

# Der österreichische Wohnimmobilienpreisindex: Methodische Verfeinerung 2017

Wolfgang Brunauer<sup>1</sup>,  
Wolfgang Feilmayr<sup>2</sup>,  
Karin Wagner<sup>3</sup>

Der österreichische Wohnimmobilienpreisindex wird mit dem dritten Quartal 2017 methodisch verfeinert und auf eine neue Datenbasis umgestellt. In diesem Beitrag werden die Auswirkungen dieser Änderung in mehrfacher Hinsicht überprüft: Zunächst wird die Doppel-Imputationsmethode mit der bisher verwendeten Dummy-Methode verglichen, danach der Effekt der Verfeinerung der Modelle von log-linearen Regressionsmodellen auf Generalisierte Additive Modelle und zuletzt die Auswirkung der Umstellung auf eine umfassendere Datenbasis analysiert. Damit soll untersucht werden, ob die Verknüpfung der bisherigen Datenreihen mit den neuen Indexreihen möglich und sinnvoll ist. Zudem wird der Frage nachgegangen, ob die gezeigten Modelle künftig für eine stärkere Regionalisierung der Indexreihen Anwendung finden können.

## 1 Vorgehensweise

Um die Effekte der methodischen Verfeinerung und der Datenumstellung voneinander zu entkoppeln, wird in folgenden Stufen vorgegangen:

- Zunächst wird auf Basis der bisher verwendeten Datengrundlage (Daten der Firma EDI-Real GmbH<sup>5</sup>) der Effekt der methodischen Umstellung vom sogenannten Dummy-Modell (Modell 0) auf das Doppel-Imputationsmodell (Modell 1) mit derselben Modellspezifikation/Variablenauswahl ermittelt.
- Danach wird der Effekt einer verfeinerten Modellspezifikation, nämlich generalisierter additiver Modelle (GAM) ermittelt (Modell 2).
- Schließlich wird der Effekt der neuen Datenbasis (JUSTIMMO<sup>6</sup>) auf die Modellergebnisse ermittelt. In Modell 3 wird dazu ein GAM-Modell angewandt und mit Doppel-Imputation ausgewertet; der einzige Unterschied zu Modell 2 ist die neue Datenbasis.

## 2 Daten und Modelle

Folgende Datenquellen werden verwendet:

- Da der Datenstand, welcher der ursprünglichen/historischen Indexkonstruktion zugrunde gelegt wurde, nicht mehr exakt reproduzierbar ist, wird der Index mit den derzeit verfügbaren Datensätzen der Firma EDI-Real neu berechnet. Das Datensample für die Modelle 0, 1 und 2 deckt den Zeitraum von Q1 07 bis Q1 17 ab. Das Datensample wird darüber hinaus noch dedupliziert und ausreißerbereinigt.
- Das Datensample für Modell 3 stammt von der Maklersoftware JUSTIMMO der B&G Consulting & Commerce GmbH. Diese Daten werden ab Q1 12 verwendet. Eine neue Datenbasis war notwendig geworden, da sich in den letzten Jahren zunehmend abgezeichnet hatte, dass die bis dahin verwendete Datengrundlage in einigen Segmenten zunehmend dünner wurde, was zu volatileren Ergebnissen führen konnte.

<sup>1</sup> DataScience Service GmbH, Wien, [wolfgang.brunauer@datascience-service.at](mailto:wolfgang.brunauer@datascience-service.at)

<sup>2</sup> DataScience Service GmbH und Department für Raumplanung, Technische Universität Wien, [wolfgang.feilmayr@tuwien.at](mailto:wolfgang.feilmayr@tuwien.at)

<sup>3</sup> Oesterreichische Nationalbank, Abteilung für volkswirtschaftliche Analysen, [karin.wagner@oenb.at](mailto:karin.wagner@oenb.at)

<sup>4</sup> Neben den Regionen Wien und Österreich ohne Wien sind künftig Publikationen für weitere Regionen geplant.

<sup>5</sup> [www.edireal.com](http://www.edireal.com)

<sup>6</sup> [www.justimmo.at](http://www.justimmo.at)

Für Modell 0 wird dabei wie folgt vorgegangen (vgl. Hill, 2011):

- Berechne für die beiden Objektarten Eigentumswohnungen (ETW) und Einfamilienhäuser (EFH; beinhaltet auch Reihenhäuser und Doppelhaushälften) für Modellperioden von jeweils einem Jahr (vier Quartale) Modelle mit folgender Spezifikation:

$$z = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + \delta_{t,1} D_{t,1} + \dots + \delta_{t,4} D_{t,4} + \varepsilon \quad (1)$$

Dabei wird das Modell jeweils mit der Modellperiode  $t$  bezeichnet, dem Endquartal einer Periode, also z. B. Q1 10<sup>7</sup> für das letzte Quartal der Modellperiode Q1 09, Q3 09, Q4 09 und Q1 10.<sup>7</sup> Diese Modelle werden jeweils getrennt für Wien und Österreich ohne Wien geschätzt. Für ETW wird weiter differenziert zwischen neuen Objekten (Alter kleiner oder gleich 3 Jahre) und gebrauchten Objekten.

In (1) wird der logarithmierte Preis  $z = \ln(p)$  modelliert durch  $n$  Objekteigenschaften  $X_i$ ,  $i \in \{1, \dots, n\}$  wie Fläche, Alter, Ausstattung, räumliche Indikatoren und die dafür geschätzten Effekte  $\beta_i$  einerseits sowie andererseits durch 0/1 codierte (Dummy-) Variablen  $D_{t,j}$ ,  $j \in \{1, \dots, 4\}$  für jede Modellperiode  $t$  mit den Effekten  $\delta_{t,j}$  für die Quartale, in denen die Preise der Objekte erhoben wurden.

- Der Preisindex für jede Modellperiode  $t$  wertet die Dummy-Effekte aus mit  $\exp(\delta_{t,j})$ . Da als Referenzquartal jeweils das erste Quartal jeder Modellperiode verwendet wird, ist  $\delta_{t,1} = 0$  und  $\exp(\delta_{t,1}) = 1$ .
- Relevant ist jeweils die Entwicklung vom vorletzten auf das letzte Quartal

einer Modellperiode  $t$ , da die Modelle sich jeweils um drei Quartale überlappen. Die Steigerungsrate für jede Modellperiode im Vergleich zum letzten Quartal der Vorperiode ist dann  $\Delta_t = 1 + \exp(\delta_{t,4}) - \exp(\delta_{t,3})$ .

- Das kumulative Produkt wird über alle Quartale gebildet, wobei das erste Quartal (das letzte der ersten Modellperiode, also Q4 07) auf 1 gesetzt wird.

Für das Modell 1 wird die sogenannte Doppel-Imputationsmethode angewandt, wie sie in Hill (2011) oder Eurostat (2013) beschrieben wird. Imputationsmethoden verwenden Laspeyres- und Paasche-Index-Formeln. Diese messen die Veränderung des Preises eines festen Warenkorbs (in unserem Fall von Immobilien) im Zeitverlauf. Hierfür ist das Vorhandensein des jeweiligen Preises jedes Elements des Warenkorbs in jeder Periode eine notwendige Voraussetzung. Im Immobilienkontext ist es nicht sinnvoll, Laspeyres- oder Paasche-Indizes auf Basis von tatsächlichen Transaktionen zu berechnen, da Immobilien selten und in unterschiedlichen Intervallen verkauft werden. Man verwendet trotzdem diese Indizes, ersetzt aber tatsächliche Transaktionsdaten mit imputierten Preisen. Damit ist es möglich, einen Preis für die Immobilie, für beide Perioden (aktuelle Periode  $t$  und Vorperiode  $s$ ) zu erhalten.

Dabei werden folgende Schritte vorgenommen:

- Berechne jeweils Modelle für  $t \in \{Q4\ 07, \dots, Q1\ 17\}$

(entspricht wie beschrieben jeweils dem letzten Quartal einer Modellperiode mit vier Quartalen) mit der Modellspezifikation (1).

<sup>7</sup> Um die Notation nicht allzu kompliziert zu gestalten, wird bei den Objekteigenschaften und räumlichen Indikatoren das Subskript  $t$  weggelassen, obwohl die Effekte selbstverständlich für jede Modellperiode variieren.

- Danach werden die Modellvorhersagen für jede Beobachtung imputiert, es wird also

$$\hat{z}_h = \hat{\beta}_1 + \sum_i \hat{\beta}_i X_{i,h} + \sum_j \hat{\delta}_j D_{j,h}$$

für alle Modellperioden berechnet (für die Immobilie  $h$ ). Da es das Ziel ist, quartalsweise Indexwerte zu berechnen, werden im zweiten Schritt alle Beobachtungen einer Modellperiode auf das letzte Quartal der Modellperiode ausgewertet. Das bedeutet, dass

$\hat{z}_h$  ersetzt wird durch

$$\tilde{z}_h = \hat{z}_h - \sum_j \hat{\delta}_j D_j + \hat{\delta}_4$$

oder in weiterer Folge vereinfacht

$$\hat{p}_{t,h}(z_{t,h})$$

- der geschätzte Preis der Immobilie  $h$  zum Zeitpunkt  $t$ .

- Danach wird mit diesen imputierten Werten der geometrische Paasche, der geometrische Laspeyres und daraus abgeleitet der Törnqvist-Index berechnet, wobei  $s = t - 1$  (Hill, 2011):

Geometrischer Paasche:

$$P_{s,t}^{GP} = \prod_{h=1}^{H_t} \left( \hat{p}_{t,h}(z_{t,h}) / \hat{p}_{s,h}(z_{s,h}) \right)^{1/H_t} \quad (2)$$

Geometrischer Laspeyres:

$$P_{s,t}^{GL} = \prod_{h=1}^{H_s} \left( \hat{p}_{t,h}(z_{s,h}) / \hat{p}_{s,h}(z_{s,h}) \right)^{1/H_s} \quad (3)$$

Törnqvist:

$$P_{s,t}^T = \sqrt{P_{s,t}^{GP} \times P_{s,t}^{GL}} \quad (4)$$

Beim geometrischen Paasche-Index werden also die Beobachtungen der „aktuellen“ Periode  $t$  ausgewertet, und zwar einmal mit dem Modell der Periode  $t$  und einmal mit dem Modell der Periode  $s$ . Aus dem Quotienten dieser Wertepaare wird das geometrische Mittel über alle Beobachtungen der Periode  $t$  gebildet. Für den geometrischen Laspeyres-Index wird analog vorgegangen, nur wird das Sample der Vorperiode  $s$  verwendet.

Um die Berechnung zu vereinfachen, wird jeweils der Logarithmus der rechten Seite der Gleichung genommen und das Ergebnis wieder entlogarithmiert. Anhand des geometrischen Paasche:

$$\ln(P_{s,t}^{GP}) = \left[ \sum_{h=1}^{H_t} \ln \left( \hat{p}_{t,h}(z_{t,h}) \right) - \sum_{h=1}^{H_t} \ln \left( \hat{p}_{s,h}(z_{s,h}) \right) \right] \times \frac{1}{H_t} \quad (5)$$

In Modell 2 wird Formel (1) ersetzt durch eine teilweise nichtlineare Spezifikation. Insbesondere werden die Quartale und die räumlichen Indikatoren mit Random Effects und die kontinuierlichen Variablen (wie etwa die Fläche und das Alter des Objekts) durch penalisierte Regressionssplines (vgl. Wood, 2006) modelliert. Wie in Brunauer et al. (2012) beschrieben, führt die Anwendung von Random Effects zu geringerer Volatilität und damit zu einem glatteren Indexverlauf, die Anwendung von penalisierten Regressionssplines zu höherer Präzision in der Schätzung. In Modell 3 schließlich wird dieselbe Modellspezifikation und Auswertemethode wie in Modell 2 verwendet, nur mit dem Datensatz von JUSTIMMO statt jenem von EDI-Real. Die Ergebnisse werden in Kapitel 3 grafisch dargestellt.

### 3 Analyse und Interpretation

Wir vergleichen nun die vier Ansätze, indem wir die Ergebnisse im Zeitverlauf abbilden. Zusätzlich zum Gesamtindex zeigen wir dabei insbesondere jene Teilindizes, die bisher in den Gesamtindex eingeflossen sind (Eigentumswohnungen und Einfamilienhäuser). Der Indexverlauf mit dem Suffix 0 beschreibt jeweils die aktuell eingesetzte Modellvariante (Dummy-Index) mittels der roten Linie, Modell 1 dasselbe log-lineare Modell, aber ausgewertet mit Törnqvist-Doppelimputation – dies in der grünen Linie. In Modell 2 wird das log-lineare Modell

Grafik 1

### Indexverläufe für Gesamtösterreich



durch ein generalisiertes additives Modell ersetzt (blaue Linie) und bei Modell 3 (orange Linie) schließlich der Datensatz von JUSTIMMO (statt EDI-Real) verwendet. Die Indexverläufe der Modelle 1 bis 3 sind jeweils so umbasiert, dass das Niveau des letzten Quartals jenem von Modell 0 entspricht.

In Grafik 1 für Gesamtösterreich zeigt sich, dass selbst die beiden Linien, bei denen die größten Unterschiede zu

erwarten wären – die rote Linie für das lineare Modell und die orange Linie für das nichtlineare Modell und die erweiterte Datenbasis – sehr ähnliche Verläufe zeigen. Das zeigt im Hinblick darauf, dass bei Modell 3 eine gänzlich andere Datenbasis herangezogen wurde, dass beide Datenbasen und Methoden zu sehr konsistenten Ergebnissen führen, und kann auch als positive Validierung der bisherigen Modellansätze betrachtet

Grafik 2

### Indexverläufe für Wien und Österreich ohne Wien

#### Wien



#### Österreich ohne Wien



werden. Über den gesamten Verlauf sieht man, dass alle vier Ansätze/Linien gleichlaufend von den Steigerungen bzw. Rückgängen sind. Der Dummy-Index zeigt den volatilsten Verlauf. Dies ist darin begründet, dass in diesem Modell noch keinerlei Imputationsme-

thoden verwendet wurden, die den Bias nicht berücksichtigter Variablen (Omitted-Variable Bias) reduzieren. In Grafik 1 – aber auch in den restlichen Grafiken – sieht man den etwas glatteren Verlauf im Vergleich zum Dummy-Index in Modell 1.

Grafik 3

### Indexverläufe für Wien - Eigentumswohnungen und Einfamilienhäuser

#### Eigentumswohnungen



#### Einfamilienhäuser



Grafik 4

### Indexverläufe für Österreich ohne Wien – Eigentumswohnungen und Einfamilienhäuser

#### Eigentumswohnungen



#### Einfamilienhäuser



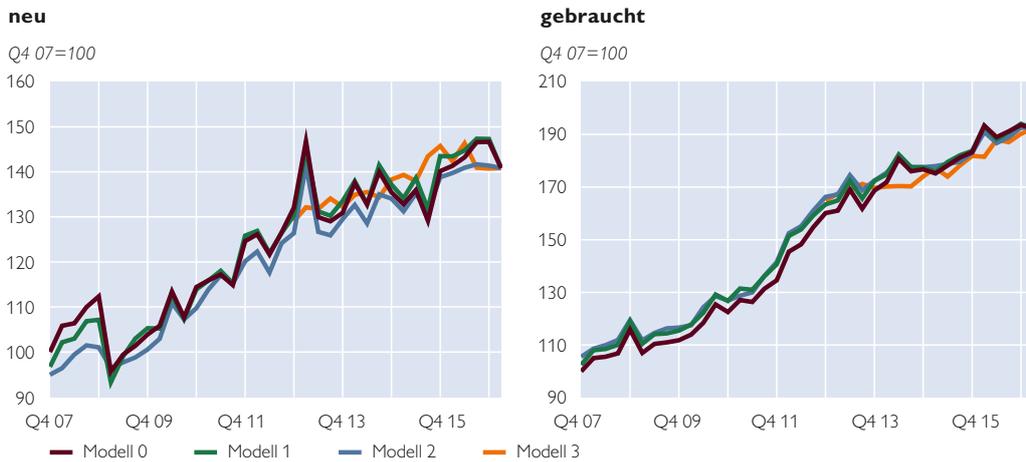
Grafik 2, linkes Panel, zeigt die Ergebnisse für Wien, während das rechte Panel die Vergleiche bei den Verläufen für Österreich ohne Wien darstellt. Es sind jeweils Wohnungen und Häuser erfasst. Wien zeigt geringere Unterschiede in den Modellen als das restliche Bundesgebiet. Der Unterschied in den Grafiken zwischen der orangen (Modell 3) und der blauen Linie (Modell 2) – also die Anwendung des gleichen Modells mit geänderter Datenbasis – ist

im rechten Panel von Grafik 2 größer. Grund dafür dürfte sein, dass die Datenmenge vor allem im restlichen Bundesgebiet bei den EDI-Real Daten geringer ist als bei den JUSTIMMO-Daten.

Grafik 3 zeigt Eigentumswohnungen (linkes Panel) und Einfamilienhäuser (rechtes Panel) für Wien. Die Einfamilienhäuser zeigen einen weit volatileren Verlauf bei allen Modellen im Vergleich zu den Wohnungen – dem weitaus

Grafik 5

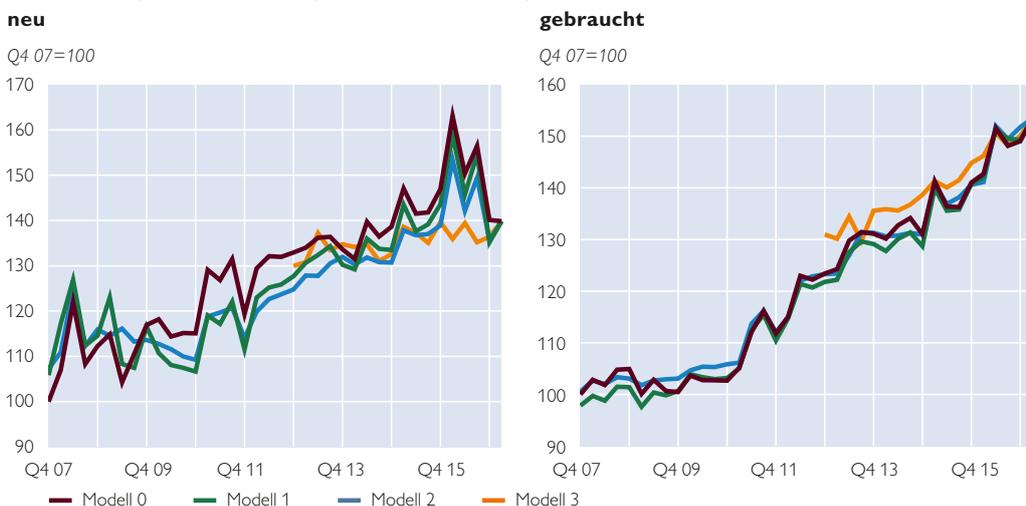
### Indexverläufe für Wien – neue und gebrauchte Eigentumswohnungen



Quelle: DSS, OeNB.

Grafik 6

### Indexverläufe für Österreich ohne Wien – neue und gebrauchte Eigentumswohnungen



Quelle: DSS, OeNB.

größten Marktsegment in der Bundeshauptstadt. Betrachtet man die vier Modelle in Restösterreich etwas genauer, so sieht man, dass die Modelle 0, 1 und 2 ähnlichere Verläufe zeigen als Modell 3. Hier wird der erwähnte Unterschied bei der Datenmenge besonders deutlich. Modell 3 mit den JUSTIMMO-Daten hat den weit glatteren Verlauf.

In Grafik 4 sind, analog zu Grafik 3, die Verläufe der Eigentumswohnungs- und Einfamilienhauspreise im restlichen Bundesgebiet zu sehen. Bei den Wohnungen ist der Zuwachs zu Jahresbeginn 2013 deutlich sichtbar. Gerade bei dieser Grafik zeigt sich im rechten Panel, dass das Modell 3 (orange Linie) das robustere Modell ist und von „dünnen“ Samplegrößen nicht beeinflusst wird.

Schlussendlich wird in Grafik 5 sichtbar, dass die Preissteigerungen 2012/13 stärker bei den neuen als bei den gebrauchten Wohnungen zu verzeichnen waren, und zwar bei allen Modellansätzen. Auch in Grafik 6 zeigt das linke Panel den volatilere Verlauf. Im rechten Panel wird beim Vergleich der blauen mit der orangen Linie wieder deutlich, wie die größere Datenmenge wirkt – es sind die Level-Verläufe sehr ähnlich, aber weniger volatil.

#### 4 Zusammenfassung und Schlussfolgerungen

Dieses Papier verfolgt drei Ziele: Die Auswirkungen der methodischen Umstellung von der sogenannten Dummy-Methode auf die Doppel-Imputationsmethode zu ermitteln, den Effekt einer Verfeinerung der Modelle von log-linearen Regressionsmodellen auf Generalisierte Additive Modelle (GAM), und schließlich die Auswirkungen der Umstellung auf eine neue Datenbasis. Tatsächlich zeigt sich ein etwas abweichender Verlauf bei Anwendung der Doppel-

Imputationsmethode im Vergleich zur Dummy-Methode, was auch darauf zurückzuführen sein dürfte, dass bei Ersterer der Omitted-Variable Bias reduziert wird. Die Anwendung von GAM führt zu einem etwas glatteren und plausibleren Verlauf und scheint sich daher positiv auf die Robustheit der Indexverläufe auszuwirken. Besonders deutlich wird dies bei Teilaggregaten mit relativ geringer Samplegröße (z.B. Eigentumswohnungen in Österreich ohne Wien). Schließlich zeigt sich, dass die Anwendung der neuen, umfassenderen Datenbasis von JUSTIMMO auch in diesen Teilaggregaten zu noch glatteren und plausibleren Indexverläufen führt. Andererseits ist die Abweichung der auf dieser Datenbasis berechneten Indizes im Vergleich zu den bisher verwendeten Daten in den Hauptaggregaten gering. Zudem stellt das Imputationsmodell zusammen mit der neuen Datenbasis eine gelungene Validierung der bisherigen Ansätze<sup>8</sup> dar. Eine Verknüpfung der bisherigen Datenreihen mit den neuen Indexreihen erscheint daher gerechtfertigt. Der Index wird künftig in besserer methodischer und datenmäßiger Qualität verfügbar sein.

Ein weiterer Vorteil der neuen Methode der Doppel-Imputation ist die Möglichkeit, stärker zu regionalisieren. Es ist daher geplant, in Zukunft neben den bisherigen Regionen Wien und Österreich (ohne Wien) auch Indizes für weitere Regionen zu berechnen. Diese bilden die notwendige Basis für Finanzmarktstabilitätsanalysen zu den Entwicklungen am Immobilienmarkt. Die Ergebnisse der regionalisierten Teilaggregate könnten zudem dafür verwendet werden, den OeNB-Fundamentalpreisindikator (Schneider, 2013) weiter zu disaggregieren und zu verfeinern.

<sup>8</sup> In Kooperation mit der TUWien.

## Literaturverzeichnis

- Brunauer, W., W. Feilmayr und K. Wagner. 2012.** A Residential Property Price Index for Austria. OeNB. Statistiken – Daten & Analysen Q3/12. 90–102.
- Eurostat. 2013.** Handbook on Residential Property Prices Indices (RPPIs).
- Hill, R. 2011.** Hedonic Price Indexes for Housing. OECD Statistics Working Papers 2011/01. OECD Publishing. Paris.
- Mundt, A. und K. Wagner. 2017.** Regionale Wohnungspreisindizes in Österreich – erste Erkenntnisse auf Basis hedonischer Modelle. Statistiken – Daten & Analysen Q1/17. 28–47.
- Schneider, M. 2013.** Are Recent Increases of Residential Property Prices in Vienna and Austria Justified by Fundamentals? In: Monetary Policy and the Economy Q4/13. OeNB. 29–46.
- Wood, S. N. 2006.** Generalized Additive Models: An Introduction to R. Chapman & Hall/CRC. Boca Raton.