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
Stylized facts (2004-2008)

- Expectations on swift housing price growth disengaged off-sett forces as stock increase and greater 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)
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)
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
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

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

Short run dynamics

\[ D\log(P_t) = \sum_{i=1}^{q} \sum_{j=1}^{n} \gamma_{ij} D \log(X_{i,t-j}) + \beta \xi_{t-1} + \varepsilon_t \]
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)
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 $X$

<table>
<thead>
<tr>
<th>If Test</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADF (1) with drift and trend</td>
<td>Rejects H0</td>
</tr>
<tr>
<td>ADF with drift</td>
<td>Does not reject H0</td>
</tr>
<tr>
<td>ADF without drift and trend</td>
<td>Rejects H0</td>
</tr>
<tr>
<td>KPSS (2) with drift and trend</td>
<td>Does not reject H0</td>
</tr>
<tr>
<td>KPSS with drift</td>
<td>Rejects H0</td>
</tr>
</tbody>
</table>

$X$ 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
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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Is</th>
<th>With a structural break in</th>
<th>In quarter</th>
</tr>
</thead>
<tbody>
<tr>
<td>House price</td>
<td>I(1)*</td>
<td>Trend</td>
<td>2 - 2004</td>
</tr>
<tr>
<td>GDP_PC</td>
<td>I(1)*</td>
<td>Intercept</td>
<td>1 - 2008</td>
</tr>
<tr>
<td>GCF_DWELL</td>
<td>I(1)**</td>
<td>Intercept</td>
<td>1 - 2008</td>
</tr>
<tr>
<td>Mortgage interest rate</td>
<td>I(1)</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>New dwellings</td>
<td>I(1)</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>

* At a 95% of level of confidence
** At a 90% of level of confidence
Empirical results: Long run (1)

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

$$\log(\text{Price}_m) = -10.28295 + 1.587379 \times \log(\text{GDPPC})$$
$$\quad \quad \quad \quad \quad \quad \quad (-15.14194) \quad (20.39477)$$
$$\quad - 0.11844 \times \log(\text{MORTG}_\text{RATE})$$
$$\quad \quad \quad \quad \quad \quad \quad (-7.234887)$$
$$\quad - 0.086308 \times \log(\text{BUILD}_\text{STRT}\_\text{FREE})$$
$$\quad \quad \quad \quad \quad \quad \quad (-10.78238)$$
$$\quad + 0.503511 \times \log(\text{GCF}_\text{DWell}) + 0.08433 \times \text{DUMMY}$$
$$\quad \quad \quad \quad \quad \quad \quad (16.99713) \quad (5.227061)$$

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

Series cointegration test (H0: $\not\exists$ cointegration):

<table>
<thead>
<tr>
<th>Statistic name</th>
<th>Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engel &amp; Granger</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tau Statistic</td>
<td>-4.7316</td>
<td>0.0388</td>
</tr>
<tr>
<td>Z-statistic</td>
<td>-34.0894</td>
<td>0.0344</td>
</tr>
</tbody>
</table>
Empirical results: Long run (2)

Jarque-Bera Probability: 0.68097
Empirical results: Short Run (1)

Results for the OLS estimation \((t\text{-ratio in brackets})\)

\[
\begin{align*}
\text{Dlog}(\text{HOUSE PRICE M2}) &= \begin{bmatrix} \text{log}[\text{HOUSE PRICE M2}(-1)] \\
-0.2878 \end{bmatrix} \\
&\quad \begin{bmatrix} (-4.8559) \end{bmatrix} \\
-1.5644 \times \text{log}[\text{GDPPC}(-1)] + 0.1398 \times \text{log}[\text{MORTG RATE}(-1)] \\
+ 0.0735 \times \text{log}[\text{BUILD STRT FREE}(-1)] - 0.5029 \times \text{log}[\text{GCF DWELL}(-1)] \\
-0.0724 \times \text{DUMMY} + 10.1785 \end{align*}
\]

\[
\begin{align*}
\text{(9.7248)} \\
-0.0404 \times \text{Dlog}[\text{MORTG RATE}(-2)] + 0.0185 \times \text{Dlog}[\text{BUILD STRT FREE}(-4)] \\
\begin{bmatrix} (-2.3352) \end{bmatrix} + (2.1213) \\
+ 0.0182 \times \text{Dlog}[\text{BUILD STRT FREE}(-2)] \\
\begin{bmatrix} (2.1802) \end{bmatrix}
\end{align*}
\]

\[R^2 = 0.8294, DW = 1.9\]
Empirical results: Short Run (2)

Actual, fitted, and residuals from estimated equation

![Graph showing actual, fitted, and residual values over time.](image-url)
Inner-sample forecasts

Static

Dynamic

![Graphs showing inner-sample forecasts for static and dynamic models. The graphs compare HOUSE_PRICF and HOUSE_PRICE_M2 over the years 1996 to 2012.]
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)
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)
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)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOG(EFFORT_NO_DEDUCT)</td>
<td>0.639301</td>
<td>0.029548</td>
<td>21.63627</td>
<td>0.0000</td>
</tr>
<tr>
<td>LOG(CONCRETE_CONSUM)</td>
<td>0.160411</td>
<td>0.009314</td>
<td>17.22280</td>
<td>0.0000</td>
</tr>
<tr>
<td>LOG(MORTG_RATE)</td>
<td>-0.332009</td>
<td>0.022267</td>
<td>-14.91011</td>
<td>0.0000</td>
</tr>
<tr>
<td>LOG(GDP_2008)</td>
<td>0.693200</td>
<td>0.060472</td>
<td>11.46309</td>
<td>0.0000</td>
</tr>
<tr>
<td>C</td>
<td>-4.298316</td>
<td>0.667614</td>
<td>-6.438330</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

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):

<table>
<thead>
<tr>
<th>Statistic name</th>
<th>Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engel &amp; Granger</td>
<td>Tau</td>
<td>-4.7603</td>
</tr>
<tr>
<td></td>
<td>Z</td>
<td>-34.725</td>
</tr>
</tbody>
</table>

BNP PARIBAS REAL ESTATE
UNIVERSIDAD COMPLUTENSE
## 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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOG(PRICE(-1))</td>
<td>-0.096204</td>
<td>0.072995</td>
<td>-1.317948</td>
<td>0.1925</td>
</tr>
<tr>
<td>0.639301292294*LOG(EFFORT_NO_DEDUCT(1))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.160410998866*LOG(CONCRETE_CONSUM(-1))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.332008749696*LOG(MORTG_RATE(-1))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.693200026756*LOG(GDP_2008(-1))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+4.29631599666</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLOG(PRICE(-4))</td>
<td>0.683577</td>
<td>0.083669</td>
<td>8.170039</td>
<td>0.0000</td>
</tr>
<tr>
<td>DLOG(CONCRETE_CONSUM(-4))</td>
<td>0.049039</td>
<td>0.016365</td>
<td>2.99613</td>
<td>0.0040</td>
</tr>
<tr>
<td>DLOG(PRICE(-3))</td>
<td>0.166395</td>
<td>0.073555</td>
<td>2.262187</td>
<td>0.0273</td>
</tr>
<tr>
<td>DLOG(MORTG_RATE(-2))</td>
<td>-0.051025</td>
<td>0.017925</td>
<td>-2.846522</td>
<td>0.0060</td>
</tr>
<tr>
<td>DLOG(GDP_2008(-1))</td>
<td>0.915952</td>
<td>0.342165</td>
<td>2.676931</td>
<td>0.0096</td>
</tr>
</tbody>
</table>

R-squared 0.787273
Adjusted R-squared 0.766000
S.E. of regression 0.011236
Sum squared resid 0.007574
Log likelihood 209.3691
Durbin-Watson stat 1.814639
Performance comparison

**Static**

**Dynamic**

\[
\hat{\epsilon}_{structural,t} = \beta_{0,t} + \beta_{1,t} (\hat{\epsilon}_{structural,t} - \hat{\epsilon}_{GASIC,t}) + \hat{\epsilon}_t
\]

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-0.000399</td>
<td>0.001123</td>
<td>-0.355012</td>
<td>0.7237</td>
</tr>
<tr>
<td>RES_EG-RES_GASIC</td>
<td>0.305778</td>
<td>0.137171</td>
<td>2.229175</td>
<td>0.0293</td>
</tr>
</tbody>
</table>
Further steps

• Assessing out-of-sample forecasts
• Impulse response analysis
• GETS methodology with OxMetrics®
• Guided regressors auto-selection
Suggestions are much appreciated!

Thank you