Reducing Revisions in Israel’s House Price Index With Nowcasting Models

Doron Sayag a,b Dano Ben-hur a
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
1. Israel House Price Index

1.1 Introduction

1.2 Hedonic model for the HPI in Israel

1.3 The basic hedonic regression equation

1.4 Determining the weight of a sale transaction

1.5 The data
1.1 Introduction

• 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.

• 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 to publish provisional indicators which are later revised.
1.1 Introduction (cont.)

Revisions in absolute size of provisional IHPIs (first, second and third)

(a) 1\textsuperscript{st} provisional price change compared to the final result

(b) 2\textsuperscript{nd} provisional price change compared to the final result

(C) 3\textsuperscript{rd} provisional price change compared to the final result

Note: figures 1(a) – 1(c) 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.
A HPI aims to measure the evolution of market prices for residential properties to be unaffected by quality changes over time.

**Main features of the IHPI:**

- Hedonic methodology.
- A two months rolling-window time dummy method.
- A stock type index > Weighted Least Squares (WLS) regression
- Three explanatory subsets of variables:
  - Dwelling physical characteristics
  - Location characteristics
  - Dummy variables
1.3 The basic hedonic regression equation

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

where the indexes $j$, $i$, $k$, $m$ and $t$ denote the transaction, sub-district, indicator variable the transaction quality measures, the 6 districts and the month, respectively.

**Note:** The above equation is based on a standard time dummy method equation:

$log p = \beta_0 + \sum_{k=1}^{K} \beta_k Z_k + \delta * Time + \varepsilon$

*But,* 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.
1.3 The basic hedonic regression equation (cont.)

\[ \log P_j = \beta_0 + \sum_{i=2}^{18} \beta_i N_i + \sum_{k=1}^{7} \eta_k Z_{jk} + \sum_{i=2}^{18} \sum_{k=1}^{6} \delta_{ik} N_i Z_{jk} + \sum_{m=1}^{6} \gamma_m D_t M_{jm} + \varepsilon_j \]

The corresponding variables are defined as follows:

- **Dwelling physical characteristics**
  - \( Z_1 \): number of rooms
  - \( Z_2 \): log of the area of the dwelling in m²
  - \( Z_3 \): a dummy variable indicating a non-standard dwelling*
  - \( Z_4 \): log of the age of dwelling
  - \( Z_5 \): a dummy variable for newly built dwellings bought "on paper" (year of construction after the year of transaction),

*single-family home, detached house, semi-detached house, penthouse, etc.
1.3 The basic hedonic regression equation (cont.)

\[ \log P_j = \beta_0 + \sum_{i=2}^{18} \beta_i N_i + \sum_{k=1}^{7} \eta_k Z_{jk} + \sum_{i=2}^{18} \sum_{k=1}^{6} \delta_{ik} N_i Z_{jk} + \sum_{m=1}^{6} \gamma_m D_t M_{jm} + \varepsilon_j \]

- **Location characteristics**
  - \( Z_6 \): socio-economic cluster of the statistical area where the transaction occurs
  - \( Z_7 \): long-term level of dwelling prices of the statistical area where the transaction occurs

- **Dummy variables**
  - \( N_i \): fixed effect for sub-district i (a total of 18 sub-districts)
  - \( M_{jm} \): dummy variable taking the value 1 if transaction \( j \) occurred in district \( m \), \( m=1,\ldots,6 \) and is 0 otherwise,
  - \( D_t \): dummy variable indicating the month of the transaction
1.3 The basic hedonic regression equation (cont.)

\[
\log P_j = \beta_0 + \sum_{i=2}^{18} \beta_i N_i + \sum_{k=1}^{7} \eta_k Z_{jk} + \sum_{i=2}^{18} \sum_{k=1}^{6} \delta_{ik} N_i Z_{jk} + \sum_{m=1}^{6} \gamma_m D_t M_{jm} + \varepsilon_j
\]

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

<table>
<thead>
<tr>
<th>District</th>
<th>Weights (base period) $\bar{w}_m^0$</th>
<th>Sub-District Code</th>
<th>Sub-District</th>
<th>Weights (base period) $\bar{w}_m^0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Jerusalem</td>
<td>16.21</td>
<td>11</td>
<td>Jerusalem</td>
<td>16.21</td>
</tr>
<tr>
<td>2. Northern</td>
<td>3.94</td>
<td>21, 22, 23, 24, 25</td>
<td>Zefat, Kinneret, Afula, Akko, Nazareth</td>
<td>0.40, 0.32, 0.84, 1.76, 0.62</td>
</tr>
<tr>
<td>3. Haifa</td>
<td>8.97</td>
<td>31, 32</td>
<td>Haifa, Hadera</td>
<td>7.23, 1.74</td>
</tr>
<tr>
<td>4. Central</td>
<td>23.77</td>
<td>41, 42, 43, 44</td>
<td>HaSharon, Petah Tiqwa, Ramla, Rehovot</td>
<td>3.95, 9.80, 2.12, 7.90</td>
</tr>
<tr>
<td>5. Tel Aviv</td>
<td>37.4</td>
<td>51, 52, 53</td>
<td>Tel Aviv, Ramat Gan, Holon</td>
<td>20.91, 9.84, 6.65</td>
</tr>
<tr>
<td>6. Southern</td>
<td>8.61</td>
<td>61, 62</td>
<td>Ashqelon, Be’er Sheva</td>
<td>5.06, 3.55</td>
</tr>
<tr>
<td>Judea and Samaria Area</td>
<td>1.10</td>
<td>74, 76</td>
<td></td>
<td>1.08</td>
</tr>
</tbody>
</table>
1.4 Base period weights in each district and sub-district for the computation of IHPI

For the computation of the IHPI, we used inflation factors for each reported transaction in a given sub-district.

\[ w_i = \frac{N_{stock,i}^0 \times \bar{P}_i^0}{N_{transactions,i}} \]

Where

- \( N_{stock,i}^0 \) is the number of dwellings in sub-district \( i \) during the base period
- \( N_{transactions,i} \) is the number of transactions reported in sub-district \( i \) in the month under consideration
- \( \bar{P}_i^0 \) is the average price of dwellings in the sub-district during the base period.
1.3 The data

3 Timelines of on-time reported transactions (2014-2018)

(a) Percentage of transactions executed in month $t$

<table>
<thead>
<tr>
<th>Months after month $t$</th>
<th>Report of transactions carried out in month $t$ (in percentages)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$</td>
<td>14</td>
</tr>
<tr>
<td>$t+1$</td>
<td>60.3</td>
</tr>
<tr>
<td>$t+2$</td>
<td>84.2</td>
</tr>
<tr>
<td>$t+3$</td>
<td>92.3</td>
</tr>
<tr>
<td>$t+4$</td>
<td>95.6</td>
</tr>
<tr>
<td>$t+5$</td>
<td>97.7</td>
</tr>
<tr>
<td>$t+6$</td>
<td>98.9</td>
</tr>
</tbody>
</table>

(b) Coverage rate of on-time reported transactions for first provisional IHPI
2. Reducing Revisions With Nowcasting Model

2.1 Revisions in House Price Index

2.2 Nowcasting model
2.1 HPI revisions

- Deng and Quigley (2008) analyzed the magnitude of HPI revisions in United States and their effects on prices in housing options markets.
  - Average quarterly revision across 238 Metropolitan Statistical Areas (MSAs) was low - about −0.125%.
  - Large-scale revisions of about 1.5% were found in about one-quarter of the MSAs, and in about 15% of the housing markets, the average revision exceeded 2%.

  - HPI revisions based on repeat-sale methods are prone to be larger and downward compared to HPIs based on hedonic methods.
2.2 Nowcasting models

- The term nowcasting is a contraction of “now” and “forecasting”.
- In its most basic form, nowcasting can be summarized as predicting the present and sometimes the recent past (Castle, Fawcett, & Hendry, 2009).
- In the last decade the use of nowcasting models has been growing rapidly, especially in economic time series, which are published with long delays.
- Giannone, Reichlin, and Small (2008) pointed out that the process of nowcasting can be formalized in a statistical model which produces predictions without the need for informal judgement.
3. Description of the nowcasting model

(a) Nowcasting the average characteristics of late-reported transaction

(b) Nowcasting the monthly average price of the late-reported transactions

(c) Nowcasting the number of late-reported transactions
The model-fitted values from the regression model to nowcast the average characteristics of late-reported transactions are as follows:

\[
\tilde{Z}_{1,l} = \gamma_0 + \sum_{k=1}^{5} \gamma_k \tilde{Z}_{k, nl} + \varepsilon_l
\]

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

The model presented in the previous equation utilizes the past behavior of 12-months in a delay of 6 months and estimates the parameters $\gamma_0$ and $\gamma_k$.

To estimate the average explained variable for the current period we used the following equation:

$$\hat{Z}_{1l} = \hat{y}_0 + \sum_{k=1}^{5} \hat{y}_k \hat{Z}_{k, nl}$$

To estimate the rest of the characteristics, we replaced the explained variable each time with the characteristic that we wished to estimate.
(a) Nowcasting of average characteristics of late-reported transactions

Provisional (known) versus nowcasted deviations for selected characteristics, 2014-2018

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

\[
\bar{P}_l = \alpha_0 + \sum_{k=1}^{5} \beta_k \bar{Z}_{k,nl} + \delta \bar{P}_{nl} + \varepsilon_l
\]
(b) Nowcasting the monthly average price of the late-reported transactions

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

![Graph showing deviations of on-time and nowcasted log prices]
(c) Nowcasting the number of late-reported transactions

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

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

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.
After learning 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:

$$\hat{N}_{F,t} = \frac{N_{t,t}}{\hat{P}_{t,t+6}}$$

where $\hat{N}_{F,t}$ 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.
(c) Nowcasting the number of late-reported transactions

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. Empirical Results

4.1 The magnitude of revisions at the district level

4.3 The magnitude of revisions at the national level
4.1 The magnitude of revisions at the district level

Revisions of price changes in six districts - nowcasting model versus traditional model

(a) Jerusalem District
(b) Northern District
(c) Haifa District
(d) Central District
(e) Tel Aviv District
(f) Southern District

- Traditional Model
- Nowcasting Model
4.1 The magnitude of revisions at the district level

Revisions of price changes in six districts - nowcasting model versus traditional model

<table>
<thead>
<tr>
<th>Revisions</th>
<th>Jerusalem District</th>
<th>Northern District</th>
<th>Haifa District</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; Provisional ΔP(%)</td>
<td>0.93</td>
<td>0.27</td>
<td>0.39</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; Provisional ΔP(%)</td>
<td>0.76</td>
<td>0.22</td>
<td>0.28</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt; Provisional ΔP(%)</td>
<td>0.39</td>
<td>0.17</td>
<td>0.16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Revisions</th>
<th>Central District</th>
<th>Tel Aviv District</th>
<th>Southern District</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; Provisional ΔP(%)</td>
<td>0.38</td>
<td>0.17</td>
<td>0.48</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; Provisional ΔP(%)</td>
<td>0.26</td>
<td>0.13</td>
<td>0.36</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt; Provisional ΔP(%)</td>
<td>0.11</td>
<td>0.10</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Notes: ΔP(%) = Price Change in percentage; Trad. = Traditional model (no nowcasting); NowC. = Nowcasting model;
4.2 The magnitude of revisions at the national level

Revisions of price changes at national level - nowcasting models versus traditional model. 2014-2018

<table>
<thead>
<tr>
<th>Revisions</th>
<th>Trad. Model</th>
<th>NowC. Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Provisional $\Delta P(%)$</td>
<td>0.25</td>
<td>0.12</td>
</tr>
<tr>
<td>2nd Provisional $\Delta P(%)$</td>
<td>0.18</td>
<td>0.09</td>
</tr>
<tr>
<td>3rd Provisional $\Delta P(%)$</td>
<td>0.10</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Notes: $\Delta P(\%)$ = Price Change in percentage; Trad. = Traditional model (no nowcasting); NowC. = Nowcasting model.
Timeliness and accuracy are considered to be the most important elements in the quality of official statistics.

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.

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.

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.
“….. yesterday is **history**, tomorrow is a **mystery**, today is a **gift**. that’s why it is called the **present**.”

- Joan Rivers
(Kungfu Panda, 2008)