Modelling residential prices with cointegration techniques and automatic selection algorithms

Ramiro J. Rodríguez – Simón Sosvila (Ph Advisor)

A presentation for

Fundamentos de Análisis Económico | Seminarios

The opinions and analyses are the responsibility of the authors and, therefore, do not necessarily coincide with those of the BNP Paribas Real Estate.

Motivation for the research

1. Explaining residential price dynamics in a context of A boom-bust business cycle
    1. Do long-term relationships hold in this environment?
    2. Does an error-correction-mechanism remains valid?
2. Testing modern selection algorithms for property markets
3. Performance comparison between structural and automatic-selected models

BNP PARIBAS REAL ESTATE

UNIVERSIDAD COMPLUTENSE MADRID

Stylized facts (2004-2008)

- Expectations on swift housing price growth disengaged off-sett forces as stock increase and grater shares of income dedicated to house acquisition
- An ever increasing trend in prices followed suit giving birth to a price bubble
- Estrangement from the long term trend of fundamental variables (i.e. house price)

BNP PARIBAS REAL ESTATE

UNIVERSIDAD COMPLUTENSE MADRID

Main findings

- Intense variations in fundamental variables during the boom-boost period shifted conventional long-term trends and relationships
- A structural shock hit the Spanish housing market in Q2 2004
- Income, mortgage rate, new dwellings and capital formation are the main determinants of prices
- Prices return to their long term path in a swift fashion
- Structural modelling outperformed our automatic selection technique for the residential market in Spain (1995-2012)

BNP PARIBAS REAL ESTATE
UNIVERSIDAD COMPLUTENSE MADRID

Reference literature

- Literature review in search of the main variables explaining prices

- Gatini, L and Hiebert P (2010)

  "Forecasting euro house prices through fundamentals"

  European Central Bank Working Papers Series Nº1249

  ✓ Parsimonious modelling

  ✓ Uses an ECM

- Cuerpo, C and Pontuch P (2013)

  "Spanish housing market: adjustments and implications"

  European Commission – ECFIN, Volume 10, Issue 8

  ✓ VAR modelling checking house price impact on real economy

BNP PARIBAS REAL ESTATE

UNIVERSIDAD COMPLUTENSE MADRID

Variables and data

An eclectic approach: DDBB with most of the variables identified the literature review

- Number of variables: 52 (6 LHS and 46 RHS candidates)
- Structure: Quarterly data
- Start date: 1995:1
- End date: 2012: 4
- Monetary variables deflated by the implicit GDP deflator
- Stock monthly variables: last point observed
- Flow variables: accumulated for the quarter
- Sources: manly public ones(ministry of public works, statistics office, Central bank)

Structural model

Long run equation

$$\log(P_t^*) = \alpha_0 + \sum_{i=1}^{q} \alpha_i \log\left(X_{i,t}\right) + \xi_t$$

Short run dynamics

$$\mathrm{Dlog}(P_t) = \sum_{i=1}^{q} \sum_{j=1}^{n} \gamma_{ij} D \log\left(X_{i,t-j}\right) + \beta \xi_{t-1} + \varepsilon_t$$

BNP PARIBAS REAL ESTATE
UNIVERSIDAD COMPLUTENSE MADRID

Variable definition

- Endogenous: Real house price per square meter

  (HOUSE_PRICE_M2)

  Appraisal based price from the National statistics Institute

- Regressors (in real terms whenever necessary):

  Gross domestic product per capita

  (GDP_PC)

  Mortgage interest rate

  (*MORTG_RATE*)

  Free market Residential buildings starts

  (*BULD_STRT_FREE*)

  Gross capital formation in dwellings

  (*GCF_DWELL*)

BNP PARIBAS REAL ESTATE

UNIVERSIDAD COMPLUTENSE MADRID

Integration tests

- Joint confirmation hypothesis (JCH) of unit root by the simultaneous use of the DF and KPSS tests (Carrión-i-Silvestre *et al.* 2001)

Example of order of integration testing of a variable X

If Test	Action
ADF (1) with drift and trend	Rejects H0
ADF with drift	Does not reject H0
ADF without drift and trend	Rejects H0
KPSS (2) with drift and trend	Does not reject H0
KPSS with drift	Rejects H0
	X is stationary

(1) In ADF the null hypothesis is that the variable is first difference stationary
(2) In KPSS the null hypothesis is that the variable is stationary

BNP PARIBAS REAL ESTATE
UNIVERSIDAD COMPLUTENSE MADRID

Integration tests results

- Constant variations in the same direction don't allow some series to lose their trend when first differentiated (Bubble conditions)
- We resort to check stationarity under structural breaks, based on Perron (1997) unit root test

Variable	Is	With a structural break in	In quarter
House price	I(1)*	Trend	2 - 2004
GDP_PC	I(1)*	Intercept	1 - 2008
GCF_DWELL	I(1)**	Intercept	1 - 2008
Mortgage interest rate	I(1)	No	
New dwellings	I(1)	No	

Dummy Variable

* At a 95% of level of confidence
** At a 90% of level of confidence

BNP PARIBAS REAL ESTATE

UNIVERSIDAD COMPLUTENSE MADRID

Empirical results: Long run (1)

Results for the FMLS - Phillips-Hansen procedure (*t*-ratio in brackets)

$$\log(\text{Price\_m2}) = \underset{(-15.14194)}{-10.28295} + \underset{(20.39477)}{1.587379} * \log(\text{GDPPC}) \underset{(-7.234887)}{-0.11844} * \log(\text{MORTG\_RATE}) \underset{(-10.78238)}{-0.086308} * \log(\text{BULD\_STRT\_FREE}) + \underset{(16.99713)}{0.503511} * \log(\text{GCF\_DWELL}) + \underset{(5.227061)}{0.08433} * \text{DUMMY}$$

$$R^2 = 0.9926 \; DW = 1.08$$

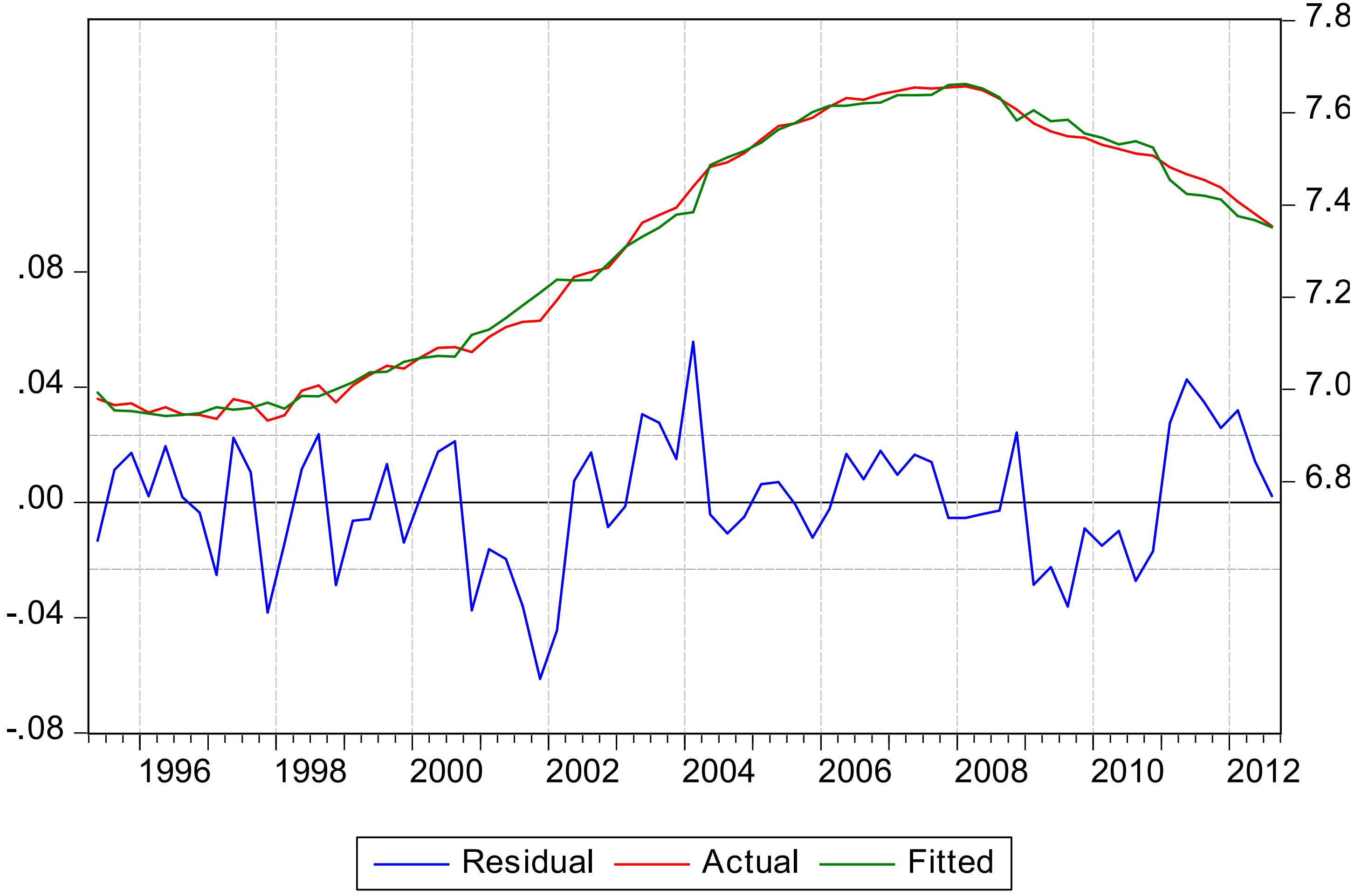
Series cointegration test (H0: ∄ cointegration):

	Statistic name	Value	P-Value
Engel & Granger	Tau Statistic	-4.7316	0.0388
	Z-statistic	-34.0894	0.0344

BNP PARIBAS REAL ESTATE

UNIVERSIDAD COMPLUTENSE MADRID

Empirical results: Long run (2)



Jarque-Bera Probability: 0.68097

BNP PARIBAS REAL ESTATE

UNIVERSIDAD COMPLUTENSE MADRID

Empirical results: Short Run (1)

Results for the OLS estimation (*t*-ratio in brackets)

$$\mathrm{Dlog(HOUSE\_PRICE\_M2)} = \underset{(-4.8559)}{-0.2878} * \{ \log[\mathrm{HOUSE\_PRICE\_M2}(-1)]$$

$$-1.5644 * \log[GDPPC(-1)] + 0.1398 * \log[MORTG\_RATE(-1)]$$

$$+ 0.0735 * \log[\mathrm{BUILD\_STRT\_FREE}(-1)] - 0.5029 \log[\mathrm{GCF\_DWELL}(-1)]$$

$$-0.0724 * \mathrm{DUMMY} + 10.1785 \} + \underset{(9.7248)}{0.6270} * \mathrm{Dlog[HOUSE\_PRICE\_M2}(-1)]$$

$$- \underset{(-2.3352)}{0.0404} * Dlog[MORTG\_RATE(-2)] + \underset{(2.1213)}{0.0185} * Dlog[BUILD\_STRT\_FREE(-4)]$$

$$+ \underset{(2.1802)}{0.0182} * Dlog[BUILD\_STRT\_FREE(-2)]$$

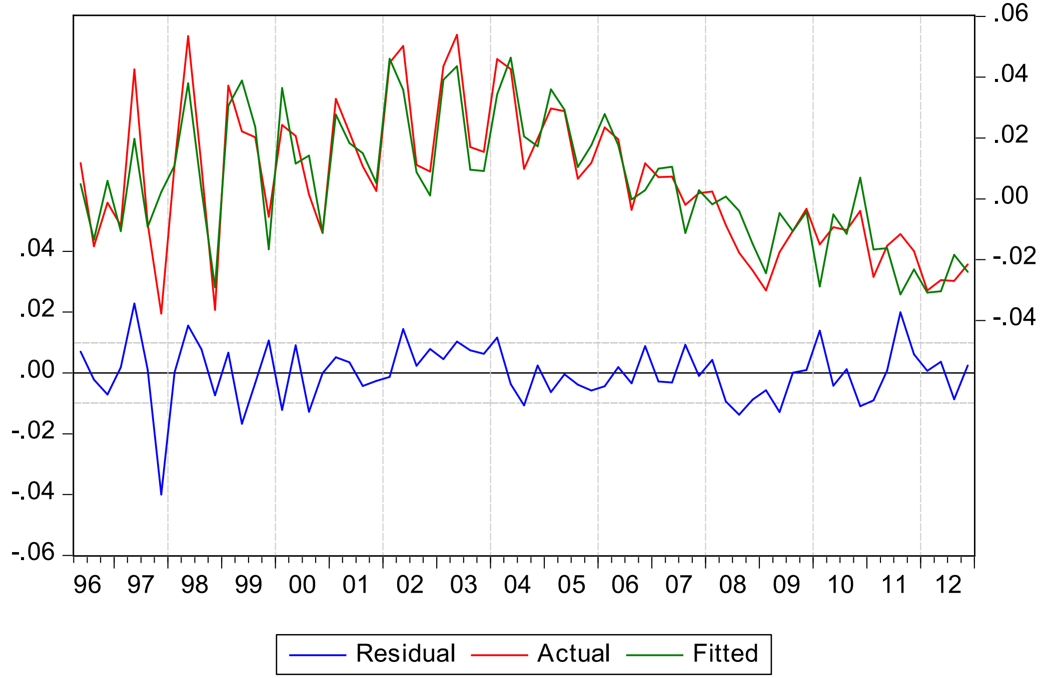
$$\mathrm{R}^2 = 0.8294, \mathrm{DW} = 1.9$$

BNP PARIBAS REAL ESTATE

UNIVERSIDAD COMPLUTENSE MADRID

Empirical results: Short Run (2)

Actual, fitted, and residuals from estimated equation

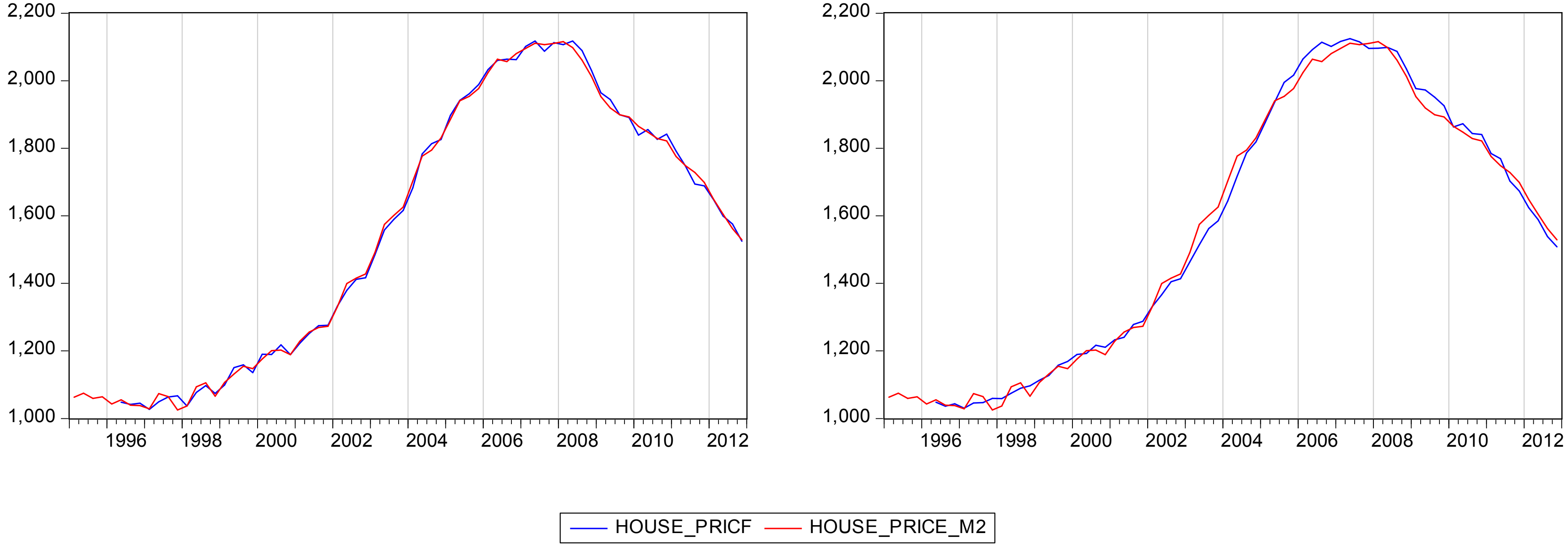


BNP PARIBAS REAL ESTATE

UNIVERSIDAD COMPLUTENSE MADRID

Inner-sample forecasts

Static　　　　　　　　　　　　Dynamic



BNP PARIBAS REAL ESTATE

UNIVERSIDAD COMPLUTENSE MADRID

Automatic model selection

- For a 46 possible candidate regressors set the are $2^{46} \approx$ 70 trillion nested models
  - ✓ Search path
- Origins in datamining (Lovell, 1983)
- GETS or LSE Approach (Hoover and Perez,1999 & Hendry and Krolzig,1999)
- Information criteria techniques (Hansen, 1999 & Perez-Amaral et al., 2003)
  - ✓ Genetic algorithm using Schwarz information criterion (GASIC) Acosta-Gonzalez & Fernandez-Rodriguez, 2007)

BNP PARIBAS REAL ESTATE
UNIVERSIDAD COMPLUTENSE MADRID

Genetic algorithm... Why?

A. Makes 200 different regressions of 5-tuple regressors (genes)
B. Compares and (with SIC) selects encompassing models (extinction and most fitted survival)
C. Combines regressors to make new regressions that encompass their parents (Origin of Species by Means of…)
D. If encompassing is possible returns to B. if not end of process (evolved current species)

BNP PARIBAS REAL ESTATE

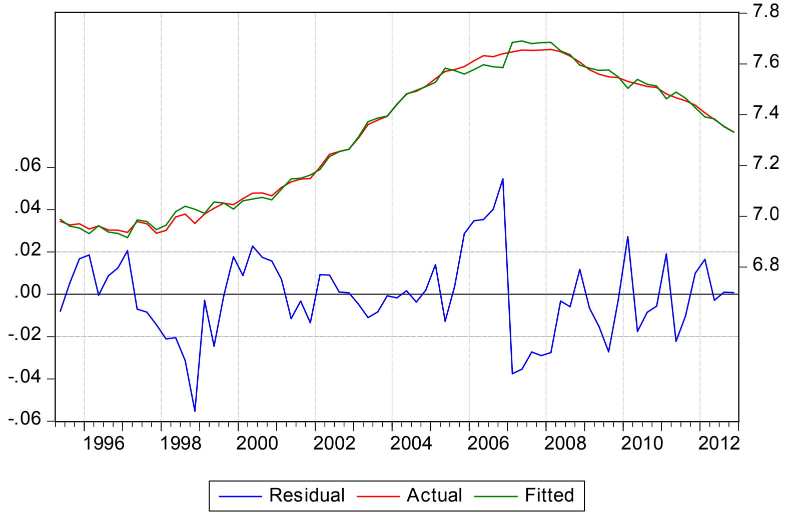
UNIVERSIDAD COMPLUTENSE MADRID

Selected model (long run)

Dependent Variable: LOG(PRICE)
Method: Fully Modified Least Squares (FMOLS)
Date: 05/31/14 Time: 12:29
Sample (adjusted): 1995Q2 2012Q4
Included observations: 71 after adjustments
Cointegrating equation deterministics: C
Long-run covariance estimate (Bartlett kernel, Newey-West fixed bandwidth =4.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(EFFORT_NO_DEDUCT)	0.639301	0.029548	21.63627	0.0000
LOG(CONCRETE_CONSUM)	0.160411	0.009314	17.22280	0.0000
LOG(MORTG_RATE)	-0.332009	0.022267	-14.91011	0.0000
LOG(GDP_2008)	0.693200	0.060472	11.46309	0.0000
C	-4.298316	0.667614	-6.438330	0.0000

R-squared	0.994379	Mean dependent var	7.315029
Adjusted R-squared	0.994038	S.D. dependent var	0.258954
S.E. of regression	0.019995	Sum squared resid	0.026387
Durbin-Watson stat	1.025831	Long-run variance	0.000740



Series cointegration test
(H0: ∄ cointegration):

	Statistic name	Value	P-Value
Engel & Granger	Tau	-4.7603	0.036
	Z	-34.725	0.03

BNP PARIBAS REAL ESTATE

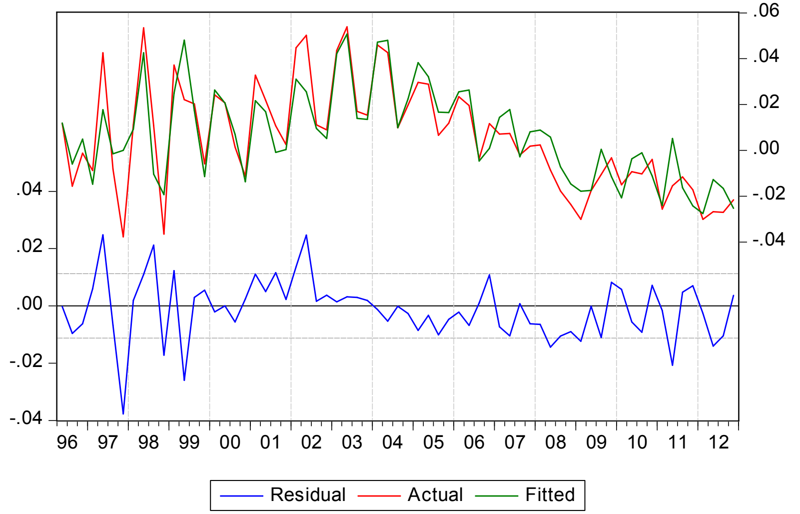
UNIVERSIDAD COMPLUTENSE MADRID

Selected model (short run)

Dependent Variable: DLOG(PRICE)
Method: Least Squares
Date: 05/31/14 Time: 12:38
Sample (adjusted): 1996Q2 2012Q4
Included observations: 67 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(PRICE(-1))-0.639301292294*LOG(EFFORT_NO_DEDUCT(-1))-0.160410998866*LOG(CONCRETE_CONSUM(-1))+0.332008749696*LOG(MORTG_RATE(-1))-0.693200026756*LOG(GDP_2008(-1))+4.2983159666	-0.096204	0.072995	-1.317948	0.1925
DLOG(PRICE(-4))	0.683577	0.083669	8.170039	0.0000
DLOG(CONCRETE_CONSUM(-4))	0.049039	0.016365	2.996613	0.0040
DLOG(PRICE(-3))	0.166395	0.073555	2.262187	0.0273
DLOG(GDP_2008(-3))	-0.800466	0.351669	-2.276190	0.0264
DLOG(MORTG_RATE(-2))	-0.051025	0.017925	-2.846522	0.0060
DLOG(GDP_2008(-1))	0.915952	0.342165	2.676931	0.0096

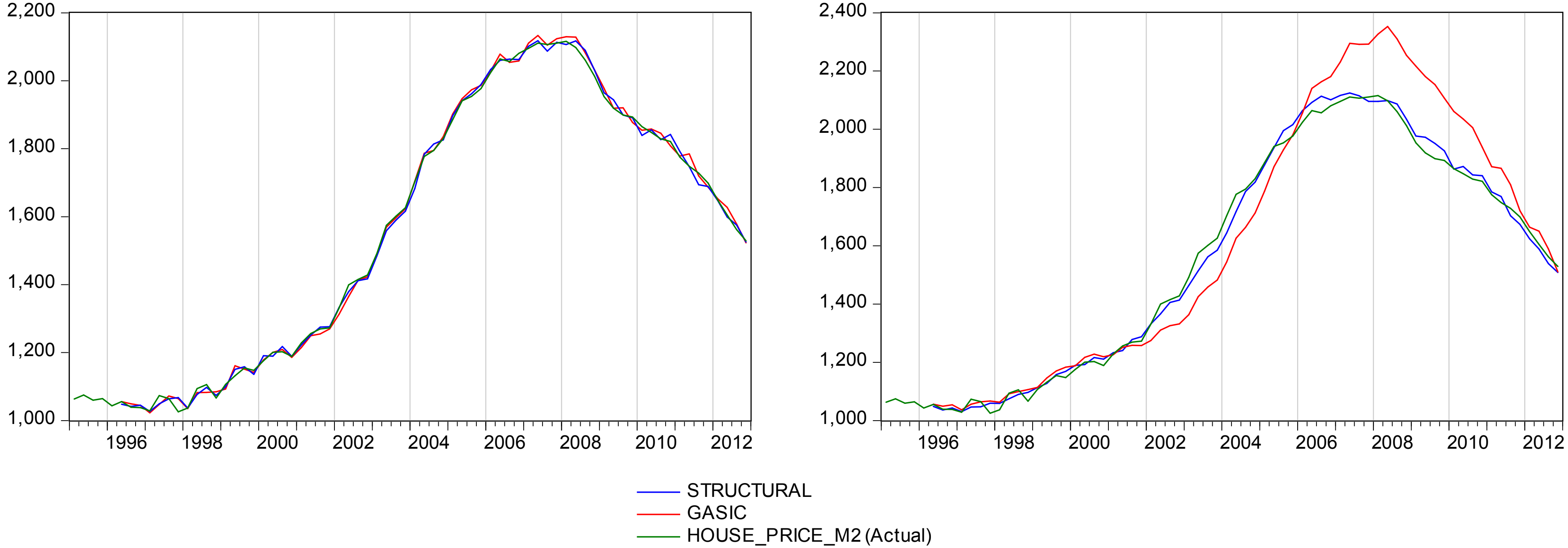
R-squared	0.787273	Mean dependent var	0.005706
Adjusted R-squared	0.766000	S.D. dependent var	0.023227
S.E. of regression	0.011236	Akaike info criterion	-6.040868
Sum squared resid	0.007574	Schwarz criterion	-5.810527
Log likelihood	209.3691	Hannan-Quinn criter.	-5.949722
Durbin-Watson stat	1.814639		



BNP PARIBAS REAL ESTATE
UNIVERSIDAD COMPLUTENSE MADRID

Performance comparison

Static | Dynamic



$$\hat{\epsilon}_{structural,t} = \beta_{0,t} + \beta_{1,t}\left(\hat{\epsilon}_{structural,t} - \hat{\epsilon}_{GASIC,t}\right) + \hat{\varepsilon}_t$$

Dependent Variable: RES_EG
Method: Least Squares
Date: 06/23/14 Time: 18:57
Sample (adjusted): 1996Q2 2012Q4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000399	0.001123	-0.355012	0.7237
RES_EG-RES_GASIC	0.305778	0.137171	2.229175	0.0293

BNP PARIBAS REAL ESTATE

UNIVERSIDAD COMPLUTENSE MADRID

Further steps

- Assessing out-of-sample forecasts
- Impulse response analysis
- GETS methodology with OxMetrics®
- Guided regressors auto-selection

BNP PARIBAS REAL ESTATE
UNIVERSIDAD COMPLUTENSE MADRID

Q&A

Suggestions are much appreciated!

Thank you

BNP PARIBAS REAL ESTATE

UNIVERSIDAD COMPLUTENSE MADRID