Reducing Revisions in Israel's House Price Index with Nowcasting Models[1]

Doron Sayag[a,b], Dano Ben-hur[a] and Danny Pfeffermann[a,c,d]

[a] Central Bureau of Statistics, Israel
[b] Bar-Ilan University, Israel
[c] Hebrew University of Jerusalem, Israel
[d] University of Southampton, UK

Abstract

National Statistical Offices must balance between the timeliness and the accuracy of the indicators they publish. Due to *late-reported transactions* of sold houses, many countries, including Israel, publish a provisional House Price Index (HPI), which is subject to revisions as further transactions are recorded. Until 2018, the Israel Central Bureau of Statistics (ICBS) published provisional HPIs based solely on the known reported transactions, which suffered from large revisions. In this paper we propose a novel method for minimizing the size of the revisions. Noting that the *late-reported transactions* behave differently from the on-time reported transactions, three types of variables are predicted monthly at the sub-district level as input data for a nowcasting model: (1) the average characteristics of the late-reported transactions; (2) the average price of the late-reported transactions; and (3) the number of late-reported transactions. These three types of variables are predicted separately, based on models fitted to data from previous months. Evaluation of our model shows a reduction in the magnitude of the revisions by more than 50%. The model is now used by the ICBS for the official publication of the provisional HPIs at both the national and district levels.

JEL Classification: C43, C51, R31

Keywords: Hedonic model, Index revision, Provisional indicator, Real estate market, Revisable statistics, Timeliness, Time Dummy Method.

1. Introduction

A major obstacle to timely statistics is often the absence of a complete, on-time dataset. In revisable statistics (unlike the consumer price index), one way to overcome the problem is by publishing provisional indicators which are later revised. A complementary step, aimed at producing more precise statistics, is the use of

[1] We thank Danny Ben-Shahar, Yoel Finkel, Larisa Fleishman, David Genesove, Can Tongur, members of the Israel Public Advisory Council for Statistics and participants in the 2019 Ottawa Group meeting in Rio de Janeiro for helpful discussions and valuable comments. Remaining errors are entirely our own. The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the Israel Central Bureau of Statistics or any other institutions with which the authors are affiliated.

nowcasting models, which have the potential to improve the quality of official statistics in terms of timeliness, high frequency, and accuracy.

Since mid-2007, house prices in Israel have risen considerably − by as much as 125% − which has made statistics on house prices of key importance. The rise in prices also increased the need for short time lag statistics on the housing market for decision-makers in different parts of the economy: individuals choosing whether to buy or sell a house, government policymakers assessing the success of housing policies, builders, central bankers, and the general public.

HPI revisions are not unique to Israel. In fact, most of the OECD countries publish provisional indices which are revised in the following months or the following quarters.[2] The importance of accurate and timely HPI was presented by Carless (2011), who found that for UK stakeholders, one of the key requirements for users of house price statistics is to be timely with minimal revisions. Moreover, some UK users specified that they need monthly house price statistics within a couple of weeks of the end of the month to which they refer.[3]

Israel's House Price Index (IHPI) has several characteristics that distinguish it from its counterpart in other countries. For instance: (1) the IHPI is published on a monthly basis but each publication reflects bimonthly price changes;[4] (2) the first provisional IHPI is released 45 days after the end of the reference month, when, on average 30% of the transactions are still missing; (3) the IHPI is subject to revisions in the three months following the initial release, which enables the inclusion of about 98% of the total transactions; and (4) until the end of 2017, the ICBS published a single national HPI, based on the traditional model (solely on the known reported transactions), which was often criticized for its large revisions.

Statistically speaking, an estimation of price change does not require full coverage of all transactions executed, if it satisfies two conditions: (i) the missing (late-reported) transactions are random and do not contain additional information; and (ii) the number of known reported transactions is sufficiently large. Otherwise, delayed-reported transactions might affect the result, as in the case of IHPI.

---

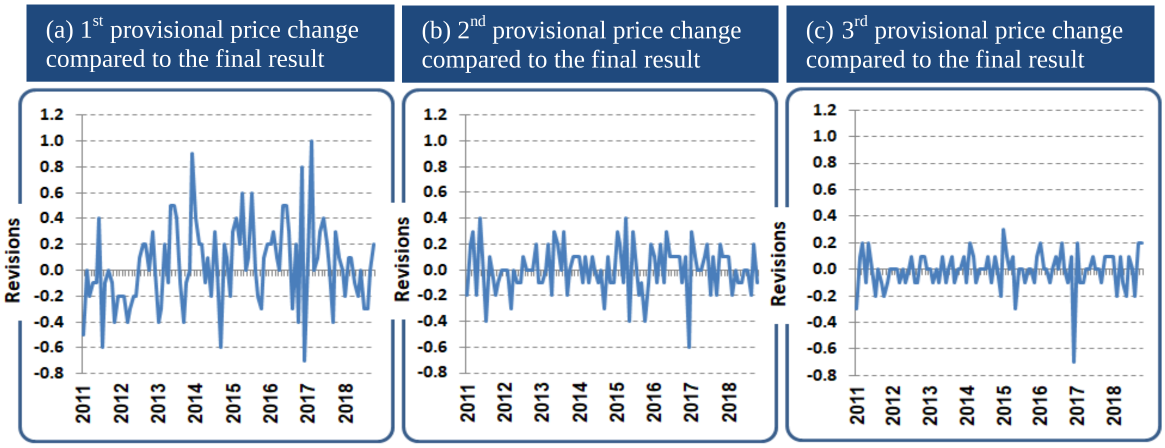
[2] Most of the OECD countries publish quarterly HPIs with a delay of 90–100 days after the reference quarter.

[3] The monthly UK HPI is determined after a long delay of 12 months period of revision.

[4] The bimonthly price change, which is an average of the price changes of two consecutive months, is used for "smoothing" the monthly changes, thus reducing the volatility of the price change between one month and the next.

Figure 1 shows the direction and magnitude of the revised provisional price changes (first, second, and third) as more transactions are reported, in comparison to the final price change that is published three months after the first provisional index is published.

Figure 1 ■ Revisions in absolute size of provisional IHPIs (first, second and third). 2011-2018



Note: Figures 1a - 1c present monthly revisions in absolute size of provisional price changes (first, second, and third). We define revisions as the final price change minus the provisional price change.

As can be seen from Figure 1, revisions occur in both directions —downward and upward—without any orderly pattern, suggesting that the deviations do not result from a systematic problem with the data. It can also be seen that the revisions of the first provisional price change are the largest, and the magnitudes of the second and third revisions become smaller, as further transactions are recorded.

In order to minimize the revisions, without changing the frequency or timeliness of HPI publication, we have developed a nowcasting model for monthly price change. To our knowledge, this is the first nowcasting model used by a national statistical office for the production and publication of HPIs. As illustrated later, the nowcasting models are used now routinely for the production of the HPI in Israel.

The rest of the paper is structured as follows: Section 2 presents background information on Israel's HPI and describes the hedonic model underlying it. Section 3 describes the databases used for computation of the HPI. Section 4 introduces the nowcasting model and Section 5 evaluates its performance. We conclude with some summary remarks in Section 6.

2. Hedonic model for the HPI in Israel

The HPI aims to measure the evolution of market prices for residential properties. Owing to high heterogeneity in residential property characteristics, the index should be unaffected by quality changes over time and reflect 'pure' price changes only. In order to construct a constant-quality index, the ICBS uses hedonic methodology based on a rolling-window time dummy method. This method (also known as the “direct method”) is viewed favorably by international guidelines and many price index specialists.[5] The data used for the computation of this method consist of newly built houses and existing resold houses during two successive months. The data contain sale prices and two explanatory subsets of variables (dwelling physical characteristics and location characteristics), enabling the employment of hedonic methodology. The log of the price is regressed against quality-measuring variables and a time dummy variable, where the pure price change over two months is derived as the exponentiated time dummy (see below).

2.1 The basic hedonic regression equation

As of 2018, the IHPI consists of six sub-indices calculated and published for the following districts: (1) the Jerusalem District; (2) the Northern District; (3) the Haifa District; (4) the Central District; (5) the Tel Aviv District; and (6) the Southern District. Equation (1) presents the model fitted at the district level:

$$\log \mathrm{P}_j = \beta_0 + \sum_{i=2}^{18} \beta_i N_i + \sum_{k=1}^{7} \eta_k Z_{jk} + \sum_{i=2}^{18} \sum_{k=1}^{6} \delta_{ik} N_i Z_{jk} + \sum_{m=1}^{6} \gamma_m D_t M_{jm} + \varepsilon_j \tag{1}$$

where the subscripts j, i, k, m and t denote the transaction, sub-district, indicator variable the transaction quality measures, the 6 districts and the month, respectively. The corresponding variables are defined as follows:

P_j is the sale price of transaction j,

N_i is a fixed effect for sub-district i (a total of 18 sub-districts)

M_{jm} is a dummy variable taking the value 1 if transaction j occurred in district m, m=1,⋯,6 and is 0 otherwise,

Z_{jk} defines 7 quality measuring variables of transaction j, where:

- Z_1 represents the number of rooms;
- Z_2 represents the log of the area dwelling in square meters;
- Z_3 is a dummy variable indicating a non-standard dwelling (single-family

[5] In terms of HPI revisions, the literature suggests that a rolling-window hedonic model that does not include all periods in a single estimation should be less sensitive to new data being added.

home, detached house, semi-detached house, penthouse, etc.);

Z_4 represents the log of the age of the dwelling;

Z_5 is a dummy variable for new dwellings bought "on paper" (year of construction after the year of transaction);

Z_6 represents the socio-economic cluster of the statistical area where the transaction occurs;[6] and

Z_7 measures the long-term level of dwelling prices of the statistical area where the transaction occurs.

D_t is a dummy variable indicating the month of the transaction, and

ε_j is a random error.

The monthly quality-adjusted price change is a weighted index (see Section 2.2 below), based on transactions executed in the two successive months under consideration.[7] Due to the logarithmic transformation of the price, the monthly price change, at the district level, is obtained as $e^{\hat{\gamma}_m}$, and the national price change is obtained by weighted aggregation over districts (see weights in Table 1 below). By re-estimating the same model each month, we enable inclusion of additional registered transactions in the estimation model.

2.2 Determining the weight of a sale transaction

The calculation of the IHPI is based on Weighted Least Squares (WLS) regression. The purpose of a weighted index is to make the reported transactions representative of the stock of dwellings in each sub-district. Weights for the IHPI take two major forms, depending on the main purpose of the index: (i) *volume* weights which equal the number of dwellings in each region, or (ii) *value* weights which equal the total value of the dwellings in the region. The weights serve to inflate each transaction such that it would reflect the relative stock (volume/value). It is generally accepted that value-weights are more appropriate for macro-economic goals, the housing stock deflator, and lender exposure. In addition, using weights based on the

[6] The Socio-economic Index was developed at the ICBS in the mid-1990s. The current index is based on the 2008 Population Census data. The variables used to construct the index reflect many aspects related to the socio-economic makeup of the population of different geographical units (such as demographic characteristics, education, unemployment rates, income, etc.). The ranking of the socio-economic cluster ranges from 1 to 20 (where 20 is the highest level).

[7] Basically, this is a standard time dummy method equation. Only here, instead of estimating separately the price change for each district, we added the interaction between districts and quality variables which enabled us to obtain the price changes of 6 districts in a single regression.

value of dwellings is more common with methods used for the computation of consumer price indexes based on the expenditure of households.

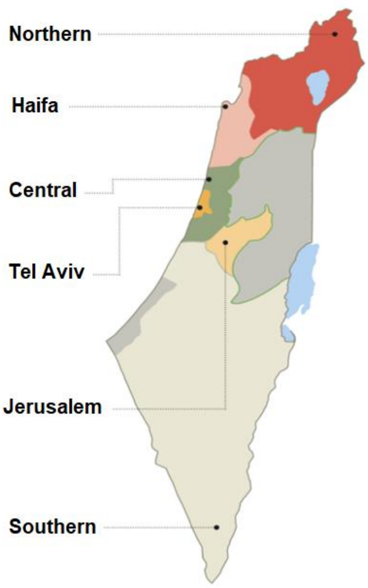
The IHPI is a stock type index, intended to measure the changes in the price component of the value of the housing stock. In order to turn a price index based on transactions into an index of price changes based on the stock value of dwellings, each transaction must be weighted according to its contribution to the total stock value of all the dwellings. The weights in the regression are calculated as the ratio of the fraction of the housing stock in the sub-district to the fraction of transactions belonging to the sub-district in the two-month period over which the monthly percentage change of prices is measured. Using the weights also allows the controlling of situations in which the numbers of transactions carried out in certain regions vary greatly from one month to the next. For the computation of the HPI at the ICBS, we assign the following value weight w_i to each reported transaction in a given sub-district:

$$w_i = \frac{N^0_{stock,i} \times \bar{P}^0_i}{N_{transactions,i}} \tag{2}$$

where $\mathrm{N}^0_{\mathrm{stock},i}$ is the number of dwellings in sub-district i during the base period, $\mathrm{N}_{\mathrm{transactions},i}$ is the number of transactions reported in sub-district i in the month under consideration and $\bar{P}^0_i$ is the average price of dwellings in the sub-district during the base period. Table 1 lists the weights at the base period assigned to the transactions in each sub-district.

Table 1

Base period weights in each district and sub-district for the computation of IHPI



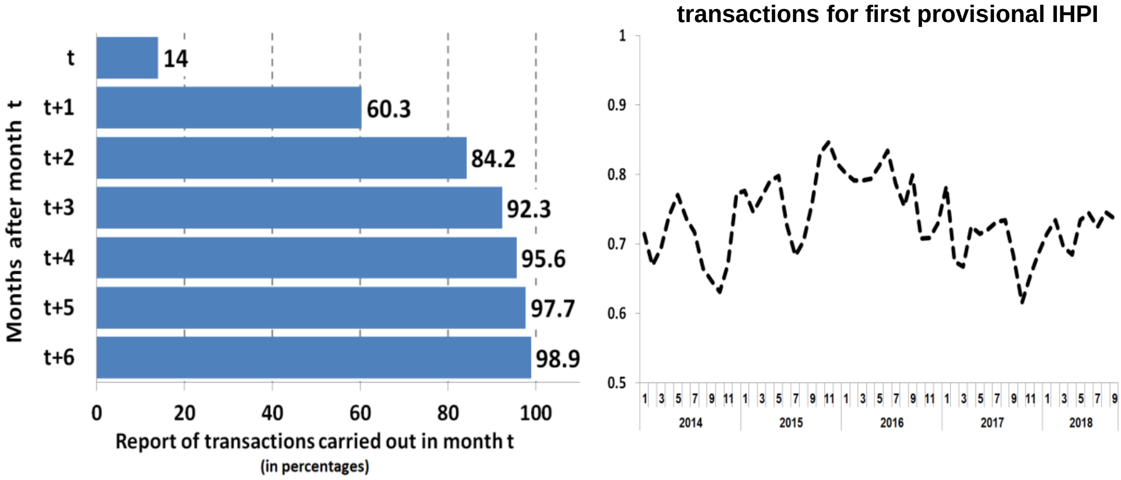
District	Weights (base period) w^0_m	Sub-District Code	Sub-District	Weights (base period) w^0_i
1. Jerusalem	16.21	11	Jerusalem	16.21
2. Northern	3.94	21	Zefat	0.40
		22	Kinneret	0.32
		23	Afula	0.84
		24	Akko	1.76
		25	Nazareth	0.62
3. Haifa	8.97	31	Haifa	7.23
		32	Hadera	1.74
4. Central	23.77	41	HaSharon	3.95
		42	Petah Tiqwa	9.80
		43	Ramla	2.12
		44	Rehovot	7.90
5. Tel Aviv	37.4	51	Tel Aviv	20.91
		52	Ramat Gan	9.84
		53	Holon	6.65
6. Southern	8.61	61	Ashqelon	5.06
		62	Be'er Sheva	3.55
Judea and Samaria Area	1.10	74, 76		1.1

The IHPIs are computed based on administrative sources with a total number of approximately 100,000 records per year at the national level. The main dataset is the CARMAN file of real estate prices, which is maintained by the Israeli Tax Authority. The data includes several characteristics of the sold dwellings: the date and the price of the sale, the location of the dwelling, its net area, gross area, number of rooms, number of floors in the building, and year of construction. Although those data are not fully suitable for statistical purposes, several modifications were made recently by the Tax Authority to satisfy statistical needs: (i) data that were submitted to the regional Property Tax Office (PTO) by handwritten form have been reported by online form since 2018. Switching from entering data manually into a computerized form has saved many typographical errors as found in previous examinations; (ii) the deadline period for reporting on a transaction to PTO was cut to 30 days instead of 40 days as it was before; and (iii) the data now contain more relevant and desired characteristics for the sake of improving IHPI quality adjustment.

While our dataset is updated on a monthly basis with new transactions, some of the transactions are reported after long delays, and thus cannot be used for calculating the first – and sometimes even the second and third provisional index. As can be seen in Figure 2(a), only 60% of the transactions are reported within 30 days of when the transaction takes place. Figure 2(b) shows the monthly percentage of transactions included in the first provisional IHPI.[8]

Figure 2 ■ Timelines of on-time reported transactions. 2014-2018

(a) Percentage of transactions executed in month t **(b) Coverage rate of on-time reported transactions for first provisional IHPI**



[8] The percentage of transactions reported on time for the first provisional IHPI is based on month t+1 and month t+2.

4. The Nowcasting Model

4.1 Background

The term nowcasting is a contraction of "now" and "forecasting" and can be summarized, in its most basic form, as predicting the present and sometimes the recent past (Castle, Fawcett, & Hendry, 2009). The technique has been a topic of long-standing interest for economists, dating as far back as the pioneering work of Mitchell and Burns (1938). The authors developed hundreds of leading, coincident, and lagging indicators of economic activity in the United States on business cycles.

In the last decade the use of nowcasting models has grown rapidly, especially in economic time series, such as Gross Domestic Product (GDP) series, unemployment rates, and tourist arrivals, which are published with long delays. Giannone, Reichlin, and Small (2008) found that the process of nowcasting can be formalized in a statistical model which produces predictions without the need for informal judgement.

Most of the models proposed in the recent literature are based on monitoring other sources of data on economic activity, characterized by high frequency and real-time data. Galbraith and Tkacz (2018), for instance, found that nowcasting Canadian GDP growth using Canadian credit, debit cards transactions, and cheques, narrows the gap between the first and final estimates by 65%. Other authors have provided some evidence of the usefulness of social media data and Google Trends data for nowcasting the "mood" of a population or events such as the presence of a flu epidemic. Other popular techniques offered in nowcasting literature are based on dynamic factor models, which extract information from a large number of indicators.

4.2 Approaches for reducing revisions in HPI

Although the use of administrative data sources in the production of official statistics is growing rapidly, methods to overcome problems such as incomplete data are limited. Specifically, the literature does not offer practical solutions regarding late-reporting observations, which are common in housing price statistics. Very few studies have investigated the magnitude of HPI revisions or the systematic bias of these revisions. Clapham et al. (2006) compared the HPI revisions in repeat-sale methods and hedonic indices for Stockholm, Sweden, over the period of 1981–1999 and found that HPI revisions based on repeat-sale methods are prone to be larger and downward compared to HPIs based on hedonic methods. Deng and Quigley (2008) analyzed the magnitude of HPI revisions in United States and their effects on prices in housing options markets.[9] They found that the average quarterly revision across 238 Metropolitan Statistical Areas (MSAs) was about -0.125%. However, the authors also found large-scale revisions of about 1.5% in absolute size in about one-quarter of the MSAs, and in about 15% of the housing markets, the average absolute revision exceeded 2%.

To better predicts the final price change in housing price markets we first studied the properties of revisions to identify reasons for large revisions of provisional HPIs. We also examined the assumption that the data generating process is the main reason that the late-reported transactions are not missing randomly. Analyzing the relationships of average characteristics and their prices, for on-time reported transactions and for late-reported transactions, indicates that the late-reported transactions behave differently from the on-time reported transactions.[10] A possible explanation for this phenomenon is different prioritizations, or more stringent examination by the Tax Authority, in cases in which the reported sale price seems unrealistic given the characteristics of a dwelling. While this process might be natural for the tax authority, it occasionally affects the provisional HPIs quite significantly. Since the known reported transactions are not representative of the completed transactions, ignoring the missing transactions might bias the HPI. Recognition of the importance of accurate provisional HPIs has led us to attempt nowcasting of the average characteristics (quality measures) of late-reported transactions and their average sale price, as well as the number of late-reported transactions.

As it turned out, adding to our hedonic regression a single record each month in each sub-district, consisting of the predicted averages, inflated by the predicted number of late-reported transactions, reduces significantly the magnitude of the revisions.

4.3 Nowcasting of average characteristics of late-reported transactions

As mentioned before, our analysis shows that late-reported transactions have different characteristics than on-time reported transactions. The model-fitted values

[9] The HPI in the United States is produced by the Federal Housing Finance Agency (FHFA) based on a repeat-sale method.

[10] We refer to transactions as "on-time reported transactions" if they are reported in time to be included in the provisional HPI under consideration.

from the regression model to nowcast the average characteristics of late-reported transactions are as follows:

$$\bar{Z}_{1,l} = \gamma_0 + \sum_{k=1}^{5} \gamma_k \bar{Z}_{k,nl} + \varepsilon_l \tag{3}$$

where the subscript "l" defines the characteristics of late-reported transactions and the double subscipting "nl" defines the characteristics of on-time reported transactions. Each observation in the dataset represents the average characteristics of the transactions that were carried out from month t - 6 until month t - 18 at the (month x district) level. The following average characteristics are nowcasted and then included in the hedonic regression.

$\bar{Z}_{1,l}$	–	average number of rooms based on late-reported transactions
$\bar{Z}_{1,nl}$	–	average number of rooms based on on-time reported transactions
$\bar{Z}_{2,nl}$	–	average area based on on-time reported transactions
$\bar{Z}_{3,nl}$	–	average socio-economic cluster based on on-time reported transactions
$\bar{Z}_{4,nl}$	–	average age of dwelling based on on-time reported transactions
$\bar{Z}_{5,nl}$	–	percentage of dwellings that are not in a residential building, based on on-time reported transactions
ε_l	–	random error

The model presented in Equation (3) utilizes the behavior of the past 12 months with a delay of 6 months and estimates the parameters γ_0 and γ_k. To estimate the average explained variable for the current period we use Equation (4):

$$\hat{\bar{Z}}_{1,l} = \hat{\gamma}_0 + \sum_{k=1}^{5} \hat{\gamma}_k \hat{\bar{Z}}_{k,nl} \tag{4}$$

To estimate the rest of the characteristics, we replaced the explained variable each time with the average that was obtained in the past for the characteristic that we wished to estimate for the late-reported transactions.

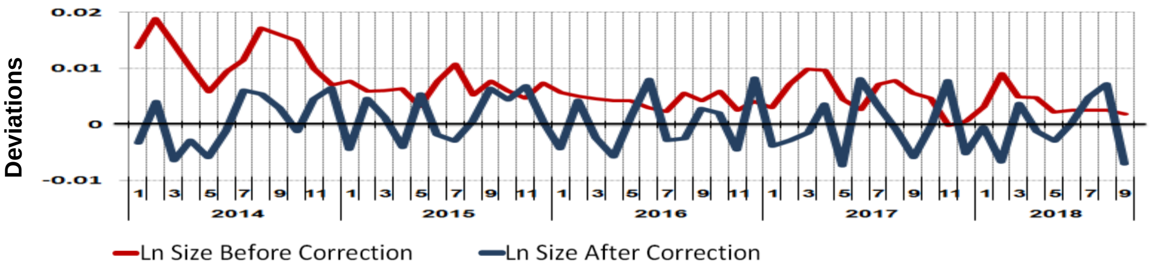
For the purposes of illustration, we present the prediction for only two characteristics: (a) area of dwelling and (b) number of rooms.

Figure 3(a) shows that the deviation between the average of the dwelling area (in log scale) of on-time reported transactions and the actual characteristics based on all the transactions (red line) is substantially larger than the deviation between the nowcasted characteristics (blue line) obtained from Equation (3) and the actual characteristics.

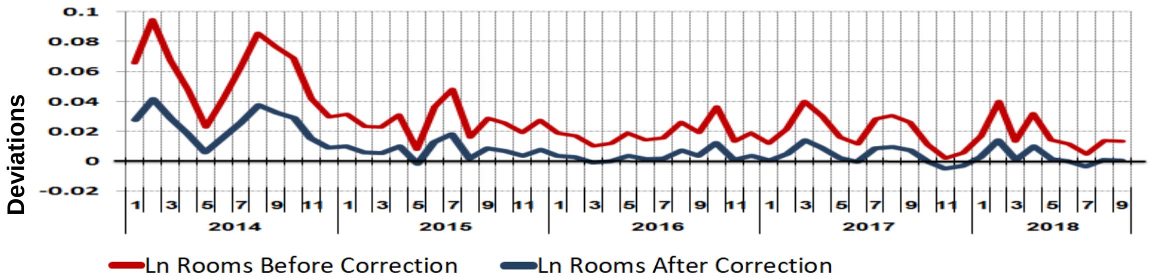
Figure 3(b) shows the reductions in deviation for average number of rooms obtained by using Equation (3).

Figure 3 ■ Provisional (known) versus nowcasted deviations for selected characteristics. 2014-2018

(a) Dwelling area



(b) Number of rooms



Note: The selected average characteristics (dwelling area and number of rooms) are plotted on the vertical axis as deviations from the actual (known 6 months later) average characteristics. The red line represents the average provisional characteristics (based on the known transactions) and the blue line represents the average nowcasted characteristics.

4.4 Nowcasting the monthly average price of the late-reported transactions

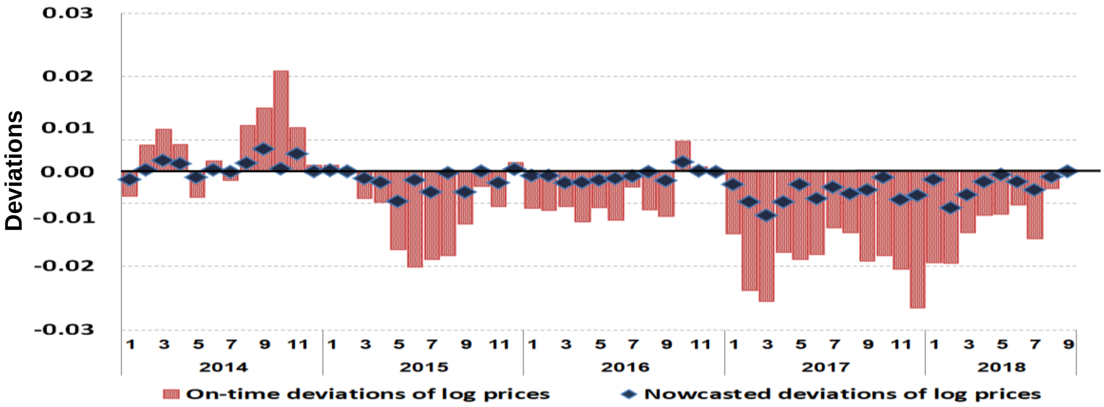
In the second stage, we nowcasted the monthly average price of the late-reported transactions at the sub-district level by fitting the following regression model:

$$\bar{P}_l = \alpha_0 + \sum_{k=1}^{5} \beta_k \bar{Z}_{k,nl} + \delta \bar{P}_{nl} + \varepsilon_l \quad (5)$$

where the dwelling characteristics $\{\bar{Z}_k\}$ are defined in Section 4.3.

Figure 4 compares the deviations of on-time and nowcasted average of the log prices of the actual transactions. As can be seen, the nowcasted deviations (blue points) are much smaller than the deviations of the on-time averages (red bars), indicating good performance of the nowcasting model.

Figure 4 ■ Deviations of on-time and nowcasted averages from the average log prices of the actual transactions

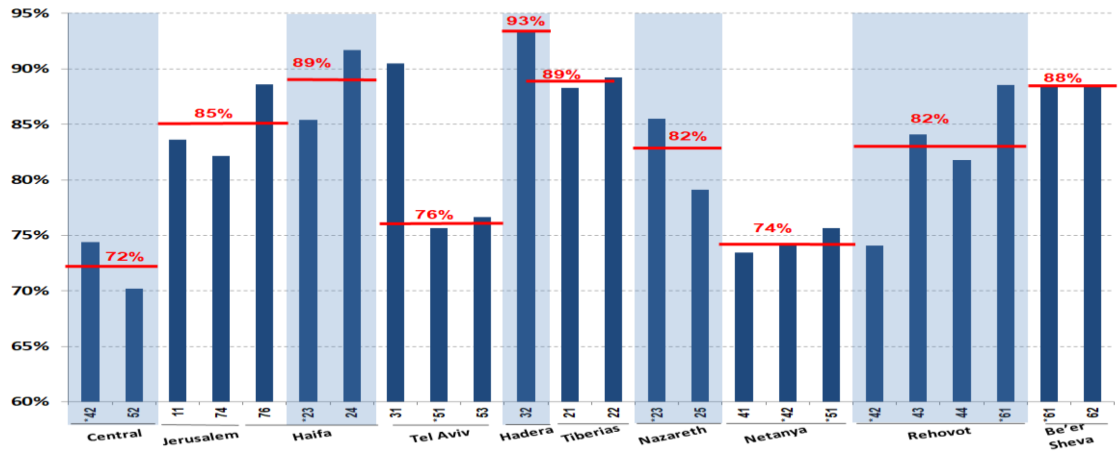


Note: The vertical axis presents the absolute size deviations from the actual log prices (known 6 months later).

4.5 Nowcasting the number of late-reported transactions

Lastly, we nowcasted the number of late-reported transactions at the sub-district level, which provides inflation factors of the imputed averages in the hedonic model, used to compute the IHPI. As mentioned in Section 2.2, the weights for the IHPI are at the sub-district level, and we therefore needed to inflate the imputed averages at the sub-district level as well. As depicted in Figure 5 the coverage rate of transactions reported on time varied substantially between sub-districts, depending on the Property Tax Office (PTO) operating in them. The PTOs are in charge of recording the transactions after approving them. Having different coverage rates in PTOs also affects the coverage rate of each sub-district because sub-district and PTO are correlated.

Figure 5 ■ Percentage Coverage of transactions reported on-time at sub-district level by regional PTO



***Notes*:** (a) The x-axis shows the regional PTO and the codes represent sub-districts, as follows: 11=Jerusalem, 21=Zefat, 22=Kinneret, 23=Afula, 24=Akko, 25=Nazareth, 31=Haifa, 32=Hadera, 41=HaSharon, 42=Petah Tiqwa, 43=Ramla, 44=Rehovot, 51=Tel Aviv, 52=Ramat Gan, 53=Holon, 61=Ashqelon, 62=Be'er Sheva, 74, 76=Two areas in Judea and Samaria. (b) Sub-districts codes with stars (23, 42, 51, 61) are sub-districts served by more than one regional PTO.

We use the following nowcasting model for predicting the number of missing (late-reported) transactions for any given sub-district, based on the past 6 months:

$$\frac{N_{t,t}}{N_{t,t+6}} = \alpha + \beta_1 \frac{N_{t,t}}{N_{t-1,t-1}} + \beta_2 \frac{N_{t-1,t-1}}{N_{t-2,t-2}} + \beta_3 \frac{N_{t-2,t-2}}{N_{t-3,t-3}} + u_t \tag{6}$$

where $N_{t,t+6}$ is the final number of transactions carried out in month t (known 6 months later) and $N_{t-k,t-k}$ is the number of transactions carried out and reported in month t - k, k = 0,1,2,3, with u_t representing a random error.

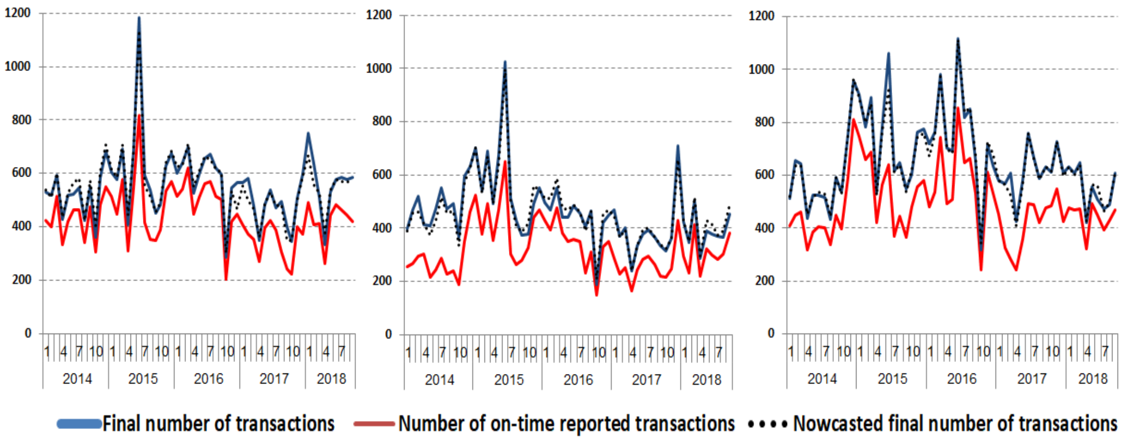
Note that in month t, the proportions $\frac{N_{t,t}}{N_{t-1,t-1}}$, $\frac{N_{t-1,t-1}}{N_{t-2,t-2}}$, and $\frac{N_{t-2,t-2}}{N_{t-3,t-3}}$, which refer to previous months, are already known. For example, in June 2018 we know the (non-final) number of transactions received for May (month t) and for April (month t – 1). We also know the number of transactions reported for month t – 2. Having learned the proportion of the on-time reported transactions out of the total number of transactions that will be obtained – $\frac{N_{t,t}}{N_{t,t+6}}$ – the final number of transactions in district *j* is predicted as:

$$\widehat{N}_{F,t} = \frac{N_{t,t}}{\widehat{P}_{t,t+6}} \tag{7}$$

where $\widehat{N}_{F,t}$ is the final number of transactions in a particular sub-district for month t and $\hat{P}_{t,t+6}$ is the predicted proportion of transactions reported on time, out of the total number of transactions as obtained from Equation (6). Figure 6 shows for three sub-districts (Jerusalem, Tel Aviv, and Petah Tiqwa) the number of transactions considered for the computation of the first provisional HPI, the final number of transactions (obtained after 6 additional months), and the final number of transactions as nowcasted by the use of Equations (6) and (7). The figure shows almost perfect prediction of the final number of transactions.

Figure 6 ■ Number of on-time reported transactions, final number of transactions, and nowcasted final number of transactions

(a) Jerusalem Sub-District **(b) Tel Aviv Sub-District** **(c) Petah Tiqwa Sub-District**



4.6 Calculating the first provisional nowcasted price change

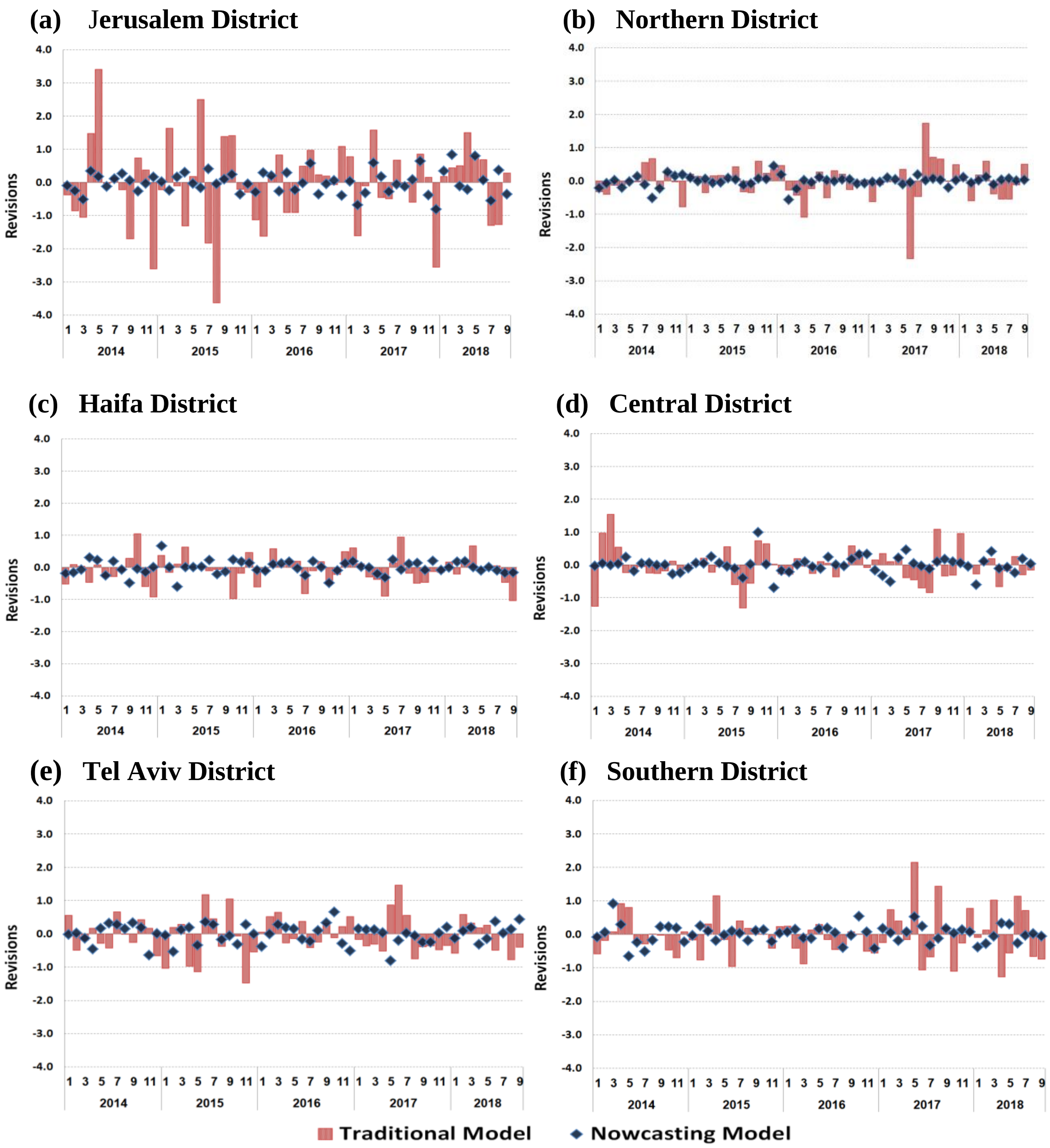
Until 2018, the hedonic regression model in Equation (1) was used for both the calculation of the provisional house price changes and for the calculation of the final house price changes. Since 2018, we have been adding to the hedonic regression for each sub-district the nowcasted average price of the late-reported transactions to the left side of the regression, and the nowcasted averages of the corresponding transaction characteristics (quality measures) to the right side, with all the averages inflated by the nowcasted numbers of the late-reported transactions (one for month t and another for month t – 1 in each of the 18 sub-districts). These records are based on the imputed characteristics, and average prices, as estimated in Equations (3) and (4), as well as the predicted numbers of the late-reported transactions, based on Equation (7). After "planting" those 36 records, the method of estimating the price change is based exactly on Equation (1).

5. Empirical Results

5.1 Examining the nowcasting model performance

In this section, we evaluate the performance of the nowcasted hedonic model in terms of the magnitude of the provisional price change revisions. We considered the period of January 2014 to September 2018 (a period of 57 months) for the evaluation. We should stress that the final price change was computed by the traditional model without the nowcasting model. Figures 7(a)-7(b) compare the revisions for each district, with and without implementing the nowcasting model.

Figure 7 ■ Revisions of price changes in six districts - traditional model versus nowcasting model. 2014-2018



Note: Revisions on the vertical axis are presented in absolute size between the first provisional price change and the final price change (known 6 months later).

As can be seen clearly in Figure 7, in all six districts, the provisional nowcasted price changes were much closer to the final price change than the price changes computed without the nowcasting. The nowcasting procedure was found to be particularly effective in the district of Jerusalem.

Table 2 displays the averages of the absolute percentage revisions in the six districts. Our nowcasting model reduced revisions of price changes at the district level by 40%

to 70%. As expected, the use of the nowcasting model was the most effective for the first provisional price change estimate.

Table 2

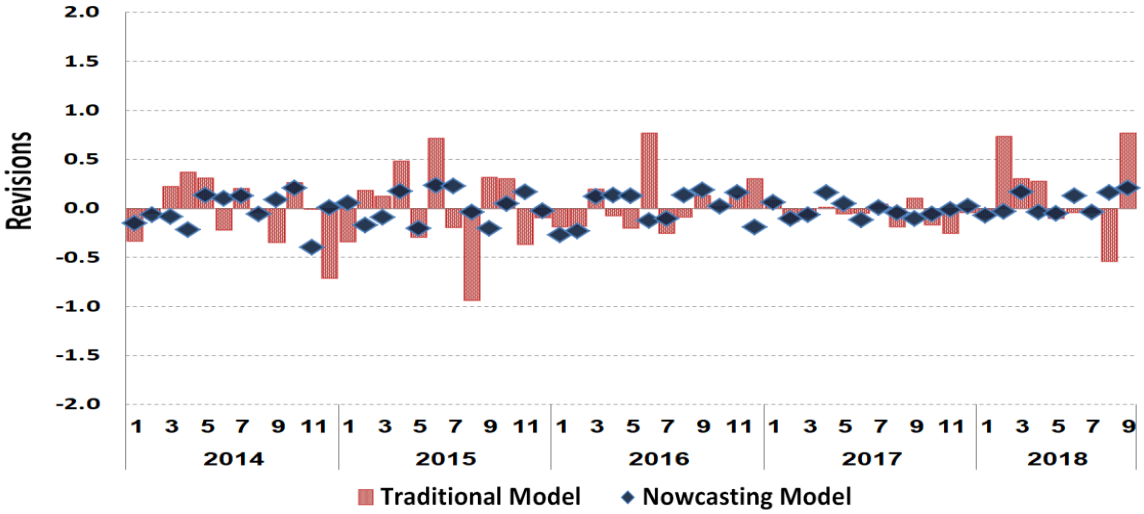
Average of absolute percentage revisions using the traditional model and using the nowcasting model in 6 districts. 2014-2018

Revisions	***Jerusalem District***		***Northern District***		***Haifa District***	
	Trad.	Nowc.	Trad.	Nowc.	Trad.	Nowc.
1st Provisional ΔP	**0.93**	**0.27**	**0.39**	**0.11**	**0.34**	**0.16**
2nd Provisional ΔP	**0.76**	**0.22**	**0.28**	**0.08**	**0.28**	**0.14**
3rd Provisional ΔP	**0.39**	**0.17**	**0.16**	**0.06**	**0.13**	**0.11**
Revisions	***Central District***		***Tel Aviv District***		***Southern District***	
	Trad.	Nowc.	Trad.	Nowc.	Trad.	Nowc.
1st Provisional ΔP	**0.38**	**0.17**	**0.48**	**0.22**	**0.51**	**0.20**
2nd Provisional ΔP	**0.26**	**0.13**	**0.36**	**0.15**	**0.42**	**0.16**
3rd Provisional ΔP	**0.11**	**0.10**	**0.19**	**0.12**	**0.20**	**0.13**

Note: ΔP = Price change; Trad. = Traditional model (no nowcasting); Nowc. = Nowcasting model.

Figure 8 presents the revisions at the national level of the two methods over the recent 5-year period. The figure shows the improvement resulting from implementing the nowcasting model in terms of predictive accuracy.

Figure 8 ■ Revisions of price changes at the national level - traditional model versus nowcasting model. 2014-2018



Note: Revisions on the vertical axis are presented in absolute size between the first provisional price change and the final price change (known 6 months later).

Table 3 displays the average of the absolute percentage revisions of the price changes at the national level. As can be seen, the average revisions of the nowcasted price changes reduced to a large extent (more than 50%) compared to the average revisions of the price changes without nowcasting.

Table 3
Average of revisions in absolute percentages - traditional model versus nowcasting model. 2014-2018

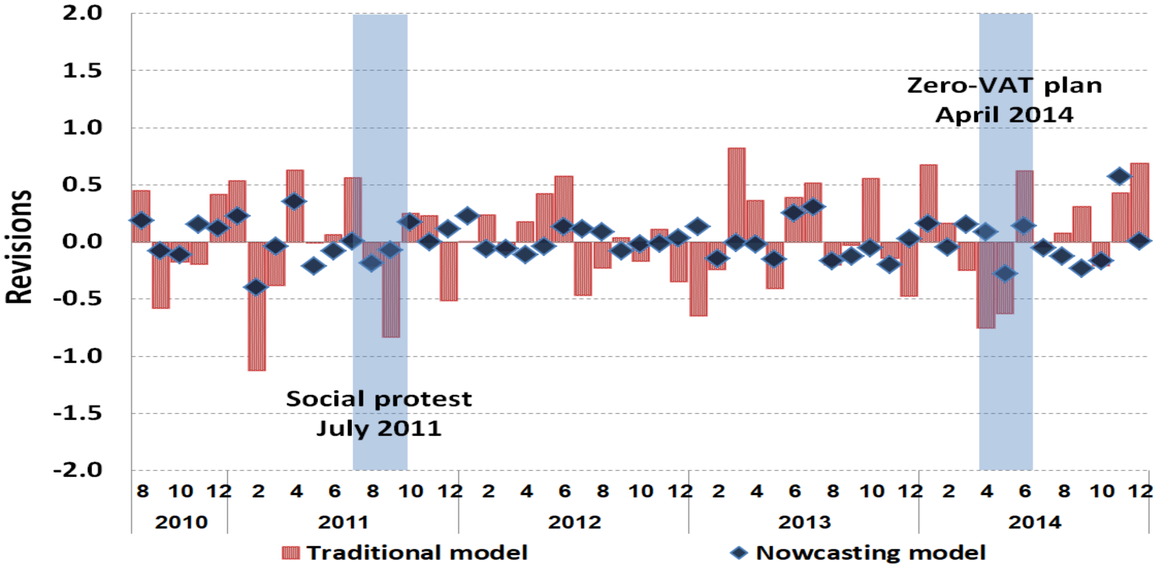
Revisions	*Trad. Model*	*Nowc. Model*
1st Provisional ΔP	**0.25**	**0.12**
2nd Provisional ΔP	**0.18**	**0.09**
3rd Provisional ΔP	**0.10**	**0.07**

Note: ΔP = Price change; Trad. = Traditional model (no nowcasting); Nowc. = Nowcasting model.

5.2. Robustness check

To further evaluate the performance of the nowcasting procedure, we examined, as robustness checks, two "unique" periods characterized by an unexpected decrease in the number of transactions and a relatively large deviation from the price trend line. The first event was the period of social protests during the summer of 2011. The second event was the period when the government declared on a zero-VAT tax for new home buyers, in April 2014. As shown in Figure 9, in both periods the provisional price changes obtained with the nowcasting procedure provide a closer estimate of the final price change (the horizontal axis) than the provisional price changes obtained by the traditional model (without the nowcasting).

Figure 9 ■ Revisions of price changes at the national level - nowcasting models versus traditional model during the periods of social protest and declaration of zero VAT



Note: Revisions on the vertical axis are presented in absolute size between the first provisional price change and the final price change (known 6 months later).

6. Concluding Remarks

Timeliness and accuracy are considered to be the most important elements in the quality of official statistics. By definition, provisional indicators based on incomplete data are subject to revisions. Nonetheless, National Statistical Offices still have an obligation to reduce the revisions to a possible minimum. Housing price indices have to account not only for quality adjustment issues, but also for occasionally long delays in obtaining complete information on sale transactions.

In this paper, we developed nowcasting models as a possible way to deal with the problem of late-reported transactions, which, as illustrated in the paper, give rise to large revisions. The revisions were found to be especially significant at lower levels of aggregation, such as districts.

The improved IHPI, which has been calculated using the nowcasting models since the beginning of 2018, has proven to be very successful in drastically reducing revisions. Evaluation of the model during the years 2014–2018 at both the district and national levels showed that nowcasted HPIs were more accurate than traditional ones. In particular, we found that the benefit from nowcasting models is most noticeable at the district level, where the number of monthly transactions is smaller, leading to higher price volatility and larger revisions. We also achieved better performance of the nowcasting models compared to traditional ones during unusual periods characterized by sharp declines in the number of transactions and by deviations from the price trend line.

We hope that our proposed model will raise awareness of the importance of reducing revisions and encourage other countries' national statistical offices, facing similar problems of late reports, to try it out with similar success. The potential usefulness of such nowcasting models is obviously not restricted to the calculation of HPIs and can be applied to other key statistics facing problems of late reports.

References

Ben-Shahar, D. and Sayag, D. (2018). The Israeli Commission Report on Prices and Rent of Dwellings. *The Public Advisory Council for Statistics Israel* (Hebrew only).

Bourassa, S. C., Hoesli, M., & Sun, J. (2006). A simple alternative house price index method. *Journal of Housing Economics*, 15(1), 80-97.

Calhoun, C. A. (1996). *OFHEO house price indexes: HPI technical description* (Working paper). Washington, DC: Office of Federal Housing Enterprise Oversight (OFHEO).

Carless, E. (2011). *Reviewing House Price Indexes in the UK*. Paper presented at the Workshop on House Price Indexes, Statistics Netherlands, The Hague, 10-11 February 2011.

Castle, J. L., Fawcett, N. W., & Hendry, D. F. (2009). Nowcasting is not just contemporaneous forecasting. *National Institute Economic Review*, 210(1), 71-89.

Clapham, E., Englund, P., Quigley, J. M., & Redfearn, C. L. (2006). Revisiting the past and settling the score: index revision for house price derivatives. *Real Estate Economics*, *34*(2), 275-302.

Clapp, J. M., & Giaccotto, C. (1999). Revisions in repeat-sales price indexes: here today, gone tomorrow*?. Real Estate Economics*, *27*(1), 79-104.

de Haan, J. and Diewert, W. E. (2013). *Handbook on residential property price indexes.* Luxembourg: Eurostat.

Deng, Y., & Quigley, J. M. (2008). Index revision, house price risk, and the market for house price derivatives. *The Journal of Real Estate Finance and Economics*, *37*(3), 191-209.

Giannone, D., Reichlin, L. & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data*. Journal of Monetary Economics*, *55*(4), 665-676.

Galbraith, J. W. & Tkacz, G. (2018). Nowcasting with payments system data. *International Journal of Forecasting*, *34*(2), 366-376.

Mitchell, W. C. & Burns, A. F. (1938). Statistical indicators of cyclical revivals. In *Business cycle indicators*, *1*, 184-260. Princeton University Press.

Sayag, D. (2012). Measuring the local house price movements and estimating the price elasticity. *Israel Economic Review*, *10*(1), 39–94.

Silver, M. (2002). *The use of weights in hedonic regressions: the measurement of quality adjusted price changes*. (Unpublished paper). Cardiff Business School.

Silver, M. (2012). *Why house price indexes differ: Measurement and analysis*. International Monetary Fund (Working Paper No. 12-125).

Silverstein, J. M. (2014, June). House price indexes: Methodology and revisions. *Research Rap Special Report*.

Appendix

A.1 Generating the variable "Price Level of Dwellings in a Statistical Area"

The explanatory variable "Price level of dwellings in a statistical area" was developed in 2018 by the ICBS for the purpose of improving the quality adjustment in the IHPI. It is intended to capture the effects of unobserved variables in a statistical area such as the level of community services, transportation accessibility, proximity to a beach, and proximity to a central business district. This variable is calculated with a six-month delay so as not to intervene with the current sales used for estimation of the price change, based on all the transactions in a 36-month moving time period. The variable is revised relatively frequently in order to provide updated information on newly built neighbourhoods and other recent developments relating to the characteristics of the statistical area. We use the following regression equation (estimated by Weighted Least Squares) for calculating this variable for the 2 936 Israeli statistical areas:

$$\log(P_j) = \beta_0 + \sum_{i=2}^{2\,936} \beta_i D_{ij} + \gamma_1 \log(S_j) + \gamma_2 V_j + \gamma_3 \log(Age_j) + \gamma_4 R_j + \sum_{j=2}^{36} \gamma_{5,J} T_j + \varepsilon_j \qquad \text{(A1)}$$

where,

- D_{ij} is a dummy variable taking the value 1 if transaction j occurred in statistical area i, i=1,….2 936 and 0 if otherwise,
- V_j is a dummy variable indicating a non-standard dwelling (detached house, semi-detached house, penthouse, etc.),
- S_j is the dwelling area,
- Age_j is the age of the dwelling,
- R_j represents the number of rooms,
- T_j is a dummy variable indicating the month of the transaction, and
- ε_j is a random error.

The coefficients $\hat{\beta}_1, \dots, \hat{\beta}_{2\,936}$ provide quality-adjusted estimators of the price levels of dwellings in the 2 936 statistical areas. Equation (A1) is calculated twice a year (in December and June) and as mentioned above, the model is fitted based on all the transactions executed in a three-year period, with a delay of 6 months. For example, the coefficients $\hat{\beta}_1, \dots, \hat{\beta}_{2\,936}$ for January 2019 to June 2019 are based on all the transactions executed between June 2015 and June 2018.