

What determines the home-price in Korean Housing Market?

Siwon Ryu, Yeong Mi Jeong

May 16, 2018

Outline

1. Motivation and Goal
2. Literature review
3. Data
4. Double selection method(Belloni at al.(2014))
5. Preliminary result
6. Reference

Motivation and Goal

- Finding factors determining home-price in Korean Housing Market using Machine Learning technique.
 - Factors:
Dependence on neighborhood, Cross-sectional dependency, Policy Change, Construction business fluctuation, Regime change, Natural disaster, Fluctuation on local business,...
- Try various ML techniques and compare each results when finding factors.
 - Support Vector machine, Elastic Lasso, LAR, and so on.

Literature review (1)

- Fan et al.(2011)
 - High-dimensional problem arises easily from VAR models. A p -dimensional time series with d lags gives $d \cdot p^2$ autoregressive parameters.
 - Let y_t^i be the home-price appreciation(HPA) in county i . Set forecasting model as follows

$$y_{t+s}^i = \sum_{j=1}^p b_{ij} y_t^j + X_t \beta_i + \varepsilon_{t+s}^i, \quad i = 1, \dots, p$$

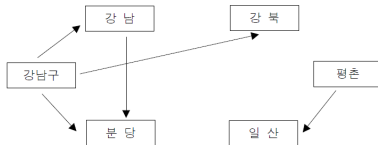
- Estimate the model of each city

$$\min_{\{b_{ij}, \beta_i\}} \sum_{t=1}^{T-s} \left(y_{t+s}^i - \sum_{j=1}^p b_{ij} y_t^j - X_t \beta_i \right)^2 + \lambda \sum_{j=1}^N w_{ij} p_{\lambda}(|b_{ij}|)$$

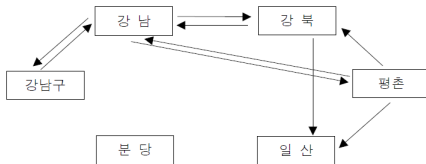
where $p(\cdot)$ is penalty, w_{ij} are weights chosen according to the geographical distances between counties.

Literature review (2)

- 왕세종(2004) : Test Geweke, and Granger causality between regions by using VAR model.
 - Pre currency crisis periods

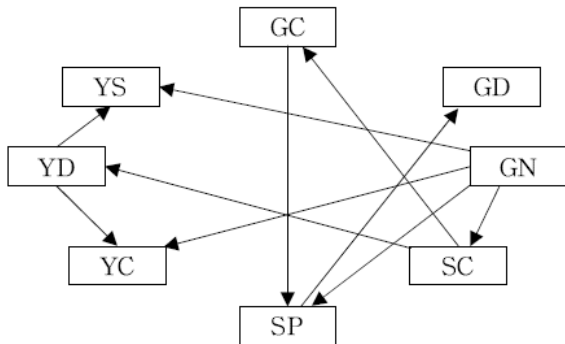


- Post currency crisis periods



Literature review (3)

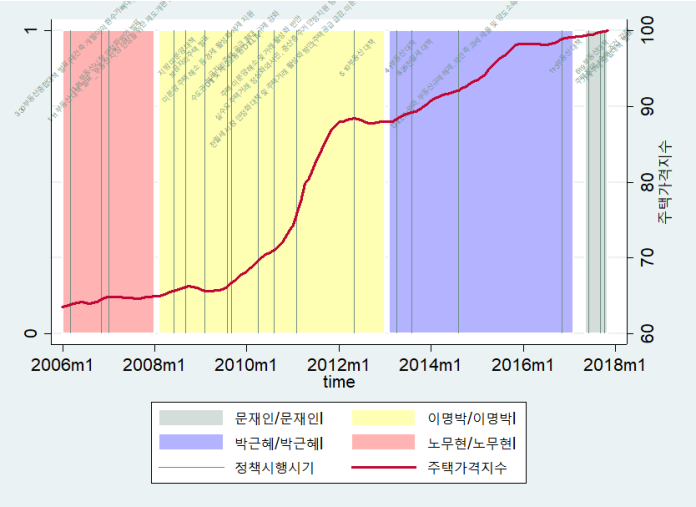
- 박해선(2014) : Test contemporaneous causality by using DAG method.



Data

- Real Apartment-Transaction Data of *Ministry of Land, Infrastructure and Transport*.
 - The data includes address, area, transaction price, transaction date.
- Home price index - *Statistics Korea*.
- Analysis period : 2006.01 ~ 2017.12
- Macro economic data : GDP, BOK basemoney rate *BOK economic statistics system*.
- Dummy variables of Policy period, and period of government.

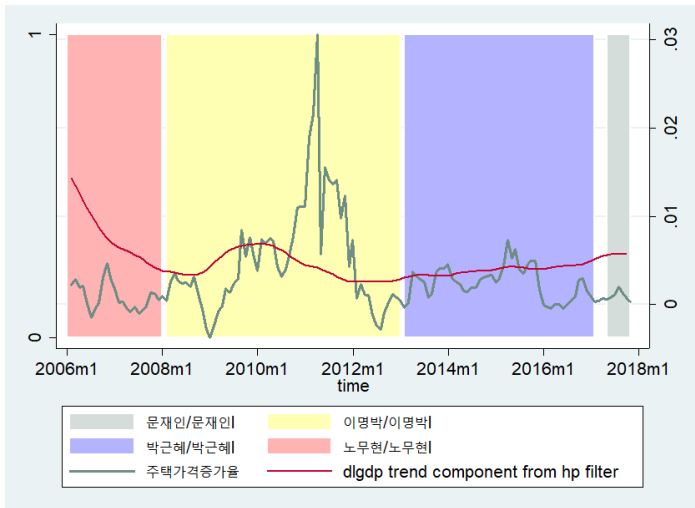
Data Plot



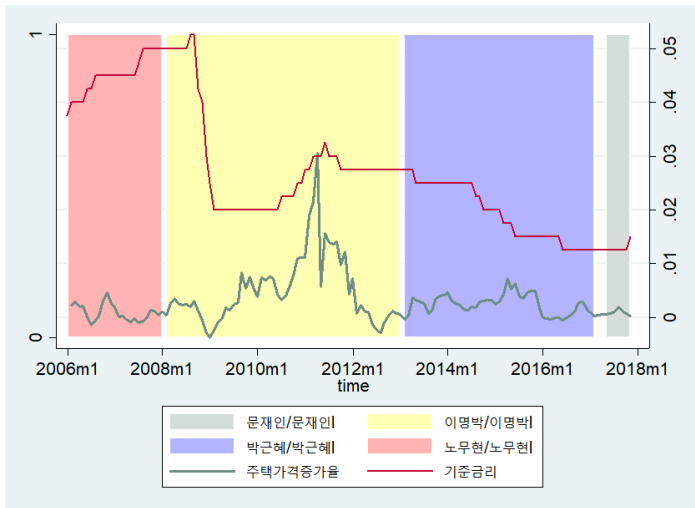
Data Plot



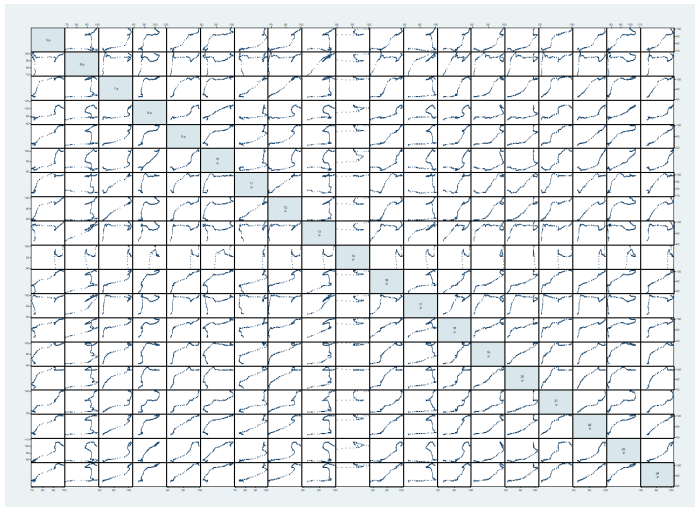
Data Plot



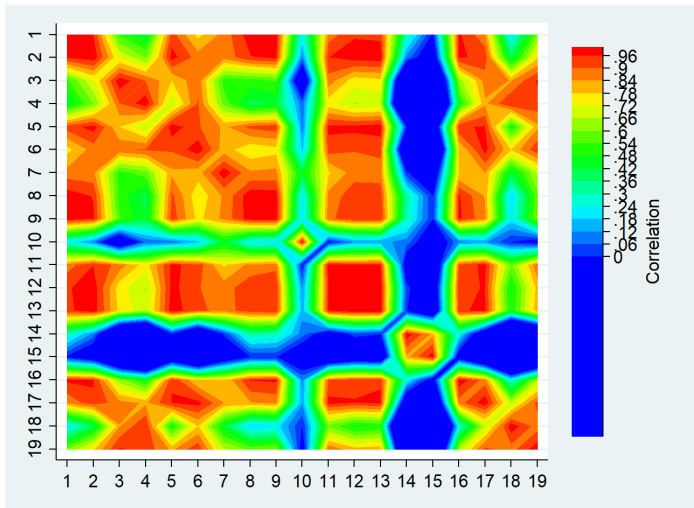
Data Plot



Data Plot



Data Plot



Double selection method(Belloni at al.(2014))

- First stage

In the first step, we select a set of control variables that are useful for predicting the treatment d_i

$$d_i = m(z_i) + v_i \rightarrow \hat{l}_1 : \text{selected control terms}$$

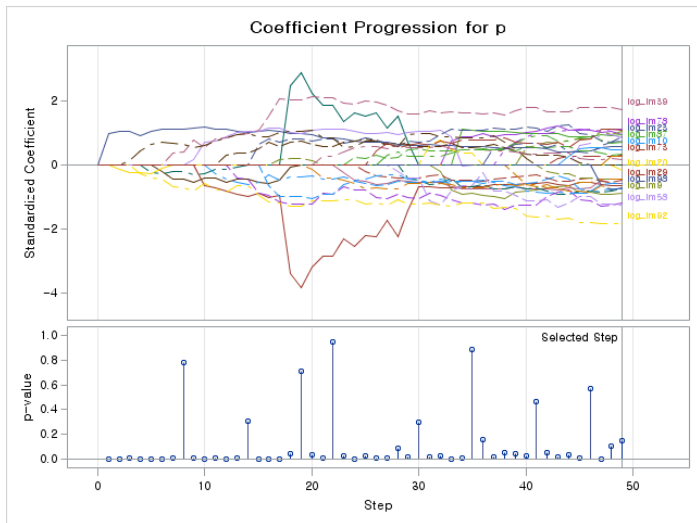
- Second stage

In the second step, we select additional variables by selecting control variables that predict y_i

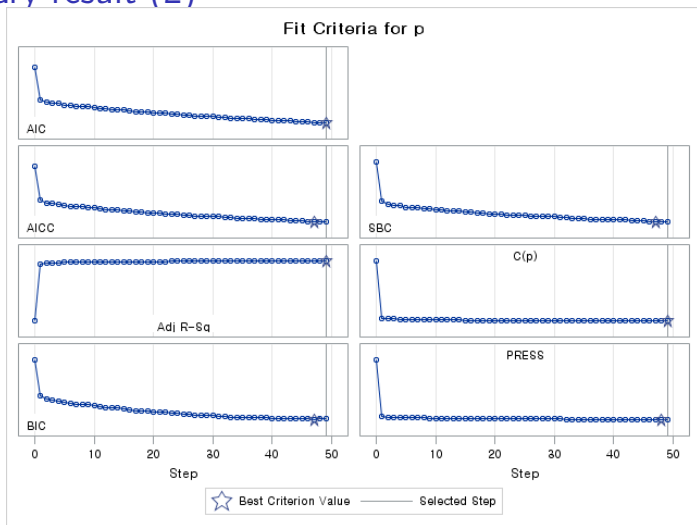
$$y_i = d_i\alpha_0 + g(z_i) + \zeta_i \rightarrow \hat{l}_2 : \text{selected control terms}$$

Preliminary result (1)

- First stage



Preliminary result (2)

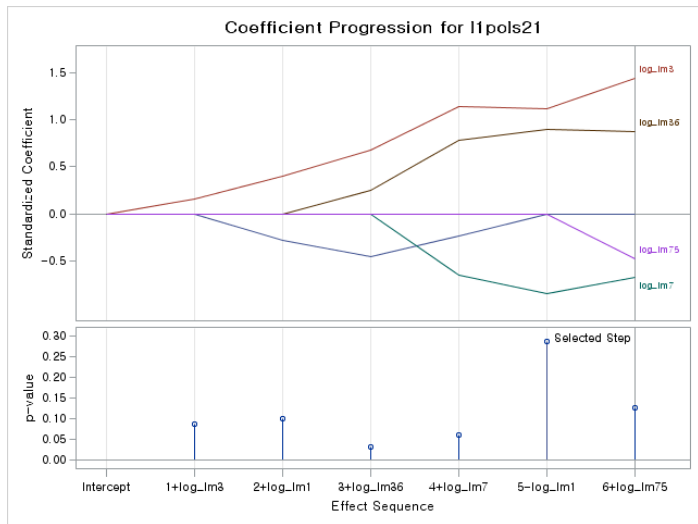


Preliminary result (3)

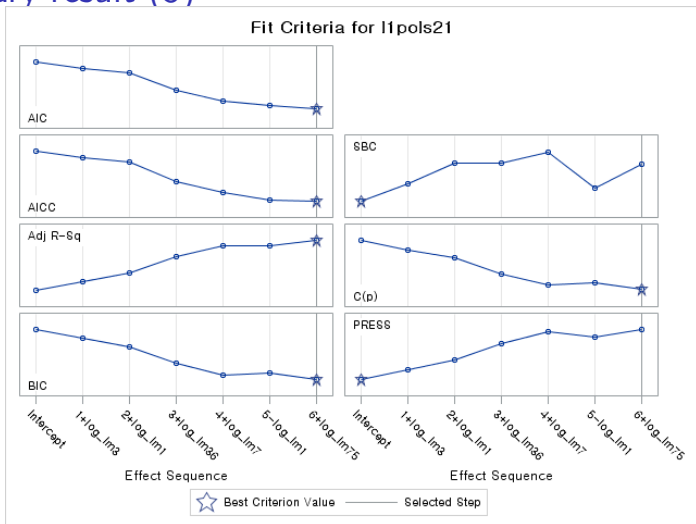
Parameter Estimates									
Parameter	DF	Estimate	Standard Error	t Value	Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	-4.149055	0.71541	-5.8	log_lm39	1	1.806263	0.221678	8.15
log_lm1	1	0.553774	0.068343	8.1	log_lm45	1	-0.489084	0.114433	-4.27
log_lm3	1	-0.869642	0.17579	-4.95	log_lm48	1	0.558536	0.212499	2.63
log_lm7	1	1.414009	0.289088	4.89	log_lm58	1	-0.537759	0.121984	-4.41
log_lm9	1	-1.193436	0.266656	-4.48	log_lm67	1	0.461882	0.155833	2.96
log_lm10	1	0.605386	0.172898	3.5	log_lm73	1	0.467901	0.131402	3.56
log_lm13	1	-0.718483	0.149074	-4.82	log_lm78	1	1.200275	0.205546	5.84
log_lm15	1	0.333051	0.197573	1.69	log_lm79	1	1.013621	0.20667	4.9
log_lm17	1	1.129545	0.164851	6.85	log_lm81	1	-0.976032	0.162091	-6.02
log_lm20	1	-0.197904	0.135321	-1.46	log_lm82	1	0.288946	0.150045	1.93
log_lm23	1	0.426274	0.114875	3.71	log_lm85	1	-0.531579	0.135944	-3.91
log_lm29	1	-0.532185	0.122584	-4.34	log_lm87	1	0.974148	0.197617	4.93
log_lm31	1	-0.774942	0.151994	-5.1	log_lm91	1	0.474927	0.124896	3.8
log_lm33	1	-1.079304	0.23172	-4.66	log_lm92	1	-1.25638	0.138379	-9.08
log_lm37	1	-0.388103	0.088038	-4.41	log_lm98	1	-0.848266	0.287318	-2.95
log_lm38	1	-0.643703	0.211903	-3.04	log_lm100	1	1.22315	0.248635	4.92

Preliminary result (4)

- Second stage



Preliminary result (5)



Preliminary result (6)

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	-3.10884	1.102335	-2.82
log_lm3	1	1.845182	0.491501	3.75
log_lm7	1	-0.957	0.445988	-2.15
log_lm36	1	0.563708	0.18215	3.09
log_lm75	1	-0.77363	0.501398	-1.54

Preliminary result (7)

- Third stage

Parameter Estimates						Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-3.92568	0.8102	-4.85	< .0001	log_lm45	1	-0.50149	0.11693	-4.29	< .0001
l1pols21	1	-0.01611	0.00871	-1.85	0.0677	log_lm48	1	0.51551	0.2312	2.23	0.0283
log_lm1	1	0.56194	0.06947	8.09	< .0001	log_lm58	1	-0.56976	0.12549	-4.54	< .0001
log_lm3	1	-0.83379	0.22179	-3.76	0.0003	log_lm67	1	0.50916	0.165	3.09	0.0027
log_lm7	1	1.43249	0.30619	4.68	< .0001	log_lm73	1	0.46143	0.13632	3.38	0.0011
log_lm9	1	-1.21129	0.29001	-4.18	< .0001	log_lm78	1	1.19552	0.20585	5.81	< .0001
log_lm10	1	0.58143	0.18478	3.15	0.0023	log_lm79	1	1.08692	0.20938	5.19	< .0001
log_lm13	1	-0.7071	0.16159	-4.38	< .0001	log_lm81	1	-0.98338	0.16923	-5.81	< .0001
log_lm15	1	0.37033	0.19976	1.85	0.0671	log_lm82	1	0.318	0.15991	1.99	0.0499
log_lm17	1	1.1582	0.16999	6.81	< .0001	log_lm85	1	-0.61665	0.14259	-4.32	< .0001
log_lm20	1	-0.2428	0.14093	-1.72	0.0884	log_lm87	1	1.02047	0.21704	4.7	< .0001
log_lm23	1	0.4085	0.11765	3.47	0.0008	log_lm91	1	0.51866	0.12998	3.99	0.0001
log_lm29	1	-0.53046	0.14384	-3.69	0.0004	log_lm92	1	-1.25151	0.14228	-8.8	< .0001
log_lm31	1	-0.75008	0.16018	-4.68	< .0001	log_lm98	1	-0.85817	0.29335	-2.93	0.0044
log_lm33	1	-0.97899	0.25738	-3.8	0.0003	log_lm100	1	1.06742	0.27684	3.86	0.0002
log_lm37	1	-0.40976	0.1234	-3.32	0.0013	log_lm36	1	-0.00996	0.10407	-0.1	0.924
log_lm38	1	-0.63789	0.23356	-2.73	0.0076	log_lm75	1	-0.11068	0.1626	-0.68	0.4979
log_lm39	1	1.84215	0.2588	7.12	< .0001						

Reference

- Sparse high-dimensional models in economics, Fan, J., Lv, J., and Qi, L. (2011), *Annual Review of Economics*.
- Dynamic Factor Models, James H. Stock, and Mark W. Watson(2010)
- 주택 가격의 지역간 상관 관계 분석 연구 : 수도권 아파트 가격을 중심으로, 왕세종, 강민석(2004)
- 주택가격의 지역간 상호의존성에 관한 연구 : 서울지역 아파트 매매가격을 중심으로, 박해선, 김승년(2014)