XLVI Reunión Anual
Noviembre de 2011
ISSN 1852-0022

ECONOMETRIC MODELLING OF COMMODITIES PRICES BY AUTOMETRICS

Ahumada, Hildegart
Cornejo, Magdalena
ECONOMETRIC MODELLING OF COMMODITY PRICES BY AUTOMETRICS*

August 2011

HILDEGART A. AHUMADA† AND MAGDALENA CORNEJO‡

† Di Tella University

E-mail: hahumada@utdt.edu

‡ Di Tella and Buenos Aires University

Abstract

Commodity prices have recently reappeared on research agendas after their steady rise over the last decade, in contrast with their previous extended decline. New explanations have been proposed for this trend emphasising the role of commonalities instead of the determinants of the relevant commodity markets. We considered both of them to explain prices of eight different commodities with significant weight in Argentina’s trade accounts. From a time series- cross section data model on an annual basis (1960-2010) we find that supply and demand determinants as well as US monetary base, US exchange rate and China’s GDP explain commodity prices.

JEL Codes: C33, F40

* This paper is part of the research project “An econometric study of commodity prices and their effects on the Argentine Economy”, supported by BID-FONCYT, PICT08 1122.
1. Introduction

Commodity prices have reappeared on research agendas after their steady rise over the last decade. Their recent behaviour contrasts sharply with their previous extended stagnation or decline, particularly when measured relative to manufactured goods. Thus, a great deal of literature, which started with the pioneering works of Singer (1950), Prebisch (1950) and Lewis (1954), tried to understand the resulting negative perspective for producer countries. The new scenario for commodity prices has had a great impact on many economies, both exporters and importers. In the case of Argentina, its current economic expansion has been mainly associated with the growth of these prices. In this respect efforts to examine the persistence of their movement and disentangle long run and short run determinants should be useful for economic policy making.

It is interesting to note that, along with the upward trend, new explanations have been proposed which emphasise the role of commonalities instead of looking at the determinants of the relevant commodity markets, as in the earlier literature. The role of the variables associated with an easy monetary policy, mainly in the US, the new use of commodities as part of the portfolio of financial investors and the growing participation in world trade of commodity consumer countries are part of the current debate. However, most of the research does not integrate all sources of explanations.

If idiosyncratic and common factors (and their different approximations) are taken into account, a large number of explanatory variables, even larger than observations, should be part of the econometric model. For such analysis, Autometrics, an algorithm for computer-automated model selection (Doornik 2009, Doornik and Hendry 2009) can help to understand the determinants of commodity prices. It is used in this paper to econometrically study the prices of eight different commodities that have a significant weight in Argentina’s trade accounts, including agricultural and mineral on an annual basis (1960-2010).

The next section reviews relevant literature. Section 3 describes the main feature of the data. Section 4 discusses a time-series cross-section data (TSCS) model and the estimation approach. Section 5 presents sensitivity results. Section 6 draws conclusions and suggests future lines of research.

2. A review of the literature

Empirical studies of commodity prices have recently been largely motivated to find the common financial and monetary determinants that can explain the upward co-movement observed over the last decade. Either aggregate indexes or factor analysis are very often employed to jointly model their behavior. These are different approaches from those of the earlier literature that focused on the time properties of individual commodity prices. Specifically, the high degree of serial correlation observed in the prices of major commodities was a time series property that cannot be explained by models of storage arbitrage (Deaton and Laroque, 1992) and this drove intense research in this direction. Furthermore, the relative weight of the trend relative to the cycle component along with the degree of volatility has been empirically studied for individual prices (see e.g. Gilbert, 2006).

The more recent developments in commodity markets have opened a debate about the causes and effects of the speculative demand for them. Associated with the view of commodity prices as assets prices the role that inventories should play has again been discussed. Krugman (2008) suggests that inventory holdings have not increased as expected

---

1 A recent reconsideration of this issue can be found in Cafiero, et al. (2011).
to sustain a speculative demand and instead an increasing demand from emerging economies for food and energy commodities put the drivers on real factors. This view has been questioned by Calvo (2008) based on the hypothesis of an inelastic demand or supply in the short-run. In such cases prices can go up without observing large inventories. For him the easy monetary policies of industrialised countries (especially the US) and the growth of sovereign wealth funds are the crucial factors. M2 or Treasury Bills are the relevant money concept to understand the increase of highly flexible prices like commodity prices, a leading indicator of future inflation. Low interest rates are the channel for such increases. Frankel (2008) also emphasizes the overshooting effect on commodity prices and explains why they should be negatively related to real interest rates. As asset prices, the present value of the expected commodity prices increase when the interest rate decrease - and vice versa - and that should be reflected on spot prices in the short-run. ²

An empirical study (Browne and Cronin, 2010) finds evidence that commodity and consumers prices are proportional to the money supply - but the commodity prices initially overshoot - by applying a cointegrated VAR approach for US data. M2, CPI, GDP and several commodity price indexes are used in this analysis.

The US exchange rate depreciation relative to other currencies has been suggested as another variable explaining price increases as most commodities are quoted in US dollars, as first discussed by Ridler and Yandle (1972), see also Gilbert (1989). Dornbusch (1985) explains this effect as an implication of the flexibility inherent in commodity prices. Many empirical works using indexes or common factors show the empirical significance of this effect, as later commented on. Furthermore, the role of exchange rates for predicting global commodity prices has been found by Chen, Rogoff and Rossi (2010).

A price index (in real terms) of commodities exported by Argentina has been built by Bastourre et al. (2008, 2010). In their first study, they investigate the drivers of a weighted price index of eight main commodities exported by Argentina employing a VECM: the real exchange rate of the US, global income, the real interest rate and a global liquidity measure (defined as the sum of the US monetary base and international reserves at the central banks of all the world) using quarterly data from 1986 to 2006. To shed further light on the factors affecting commodity prices, they estimate a STAR model (for a nonlinear adjustment) on a monthly basis considering the world industrial production index, the real exchange rate of the US, the real international rate and the real Dow Jones index as the main determinants of "fundamentals".

Other econometric approaches have been used to avoid the use of commodity price indexes, Deaton (1999, see also Byrne, Fazio and Fiess, 2010) has stressed the limitations of fixed aggregation since even a common shock may have different impact from good to good. Instead, Byrne et al. (2010) employ a Factor augmented VAR approach to empirically study real commodity prices, real interest rates and risk (measured as standard deviation of monthly stock prices) for a set of 24 non-fuel commodities (controlling by real oil prices and the growth rate of real US GDP) over more than 100 years to concentrate on fundamentals and eliminate the influence of speculation. They find a negative response of the principal component of prices, determined by Bai and Ng approach (see Smith and Fuertes, 2010), to both interest rates and risk during the first five or four years after the shock. It should be noted that the principal component corresponds to the whole sample of more than a century. Lombardi et al. (2010) also applied a FAVAR approach to fifteen non-oil commodity prices on a quarterly basis 1970-2008 identifying common trends captured by the food and metals factors. They find effects of industrial production and exchange rates but no strong effects of oil prices and interest rates on individual commodity prices.

² An argument against the role of speculation is the fact that commodities without futures markets show similar behavior as those with such markets (see Frankel and Rose, 2010).
Vansteenkiste (2009) also studies common factors by finding the dynamic principal components of 32 non-fuel commodities on a quarterly basis over 1957-2008. The common factor, considered as an unobserved component, shows that commodities have become more correlated recently although their correlation was stronger in the past (in the 1970s). She also finds that this common factor can be explained by the US dollar effective exchange rate, the US real interest rate, oil and fertilizer prices and since the 1990s, by global industrial activity. PCGets was used to select the explanatory variables which were instrumented by their lags to avoid endogeneity issues. In this study different behaviour over time is analysed and the common factor is assumed dynamic, but the autoregressive AR(1) parameter of its transition equation is still assumed to be time invariant.

Recently, Frankel and Rose (2010) have attempted to explain real commodity prices allowing for microeconomic as well as macroeconomic determinants. They used several regression models and also provide results from an annual panel data of eleven commodities, a different approach from that previously reviewed. A main contribution of theirs lies in the fact that they employ the inventory levels as an explanatory variable. They find that they have a negative sign indicating how the decision whether to hold the commodity for another period has a negative impact on the current price. Nevertheless, the real interest rate included in their estimations does not appear to have a significant effect. Although this paper is aimed at examining cyclical components, they employ annual data to estimate either levels or difference models without distinguishing short run and long run effects in the models.\(^3\)

Regarding short run and long run effects, it should be noted that Deaton and Laroque (1992, 1995) have offered theoretical explanations of the short run dynamics of commodity prices based on the role of inventories, introducing speculative inventories to try matching the time-properties of the data. However, in a second stage, Deaton and Laroque (2003) focus on the long run determinants, commodity supply and demand should be cointegrated. Supply is thought to be infinitely elastic but to react with long lags while demand responds to world income apart from prices. For mineral products, they address the effect of interest rates on the growth of their prices following early developments as in Hotelling’s model.

To sum up we found a large set of potential explanatory variables of commodity prices that include monetary, financial and exchange rate variables along with the supply and demand determinants of each specific good. As explained in Section 4, different proxies for the common variables previously suggested and the commodity productions and inventories are included in the econometric analysis to jointly evaluate their long run relationship and short run dynamics.

### 3. Data description

The study is focused on the real price (deflated by US CPI index) of eight commodities - aluminum, copper, gold, petroleum, beef, corn, soybeans and wheat\(^4\), which have a significant weight in Argentina’s trade accounts (see Appendix 1 for data definitions and sources). They are very different kind of commodities. Some of them are traditional Argentine productions (beef, corn, and wheat) that have been, to different degrees, competing with soybean in production. Soybean and its derived products have been converted in the main source of external reserves. Corn and soybean have also experienced a growing demand as

---

\(^3\) Preliminary cointegration results in Frankel and Rose (2010) suggest that the three variables that are most consistently significant are the spread, the volatility and inventories. They find cointegration in commodity-specific models, but have weaker results for panel cointegration.

\(^4\) They are part of the set studied by Frankel and Rose (2010) except aluminum and beef and they also include five other commodities. They are different from those used by Bastourre et al. (2008, 2010) for Argentina since they aggregate metals, include derived product of soybeans and exclude oil.
bio-fuel. Gold is a new and growing product among the metal exports of Argentina. Regarding petroleum and derived products, Argentina was transformed from being a net exporter to a net importer over the period studied while remaining as an exporter of some derived products. Because of their special characteristics the inclusion of gold and oil in the set of commodities may be the more controversial—in fact, many studies have excluded them-. However, we prefer to include them but also evaluate how different from the rest their behaviors are.

Figure 1 shows the joint behavior of the log of real price indexes over 1960-2010. We draw lines for 1974 to observe the similar behavior after the first oil crisis and in 2000s when the upward co-movement seems to start.

Figure 1. Real price indexes from 1960 to 2010 (in logs)
For the eight commodities Figure 2 allows us to observe real prices, production and inventories. From a long run perspective quantities have shown an upward trend but inventories have had different patterns being more close to production for agricultural products as expected from being less storable. A negative relationship between the log difference of real prices and the log difference of inventories can be often observed in the next figure.

---

5 We have attempted to use world data whenever available for the whole sample to obtain a large T panel. We use world production except for petroleum before 1973 (See Appendix 1).
The next table shows the contemporaneous correlations. For the log of real prices, there are high correlations within groups like agricultural and metals (aluminum and copper) but also between groups, the highest of them correspond to gold and petroleum. For log differences, most of the correlations are reduced after this transformation but some of them remain high in the case of within agricultural and mineral groups. A similar pattern of correlations is observed for the residuals of a VAR(2) model of real prices in log levels. Therefore, the additional information given by lagged real prices of commodities (other than the own) does not have much effect with regard to explaining real prices on a yearly basis.
Table 1. Matrix correlations of commodity real prices in log levels and log differences, the residuals of a VAR(2) and from the Eq. 2.

Log levels:

<table>
<thead>
<tr>
<th></th>
<th>al</th>
<th>co</th>
<th>go</th>
<th>pe</th>
<th>bf</th>
<th>cn</th>
<th>so</th>
<th>wh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aluminum</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Copper</td>
<td>0.6688</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gold</td>
<td>-0.0997</td>
<td>-0.1241</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Petroleum</td>
<td>-0.0893</td>
<td>-0.0604</td>
<td>0.8608</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beef</td>
<td>0.7795</td>
<td>0.6227</td>
<td>-0.1594</td>
<td>-0.1847</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corn</td>
<td>0.7771</td>
<td>0.6634</td>
<td>-0.1786</td>
<td>-0.1363</td>
<td>0.8420</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soybeans</td>
<td>0.7947</td>
<td>0.6538</td>
<td>-0.1133</td>
<td>-0.1120</td>
<td>0.8680</td>
<td>0.9657</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>Wheat</td>
<td>0.7710</td>
<td>0.6716</td>
<td>-0.0735</td>
<td>-0.0399</td>
<td>0.8155</td>
<td>0.9716</td>
<td>0.9509</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Log differences:

<table>
<thead>
<tr>
<th></th>
<th>al</th>
<th>co</th>
<th>go</th>
<th>pe</th>
<th>bf</th>
<th>cn</th>
<th>so</th>
<th>wh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aluminum</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Copper</td>
<td>0.5903</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gold</td>
<td>0.3932</td>
<td>0.3885</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Petroleum</td>
<td>0.2493</td>
<td>0.3484</td>
<td>0.3983</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beef</td>
<td>0.1634</td>
<td>0.4821</td>
<td>0.1754</td>
<td>0.0027</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corn</td>
<td>0.3052</td>
<td>0.4389</td>
<td>0.3839</td>
<td>0.3699</td>
<td>0.1325</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soybeans</td>
<td>0.2862</td>
<td>0.3865</td>
<td>0.3224</td>
<td>0.1233</td>
<td>0.6991</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wheat</td>
<td>0.2447</td>
<td>0.4763</td>
<td>0.4472</td>
<td>0.3835</td>
<td>0.2495</td>
<td>0.8220</td>
<td>0.6126</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

VAR(2):

<table>
<thead>
<tr>
<th></th>
<th>al</th>
<th>co</th>
<th>go</th>
<th>pe</th>
<th>bf</th>
<th>cn</th>
<th>so</th>
<th>wh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aluminum</td>
<td>0.1570</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Copper</td>
<td>0.5419</td>
<td>0.2160</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gold</td>
<td>0.4173</td>
<td>0.2779</td>
<td>0.1423</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Petroleum</td>
<td>0.3110</td>
<td>0.2316</td>
<td>0.2373</td>
<td>0.2224</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beef</td>
<td>0.2907</td>
<td>0.5437</td>
<td>0.2209</td>
<td>0.0972</td>
<td>0.1085</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corn</td>
<td>0.3047</td>
<td>0.2996</td>
<td>0.2802</td>
<td>0.2810</td>
<td>0.2834</td>
<td>0.1484</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soybeans</td>
<td>0.4124</td>
<td>0.4186</td>
<td>0.4194</td>
<td>0.2442</td>
<td>0.5429</td>
<td>0.7878</td>
<td>0.1682</td>
<td></td>
</tr>
<tr>
<td>Wheat</td>
<td>0.3678</td>
<td>0.4480</td>
<td>0.2740</td>
<td>0.2730</td>
<td>0.2837</td>
<td>0.8696</td>
<td>0.7454</td>
<td>0.1489</td>
</tr>
</tbody>
</table>

Correlation matrix of residuals Equation (2):

<table>
<thead>
<tr>
<th></th>
<th>al</th>
<th>co</th>
<th>go</th>
<th>pe</th>
<th>bf</th>
<th>cn</th>
<th>so</th>
<th>wh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aluminum</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Copper</td>
<td>0.2587</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gold</td>
<td>0.1423</td>
<td>-0.0840</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Petroleum</td>
<td>-0.0038</td>
<td>-0.0059</td>
<td>-0.1595</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beef</td>
<td>-0.1844</td>
<td>0.2343</td>
<td>0.0644</td>
<td>-0.0345</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corn</td>
<td>-0.0203</td>
<td>0.0549</td>
<td>-0.1052</td>
<td>-0.0037</td>
<td>-0.2139</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soybeans</td>
<td>0.1930</td>
<td>0.2657</td>
<td>0.0286</td>
<td>-0.1584</td>
<td>-0.2406</td>
<td>0.3380</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>Wheat</td>
<td>-0.1197</td>
<td>0.0229</td>
<td>-0.0398</td>
<td>-0.0590</td>
<td>-0.0528</td>
<td>0.6089</td>
<td>0.2551</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Before the TSCS data model is presented, some exploratory results from a system estimation approach (not reported) need to be commented on. First, starting with a VAR(2) unrestricted model for nominal prices over 1960-2010, automatic selection (at 5%) indicates that only the petroleum and gold prices, both lagged one year, remain as regressors in the system of prices. That is, petroleum and gold have Granger- caused commodity prices in nominal terms during the sample. However, this result is not found in real terms. Then, for
real prices all the monetary and financial variables described in the next section were included in their system as regressors. In this case, only an effect of the log of US monetary base (t and t-2) on real prices was automatically selected. The LR solution shows a positive effect for gold and petroleum. The role of the US monetary base is also discussed in the following sections.

4. A time series-cross section data model for commodity prices

In this section we develop a TSCS model to econometrically study a set of commodity prices relevant to the Argentine economy. The approach is similar to that followed by Frankel and Rose (2010) for panel data but we jointly take into account many of the explanatory variables suggested in the literature we reviewed in Section 2. We also allow for long run and short run effects following a general to specific approach that includes levels as well as differences (as in Bardsen, 1989 see also Banerjee, et al. 1993). Because we pool time series and cross section data, we allow for commodity and time heterogeneity by fixed effects, automatically selected, at the same time. Interactions between time and commodity effects are considered as well. They are tested by Impulse Saturation (IS), which adds a 1-0 dummy variable by each observation as regressor (see Hendry, Johansen and Santos, 2008). IS is an application of the case of more variables than observations, which is embodied in Autometrics\textsuperscript{6}. This is an automatic model selection algorithm to select the dominant congruent model (according to a set of diagnostic statistics) and not only a best fit. After a multiple-path search, a tree search is used to discard paths rejected as reduction of the initial model. Even more variables than observations can be handled as all possible combinations are considered part of the information set.

It should be noted that our model can be considered as a TSCS model (see Beck and Katz, 2009) since it has (fixed) \(N\) equals to 8 commodities and (asymptotic in) \(T\) equals to 48 observations after lags and difference transformation. Although the model we analyse is quite general in so far as explanatory variables and dynamic structure are concerned, we are initially assuming -as usual when pooling observations- that there is no cross dependence of prices unless by commodity invariant regressors and time effects. Those variables cannot be included in the regression together but the automatic model selection can identify the appropriate specification allowing for all their combinations. Besides, because of the type of model, we are not considering the effect of different lagged prices apart from the own for each commodity; the VAR(2) results commented in last section suggested that it would be not a problem to the extent that prices are measured in real terms.

Furthermore, we assume equal coefficients for all commodities and over time. However, in Section 5.2, we test parameter homogeneity and constancy hypotheses by Autometrics including multiplicative dummies by each commodity, and over time after 1974 and 2000 (see Figure 1). We also assume in this model all the explanatory variables as exogenous, an issue discussed in Section 5.3.

Then, the general unrestricted model (GUM) of real commodity prices to start model selection has the following form,

\[
\Delta \rho_{it} = \alpha_i + \gamma_i + \mu_i \text{ trend}_t + \rho_1 \Delta \rho_{it-1} + \rho_2 \Delta \rho_{it-2} + \delta \Delta \rho_{it-1} + x'_{it-1} \beta + x'_{it-1} \theta + z'_{it-1} \lambda + \\
+ \Delta x'_{it-s} \varphi + \Delta x'_{it-s} \varphi + \Delta z'_{it-s} \tau + \Delta z'_{it-s} \tau + \varepsilon_{it}
\]

where \(i = 1, \ldots, N; \ t = 1, \ldots, T; \) and \(s = 0,1,2\) indicate lags. The vector \(x\) denotes the variables corresponding to the commodity markets with either time \((t)\) and commodity \((i)\) variation.

\textsuperscript{6} We use Oxmetrics 6.2. See Doornik and Hendry (2009).
(lagged real prices, volatility, production and inventories) or only time variation \((t)\) such as those proxies for demand (US, OECD, World, India and China real GDP; industrial production of advanced economies, the US and India; trade participation measured by exports plus imports of China, India and BRIC countries relative to the world’s commercial trade; \(z\) denotes the only time variant explanatory variables related to monetary aggregates (the US monetary base, M2 and global liquidity), interest rates (1-year treasury constant maturity rate and federal fund rate), the US nominal effective exchange rate and the Dow Jones Index as a substitute asset in portfolios. Apart from these variables we allow for commodity trends, time and individual effects and their possible interactions.

Using the above depicted information set\(^7\) the following equation was selected to explain commodity prices,\(^8\)

\[
\Delta p = 5.630 - 0.184 \Delta p_{t-1} - 0.211 Iq_{t-1} - 0.364 \Delta USE_{t-1} + 0.126 \Delta chinagdp_{t-1} \\
+ 0.106 \Delta p_{t-1} - 0.127 \Delta p_{t-2} + 0.868 \Delta loecdgd_{t-1} + 0.200 \Delta chinagdp_{t-1} \\
+ 0.249 \Delta LUSMB_{t-1} - 0.425 \Delta LUSE_{t-1} - 0.100 \Delta lnv_{t-1} + 5.568 \Delta p_{t-1}*y_{1974} \\
+ 0.004 \text{trend}*go + 0.005 \text{trend}*pe - 0.006 \text{trend}*bf - 0.003 \text{trend}*wh \\
+ 0.354 y_{1973} + 0.165 y_{1983} - 0.211 y_{2009} - 0.560 al_{73} + 0.376 co_{64} \\
- 0.628 co_{06} + 0.346 or_{74} + 0.561 or_{80} - 0.388 pe_{73} + 0.489 pe_{79} \\
- 0.593 pe_{86} - 0.472 pe_{88} - 0.417 pe_{98} - 0.442 pe_{09} - 0.339 cb_{74} \\
- 0.112 co - 1.830 go + 1.608 pe - 2.529 bf - 2.371 cn - 1.089 so - 0.684 wh \quad (2)
\]

\[
\begin{align*}
\Delta p & = 5.630 & \text{coefficient} & \pm 0.184 & \Delta p_{t-1} & \pm 0.211 & Iq_{t-1} & \pm 0.364 & \Delta USE_{t-1} & \pm 0.126 & \Delta chinagdp_{t-1} & \\
& & \text{standard error} & & \text{standard error} & & \text{standard error} & & \text{standard error} & & \text{standard error} & \\
& & R^2 = 0.664 & & \hat{\sigma} = 0.119 & & F_{pr}(2,343) = 0.968 & \text{[0.381]} & F_{arch}(1,382) = 3.042 & \text{[0.082]} & X^2_{nd}(2) = 1.046 & \text{[0.593]} \\
& & & & & & F_{het}(40,330) = 1.471 & \text{[0.038]} & F_{X_{het}}(139,231) = 1.154 & \text{[0.168]} & & \\
& & & & & & F_{reset}(2,343) = 0.974 & \text{[0.379]} & N = 8 & & T = 48 (1963 – 2010) \\
\end{align*}
\]

\(\text{\footnotesize \# We had occasionally to limit the information set due to colinearities and reselect variables which turned out to have signs different from expected.} \)

\(\text{\footnotesize \# Equation (2) reports diagnostic statistics for testing residual autocorrelation (AR), autoregressive conditional heteroscedasticity (ARCH), normality (nd), heteroscedasticity (HET and X-HET which uses squares and cross terms of the original regressors) and RESET (RESET). See Doornik and Hendry (2009) for details and references.} \)
Additionally we computed the following statistics:\(^9\)

\[
\begin{align*}
AC(1) \text{ DPD: } & -0.656 \ [0.512] \\
AC(2) \text{ DPD: } & -1.379 \ [0.168] \\
\text{Breusch-Pagan LM test of independence}\(^{10}\): & \chi^2 (28) = 49.853 \ [0.007]
\end{align*}
\]

Equation (2) is theory consistent and passes most of the diagnostic tests. However, since homoscedasticity (at 5\%) and no cross section dependence can be rejected we report, underneath the equation, the OLS standard errors in parentheses, the robust standard errors (from DPD) in brackets and Driscolly-Kraay\(^{11}\) standard errors in curly brackets. The effects of the all explanatory variables are maintained. It should be noted that the high correlations that appear not to be accounted by the model are those of the agricultural products, e.g. corn-wheat, as shown in table 1.

The estimated equation encompassed LR and SR effects. First we focus on the LR effects derived from the one lagged price, exchange rate, production and China’s GDP (shown in Figure 4), apart from different trends for some commodities. Since we can reparameterise the LR effects of Equation (2) as an Equilibrium Correction (EqC) term, the solved LR solution implicit in this equation is,

\[
lp_t = \text{constant} - 1.146 LQ_{t-1} - 1.975 LUSE_{t-1} + 0.683 \text{Ichinagdp}_t + 0.022 \text{trend}^{*}go \\
+ 0.030 \text{trend}^{*}pe - 0.034 \text{trend}^{*}bf - 0.019 \text{trend}^{*}wh + 1.923 y_{1973} + 0.895 y_{1983} \\
- 1.148 y_{2009}
\]

which shows the expected sign of the main explanatory variables. The adjustment coefficient of this EqC term in Equation (2) is 0.18 in the first year, significant at all traditional levels and using the different SE.

However, we have assumed that in the model of Equation (2) all variables are stationary, either by differentiation or cointegration, and therefore traditional inference can be performed (see Sims, Stock and Watson, 1990). Although it has been suggested that the problem of spurious correlation is less serious for models with cross section dimension of the data, standard univariate unit root tests (see Appendix 2) indicate that some of the variables, specifically real commodity prices would be I(1). In this case non-standard critical values are required to reject the hypothesis of no cointegration in models parameterized as in Equation (2).\(^{12}\) As explained in Appendix 3, critical values for the t- ratio statistic have been calculated by Monte Carlo for time series and time-series panel data without cross-sectional dependence. However, they should be obtained by a bootstrapping approach when cross dependency is present as in the case analysed. According to the empirical distribution obtained from the bootstrapping of Appendix 3, the null of no cointegration is rejected at

---

\(^9\) Although the diagnostic tests of Autometrics are set for time series we used the default setting for TSCS as well, but after estimation we test for autocorrelation using the statistics of Panel Data Models in Oxmetrics (DPD). For automatic selection the LM AR(2), for example, is calculated using only a few observations (24/384) “contaminated” by the TSCS structure of our data (\(N=8; T=48\)). Although it may be biased in favor of the null when selecting models it has proved to be a more useful approach than omitting the time series statistics from the diagnostic tests.

\(^{10}\) Breusch and Pagan (1980) LM test was obtained from Stata.11 to detect whether the residuals of the estimated fixed-effect model were cross-sectional independent or not.

\(^{11}\) By relying on large T asymptotics, Driscoll and Kraay (1995) proposed a correction of standard errors for spatial correlation in panel data, which are those used in Equation (2). They were computed in Stata.11.

\(^{12}\) Statistics of the same type in Appendix 2 indicate that the EqC term of Equation (3) is apparently not an integrated variable.
traditional levels (the lower 1% and 5% quantiles of the t-ratio distribution plotted in Figure A1e) are -6.53 and -5.78, respectively).

In equation (2) and (3) we can observe the effects of variables entering the LR. We found a negative elasticity of real commodity prices (in dollars) with respect to the level of the nominal US exchange rate (as well as the negative effect of its last year change). The magnitude of the LR elasticity is larger than unity in absolute value. It appears to be higher than expected from simple theoretical models, which assume exchange rates in real terms. However, our finding corroborates the fact that commodities prices are more flexible than those in the US CPI. Overshooting was also found in the empirical literature using different approaches.

We can also observe the negative elasticity of real commodity prices with respect to their production. Thus, the normalization of the model prices adjusting to excess supply appears to be compatible with the data. Demand is in the LR depending on the evolution of China GDP. Figure 4 allow us to observe this LR supply/demand relationship.

Figure 4. Production and China real GDP (in logs)

Although at first sight it may be a bit surprising that other GDPs do not enter the LR (the estimated coefficient of $loecdgdp$ was negative) the next figure shows how similar the behavior of the OECD and China’s GDP have been over the sample. However, the results indicate that China’s performance have additional effects on commodity prices in the LR, particularly since the 1990s. We can note that for the SR both growth rates are significant but in this case the effect of the OECD is larger.

---

13 We could not find a measure of the US real exchange rate since 1960. The effects of inappropriate US Exchange rate index are discussed in Gilbert (1989).
An interesting result we found is that both production and inventories are detected as significant in Equation (2) but the role of inventories is only for the SR, see Figures 2 and 3. Since it is measured as $\Delta_2\text{linv}_t = (\Delta\text{linv}_t + \Delta\text{linv}_{t-1})$, short run movements of excess supply over demand (in the last two years relative to the 2-year lagged level) negatively affect real prices. Again the negative sign suggests that real prices adjust to inventory changes and not vice versa, an issue further considered in section 5.3.

Given the LR effects of production and the proxy for demand inventory changes may be indicating nonlinearities for the SR effects of excess supply. The role of production and demand (worldwide instead of just from China) in the LR and inventories in the SR has been previously suggested by Deaton and Laroque (1992, 1995, 2003). They are nested in Equation (2).

Taking into account monetary effects we found that the one-year lagged growth of the US monetary base has a positive sign, about 0.25. However, similarly to Frankel and Rose (2010) the selected model does not include interest rates, a result discussed in the next section. We can note that we also find that the US exchange rate depreciation has a negative effect after one period, which may be also associated with the US monetary policy.

5. Sensitivity results

In this section we discuss some interesting results obtained from small modifications of the basic setup of the model selected. They involve the role of the interest rates (in Section 5.1) and the assumption of equal coefficients across commodities and over time (in Section 5.2). Then, exogeneity issues are analysed (in Section 5.3).

5.1. The effect of the interest rate on commodity prices

In our model there is no effect from the interest rates. Instead, we found that the US monetary policy may have a short run effect due to the growth rate of one-year lagged monetary base and maybe through its effect on the exchange rate depreciation, which should be determined by both money and real factors. Could the effect of inventories be already capturing those of the interest rates? Since the channel of interest rates has been widely suggested in the literature, as reviewed in Section 2, we reconsidered -given Equation (2) estimates- the effect of the 1-year Treasury maturity constant rate and the Federal fund rate in logs and their two lags (keeping fixed the main explanatory variables of Equation (2) in Autometrics). We found that the interest rate is now significant -as it is changes in inventories- but the growth of
money base is not any longer. Excluding the effects of the monetary base and the gold and petroleum trends the estimated equation is,

\[
\Delta p_t = 5.424 - 0.128 \Delta p_{t-1} - 0.235 \Delta q_{t-1} - 0.286 \Delta \text{LUSE}_{t-1} + 0.177 \Delta \text{chinagdp}_{t-1} + 0.147 \Delta \text{chinagdp}_t - 0.441 \Delta \text{LUSE}_{t-1} - 0.106 \Delta \text{linv}_t + 5.588 \Delta \text{lp}_{t-1} \cdot y_{1974} - 1.248 \text{li}_{t-1} + 0.061 \Delta \text{lp}_{t-1} - 0.168 \Delta \text{lp}_{t-2} + 0.604 \Delta \text{loecdgdp}_t
\]

\[
+ 0.147 \Delta \text{chinagdp}_t - 0.441 \Delta \text{LUSE}_{t-1} - 0.106 \Delta \text{linv}_t + 5.588 \Delta \text{lp}_{t-1} \cdot y_{1974} - 0.007 \text{trend}^* \text{bf} - 0.004 \text{trend}^* \text{wh} + 0.371 y_{1973} + 0.166 y_{1983} - 0.282 y_{2009} - 0.552 a_{73} + 0.407 c_{o64} + 0.599 c_{o06} + 0.360 o_{74} + 0.547 o_{80} - 0.393 p_{e73} + 0.468 p_{e79} - 0.607 p_{e86} - 0.471 p_{e88} - 0.363 p_{e98} - 0.401 p_{e09} - 0.349 c_{b74} - 0.137 c_{o} - 2.004 g_{o} + 1.864 p_{e} - 2.823 b_{f} - 2.632 c_{n} - 1.217 s_{o} - 0.747 w_{h}
\]

\[
\text{R}^2 = 0.655 \quad \hat{\theta} = 0.120
\]

In Equation (4), apart from some differences in the deterministic components, the effects of all the explanatory variables are maintained even the log differences of inventories, except for \(\Delta \text{lp}_{t-1}\). Therefore, the interest rate denoted as \(\text{li}_{t-1}\) appears in Equation (4) as the channel of the US monetary policy, which is instead measured by the growth rate of the monetary base in Equation (2).

5.2. Testing individual and time heterogeneity

The estimated model assumes that individual and time heterogeneity are captured by commodity trends, commodity and year fixed effects and their interactions. However, in this section we present the results of testing the hypothesis of equal coefficients (elasticities and marginal effects) across commodities and over time in Equation (2). To perform this evaluation we include multiplicative dummies of all the variables in Equation (2) for each of the 8 commodities and after 1974 and 2000 (see Figure 1). No different effects are detected.

\[
\text{F}_{\text{het}}(36,334) = 1.350 \quad \text{F}_{\text{Xhet}}(113,257) = 1.126 \quad T = (48) 1963 - 2010.
\]

AC(1) DPD: -0.424 [0.672]  AC(2) DPD: -1.548 [0.122]
after 1974 or 2000, but regarding commodity heterogeneity we found a different effect of the China demand on petroleum prices\textsuperscript{15}, that is,

\[ \Delta l_p = 6.227 - 0.217 l_{p,t-1} - 0.229 l_{q,t-1} - 0.396 LUSE_{t-1} + 0.121 l\text{chinagdp}_{t-1} \]

\[ + 0.166 l\text{chinagdp}_{t-1}^* pe + 0.118 \Delta l_p - 0.122 \Delta l_p_{t-2} + 0.873 \Delta l\text{oecdgdp}_t \]

\[ + 0.188 \Delta l\text{chinagdp}_t - 0.419 \Delta LUSE_{t-1} + 0.245 \Delta LUSMB_{t-1} - 0.095 \Delta l\text{linv}_d \]

\[ + 5.320 \Delta l_p_{t-1} y1974 + 0.005 trend*or - 0.007 trend*bf - 0.004 trend*wh + 0.351 y1973 \]

\[ + 0.168 y1983 - 0.200 y2009 - 0.560 al73 + 0.359 co64 + 0.633 co06 + 0.343 or74 \]

\[ + 0.575 or80 - 0.402 pe73 + 0.442 pe79 - 0.600 pe86 - 0.477 pe88 - 0.378 pe98 \]

\[ - 0.535 pe09 - 0.328 cb74 - 0.116 co - 1.987 go + 1.534 pe - 2.744 bf - 2.573 cn \]

\[ - 1.181 so - 0.744 wh \]

\[ R^2 = 0.674 \]

\[ \hat{\rho} = 0.117 \]

\[ F_{ar}(2,343) = 1.374 [0.255] \]

\[ F_{arch}(1,382) = 2.767 [0.097] \]

\[ \chi^2_{nd}(2) = 1.469 [0.480] \]

\[ F_{het}(40,330) = 1.262 [0.142] \]

\[ F_{het}(139,231) = 1.208 [0.103] \]

\[ F_{RESET}(2,343) = 0.554 [0.575] \]

\[ N = 8 \]

\[ T = 48 (1963 – 2010). \]

\[ AC(1) DPD: -0.919 [0.358] \]

\[ AC(2) DPD: -1.374 [0.169] \]

\[ Breusch-Pagan LM test of independence: \chi^2 (28) = 47.587 [0.012] \]

Equation (5) estimates are quite similar to those of Equation (2), the main effects remains except for the effect of the petroleum trend (trend*pe replaced by lchinagdp_{t-1}^*pe ); also homocedasticity is not rejected at 5%. The next figure may explain the stronger LR effect of China GDP on the real petroleum price. During the sample China passed from being a net petroleum exporter to a net importer.

\textsuperscript{15} Automatic selection was performed using a 1% target size.
Therefore, Equation (5) improves the estimates of Equation (2) after testing heterogeneity. Cointegration can be assumed for this equation too as, according to the empirical distribution obtained from the bootstrapping of Appendix 3, the null of no cointegration is rejected at traditional levels (the lower 1% and 5% quantiles of the t-ratio distribution plotted in Figure A1f) are -5.72 and -5.86, respectively. The next subsection discusses exogeneity using the estimates of Equation (5).

5.3. Testing exogeneity

5.3.1 Commodity inventories

Since we found a contemporaneous effect of $\Delta_2 \text{linv}_t = (\Delta \text{linv}_t + \Delta \text{linv}_{t-1})$ on prices we test by instrumental variables (IV) if Equation (5) can be considered as a valid conditional model on this variable. We employed as instruments: the log difference of the first lag of inventories ($\Delta \text{linv}_{t-1}$), the log difference of the first lag of China’s GDP ($\Delta \text{chinagdp}_{t-1}$) and the log difference of the first lag of the OECD’s GDP ($\Delta \text{oecdgdp}_{t-1}$).

The IV estimation is,

$$\Delta p_t = 6.116 - 0.211 \Delta p_{t-1} - 0.223 \Delta q_{t-1} - 0.398 \Delta \text{LUSE}_{t-1} + 0.120 \Delta \text{chinagdp}_{t-1}$$

$$(0.872) (0.028) (0.040) (0.090) (0.024)$$

$$+ 0.162 \text{chinagdp}^* \text{pe} + 0.112 \Delta \text{lp}_{t-1} - 0.123 \Delta \text{lp}_{t-2} + 0.871 \Delta \text{oecdgdp}_{t-1}$$

$$(0.034) (0.038) (0.038) (0.161)$$

$$+ 0.191 \Delta \text{chinagdp}_{t-1} + 0.244 \Delta \text{LUSMB}_{t-1} - 0.395 \Delta \text{LUSE}_{t-1} - 0.115 \Delta_2 \text{linv}_t$$

$$(0.068) (0.092) (0.146) (0.026)$$

$$+ 5.351 \Delta \text{lp}_{t-1} \cdot y_{1974} + 0.005 \text{trend}^* \text{go} - 0.006 \text{trend}^* \text{bf} - 0.003 \text{trend}^* \text{wh}$$

$$(0.776) (0.002) (0.007) (0.793)$$

$$+ 0.353 y_{1973} + 0.163 y_{1983} - 0.194 y_{2009} - 0.568 a_{173} + 0.360 c_{064}$$

$$(0.052) (0.047) (0.050) (0.129) (0.120)$$

$$+ 0.641 c_{06} + 0.346 o_{74} + 0.575 o_{80} - 0.403 p_{73} + 0.441 p_{79}$$

$$(0.121) (0.122) (0.121) (0.130) (0.120)$$

$$- 0.603 p_{86} - 0.474 p_{88} - 0.379 p_{98} - 0.449 p_{09} - 0.540 c_{b74}$$
Equation (6) shows similar results to Equation (5), in particular the estimated coefficient of $\Delta linv_t$ is only slightly higher than its OLS estimates. Therefore, it suggests that commodity prices can be modelled conditioning on inventory changes.

5.3.2 Commodity production

The estimated equations have been developed as conditional EqC models which require weak exogeneity of the variables entering the LR relationships. For inference about the LR parameters, only commodity prices should adjust to deviations from this LR relationship but commodity productions may adjust too. In order to evaluate this possibility we also estimated a model for production using the same information set and including the EqC term estimated in Eq(5). The next equation show the results,

$$
\Delta lq_t = 0.533 + 0.010 EqCM - 0.166 \Delta lq_{t-1} - 0.025 \Delta linv_{t-1} - 0.014 \Delta lp_{t-1} \\
+ 0.005 \Delta lp_{t-2} - 0.002 lp_{t-1} - 0.018 lq_{t-1} - 0.002 lchinagdp_{t-1} \\
+ 0.017 LUSE_{t-1} - 0.178 al82 - 0.213 so74 - 0.207 so76 + 0.147 so77 \\
- 0.210 so80 - 0.269 so83 + 0.215 so86 - 0.039 co - 0.326 go \\
+ 0.140 pe - 0.406 bf - 0.342 cn - 0.142 so - 0.132 wh
$$

To take into account a different normalisation, quantities conditional on prices, levels of the variables entering the LR were also included. As shown in Equation (7) none of them was significant.
The EqC term is not significant. Only $\Delta lq_{it1}$ and $\Delta linv_{it1}$ were significant and their negative signs suggest that as production and inventories grow in the previous year, current production decreases. Therefore, production seems to be weak exogeneous and a valid conditional model can be obtained for commodity prices.

6 Conclusions

This paper has studied the main explanatory variables of real prices for eight different commodities, which have a significant weight in Argentina’s trade accounts. Using a time series - cross section model we considered a large information set that included monetary and financial variables along with the supply and demand determinants of different goods. Their effects in the short run and long run have jointly been modelled. Given the kind of data used, we also allow for time and commodity heterogeneity by including commodity trends, individual and year fixed effects and their interactions. Following a general to specific approach, Autometrics, an algorithm for computer-automated model selection helped us to choose an appropriate model for commodity prices. It was also useful for testing other misspecification like heterogeneous coefficients across commodities and over time.

We found that the US exchange rate, commodity productions (from the supply side) and China’s GDP (from the demand side) has effect on real prices in the long run. Apart from their growth rates, those of the OECD’s GDP and inventories have short run effects. Regarding monetary variables, we also found a significant positive effect of changes in the US monetary base on the change in commodity prices. The role of interest rates has also been analysed. Interest rates may be the channel by which the US monetary base affects commodity prices.

Testing for commodity and time heterogeneous coefficients, only a different effect of China on prices was detected for petroleum in the long run. Therefore, taking into account this effect we found an econometric model which explains real commodity prices for both before and after the 2000s. According to the tests examined, it is a valid conditional equilibrium correction model for the cross section-time series data. Although the commodities we studied are the relevant for the Argentine economy they include agricultural, minerals and even petroleum. Once heterogeneity is modelled as in this paper the empirical analysis can be easily extended to a larger commodity set in a future research.
Appendix 1: Data definitions and sources

All data are yearly over 1960-2010. Data on prices, productions and inventories are across eight commodities: aluminum (al), copper (co), gold\(^{17}\) (go), petroleum (pe), beef (bf), corn (cn), soybeans (so) and wheat (wh). The variables in the reported equations (where \(l\) indicates the corresponding logarithms) are described in the next table,

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p)</td>
<td>Real price</td>
<td>IMF and World Gold Council</td>
</tr>
<tr>
<td>(q)</td>
<td>Production</td>
<td>USDA, USGS and EIA</td>
</tr>
<tr>
<td>(inv)</td>
<td>Inventories</td>
<td>USDA, USGS, EIA, World Gold Council, FAO</td>
</tr>
<tr>
<td>chinagdp</td>
<td>China’s real GDP</td>
<td>IMF</td>
</tr>
<tr>
<td>oecdgdp</td>
<td>OECD’s real GDP</td>
<td>World Bank – WDI</td>
</tr>
<tr>
<td>(i)</td>
<td>1-year treasury constant maturity rate</td>
<td>St Louis Fed</td>
</tr>
<tr>
<td>USE</td>
<td>US nominal effective exchange rate</td>
<td>St Louis Fed</td>
</tr>
<tr>
<td>USMB</td>
<td>US monetary base</td>
<td>Federal Reserve</td>
</tr>
</tbody>
</table>

**Prices:** nominal prices were deflated by US CPI (IMF), 1960=100. Their volatilities were calculated using monthly data as the standard deviations of the previous 12-months prices. Petroleum prices were measured as simple averages between WTI and Brent.

**Production:** world production was used for all commodities 1960 = 100. Since world data was not available for petroleum between 1960 and 1972 this series was backward extended using US data.

**Inventories:** world inventories, except copper and aluminum for which U.S data are used, 1960=100. Beef inventories were measured considering live cattle stocks (millions of heads).

**Gross Domestic Products / Industrial Production / Trade participation:** US, OECD, China and India gross domestic products in real terms. Trade shares of China, India and BRIC (Brazil, Russia, India and China) countries were measured by exports plus imports relative to the world. Industrial production (index 2005=100) of advanced economies, US and India were obtained from the IMF.

**Monetary Aggregates:** US monetary base, US M2 and global liquidity (measured as the US monetary base plus world total reserves).

**Interest Rates:** the 1-year US Treasury constant maturity rate and the Federal fund rate.

**Exchange Rate:** the US nominal effective exchange rate (2005=100).

\(^{17}\) Gold for not monetary uses is considered.
### Appendix 2: Unit root tests

#### Time and commodity-varying variables:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Levin-Lin-Chu</th>
<th>Fisher</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(c + \text{trend})</td>
<td>(c)</td>
</tr>
<tr>
<td>Lp</td>
<td>-0.776***</td>
<td>-1.617*</td>
</tr>
<tr>
<td>LQ</td>
<td>-2.560***</td>
<td>-3.462***</td>
</tr>
<tr>
<td>LINV</td>
<td>-2.557***</td>
<td>-3.310***</td>
</tr>
<tr>
<td>D.Lp</td>
<td>-10.930***</td>
<td>-11.822***</td>
</tr>
<tr>
<td>D.LQ</td>
<td>-7.673***</td>
<td>-8.176***</td>
</tr>
<tr>
<td>D.LINV</td>
<td>-8.538***</td>
<td>-8.495***</td>
</tr>
</tbody>
</table>

Notes: Levin-Lin-Chu (2002) unit root test assumes a common autoregressive parameter for all panels and requires that the number of periods grow more quickly than the number of panels \(N/T\) tends to zero; the adjusted t statistic is reported. The Fisher-type (Choi 2001) test performs a unit root test based on augmented Dickey-Fuller tests and Phillips-Perron tests and requires the number of periods tend to infinity, both were performed with 2 lags length, the inverse chi-squared statistic is reported.

#### Time-varying variables:

<table>
<thead>
<tr>
<th>Variable</th>
<th>ADF</th>
<th>PP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(T_1)</td>
<td>(T_\mu)</td>
</tr>
<tr>
<td>Ichinagdp</td>
<td>-0.558</td>
<td>0.977</td>
</tr>
<tr>
<td>LUSE</td>
<td>-1.778</td>
<td>-1.922</td>
</tr>
<tr>
<td>Li</td>
<td>-1.651</td>
<td>-1.267</td>
</tr>
<tr>
<td>D.Ichinagdp</td>
<td>-5.238***</td>
<td>-5.352***</td>
</tr>
<tr>
<td>D.LUSE</td>
<td>-4.309***</td>
<td>-4.308***</td>
</tr>
<tr>
<td>D.li</td>
<td>-5.216***</td>
<td>-5.073***</td>
</tr>
</tbody>
</table>

Notes: The null hypothesis for the ADF and PP tests is that of unit root against the alternative of stationarity.

#### Long-run term:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Levin-Lin-Chu</th>
<th>Fisher</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(c + \text{trend})</td>
<td>(c)</td>
</tr>
<tr>
<td>EqC</td>
<td>-11.851***</td>
<td>-10.844***</td>
</tr>
</tbody>
</table>

Notes: The Fisher-type tests were performed with 2 lags length.

---

18 * indicate significance at the 0.10 level, ** indicate significance at the 0.05 level, and *** indicate significance at the 0.01 level.
Appendix 3: Bootstrapping for testing cointegration with cross-sectional dependence

When the variables which are part of the long run relationship are I(1) testing for cointegration requires non-standard critical values. In the EqC form of the conditional autoregressive distributed lag model, cointegration can be tested by the significance of the lagged explained variable (or $lp_{it}$ in the case of Equation (1)) using a t-statistic for its coefficient estimates. This is analogous to the “PcGive unit root test” in Hendry (1989, p.149) for time series, whose critical values have been calculated by Monte Carlo simulation by Banerjee et al. (1993) and Ericsson and Mackinnon (2002) (see also Hendry and Doornik, 2009). For panel data, Westerlund (2007) proposes related statistics and derived critical values assuming no cross-sectional dependence for large $T$\textsuperscript{19}. When such dependence is allowed he suggests a bootstrap approach.

For the cross section time series model ($N=8$, $T=48$) of this paper the hypothesis of no cross sectional dependence has been rejected, indicating that correlations between pairs of individual residuals of agricultural prices are not zero. Therefore, the null of no cointegration (a zero coefficient for $lp_{it}$) should be tested using critical values obtained by bootstrapping. Thus, similarly to Westerlund (2007), we obtained the model for $\Delta lp_t$ under the null of no cointegration but residual resampling was performed as suggested by Hansen (2000) and Hansen and Seo (2002) for heteroskedasticity. For this bootstrap, residuals ($w_{it}$) are obtained as follows,

$$w_{it} = e_{it} * u_{it} \text{ where } u_{it} \sim iid N(0,1)$$

where the regression residuals ($e_{it}$) are held fixed at their sample values across replications.

While using a Hansen’s approach for the case of panel data $w_{it}$ with $u_{it} \sim iid N(0,1)$ allows for heteroskedastic errors, that is,

$$E\left[(NT)^{-1} \sum_{i,t} w_{it}\right] = E\left[(NT)^{-1} \sum_{i,j} e_{it} u_{it}\right] = (NT)^{-1} \sum_{i,j} e_{it} E[u_{it}] = 0 \quad \text{(A1)}$$

$$E\left[(NT - K)^{-1} \sum_{i,j,t,s} w_{it} w_{js}\right] = E\left[(NT - K)^{-1} \sum_{i,j,t,s} (e_{it} u_{it})(e_{js} u_{js})\right]$$

$$= (NT - K)^{-1} \sum_{i,j,t,s} (e_{it} e_{js}) E[u_{it} u_{js}] = \sigma_{it}^2 \quad \text{(A2)}$$

when $i=j$ and $t=s$ and zero otherwise, a small modification of the distribution of $u_{it}$.

$u_{it} \sim iid N(1,1)$ also allows for cross-sectional dependence of the regression residuals. A1 (provided a constant is included in the regression) and A2 are maintained and

\textsuperscript{19}Reported simulation results start at $N=10$ and $T=100$. 

21
\[ E \left[ (NT - K)^{-1} \sum_{i,j} \sum_{t,s} w_{it} w_{js} \right] = E \left[ (NT - K)^{-1} \sum_{i,j} \sum_{t,s} (e_{it} u_{st})(e_{js} u_{ts}) \right] \]

\[ = (NT - K)^{-1} \sum_{i,j} \sum_{t,s} (e_{it} e_{js}) E[u_{it}u_{js}] \]

\[ = (NT - K)^{-1} \sum_{i,j} \sum_{t,s} (e_{it} e_{js}) E[u_{it}]E[u_{js}] = \hat{\gamma}_{ij} \quad \text{(A3)} \]

for \( i \neq j, t=s \) and zero otherwise, where \( \hat{\gamma}_{ij} \) are the residual covariances. Therefore pairs of cross sectional dependence are conserved.\(^{20}\)

In each case 1000 replications were performed. Before obtaining the distribution under the null of no cointegration, a simulation analysis was made under the alternative of cointegration assuming the estimated \( \Delta p_t \) of Equation (2) and (5) as the DGP and resampling residuals. To evaluate the effect of cross sectional dependence on the distribution of the t-ratio Figures A1 a) and c) for Equation (2) and b) and d) for Equation (5) show the t-statistics under the alternative hypothesis for \( w_{it} = e_{it}^* u_{it} \) where \( u_{it} \sim iid \mathcal{N}(1,1) \) (c) and d)) and \( w_{it} = u_{it} \) where \( u_{it} \sim iid \mathcal{N}(0,\sigma^2) \), assuming \( \sigma^2 = \hat{\sigma}^2 \) (a) and b)). They allow us to compare the effect of the cross-sectional dependence of the estimated residuals for a finite sample with the hypothetical case of no cross sectional dependence under cointegration.

Figures e) and f) correspond to the null hypothesis of no cointegration. The DGP, as suggested in Westerlund (2007) was obtained by estimating \( \Delta p_t \) without the variables in levels. We retain only significant variables at traditional levels and include the deterministic components required for adequate model specification as an indicator they enter the DGP.\(^{21}\)

The estimated equation assumed to be the DGP is,

\[ \Delta p_t = -0.085 - 0.189 \Delta p_{t-2} + 0.339 \Delta \text{chinagdp}_t + 0.696 \Delta \text{ocecdgdp}_t \]

\[ + 0.377 \Delta \text{LUSMB}_{t-1} - 0.547 \Delta \text{LUSE}_{t-1} - 0.135 \Delta y_{1973} - 0.315 y_{1973} \]

\[ (0.012) \quad (0.040) \quad (0.073) \quad (0.012) \]

\[ [0.011] \quad [0.014] \quad [0.035] \quad [0.185] \]

\[ (0.013) \quad (0.042) \quad (0.158) \quad (0.184) \]

\[ + 0.530 al73 + 0.686 co06 + 0.425 go74 + 0.629 go80 - 0.730 pe86 \]

\[ (0.147) \quad (0.139) \quad (0.138) \quad (0.138) \]

\[ [0.079] \quad [0.008] \quad [0.012] \quad [0.011] \]

\[ [0.030] \quad [0.015] \quad [0.042] \]

\[ - 0.500 pe88 - 0.556 pe09 + 5.731 \Delta p_{t-1}^*pe74 + 0.060 pe \quad \text{(A3)} \]

\[ (0.143) \quad (0.140) \quad (0.900) \quad (0.022) \]

\[ [0.017] \quad [0.019] \quad [0.077] \quad [0.003] \]

\[ [0.040] \quad [0.034] \quad [0.209] \quad [0.024] \]

\(^{20}\) Other residual covariances, say for \( i \neq j, t=s \) are also maintained.

\(^{21}\) An empirical problem is to determine which deterministic components should be included in the EqC term and which in the short run dynamics. It is a difficult issue for time-series but even more difficult for CS-TS data. The distributions plotted in figure e) and f) are conditional to the deterministic components included in the assumed DGP.
\[ \hat{\theta} = 0.137 \quad R^2 = 0.524 \]

\[
\begin{align*}
F_{ar}(2,365) &= 1.069 [0.345] & F_{arch}(1,382) &= 1.537 [0.216] & F_{het}(14,361) &= 1.878 [0.028] \\
F_{Xhet}(29,346) &= 1.894 [0.004] & \chi^2_{nd}(2) &= 4.600 [0.100] & F_{RESET}(2,365) &= 2577 [0.077] \\
N &= 8 & T &= 48 (1963 – 2010) \\
AC(1): 0.121 [0.904] & AC(2): -1.898 [0.058] \\
& \text{Breusch-Pagan LM test of independence: } \chi^2(28) = 84.627, Pr = 0.0000
\]

From the estimated \( \Delta \rho_t \) (fixed across replications) and \( w_t = e_t \cdot u_t \) where \( u_t \sim iid N(1,1) \) and \( e_t \) obtained from the last equation, we generate \( \Delta \rho_t \) and \( \rho_t \) to generate the DGP under the null (starting at zero for each group). The series were used to estimate the models of Equation (2) and (5) and calculate the t-values of \( \rho_{t,v} \) under the null. The distribution is reported in Figure e) and f) and commented in Section 4.

**Figure A1.** Bootstrap distribution of the t-ratios under cointegration, without and with cross sectional dependence and under the null of no cointegration and cross sectional dependence for Equation (2) and (5).
References


