הלשכה המרכזית לסטטיסטיקה
Central Bureau of Statistics
دائرة الإحصاء المركزية

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

16th Meeting of the Ottawa Group l Rio de Janeiro, Brazil, 8-10 May 2019 l Session 10

1. Israel House Price Index

1.1 Introduction

1.2 Hedonic model for the HPI in Israel

1.3 The basic hedonic regression equation

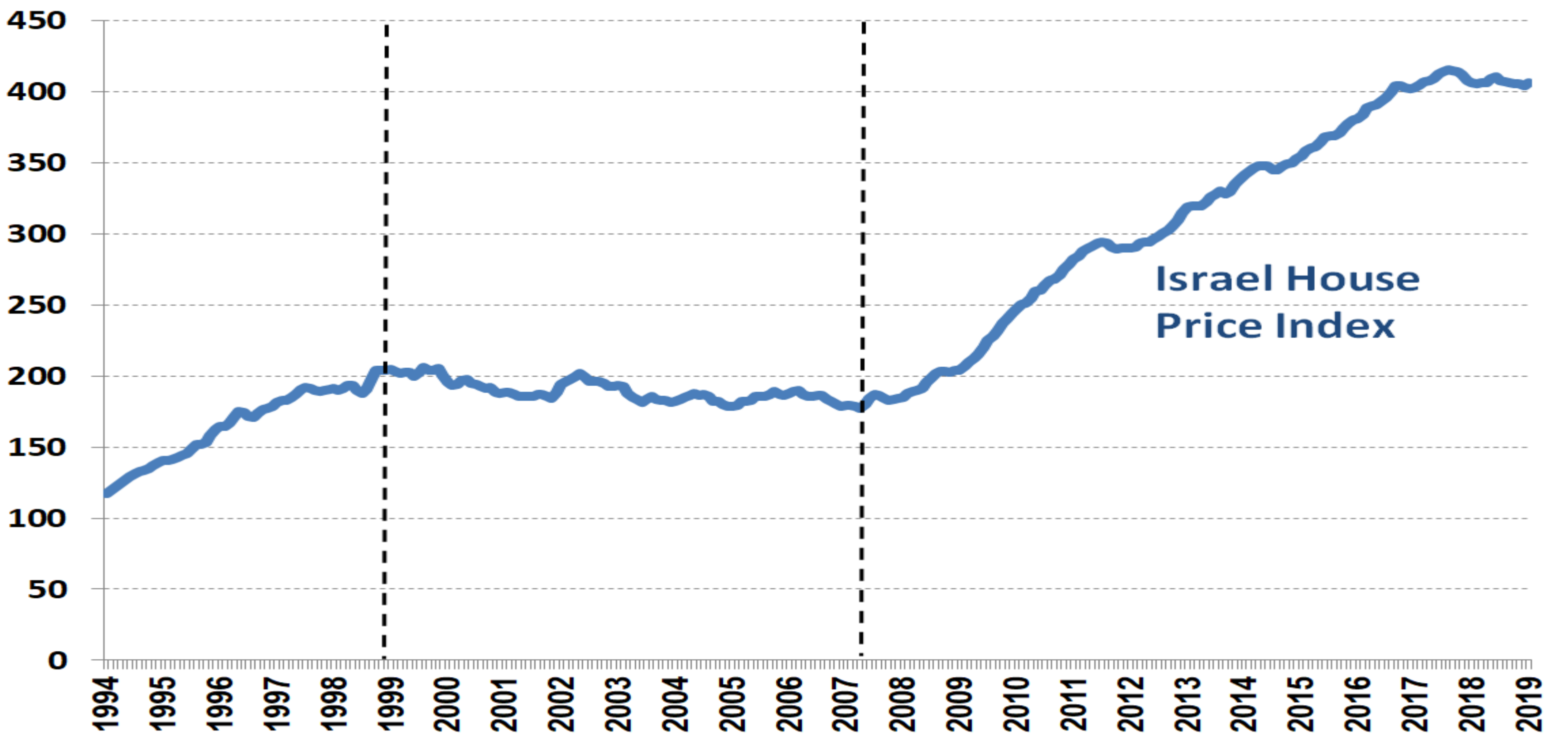
1.4 Determining the weight of a sale transaction

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.

1 **Evolution of the Israel House Price Index. (1994-2019)**



1.1 Introduction (cont.)

הלשכה המרכזית לסטטיסטיקה
Central Bureau of Statistics
دائرة الإحصاء المركزية

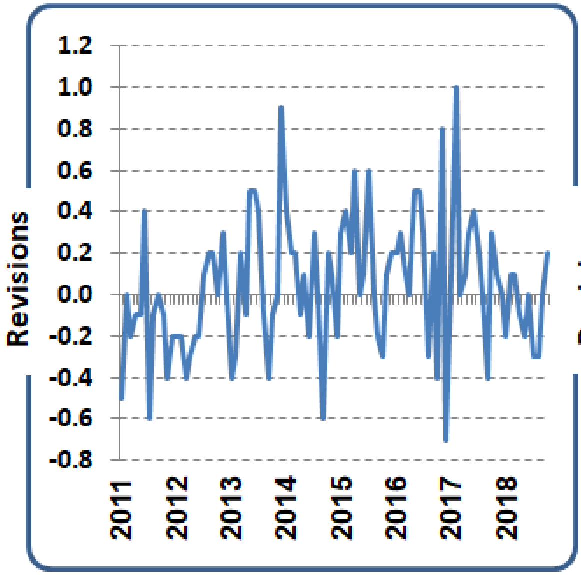
- The rise in prices also increased the need for short time lag statistics on the housing market for decision-makers in different parts of the economy: individuals choosing whether to buy or sell a house, government policymakers assessing the success of housing policies, builders, central bankers, and 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.

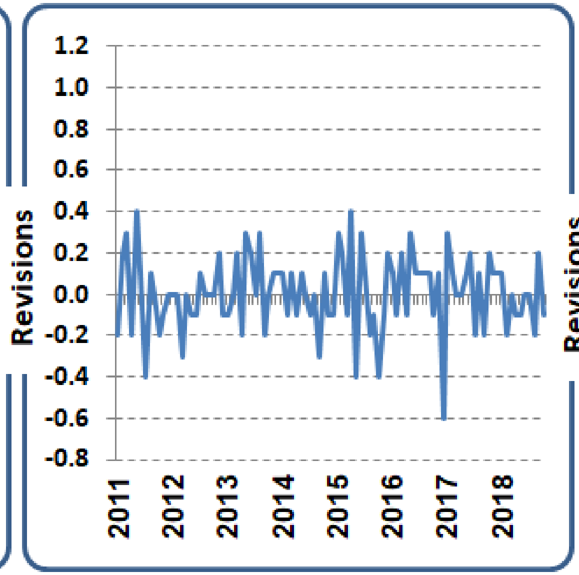
1.1 Introduction (cont.)

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

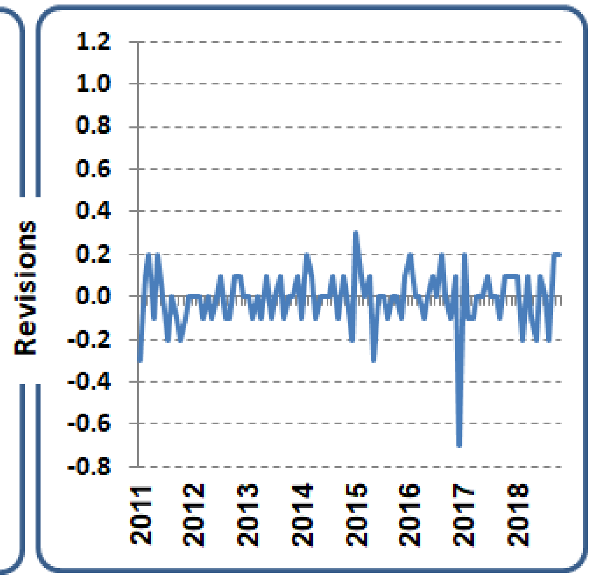
(a) 1st provisional price change compared to the final result



(b) 2nd provisional price change compared to the final result



(C) 3rd 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.

1.2 The hedonic model for HPI in Israel

הלשכה המרכזית לסטטיסטיקה
Central Bureau of Statistics
دائرة الإحصاء المركزية

A HPI aims to measure the evolution of market prices for residential properties ➡ 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

הלשכה המרכזית לסטטיסטיקה
Central Bureau of Statistics
دائرة الإحصاء المركزية

1.3 The basic hedonic regression equation

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

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.)

הלשכה המרכזית לסטטיסטיקה
Central Bureau of Statistics
دائرة الإحصاء المركزية

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

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,…,6 and is 0 otherwise,

D_t dummy variable indicating the month of the transaction

1.3 The basic hedonic regression equation (cont.)

הלשכה המרכזית לסטטיסטיקה
Central Bureau of Statistics
دائرة الإحصاء المركزية

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

- 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

District	Weights (base period) w_m^0	Sub-District Code	Sub-District	Weights (base period) w_i^0
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		1.08

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.

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

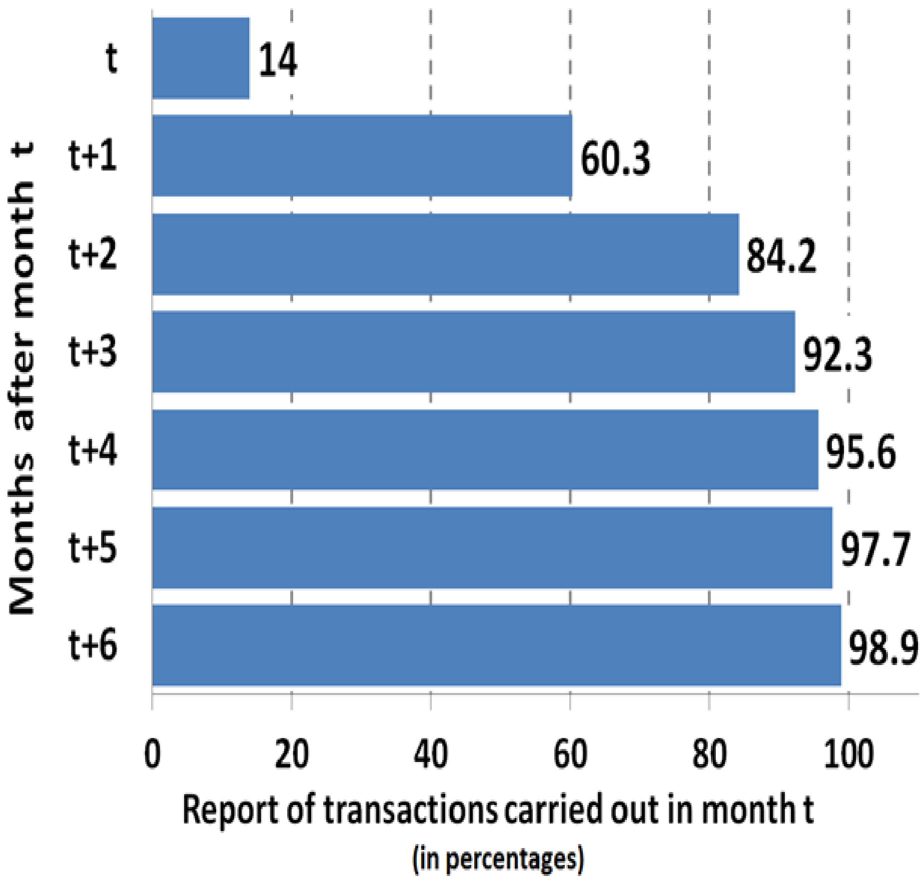
Where

$\mathrm{N}^0_{\mathrm{stock},i}$	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
$\bar{P}^0_i$	is the average price of dwellings in the sub-district during the base period.

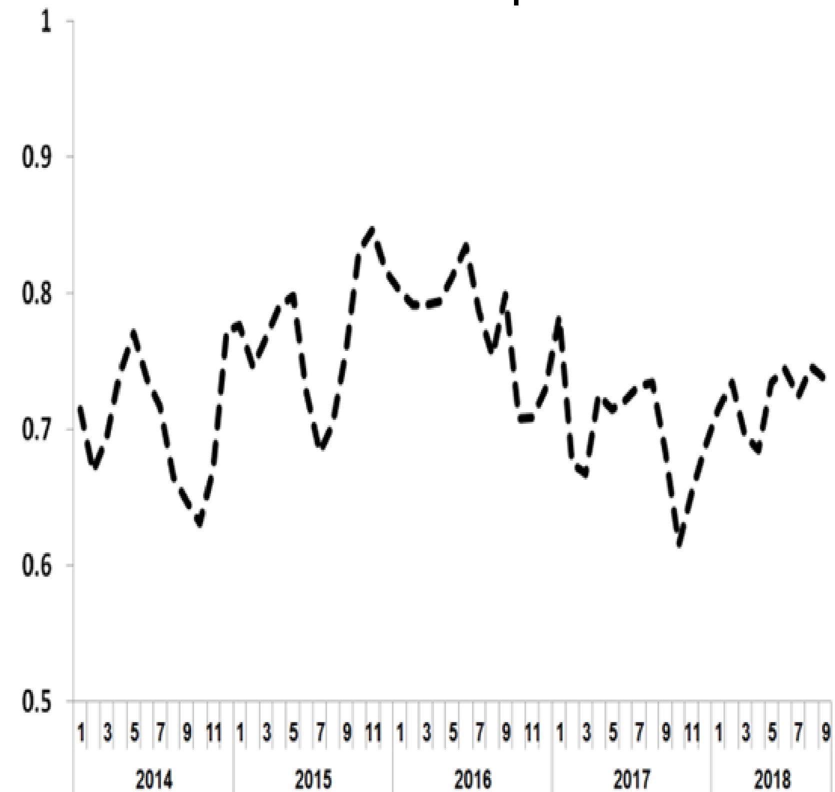
1.3 The data

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

(a) Percentage of transactions executed in month t



(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%.

- Clapham, Englund, Quigley, and Redfearn (2006) compared the HPI revisions in repeat-sale methods and hedonic indices for Stockholm, Sweden, over the period of 1981–1999.
  - 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

הלשכה המרכזית לסטטיסטיקה
Central Bureau of Statistics
دائرة الإحصاء المركزية

- 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

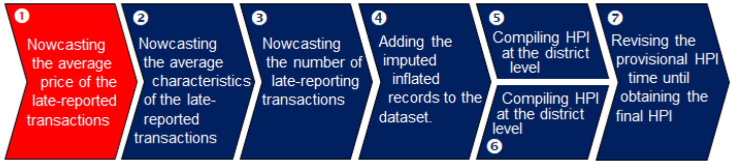
(a) Nowcasting the average characteristics of late-reported transaction

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

(c) Nowcasting the number of late-reported transactions

(a) Nowcasting of average characteristics of late-reported transactions

הלשכה המרכזית לסטטיסטיקה
Central Bureau of Statistics
دائرة الإحصاء المركزية



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

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

late-reported transactions

on-time reported transactions

$\bar{Z}_{1,\boldsymbol{l}}$ – average number of rooms based on **late-reported transactions**

$\bar{Z}_{1,\boldsymbol{nl}}$ – average number of rooms based on **on-time reported transactions**

$\bar{Z}_{2,\boldsymbol{nl}}$ – average area based on **on-time reported transactions**

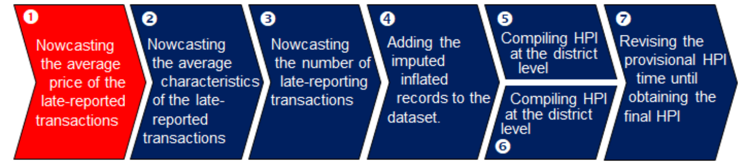
$\bar{Z}_{3,\boldsymbol{nl}}$ – average socio-economic cluster based on **on-time reported transactions**

$\bar{Z}_{4,\boldsymbol{nl}}$ – average age of dwelling based on **on-time reported transactions**

$\bar{Z}_{5,\boldsymbol{nl}}$ – percentage of dwellings that are not in a residential building, based on **on-time reported transactions**

$\varepsilon_{\boldsymbol{l}}$ – 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 γ_0 and γ_k.

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

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

To estimate the rest of the characteristics, we replaced the explained variable each time with the characteristic that we wished to estimate.

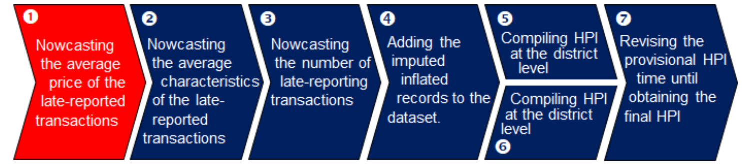
(a) Nowcasting of average characteristics of late-reported transactions

הלשכה המרכזית לסטטיסטיקה
Central Bureau of Statistics
دائرة الإحصاء المركزية

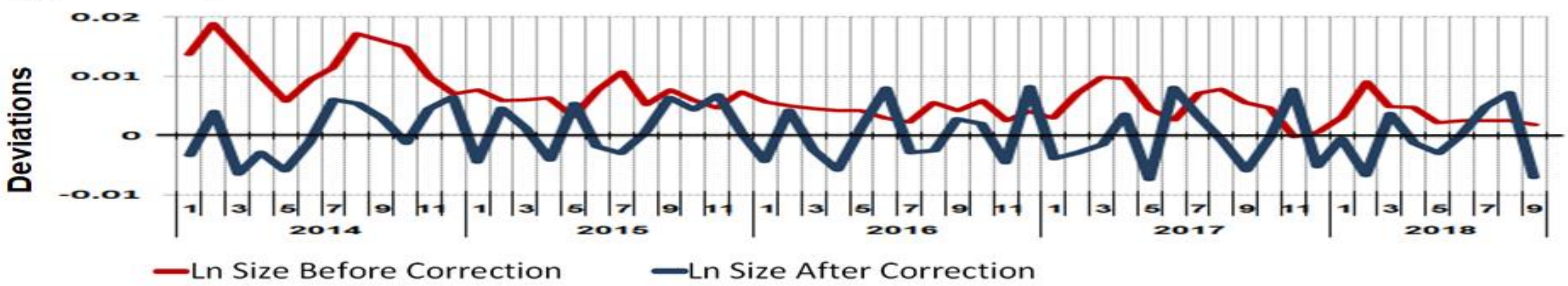
4

Provisional (known) versus nowcasted deviations for selected characteristics.

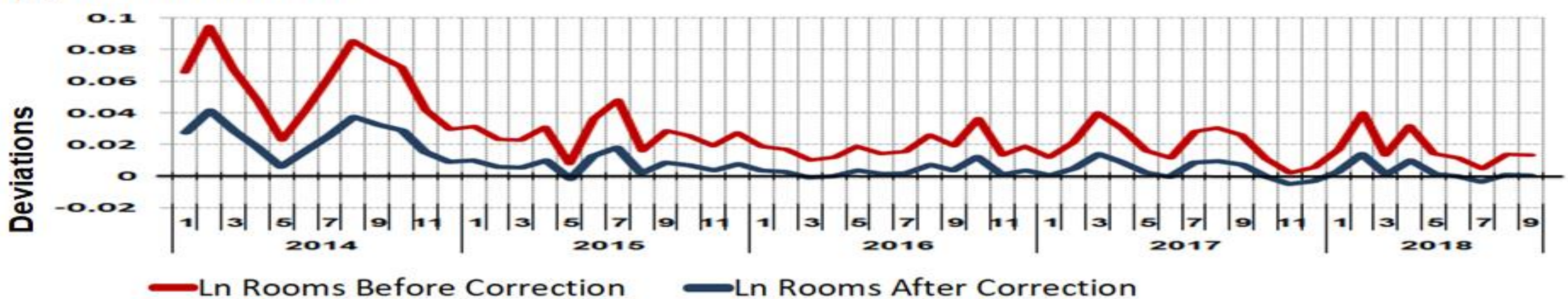
2014-2018



(a) Dwelling area

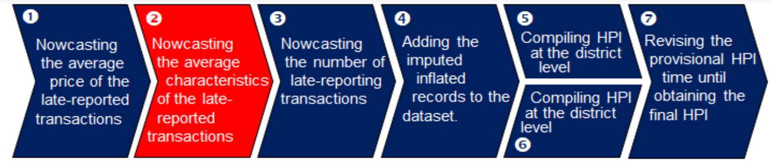


(b) Number of rooms



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.

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

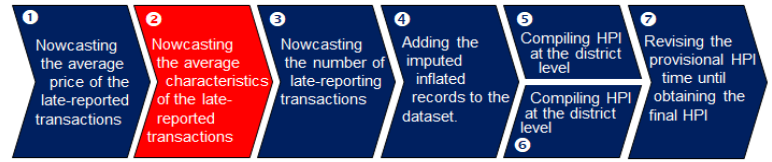


In the second stage, we nowcasted the monthly average price of the late-reported transactions at the sub-district level by fitting the following regression model:

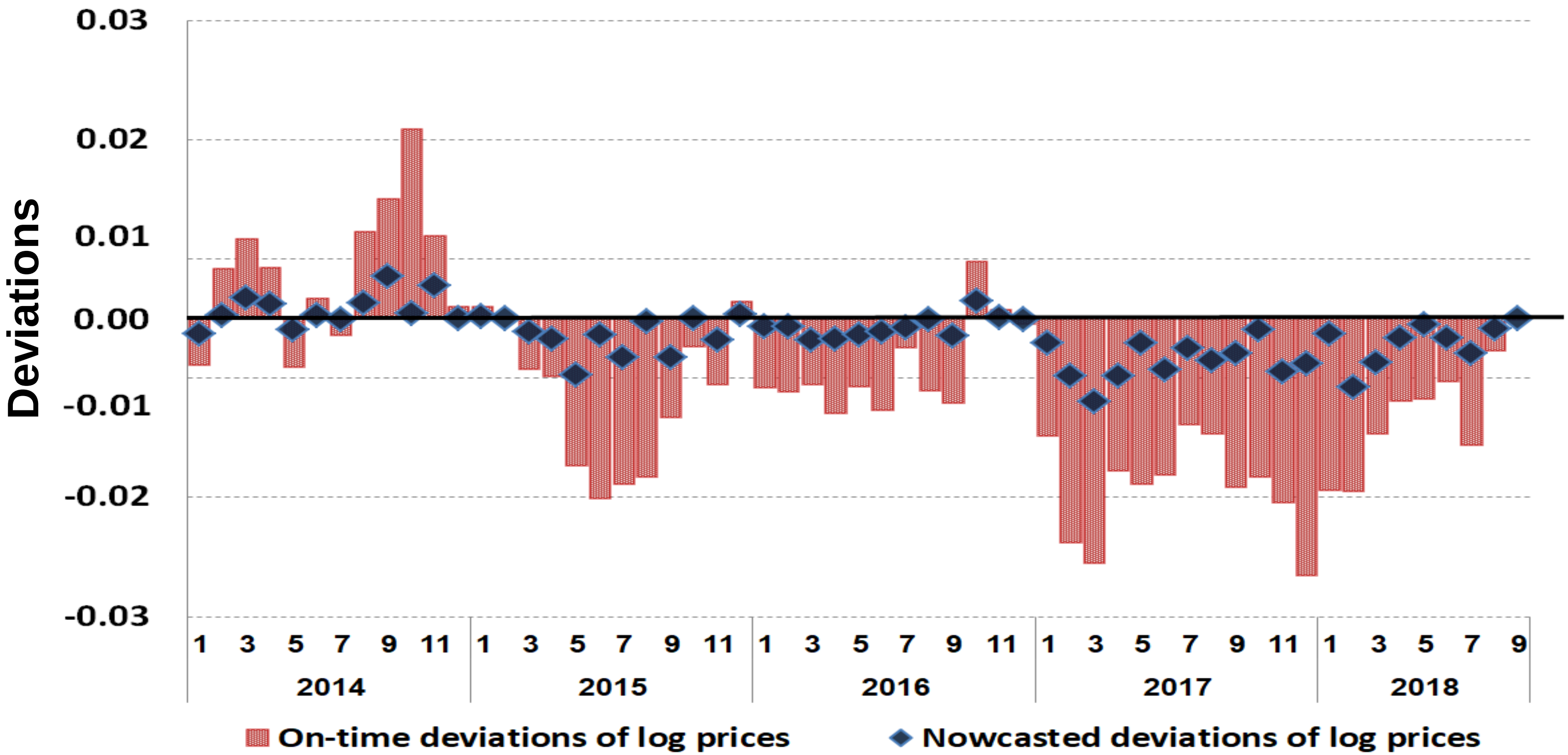
$$\bar{P}_{l} = \alpha_0 + \sum_{k=1}^{5} \beta_k \bar{Z}_{k,nl} + \delta \bar{P}_{nl} + \varepsilon_{l}$$

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

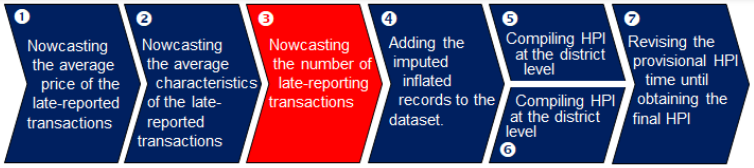
הלשכה המרכזית לסטטיסטיקה
Central Bureau of Statistics
دائرة الإحصاء المركزية



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



(c) Nowcasting the number of late-reported transactions

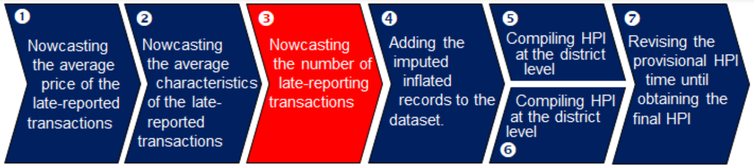


We used the following nowcasting model for predicting the number of late-reported transactions for any given sub-district, based on the past 6 months:

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

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.

(c) Nowcasting the number of late-reported transactions



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:

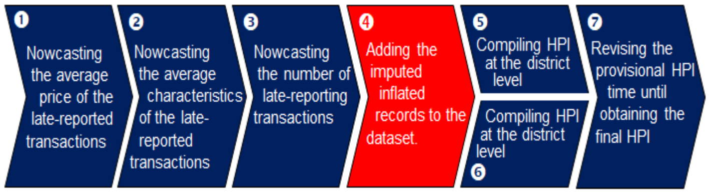
$$\widehat{\mathrm{N}}_{F,t} = \frac{\mathrm{N}_{t,t}}{\hat{P}_{t,t+6}}$$

where $\widehat{\mathrm{N}}_{\mathrm{F},t}$ 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.

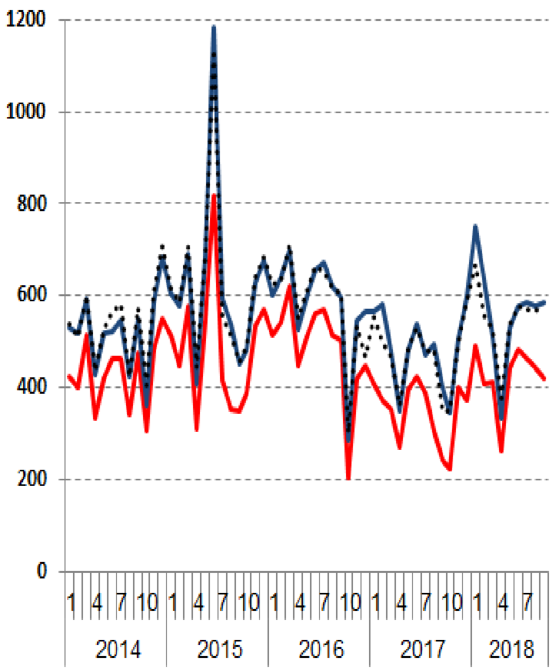
(c) Nowcasting the number of late-reported transactions

הלשכה המרכזית לסטטיסטיקה
Central Bureau of Statistics
دائرة الإحصاء المركزية

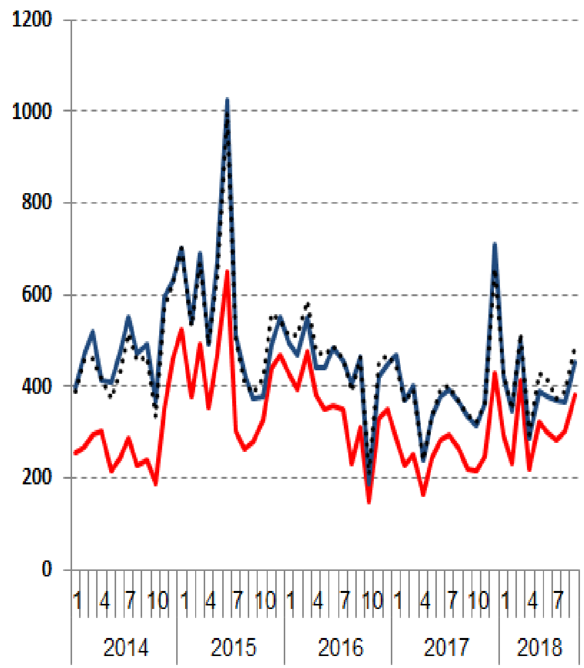
6 Number of on time-reported transactions, final number of transactions and nowcasted final number of transactions



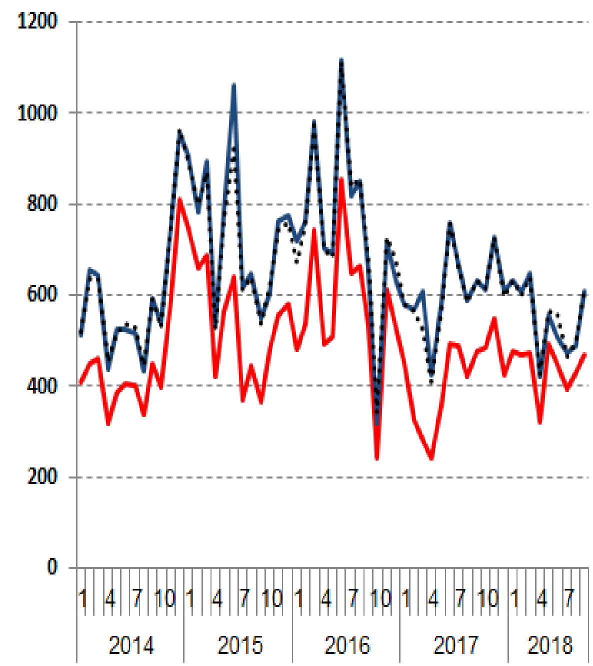
(a) Jerusalem Sub-District



(b) Tel Aviv Sub-District



(c) Petah Tiqwa Sub-District



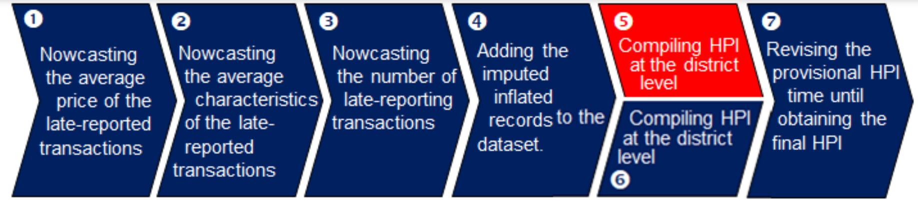
4. Empirical Results

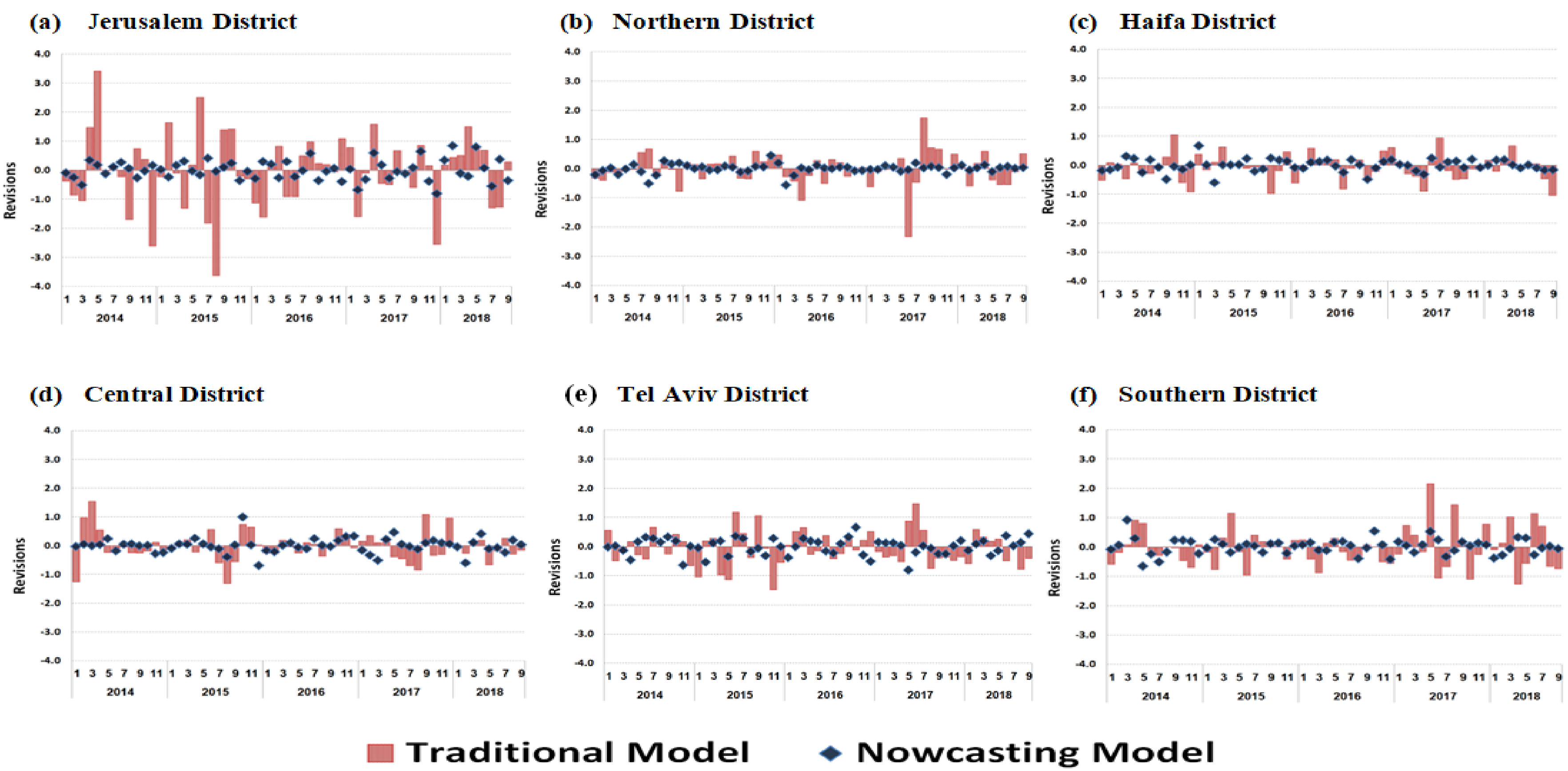
4.1 The magnitude of revisions at the district level

4.3 The magnitude of revisions at the national level

4.1 The magnitude of revisions at the district level

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

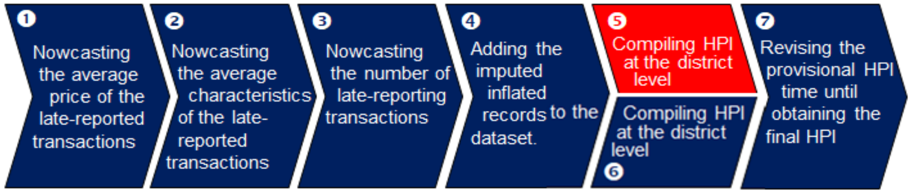




4.1 The magnitude of revisions at the district level

Central Bureau of Statistics
הלשכה המרכזית לסטטיסטיקה
دائرة الإحصاء المركزية

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



Revisions	*Jerusalem District*		*Northern District*		*Haifa District*	
	Trad.	NowC.	Trad.	NowC.	Trad.	NowC.
1st Provisional $\Delta P_{(\%)}$	**0.93**	**0.27**	**0.39**	**0.11**	**0.34**	**0.16**
2nd Provisional $\Delta P_{(\%)}$	**0.76**	**0.22**	**0.28**	**0.08**	**0.28**	**0.14**
3rd Provisional $\Delta P_{(\%)}$	**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 $\Delta P_{(\%)}$	**0.38**	**0.17**	**0.48**	**0.22**	**0.51**	**0.20**
2nd Provisional $\Delta P_{(\%)}$	**0.26**	**0.13**	**0.36**	**0.15**	**0.42**	**0.16**
3rd Provisional $\Delta P_{(\%)}$	**0.11**	**0.10**	**0.19**	**0.12**	**0.20**	**0.13**

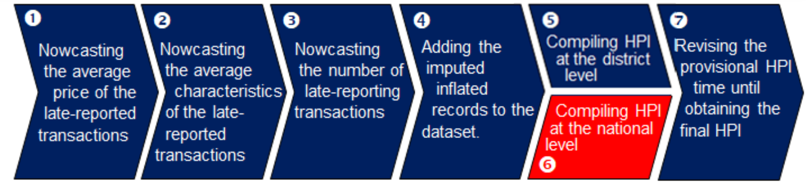
Notes: $\Delta P_{(\%)}$ = Price Change in percentage; Trad. = Traditional model (no nowcasting); NowC. = Nowcasting model;

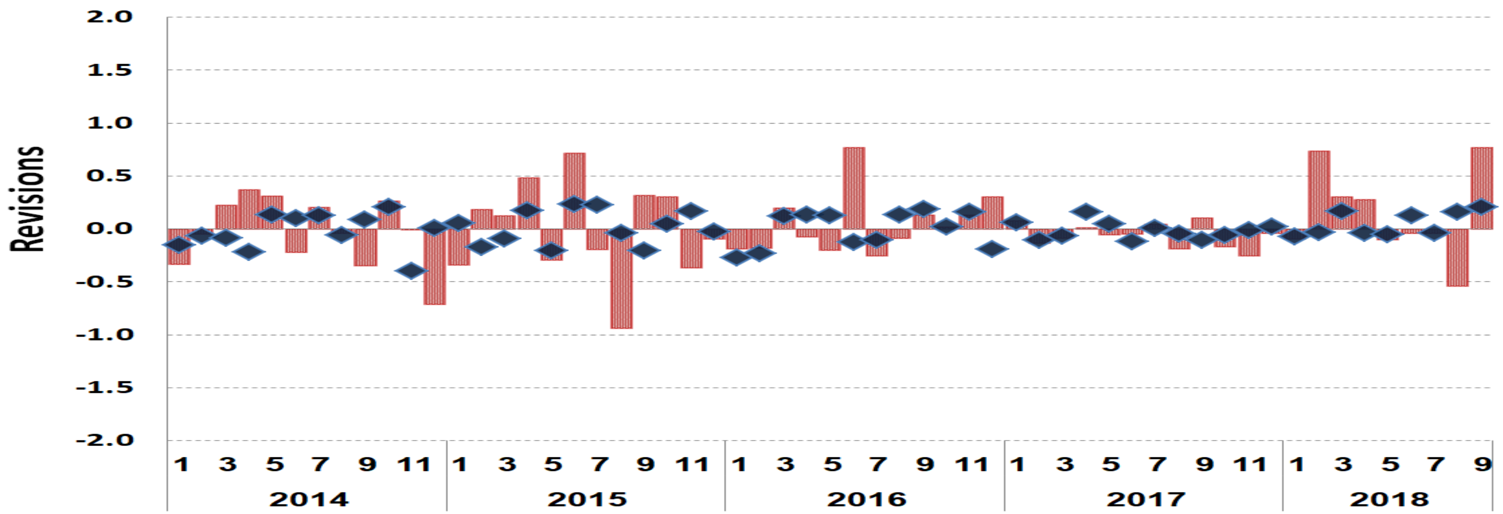
4.2 The magnitude of revisions at the national level

הלשכה המרכזית לסטטיסטיקה
Central Bureau of Statistics
دائرة الإحصاء المركزية

9

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





Revisions	*Trad. Model*	*NowC. Model*
1st Provisional $\Delta P_{(\%)}$	**0.25**	**0.12**
2nd Provisional $\Delta P_{(\%)}$	**0.18**	**0.09**
3rd Provisional $\Delta P_{(\%)}$	**0.10**	**0.07**

Notes: $\Delta P_{(\%)}$ = Price Change in percentage; Trad. = Traditional model (no nowcasting); NowC. = Nowcasting model;

Concluding Remarks

- 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)