

Neuronale Netzwerke vs hedonische Bewertungsverfahren von Schweizer Wohnimmobilien

Swiss Real Estate Congress, 20. Mai 2022

Dr. Rudolf Marty, Senior Researcher

Swiss Real Estate Institute, rudolf.marty@swissrei.ch



Zu untersuchende Fragen

- Forschungsfrage I: Wie relevant sind die im SRED-Datensatz vorliegenden Objektmerkmale und ausgewählte Umgebungsvariablen für die Prognosen der Transaktionspreise?
- Forschungsfrage II: Lässt sich die Prognosegüte eines log-linearen hedonischen Modelles mit Hilfe einer Methode des Machine Learning (d.h. mittels eines künstlichen neuronalen Netzwerkes) verbessern?

Das log-lineare hedonische Modell

- Das log-lineare Modell beschreibt den Logarithmus des Preises der i -ten Immobilie, $\ln(P_i)$, als Linearkombination der Logarithmen von k (quantitativen) Objekt- bzw. Umgebungsvariablen $\ln(X_{j,i})$:

$$\ln(P_i) = \sum_{j=1}^k \beta_j \ln(X_{j,i}) + u_i \quad (1)$$

mit:

- P_i : Transaktionspreis des i -ten Objektes
- $X_{j,i}$: j -tes Objektmerkmal des i -ten Objektes
- β_j : Sensitivität des j -ten Merkmals in Bezug auf Transaktionspreis (Elastizität)
- u_i : unsystematischer Einfluss auf den Preis des i -ten Objektes, unkorreliert mit Preis des k -ten Objektes

Das künstliche neuronale Netzwerk (ANN) für Objektpreise

- Ein ANN-Modell beschreibt den Logarithmus des Preises der i -ten Immobilie, $\ln(P_i)$, als Linearkombination von M nicht-beobachtbaren "versteckten Schichten" ("hidden layers") $g(v_m)$, $m=1, \dots, M$:

$$\ln(P_i) = g_0 + \sum_{m=1}^M g_m \left(\sum_{j=1}^k \alpha_j \ln(X_{j,i}) + \alpha_0 \right) \quad (2)$$

mit:

- g_m , $m=1, \dots, M$: nicht-lineare Aktivierungsfunktion (z.B. sigmoid-Funktion $g_m = 1/(1 + v_m)$)
- $X_{j,i}$, $j = 1, \dots, M$: j -tes Objektmerkmal des i -ten Objektes
- α_j , $j=0, \dots, k$: zu schätzende Parameter ("weights" und "bias"-Parameter) mittels Algorithmus (z.B. gradient descent)

Verwendete Daten

- Eigentumswohnungen: Transaktionspreise und 10 (quantitative und qualitative) Objektmerkmale, gesamte Schweiz, N=156'000, Periode Transaktion: 2000 Q1-2020 Q4; Quelle: SRED-Datenbank
- Einfamilienhäuser: Transaktionspreise und 11 (quantitative und qualitative) Objektmerkmale, gesamte Schweiz, N=118'000, Periode: 2000 Q1 - 2020 Q4; Quelle: SRED-Datenbank
- Kantonale Arbeitslosenquoten (Umgebungsvariable), Periode: 2000 Q1 - 2020 Q4; Quelle: SECO

log-lineares hedonisches Modell, schrittweise Regression

Eigentumswohnung			Einfamilienhaus		
Merkmal	PLZ	Bezirk	Merkmal	PLZ	Bezirk
1a Postleitzahl (PLZ)	.77/-1.41	-	1a PLZ (Min./Max.)	1.0/-1.66	-
1b Bezirk (Min./Max.)	-	.62/-.44	1b Bezirk	-	0.76/-.67
2 JahrTransaktion	0.04 ¹	0.04 ¹	2 JahrTransaktion	0.03 ¹	0.03 ¹
3 log(Nettowohnfläche)	0.84 ¹	0.88 ¹	3 log(Kubatur)	0.42 ³	0.45 ¹
4 AusbauObjekt	0.08 ¹	0.10 ¹	4 log(Grundstückfl.)	0.18 ¹	0.16 ¹
5 QualitätMikrolage	0.09 ¹	0.11 ¹	5 QualitätMikrolage	0.10 ¹	0.11 ¹
6 ZustandObjekt	0.07 ¹	0.07 ¹	6 AusbauObjekt	0.08 ¹	0.09 ¹
7 log(ZahlNasszellen)	0.09 ¹	0.12 ¹	7 AlterObjekt	-0.11 ¹	-0.09 ¹
8 AlterObjekt	-0.04 ¹	-0.02 ¹	8 ZustandObjekt	0.05 ¹	0.04 ¹
9 ErstZweitDomizil	0.09 ¹	0.17 ¹	9 log(ZahlNasszellen)	0.09 ¹	0.12 ¹
10 log(ZahlGaragenp.)	0.06 ¹	0.06 ¹	10 log(ZahlZimmer)	0.09 ¹	0.09 ¹
11 log(ZahlZimmer)	0.07 ¹	0.01 ¹	11 ArbQuoteKanton	-0.02 ¹	-0.02 ¹
12 ArbQuoteKanton	-0.01 ¹	-0.01 ¹	13 log(ZahlGaragenp.)	0.03	0.04 ¹
13 -	-	-	13 ErstZweitDomizil	0.03 ¹	0.08 ¹
R ²	0.8971	0.8386	R ²	0.8375	0.7739
¹ bzw. ² : Signifikanz 1 bzw. 10 Prozent			Quelle: SRED, Seco		

Schätzung ANN für Objektpreise

- 1. Schritt: Standardisierung des Datensatzes X mit Funktion $X_s = (X - X_{min}) / (X_{max} - X_{min})$ (X_s liegt zwischen -1 und 1)
- 2. Schritt: Festlegung Zahl der "Knots" und Aktivierungsfunktion ("soft max") - Zahl der zu schätzenden Parameter:
Objektmerkmale mal Knots (bei EFH: 2999 mal 6: knapp 18'000 Parameter , bei EGTW: 2430 mal 6: 14'480 Paramter)
- 3. Schritt: Algorithmus, der Summe der Abstandsquadrate hinsichtlich der zu schätzenden Parameter (d.h. hinsichtlich der "biases" und "weights") minimiert
- Bemerkungen: Rechenaufwand für 3. Schritt nicht zu unterschätzen!

Prognosevergleich Hedon-Modell vs. ANN (1)

ANN und PLZ: generieren genauere Preisschätzungen als hedonisches Modell mit Bezirken

■ Prognosefehler-Statistiken mit Testdatensatz::

Teststatistik	ANN				log-lin. hedonisches Modell			
	EGTW, N=38'925		EFH, N=29'936		EGTW, N=38'925		EFH, N=29'936	
-	PLZ	Bezirk	PLZ	Bezirk	PLZ	Bezirk	PLZ	Bezirk
<i>RMSE</i> ¹	0.2189	0.1732	0.1934	0.2385	0.2004	0.2343	0.2250	0.2546
<i>MAE</i> ²	0.1302	0.1647	0.1380	0.1719	0.1516	0.1785	0.1617	0.1866
Innerhalb10P ³	0.4971	0.4039	0.4920	0.4046	0.4267	0.3698	0.4267	0.3644
Innerhalb20P ⁴	0.7940	0.6950	0.7793	0.6894	0.7311	0.6521	0.7188	0.6469

¹: Wurzel des Durchschnitts der quadrierten Prognosefehler

²: Durchschnitt der absoluten Prognosefehler

³: Häufigkeit[|relativer Prognosefehler|] < 10 Prozent

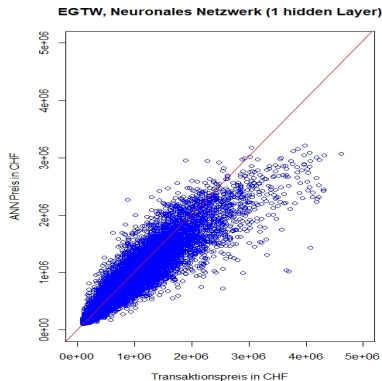
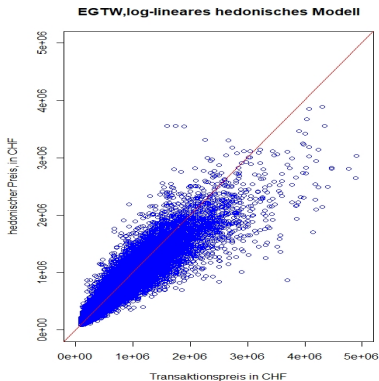
⁴: Häufigkeit[|relativer Prognosefehler|] < 20 Prozent

Quelle: SRED, Seco

Prognosevergleich Hedo-Modell vs. ANN (2)

Hedonisches Modell: unverzerrte Prognose, ANN-Modell: leicht verzerrte Prognose

- Regression Transaktionspreise auf theoretische Preise, Testdaten



Fazit: Präzisere Preisprognosen durch Verwendung von PLZ als Standortvariable und ANN-Verfahren

- Relevanz der Objektmerkmale in Bezug auf Transaktionspreis:
 - Erstaunlicherweise Transaktionsjahr zweitwichtigstes Objektmerkmal (Grund: Preismomentum)!
 - Umgebungsvariable: kantonale Arbeitslosenquote relativ wenig relevant hinsichtlich der Transaktionspreise
 - Standortvariable: hat mit Abstand höchsten Erklärungsbeitrag hinsichtlich Transaktionspreis, muss aber genau erfasst werden (Verwendung von PLZ statt Bezirk kann Teststatistiken "Innerhalb10Prozent" bzw. "Innerhalb20Prozent" bis zu 10 Prozentpunkte erhöhen).
- Methodik: Verwendung von ANN-Verfahren statt hedonischer Methode führt zu 5-7 Prozentpunkte höheren Teststatistiken "Innerhalb10Prozent" bzw. "Innerhalb20Prozent".

Fragen und Diskussion

- Fragen?
- Besten Dank für Ihre Aufmerksamkeit!

Literatur

-  Hastie, T., Tibshirani, R., Friedman, J. (2008). *"The Elements of Statistical Learning: Data Mining, Inference and Predictions"*. Stanford: Springer Series in Statistics.
-  Sconamiligo, D., Bourasse, S., Hoesli, M. (2019). *"Estimation and updating methods for hedonic valuations"*. *Journal of European Real Estate Research Vol 12 No 1, S. 134-150.*
-  Sirmans, G., Macpherson, D. (2015). *"The composition of hedonic pricing models"*. *Journal of Real Estate Literature Vol 13 No 1, S.3-43.*
-  Stubnova, M., Urbanikova, M., Hudakova, J., Papcunova, V. (2020). *"Estimation of Residential Property Market Price: Comparison of Artificial Neuronal Network and Hedonic Pricing Model"*. *Emerging Science Journal, S. 530-538.*