td>1965-74</td>
<td>0.2</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.4</td>
<td>0.1</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.3</td>
</tr>
<tr>
<td>1975-84</td>
<td>0.2</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.4</td>
<td>0.2</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.4</td>
</tr>
<tr>
<td>1985-89</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.2</td>
<td>0.1</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.3</td>
</tr>
<tr>
<td>1990-99</td>
<td>0.2</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.4</td>
<td>0.3</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.4</td>
</tr>
<tr>
<td>2000+</td>
<td>0.8</td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.4</td>
<td>0.8</td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.4</td>
</tr>
<tr>
<td>Multi-storey bldg.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of dwellings in building</td>
<td>24.0</td>
<td>16</td>
<td>326</td>
<td>1</td>
<td>25.8</td>
<td>21.9</td>
<td>15</td>
<td>246</td>
<td>1</td>
<td>24.3</td>
</tr>
<tr>
<td>Property tax rate (NIS/sq.m.)</td>
<td>49.6</td>
<td>42.2</td>
<td>107.1</td>
<td>25.9</td>
<td>19.1</td>
<td>51.1</td>
<td>45.9</td>
<td>105.4</td>
<td>25.6</td>
<td>18.2</td>
</tr>
<tr>
<td>Proximity to center of locality (km.)*</td>
<td>2.1</td>
<td>1.8</td>
<td>10.4</td>
<td>0.0</td>
<td>1.6</td>
<td>2.4</td>
<td>1.9</td>
<td>10.4</td>
<td>0.0</td>
<td>1.8</td>
</tr>
<tr>
<td>Proximity to Tel Aviv (hundreds of km)</td>
<td>0.5</td>
<td>0.4</td>
<td>3.3</td>
<td>0.0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.3</td>
<td>3.2</td>
<td>0.0</td>
<td>0.4</td>
</tr>
<tr>
<td>Fronting sea***</td>
<td>0.1</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.3</td>
<td>0.1</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.2</td>
</tr>
<tr>
<td>Avg. income in CT (thousands of NIS)</td>
<td>90.0</td>
<td>83.7</td>
<td>442.8</td>
<td>36.1</td>
<td>32.2</td>
<td>98.7</td>
<td>91.3</td>
<td>268.8</td>
<td>36.2</td>
<td>34.4</td>
</tr>
<tr>
<td>No. of residents in locality</td>
<td>4084.6</td>
<td>4022</td>
<td>14172</td>
<td>22</td>
<td>1385.3</td>
<td>4255.1</td>
<td>4127.0</td>
<td>14172.0</td>
<td>330.0</td>
<td>1489.0</td>
</tr>
</tbody>
</table>

* Proximity of center of CT where property is located to center of locality (by air).
** Proximity of center of CT where property is located to center of Tel Aviv (by air).
*** 1 for CT fronting Mediterranean Sea; 0 otherwise.
Figure captions:

Figure 1: District division of Israel

Figure 2: Geographical distribution of the mean price per square meter

Figure 3: Distribution of Model (A2) coefficients across centiles of log price per square meter

Figure 4: Distribution of the accuracy indices by CT’s socio-economic cluster
Figure 1: District division of Israel
Figure 2: Geographical distribution of the mean price per square meter
Figure 3: Distribution of Model (A2) coefficients across centiles of log price per square meter
Figure 4: Distribution of the accuracy indices by CT’s socio-economic cluster