UNIVERSITÀ POLITECNICA DELLE MARCHE
Facoltà di Economia G.Fuà

DIPARTIMENTO DI SCIENZE ECONOMICHE E SOCIALI (DISES)
Ph.D. degree in Economics - XXXII cycle

Doctoral Thesis

The dynamics of commodity prices: common movement and latent factors

Author:
Chiara Casoli

Supervisor:
Prof. R. Esposti

Contents

List of Figures

List of Tables

Ringraziamenti

Questa tesi è il frutto di tre anni di duro lavoro, in cui sono cresciuta non solo professionalmente ma anche e soprattutto introspettivamente. Non avrei raggiunto questo traguardo senza il supporto di tante persone, che mi hanno affiancata durante tutto questo tragitto. Grazie a chi già c'era, e a chi ho incontrato solo adesso. Alla famiglia, a Giovanni, agli amici, ai colleghi, ai professori.

Un ringraziamento speciale ai prof. Esposti e Lucchetti, per il supporto e per aver creduto in me (anche e soprattutto quando non ci credevo io), ai coinquilini fantasma di Villa Maratta Emanuele e Thomas, ma soprattutto ai miei due inseparabili "papermates" Luca e Francesco.

A tutti voi, **grazie**.

TRUTH, n. An ingenious compound of desirability and appearance. Discovery of truth is the sole purpose of philosophy, which is the most ancient occupation of the human mind and has a fair prospect of existing with increasing activity to the end of time.

The Devil's Dictionary
AMBROCE BIERCE

Introduction

Many economists tried to shed light on the nature and determinants of commodity markets. Understanding their intrinsic characteristics could have important policy implications especially for less developed countries, since they depend heavily on export of primary goods. More than this, commodities enter in the production chain of almost all other goods. For this reason, there is a huge amount of interest on commodity prices and their dynamics, ranging from the development of *ad hoc* theoretical models to the analysis of the corresponding stochastic processes. Analysing commodity prices is challenging and controversial; commodities are indeed goods with specific and unique features which make them different from all other assets, comprehending their possibility to be stored. It follows that their prices are determined and behave in a particular and distinctive way.

Literature has devoted attention to the topic from several perspectives. For what concerns the main trends in a long-run perspective, whereas for many decades the focus has been on the existence of a declining trend of primary commodity prices, later on attention moved towards the possibility of a new period of scarcity of resources, spread by the recent peak of the mid-2000s. Another stylised fact often investigated is the tendency of different commodity prices to share the same fluctuations. Even if this common movement may be associated with an interdependency among prices of different kinds of commodities or caused by some common drivers that influence some independent processes, there is evidence that one of these - or a combination of both - is in act. Analysing co-movement is important because of the need to understand - if evidence for this common movement is actually found - which of the above cited mechanisms are in facts the responsible, and because it is not yet fully consolidated in literature if this common movement regards the short-term dynamics or tends to persist over a long-term horizon.

This work examines the topic of co-movement of commodity prices belonging to different categories, with the aim of give an answer to two main questions: *is there actually a co-movement among the different commodity prices?* and if yes: *with which features? How does this common movement behave? Does it imply short- or long-term common dynamics?* These aspects are not trivial since they can offer important tools for understanding the intrinsic nature of commodity markets. A long-term common movement clearly has more to say about the trend of commodity prices: if it is declining, then the hypothesis of the Prebisch (1962) and Singer (1950) of declining prices (PSH) with respect to manufactures is confirmed; if it is rising, then the worry of the new resource scarcity era has one supporting point. This research question opens the field for the two main objectives of this work: from the one hand, trying to encompass co-movement (split into short- and long-term and represented by two corresponding latent factors) into a market fundamentals model, and from the other, trying to empirically assess the same challenge with the estimation of these unobservable variables.

This work is developed into three Chapters focusing on different aspects of commodity prices, with the common aim to determine *if* these prices co-move and *how*. Chapter 1 provides a literature review of commodity price dynamics, comprehending the PSH, the stochastic processes analysis, the study of the main price drivers and the common movement. It also provide the first results regarding the univariate features of the series, concentrating on a set of commodity prices selected from the IMF public database. Chapter 2 consists in the first attempt, to our knowledge, of developing a multi-commodity market model with latent factors capturing the co-movement. Main inspirations come from the model of Gilbert (1995) and from the factor structure imposed to commodity prices by Schwartz (1997); Schwartz and Smith (2000). Unfortunately, the model has proven to be not estimable due to the well known convergence problems of the Kalman filtering techniques. Nevertheless, we believe that further research could substantially improve the model - or take inspiration from it - to reach more solid solutions and to impose appropriate parameter restrictions in order to recover the structural interpretation (this cannot be done if estimation with parametric techniques cannot be performed). This, however, has opened the way for the development of a suitable estimation method, which presentation is the core of Chapter 3 and constitute the main contribution of the entire work. Since we have not been able to estimate the theoretical model with standard techniques, we have tried to look at the co-movement matter from the opposite perspective: instead of starting from theory, letting "data speak as freely as possible" (Barigozzi and Luciani, 2017). The proposed methodology combines a decomposition in Transitory and Permanent components as in Gonzalo and Granger (1995) and Dynamic Factor Models estimation, (Stock and Watson, 2011; Bai and Ng, 2004; Doz et al., 2012). The new algorithm is also exploited to estimate the price equation obtained with the theoretical model of Chapter 2, filling part of the related gap, but in Chapter 3 it is applied to data with no constraints deriving from any sort of modelling choices.

What emerges from the present work is that commodity prices are to be considered as non-stationary, and share a common movement that is in part explained by interdependencies among commodity prices belonging to the same group (meaning that each price influences its complements and substitutes and compete for the same fixed resources for production), and in part is instead driven by the same common factors. Specifically, the permanent, or long-run, common movement counts more with respect to the transitory or short term common movement, which is rather marginal, in determining the price of each commodity. By looking at the long-run extracted factors, however, it is impossible to determine if there is in act a clear tendency of prices to rise or decline; rather, there are categories of commodities in which trends have a precise directions. For instance, food prices seem to share a declining trend over the time span, whereas livestock commodity prices are subject to a slightly upsurging trend. This could mean that for some categories the demand/supply pressure is becoming higher, and in particular the explanation holds for livestocks, which are facing a rapid growth in the demand side.

Chapter 1

On the commodity price dynamics: trends, drivers and common movement

1.1 Introduction

Recent commodity prices surge in both levels and volatilities, occurred during the mid-2000s, has spread the worry of a new resource scarcity era, based on the assumption that the world is running out of raw materials. This picture shows a scenario in which the increasing burden of demand for different kinds of primary resources over a more and more constrained supply would generate an upward pressure on prices of many commodities, causing what as been called a *Great paradigm shift* (Grantham, 2011). According to this point of view, the days of declining prices are at the end, and we are entering in a new phase in which commodity prices will stay at high levels.

It is worth noting that after some years of turbulence, international prices of many primary commodities are declining again, and the most recent outlooks (FAO, 2017; IMF, 2016; World Bank, 2017) forecast a situation of moderation with stable or declining price series for the majority of commodities, thus opening new possibilities and posing several doubts about a paradigm shift.

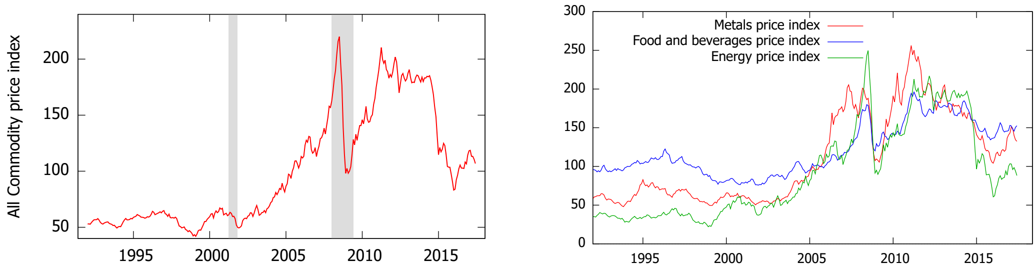
Anyway, monitoring the fluctuations of commodity prices is important for several reasons, among which the high dependence of some developing countries on the production and export of primary goods, the impact of some specific commodities on global economic activity (such as oil or gold), the increasing presence of commodity assets in portfolio allocation choices and, more crucially, the linkages between primary commodity access and poverty (i.e. agricultural prices and food security). With reference to the latter issue, agricultural commodity prices are especially scrutinised due to the negative effects that bubbles, instabilities or particularly high levels cause in terms of access to food for the poorest countries and households; the main challenge in this sense is given by the increasing pressure of rapid demographic and economic growth over food supply, which has grown more slowly with respect to global demand.

It is important to investigate the common behaviour of primary commodity prices across different markets, which for some reasons seem to share the same features, and to establish which are these reasons. Nevertheless, it is clear that before moving to the common movement analysis, some words on single series behaviour are necessary. Early literature on commodity prices has focused on a declining trend hypothesis for commodity prices (the PSH). Immediately after the introduction of the PSH, another strand of literature started to focus on the mechanisms of price formation starting from Gustafson (1958) and the introduction of the storage model, while an expanding branch started working on the empirical counterpart; as econometric techniques got more sophisticated, the quality and level of detail and accuracy of these studies has improved. Empirical works on commodity prices could be categorised into those studying the stochastic properties of time series and the existence of the long-run declining trend prophesied by the PSH, others which refer to the complex framework of interactions between spot and future prices, the determination of drivers responsible for price fluctuations, another branch examining spillover effects of shocks from some prices to others, and, more recently, analyses of common movements, exploiting cointegration first, and latent factors then.

Attention on commodity prices has been motivated on the one hand by the need of constructing appropriate policies and on the other by the challenging desire to understand the evolving dynamics of the series, which alternate periods of price stability to peaks of high levels and high volatilities. As mentioned before, commodity prices evolution is crucial especially for developing economies and for the relevance in terms of possible signal for resource scarcity periods. Upward long-term trends would indeed be an indicator of demand-side pressure, whereas periods of abundance would be matched with declining trends.

The aim of this Chapter is to present an overview of different strands of literature aiming at assessing the stochastic properties of commodity prices, the relative drivers and the common movement of different price series. With reference to the stochastic properties, some commodity prices are analysed - including a set of energy, metals and agricultural goods - by use of a battery of tests, in order to add a contribution to the study of the time series properties of commodity prices. This task is also necessary for mapping the different commodity prices depending on the stochastic processes in order to move to the multivariate framework and the common movement analysis, carried out in next Chapters.

The rest of the Chapter is organised as follows: Section 1.2 presents the stylised facts, a brief recap of literature review on commodity prices in an univariate perspective and some diagnostics on the dataset of reference. Section 1.3 shifts from the literature on the movements of price series to the causes, providing a review of the main drivers and determinants. In section 1.4 the



Note: Grey bars indicate economic recessions according to NBER.
Source: OECD and FAO (2017) indeces.

Figure 1.1: IMF Price Indeces for commodity groups, 2005 = 100

principal research question of this study is introduced, aiming at assessing the presence of a common (especially long-run) movement among different categories of commodities; the concept of co-movement of prices is analysed within the literature together with related possible explanations. Section 1.5 concludes and introduces next Chapters topics.

1.2 Price dynamics

The study of commodity price dynamics can be dated backward to the Prebisch-Singer hypothesis - and its empirical validation - that starts from the assumption according to which prices of primary commodities tend to decline over the long term relatively to prices of manufactured goods (Prebisch, 1962; Singer, 1950). Thenceforth, literature on empirical investigation of commodity price dynamics has expanded enormously, till encompassing the study of trend nature, stochastic behaviour, volatility, spillover effects and common movement.

1.2.1 Stylised facts

While during the last decades primary commodity prices where exhibiting a stable path and moderate volatility, there has been a recent change of direction starting from the half of 2000s, as depicted in figure 1.1.

All categories of commodities show an upsurge occurred before the economic fall of 2008, but in the last periods markets seem to have stabilized showing less turbulence. Evidence for this fact is found even if considering the majority of the singular series, but the impact is stronger when considering some aggregation of commodity sets, because a tendency for common movement is highlighted.

World Bank (2017), providing some short-run commodity snapshots, forecasts a slight increase for all but agricultural prices, which markets are now well supplied thanks to last seasons' favourable weather conditions. In particular, energy markets are projected to experience growing prices due to ongoing rebalances: demand for crude oil has been indeed increasing dragged by the

United States and Europe' consumption, whereas supply is on track. Coal prices are increasing driven by China's environmental policies and demand, while on the contrary natural gas price will exhibit a modest growth. For what concerns metals and minerals, prices, which surged sharply during last year, are projected to ease slightly during a short-time horizon. There is uncertainty on predicting future prices of livestock products, but with general consensus that growth in meat production is expected to be less strong in the future with respect to the last decade, together with global - and especially Chinese - demand. Actually, the declining price for livestock commodities is to be attributed to stocks and relatively low feed costs, but the demand/supply pressure issue is not yet solved.

Moving to a long-term perspective, commodity prices' pattern will depend on the understanding of which are the long run fundamentals driving and shaping the series; it is not yet clear what we should expect to happen, since there is no consensus about which are the fundamentals, among all the mechanisms, with the biggest impact of price formation.

1.2.2 The declining trend hypothesis

As mentioned, the investigation of primary commodity prices movement on a long-run perspective can be dated back to Singer (1950) and Prebisch (1962), according to which the series would exhibit a deterioration with respect to manufacture prices, due to the gradual shift towards more diversification and specialization among developing countries, as economies get richer, and the income inelasticity of primary commodities demand. Lipsey (1994) stresses that, whereas the classical view focuses on inevitable increasing costs, it is possible that the declining trend hypothesis is sensitive to many aspects in which prices are measured and collected.[1]

With reference to commodity prices, several contributions have investigated the secular decline of them, finding heterogeneous and sometimes contradicting results. Even if the presence for a declining trend for the series of primary commodity prices is originally found by the two authors, the consideration of different time spans, techniques or the commodities included in the sample has led many other studies to conclude that there is no clear sign of this downward trend.

Grilli and Yang (1988) provide an empirical investigation for the PSH by constructing an index covering prices of different types of primary commodities for the period 1900-86, and another index of unit values of manufactures exported by industrial countries (and some other variants and subgroup indices). The ratio between the two falls, on trend, by 0.5 percent a year, and by 0.6

[1] Lipsey (1994) focuses on the denominator side, i.e. the manufactured good prices, and concludes that these prices may have been overstated, especially due to the treatment of quality change. On the contrary, we will mainly focus on the numerator side (primary good prices).

per cent when fuel commodities are excluded. More specifically, the decline is much stronger for metal and agricultural (non-food) commodity prices (respectively, 0.82 and 0.84 per cent p.a.) than for food prices (0.36 per cent p.a.); on the contrary, beverage relative prices increase substantially over the covered time. Support for the PSH is also found by Reinhart and Wickham (1994), which attribute the continuous decline from the 1980s to a secular deterioration. Cuddington (1992) considers, instead of an index, 26 commodity annual prices individually, from 1900 to 1983, finding that 16 of the 26 series have to be considered trend-less, whereas of the remaining commodities, five have negative and statistically significant trends and five exhibit positive trends. In a more recent work, Yamada and Yoon (2014) show that the secular deterioration hypothesis holds sometimes, but not always, and only for a sub-group of commodities, and moreover this deterioration has become weaker recently. When reviewing the PSH, Baffes and Etienne (2016) stress that half of the empirical studies find support and the other half reject it. They moreover find that income has a negative effect on real food prices, and that the channel through which income influences food commodity prices in the long-run are the manufacturing prices (the denominator).

A crucial consideration about this debate regards the nature of the price series as *trend stationary* or *difference stationary.* A trend stationary series is a series that is stationary around a deterministic trend; a difference stationary series, instead, is an integrated series of order d, which is stationary if differenced d times (Nelson and Plosser, 1982). As pointed out by Bleaney and Greenaway (1993), the distinction is crucial when analysing trends and their nature. The debate on the consideration of commodity prices as stationary or integrated series is still open, but will be further examined in next Section.

Moreover, several other works testing for the PSH conclude that there is much evidence for structural breaks, rather than a secular declining trend, suggesting that there may be some critical events which causes fundamental shifts in the markets of commodities. Among studies focusing on these breaks in the data, concluding that the PSH cannot be considered as an universal phenomenon, the work by Bleaney and Greenaway (1993), for instance, updates the Grilli-Yang series up to 1991, showing that a one-for-all drop in prices' data after 1980 is preferred to a declining trend, while for the case of metal commodities a random walk hypothesis seems more plausible. Also Cuddington et al. (2002) prefer the explanation of one or more abrupt shifts downward, with the most evident case occurred in 1921, with total absence of positive or negative trends after or before. In addition to the break, real primary prices are to be considered $I(1)$, whereas Kellard et al. (2002) consider 23 of the 24 selected commodities as trend stationary. The most relevant conclusion of their study is that the pervasiveness of the PSH should be considered a function of some a priori decision criteria, as their various specifications lead to different results. Also in this case there is evidence for structural breaks, more precisely two, and when accounting for it, 16 series present a significant negative deterministic trend. Structural breaks are found even if considering very long time series, dating back to 1650 (Arezki et al., 2014; Harvey et al., 2017); in the former case the shifting points coincide with crucial events such as for instance the Industrial Revolution, whereas in the latter study, the path of commodity prices are partitioned into four regimes, and a long-run downward trend is estimated in all but in the second regime, ranging from 1820 to 1870.

Cuddington and Urzúa (1989) introduce interest in cyclical movements of primary commodity prices, with the belief that for policy considerations, the amplitude, duration and shape of the cycles are of the same importance as long-run trends. There is also in this case little support for the PSH when allowing for a structural shift after 1920, neither if considering a trend stationary nor a difference stationary model. 18 complete cycles in industrial commodity prices are found, from 1862 to 1999, by Cashin and McDermott (2002), who, exploiting *The Economist*'s index, find a downward trend of 1.3 percent per year with little support for a break over the last 140 years. The most relevant conclusions of this work could be summarised as follows: first of all, there has been, since the 1990s, an increment in the variability of price movements, while the duration of cycles is reduced and frequency has become larger starting from the collapse of Bretton Woods regime[2]; secondly, volatility implies more difficulty in detecting trend's significance in various sub-periods; finally, long-run trends are small in comparison with annual variability in prices. Evidence for cycles' relevance is also found in Erdem and Ünalmış (2016), aiming at analysing super-cycles (defined as movements with a 20 to 70 years duration) in oil prices. Not only this super-cycles do exist, but also they show no sign of moderation, with the last registered peak in 2012, date from which prices are in a downward pattern. As in most recent studies, trends and cycles are analysed exploiting new techniques allowing for gradual evolution of long-run trend over time, rather than assuming it constant, such as band-pass filters (see also Cuddington and Nülle (2014)), and incorporating the idea of co-movement among different commodity prices.

The existence of a declining trend has proven to be controversial. Not only different time spans and different considered prices lead to different results, but also different methodologies may lead to opposite conclusions. If on the one hand, the existence of this declining trend has been economically motivated with the increasing costs of manufactures (with respect to those of primary goods), on the other hand the resource scarcity issue prophesies that prices will inevitably increase as exhaustible assets, such as commodities, become more scarce. Slade (1982) highlights that whereas theoretical models such as the one of Hotelling (1931) predict exponentially increasing prices, empirical studies generally find or discuss a relative decline in natural resource prices. She motivates, by studying mineral commodities, the controversial issue by suggesting that other than exhaustion, another key aspect for price patterns is

[2]The most plausible explanation is given by the US dollar fluctuations in a flexible exchange rate regime, being dollar the reference currency at which commodities are priced.

the technological development. A possible solution may be that whereas during the past decades technological improvements had more weight that exhaustion, the contrary has happened more recently. Even if this reasoning could also be applied for agricultural commodities produced in developed countries, for poor countries the question is controversial: lower technological improvement goes along with a greater availability of resources, namely land and labour (Deaton and Laroque, 2003). Recently, it has been pointed out both that the change of path - from a downward long-run trend to an increasing one - do exist and it may be driven by the rapid and sustained growth of the Chinese economy. Farooki and Kaplinsky (2013), for instance, note that whereas in the case of the two commodity price booms of 1950s and 1970s, prices fell rapidly to standard levels, in this case the boom may be permanent because of the different economic features of the system at the considered times. They state that the two preceding booms where a consequence of interruptions to supply combined with rapidly growing demand expectations; on the contrary, China is, differently from the developed economies, "still at an early stage of its commodity-intensive growth path", an in addition it is pushing substantially the demand for commodities through the channel of growing consumption of meat.

Whether a declining trend or an upsurging one is prevailing is a challenging open question, which for sure need further investigation in the forthcoming years, and which of the two will prevail essentially depends on which of these economic facts will be dominant.

1.2.3 Price series properties

What emerges from the huge amount of analyses cited above is a general lack of consensus about the validity of a declining trend assumption, and the stationarity assumption of commodity prices is still under investigation; by the way, large part of literature on the topic agree that the high rate of persistence exhibited by the series is more in line with a random walk hypothesis, causing so additional uncertainty to prediction of future prices' movements and patterns. It is important to stress that the presence or absence of unit roots highly depends on the transformation of prices from nominal to real, and the effects of this manipulation may not be trivial. In addition to the just mentioned source of uncertainty, price volatility, which could be defined as unexpected prices' movements, contributes to difficulty of predictability. Apart from evidence of a volatility increment in commodity markets (Cashin and McDermott, 2002), presence of volatility clusters has been modelled exploiting the introduction of *autoregressive conditional heteroskedasticity*, or ARCH, and GARCH models (Engle, 1982). Possible volatility spillovers across markets, such as the transmission of volatility from oil price to other commodity prices as agricultural ones, has been deeply analysed (see, for instance, Mensi et al. (2013); Du et al. (2011)) with explanations ranging from speculation to interdependences of commodity markets.

This Section reports a battery of tests carried out to a set of 38 commodity prices, with the aim of mapping series with respect to their stochastic characteristics. This univariate analysis is essential not only to contribute to the academic debate on the topic, but also it provides some preliminaries results that will be essential in the multivariate scenario of the following Chapters. The starting point is performing tests for stationarity of the data, in order to distinguish the $I(0)$ series from the $I(1)$ ones; then, another question that may arise regards the possibility of integration neither of order 1 or 0, but of a fraction of value d between 0 and 1, or the case of a $d > 1$ (the explosive root situation). Finally, some considerations about volatility will be provided, together with the results of ARCH tests. This exercise is aimed at reaching some preliminary conclusions about prices behaviour, other than comparing results to the literature ones. However, being the topic enormously vast, the carried out tests could not be considered as exhaustive, and a more deep diagnostic would be required; for the scope of this work, it will be sufficient to provide only the major considerations about stochastic properties of the series. To balance the problem, the main diagnostics on price series have been carried out in both nominal and real terms, without deep discrepancies in the results. The log-transformed series, instead, should preserve the main properties of the original series.

Data description

The analysis exploits monthly spot prices from the IMF public database of primary commodity prices, focusing on series from different categories, specifically energy, metals, food (livestock products, crop commodities, beverages) and other agricultural raw materials. In particular, 38 series are selected, covering a time span from January 1980 to December 2018.[3] The selection of this huge set of commodities is made with two main considerations. The first and most important has to do with the empirical strategy that will be developed through next Chapters, and is that we want to split (see Chapter 3) the common movement that may arise among related commodities (for instance belonging to the same market) from the general co-movement among commodity prices of different groups. For this reason, each group should include a sufficient number of prices, and there should be as many groups as possible. The selected 38 prices cover different categories but comprehend series which can substantially differ. More than this, the economic reason wants that we cover each market which has a relevance on a global production and transformation chain.

[3]The complete list of the used prices, the relative description and the summary statistics are reported in Appendix A. The selected commodities are all those present in the IMF database being in *spot* prices and not controlled, such as iron ore, with availability starting from January 1980.

All prices have been indexed by imposing the value of January 2000 equal to 100, in order to get rid of different units of measure, deflated using the consumer price index provided by the FED database and expressed in logarithms. Peterson et al. (2000) provide some considerations about the issues arising when using real prices instead of nominal prices, underlying that the choice of the deflator could change the time series properties of the commodity price nominal series. Although the awareness of this fact, there is the counter-factual that almost all the existing studies work with deflated prices, and relative prices are also the natural choice for providing some economic interpretation. necessary also in the implementation of the model presented in Chapter 2.

The $n \times 1$ vector Y_t of log real prices contains all the y_t series, with $T = 468$ observations and $n = 38$.

Univariate tests

This Section exposes a battery of tests carried out to the dataset, with the aim of mapping commodity prices with similar characteristics together. Of course the first hypothesis under investigation is stationarity of the commodity prices. Then, another question that may arise regards the possibility of integration neither of order 1 or 0, but of a fraction of value d between 0 and 1. Furthermore, there is also the case of $y_t \sim I(d)$ with $d > 1$, which denotes an explosive root and is in general associated with bubble behaviour. Finally, some considerations about volatility will be provided, together with the results of ARCH tests. This analysis is aimed at reaching some preliminary conclusions about prices behaviour, other than comparing results to the literature ones. However, being the topic enormously vast, the carried out tests could not be considered as exhaustive, and a more deep diagnostic would be required; for the scope of this work, it will be sufficient to provide only the major considerations about stochastic properties of the series.

Understanding whether or not a time series exhibits stationarity is crucial - as mentioned - both from a theoretical and an empirical perspective; on the one hand, implications of non-stationarity include the impossibility for the series to revert, once affected by a shock, to their long-run equilibria; the existence of a long-run equilibrium is not to be taken for granted, but it is supposed to exist in price theory. The kind of persistence characterising the processes states that spot prices should be supposed to revert back, in the long-run, to the underlying costs of production, which series however may also be not constant over time (Maples and Brorsen, 2017). Moreover, has previous Section has explored, there is also economic support for non-stationarity of the series. On the other hand, from an empirical point of view, failing to account for non-stationarity when unit roots are actually present could bring to misleading results in the choice of the stochastic process representation, may generate spurious regressions in the multivariate side, and may lead to misspecification if the system is in facts cointegrated.

The ADF (Dickey and Fuller, 1979), PP (Phillips and Perron, 1988), and KPSS (Kwiatkowski et al., 1992) tests are performed to all real log-price series, allowing for various specifications, in both levels and differences. If a series is non-stationary in levels, but is stationary when differenced once, then $y_t \sim I(1)$, where y_t is the considered process. Note that whereas the ADF and the PP tests assume the null hypothesis of integration, the KPSS procedure tests the null hypothesis of stationarity. Tables 1.1, 1.2 and 1.3 report the mentioned tests for the 38 log real commodity prices and their first differences, for various deterministic components. As it is shown, for the majority of the series the hypothesis of stationarity cannot be confirmed, and it is more plausible to conclude that the stochastic process which has generated the series is non-stationary. Specifically, and considering a threshold of significance of the p-value at 0.05, the KPSS tests reject the null hypothesis of stationarity for all the series (with the exception of barley, deterministic component of only constant). The ADF and PP conclude that the series $\sim I(0)$ for all the deterministic specifications are aluminium and tea prices. Another price that is considered stationary according to the results, even if depending on the deterministic component chosen, is poultry. Other prices are in an ambiguous situation, being borderline or with contradicting results according to the considered tests: these include hard logs, soft sawnwood, sunflower oil, lamb and silver. Importantly, when differenced, all the series result to be stationary, meaning that the order of integration for those non-stationary in levels is 1.

This not well defined scenario opens the question of whether some series may actually be not $I(0)$ or $I(1)$, but rather $I(d)$ with $0 < d < 1$. In fact, it could be possible that standard unit root tests fail to detect the actual order of integration, by considering only integers. The concept of fractionally integrated processes, or exhibiting long memory, has to do with hyperbolically decaying autocorrelations; Baillie (1996) provides an extensive review of these kinds of time series processes. Summarising, a process of this kind, y_t is said to be integrated of order d, if

$$(1 - L)^d y_t = u_t$$

holds, with $-0.5 < d < 0.5$ and u_t being a stationary, ergodic process. Whereas the case of $d = 0$ goes back to the short memory case of a standard stationary and invertible process, if $d \geq 0.5$, y_t is non stationary, even if still mean reverting, and possesses infinite variance. There exists several procedures for the estimation of d, which could be exploited to test for fractional integration of series. The Geweke and Porter-Hudak (1983) procedure, also known as the GPH test for long memory, performs a semiparametric log-periodogram regression and is built under the null hypothesis of $d = 0$. An improvement of the GPH has been made by Robinson (1995), which log-periodogram regression allows the intercept and slope to vary for different series; Phillips et al. (1999)'s modification of the GPH test allows to address the case of $d = 1$, in addition

Table 1.1: Unit root ADF tests of log commodity prices. Test statistics and p-values in parentheses.

	constant	constant and trend	constant and quadratic trend
Aluminium	-3.6364	-4.2120	-4.1945
	(0.0051)	(0.0042)	(0.0176)
Barley	-2.9794	-3.0020	-2.9028
	(0.0369)	(0.1316)	(0.3444)
Beef	-2.5180	-2.1509	-2.8466
	(0.1111)	(0.5166)	(0.3739)
Coal	-2.5922	-2.6337	-3.0430
	(0.0945)	(0.2651)	(0.2759)
Cocoa	-3.0791	-2.8095	-2.5766
	(0.0282)	(0.1937)	(0.5248)
Coffee	-2.2290	-2.5392	-3.2590
	(0.1960)	(0.3090)	(0.1869)
Rapeseed oil	-2.9553	-2.9127	-2.7454
	(0.0393)	(0.1583)	(0.4290)
Copper	-1.9859	-2.4344	-2.3466
	(0.2933)	(0.3615)	(0.6541)
Cotton	-3.1368	-3.7775	-4.4099
	(0.0240)	(0.0176)	(0.0088)
Hides	-1.6329	-3.2547	-3.9494
	(-3.9494)	(0.0740)	(0.0363)
Lamb	-2.5691	-3.7037	-3.9654
	(0.0995)	(0.0220)	(0.0347)
Lead	-2.4523	-3.2849	-2.8582
	(0.1275)	(0.0687)	(0.3677)
Soft logs	-2.1776	-2.2394	-2.7650
	(0.2147)	(0.4669)	(0.4182)
Hard logs	-3.2750	-3.2756	-3.4521
	(0.0161)	(0.0703)	(0.1254)
Maize	-2.7970	-2.7847	-2.9735
	(0.0587)	(0.2029)	(0.3090)
Nickel	-2.9348	-2.9954	-2.9189
	(0.0414)	(0.1334)	(0.3362)
Crude oil (1)	-2.0692	-2.4029	-1.8267
	(0.2574)	(0.3780)	(0.8758)
Crude oil (2)	-1.9226	-2.2964	-1.7171
	(0.3221)	(0.4354)	(0.9047)
Crude oil (3)	-2.6988	-2.8543	-2.5859
	(0.0742)	(0.1778)	(0.5195)
Olive oil	-2.7938	-2.9373	-3.2710
	(0.0591)	(0.1506)	(0.1825)
Swine	-1.9873	-3.2855	-4.7359
	(0.2927)	(0.0686)	(0.0028)
Poultry	-1.9998	-3.2644	-6.0527
	(0.2872)	(0.0722)	(0.0000)
Rice	-2.9057	-2.7905	-2.6000
	(0.0447)	(0.2007)	(0.5115)
Rubber	-2.7787	-2.7848	-2.2942
	(0.0613)	(0.2028)	(0.6818)
Salmon	-2.4852	-2.1685	-3.7585
	(0.1191)	(0.5067)	(0.0606)
Hard sawnwood	-2.4799	-2.5684	-3.2493
	(0.1204)	(0.2951)	(0.1904)
Soft sawnwood	-2.7981	-2.7880	-4.6884
	(0.0585)	(0.2017)	(0.0033)
Shrimps	-2.0826	-3.5334	-3.6828
	(0.2519)	(0.0359)	(0.0733)
Sunflower oil	-3.1946	-3.1962	-3.2083
	(0.0203)	(0.0851)	(0.2059)
Tea	-3.4624	-3.9080	-4.7834
	(0.0090)	(0.0117)	(0.0023)
Tin	-2.4899	-2.3392	-2.5578
	(0.1179)	(0.4120)	(0.5356)
Uranium	-2.2165	-2.2138	-1.6540
	(0.2005)	(0.4813)	(0.9186)
Wheat	-2.8445	-2.9978	-3.1097
	(0.0522)	(0.1327)	(0.2462)
Wool	-2.2269	-2.2763	-3.2014
	(0.1968)	(0.4464)	(0.2085)
Zinc	-3.0804	-3.1354	-3.4702
	(0.0281)	(0.0980)	(0.1205)
Gold	-1.5147	-2.1219	-1.0412
	(0.5264)	(0.5330)	(0.9849)
Silver	-2.2799	-2.4305	-1.6370
	(0.1786)	(0.3635)	(0.9220)
Platinum	-2.9695	-3.4142	-2.5309
	(0.0378)	(0.0494)	(0.5510)

Table 1.2: Unit root ADF tests of log differenced commodity prices. Test statistics and p-values in parentheses.

	constant	constant and trend	constant and quadratic trend
Aluminium	-7.6048	-7.6137	-7.6452
	(6.9343e-012)	(3.4580e-011)	(1.5914e-014)
Barley	-18.965	-18.944	-18.959
	(7.4914e-036)	(6.4229e-043)	(4.0093e-042)
Beef	-5.5844	-5.7461	-5.8070
	(1.1203e-006)	(4.6536e-006)	(2.2576e-005)
Coal	-5.8791	-7.7476	-7.7586
	(2.3279e-007)	(1.2965e-011)	(1.1921e-015)
Cocoa	-14.913	-14.974	-15.062
	(6.6565e-035)	(7.0971e-043)	(0.0000)
Coffee	-8.6432	-8.6806	-8.7010
	(5.9617e-015)	(8.8394e-015)	(1.7288e-030)
Rapeseed oil	-23.623	-23.617	-23.654
	(9.1574e-036)	(1.1563e-048)	(6.5634e-048)
Copper	-5.9826	-5.9774	-6.0207
	(1.3188e-007)	(1.3164e-006)	(6.3461e-006)
Cotton	-6.1258	-6.1673	-6.1775
	(5.9308e-008)	(4.4676e-007)	(2.2386e-006)
Hides	-15.645	-15.665	-15.668
	(4.7529e-037)	(1.9604e-046)	(0.0000)
Lamb	-10.825	-10.822	-10.895
	(7.1127e-022)	(3.3169e-023)	(5.5846e-159)
Lead	-17.453	-17.523	-17.640
	(2.9837e-034)	(8.1469e-040)	(3.2086e-039)
Soft logs	-29.881	-29.848	-29.822
	(1.1191e-024)	(2.4792e-044)	(2.3477e-043)
Hard logs	-10.718	-10.718	-10.771
	(1.5836e-021)	(9.2469e-023)	(3.6703e-145)
Maize	-16.226	-16.220	-16.221
	(2.0476e-032)	(1.6124e-036)	(1.3692e-035)
Nickel	-15.239	-15.223	-15.251
	(1.2790e-030)	(9.4606e-034)	(7.3606e-033)
Crude oil (1)	-10.172	-10.203	-10.333
	(9.3268e-020)	(1.2818e-020)	(9.2862e-105)
Crude oil (2)	-10.629	-10.661	-10.782
	(3.0913e-021)	(1.6117e-022)	(2.5652e-146)
Crude oil (3)	-15.610	-15.603	-10.204
	(2.5129e-031)	(7.9295e-035)	(4.1331e-095)
Olive oil	-17.373	-17.355	-17.366
	(3.8101e-034)	(2.0610e-039)	(1.4759e-038)
Swine	-5.6663	-5.6622	-5.6503
	(7.2896e-007)	(7.2502e-006)	(5.2269e-005)
Poultry	-6.0452	-6.2494	-6.2561
	(9.3200e-008)	(2.7677e-007)	(1.2737e-006)
Rice	-14.172	-14.193	-14.244
	(1.1999e-032)	(6.2288e-039)	(0.0000)
Rubber	-16.369	-16.371	-16.480
	(1.1854e-032)	(6.4080e-037)	(2.7367e-036)
Salmon	-6.3794	-6.5694	-6.5648
	(1.3889e-008)	(3.9895e-008)	(1.0038e-007)
Hard sawnwood	-13.993	-13.978	-14.023
	(4.3141e-032)	(7.3012e-038)	(0.0000)
Soft sawnwood	-11.896	-11.885	-11.921
	(2.2872e-025)	(6.8421e-028)	(0.0000)
Shrimps	-17.672	-17.667	-17.648
	(1.5705e-034)	(3.7381e-040)	(3.0813e-039)
Sunflower oil	-14.798	-14.783	-14.795
	(1.4695e-034)	(6.6585e-042)	(0.0000)
Tea	-7.5000	-7.5152	-7.5365
	(1.3748e-011)	(7.0390e-011)	(1.5371e-013)
Tin	-11.282	-11.401	-11.462
	(2.2972e-023)	(1.0126e-025)	(5.4344e-239)
Uranium	-7.8283	-7.8939	-8.0465
	(1.5812e-012)	(4.3532e-012)	(4.9407e-019)
Wheat	-17.002	-17.001	-16.998
	(1.2496e-033)	(1.5309e-038)	(1.2255e-037)
Wool	-16.392	-16.517	-16.510
	(1.0877e-032)	(2.6497e-037)	(2.2850e-036)
Zinc	-16.292	-16.293	-16.281
	(1.5858e-032)	(1.0296e-036)	(9.4100e-036)
Gold	-7.0957	-7.4430	-15.860
	(1.8207e-010)	(1.1782e-010)	(0.0000)
Silver	-13.060	-13.113	-13.222
	(3.8683e-029)	(1.2263e-033)	(0.0000)
Platinum	-16.569	-16.591	-16.883
	(5.6516e-033)	(1.6985e-037)	(2.4149e-037)

to the standard null hypothesis, providing a test statistic for both cases. All the three mentioned tests have been performed to the 38 price series; whereas with the GPH and Robinson tests it is impossible to assess whether a series is $I(1)$ - since they only accept or reject the null of $d = 0$, with the Phillips modification results comprehend more information, and thus we will focus on them. Table 1.4 shows that all the series cannot be considered $I(0)$ processes, reporting the results for the Robinson (1995); Phillips et al. (1999) tests, with the only exception of the hard logs one. Curiously, the standard unit root tests lead to ambiguous results about stationarity of thi series. All the other log real prices are to be considered $I(1)$, with the exceptions of aluminium, rapeseed oil, lamb, swine, poultry, shrimps, tea and platinum, which are fractionally integrated.

Since tests for long memory could be considered as more sophisticated with respect to the standard unit root tests, commodity prices would be likely considered as non-stationary, with only few reported exceptions.

The results are consistent with those of Ardeni (1989); Schroeder and Goodwin (1991); Myers (1994) and Harri et al. (2009), which find that commodity prices are to be considered as non-stationary. Some cases of evidence for stationarity include some the yet cited works on the PSH or other studies on the storage model (see, for instance, Deaton and Laroque (1992)).

Finally, one last consideration about the order of integration of commodity prices regards the opposite situation, i.e. the presence of explosive roots: it may be possible that some price series exhibit roots which are indeed larger than one; eventually, this is more likely to occur during periods of turmoil and financial exuberance, for very short time spans. If there is evidence for some bubbles within the markets, these could be reflected indirectly in explosive behaviour of the price series, as argued in Hall et al. (1999). One of the many methods to test for this phenomenon has been introduced by Phillips et al. (2011): they propose to estimate a model of the kind of

$$y_t = \mu + \delta y_{t-1} + \sum_{j=1}^{J} \phi_j \Delta y_{t-j} + u_t, \tag{1.1}$$

with $u_t \sim NID(0, \sigma^2)$. The estimation is repeated recursively using subsamples of data, incremented each time by a fixed number of observations. The null hypothesis is the standard case of ADF unit root ($\delta = 1$), whereas the right-tailed alternative is $H_1 : \delta > 1$. The largest statistic of the forward recursive regressions is then compared with the critical values[4] simulated by the authors. A generalisation of this procedure is obtained with the GSADF test (Phillips et al., 2015), which simply solves the weakness of the SADF in detecting an episode of market turbulence in a long time span of a series; the *generalised* superior ADF, indeed, uses flexible window widths in the im-

[4]For this reason, the test is also known as Superior ADF_r ($SupADF_r$), or, more briefly, SADF.

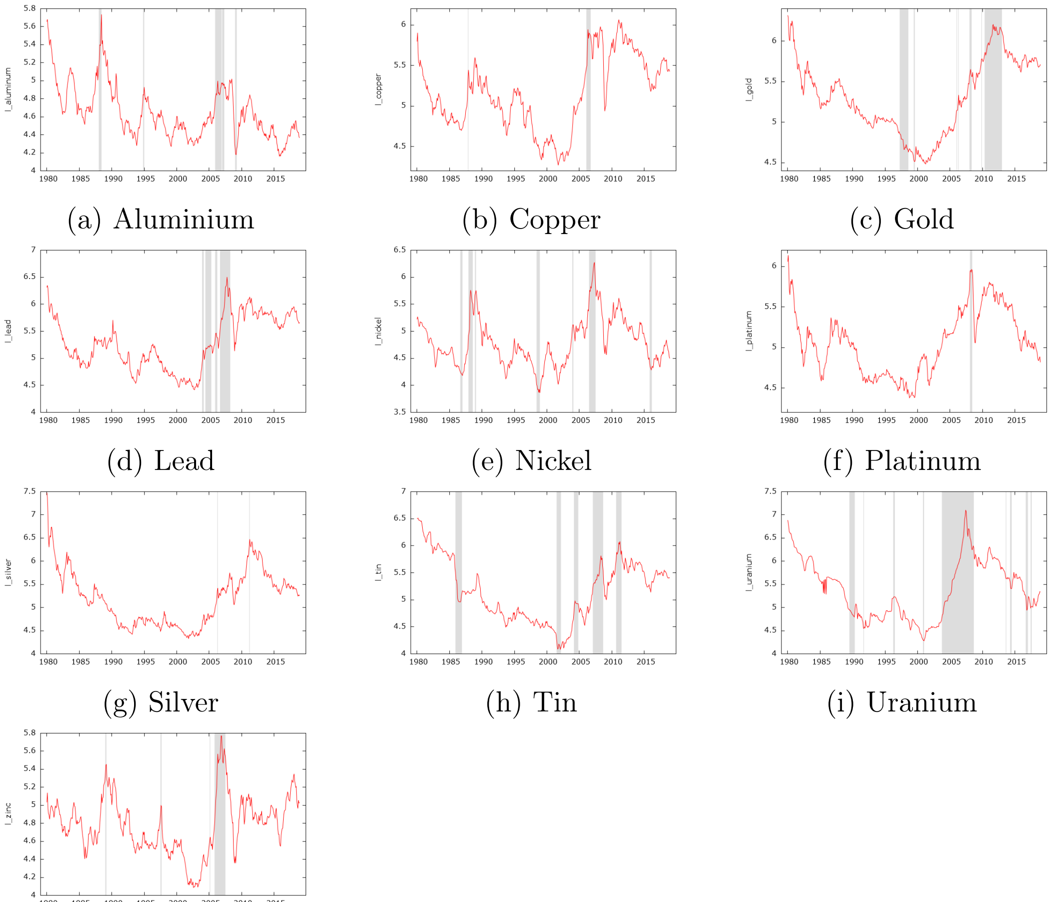
plementation, making both the starting and the ending points of the rolling variable, covering more sub-samples of the data and gaining in flexibility; given its characteristics, the GSADF test is particularly useful to have a more precise idea of *when* a bubble, if present, actually occurs, with respect to the SADF. Moreover, in this way it is possible to test also for multiple bubbles (bubbles in this case are simply detected as the sub-periods in which there are explosive roots).

The results of the performed tests to the 38 series could be summarised graphically as reported in Figures 1.2, 1.3, 1.4, 1.5 and 1.6, showing the periods of bubble-like episodes for each price series, if any. Prices have been divided according to the category, in metals, energy commodities, livestock products, raw materials and food prices, respectively. 35 out of 38 of the considered commodity prices manifest an explosive behaviour in at least one sub-period. The prices without evidence of explosive roots are beef, swine, soft sawnwood, rapeseed oil and sunflower oil prices.

To conclude, it is possible to state that the majority of commodity prices, regardless the typology of product, are non-stationary over the considered time span, and exhibit explosive behaviour for some short intervals.

At the end, even since now we have examined data paying attention to *levels*, volatility is another aspect that matters, as the jump of the last decade in all price series regarded also a variability increase. Volatility, which could be thought at as an expression of market uncertainty, tend to increase in case of turbulence, while it follows a more predictable path during more stable periods. Despite the importance of the topic, we will focus only on literature results with only some preliminary consideration, as an analysis of volatility of commodity prices is beyond the scope of this work. The phenomenon of volatility clusters could be detected even by looking at the residuals of the series, once a proper $ARMA$ specification is chosen for the process and the model is estimated. ARCH tests (Engle, 1982), performed to the log-differenced prices, confirm the hypothesis that volatility follows an autoregressive structure that decays slowly over time, for the majority of series. This opens the possibility that the explosive behaviour detected above by using the appropriate bubble tests, may actually be a manifestation of extreme variability episodes, which are likely to display in periods of market turmoil. Table 1.5 reports the ARCH test results on the $ARMA$ model implemented for each series, appropriately differenced to induce stationarity.[5] There are only five cases in which an ARCH effect is not detected at a 5% of significance: cocoa, maize, olive oil, sunflower oil and wool returns, whereas nickel and zinc ones exhibits an ARCH structure at a 10% level of significance. For the remaining series it is evident that volatility matters. Note that for some series, the $ARMA$ specification is still not sufficient to ensure absence of autocorrelation of residuals (coal, coffee, hard logs, swine, poultry, salmon, tea and gold).

[5]The $ARMA$ parameters p and q have been selected as the best model following the Hannan-Quinn criterion.



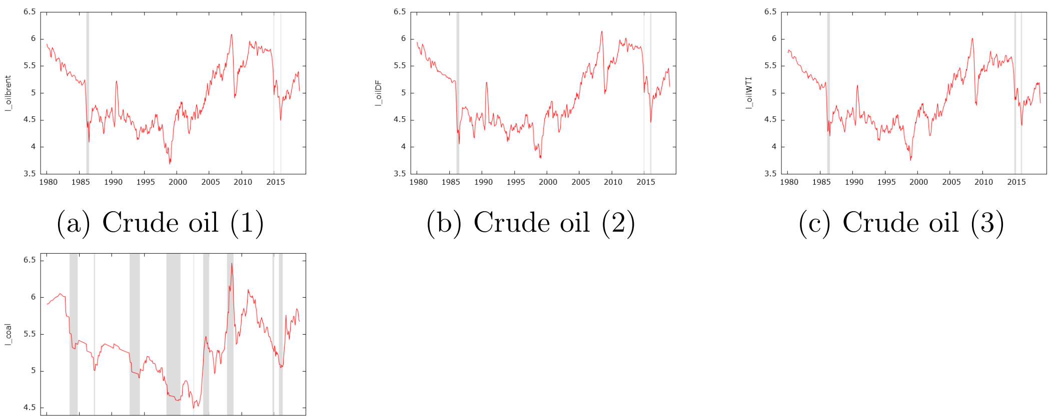
(a) Aluminium (b) Copper (c) Gold

(d) Lead (e) Nickel (f) Platinum

(g) Silver (h) Tin (i) Uranium

(j) Zinc

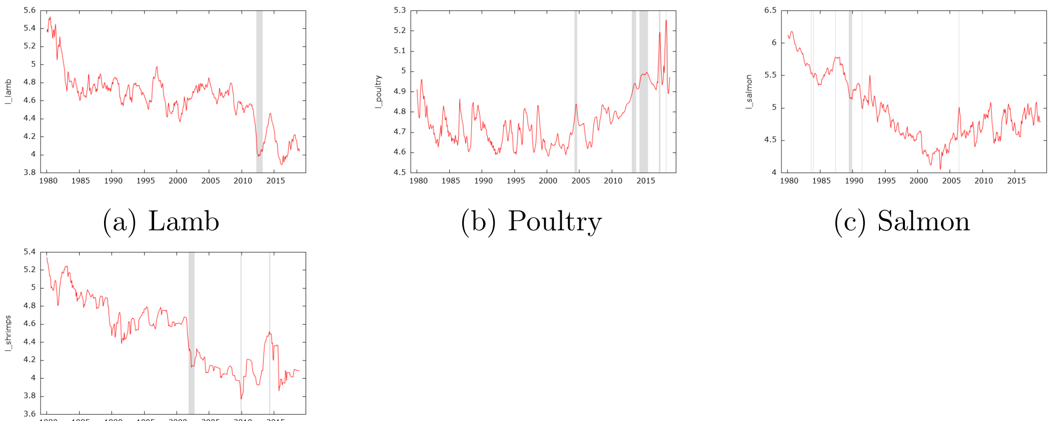
Figure 1.2: Explosive roots multiple bubble tests (Phillips et al., 2015), metals. Episodes of bubbles in shaded areas.



(a) Crude oil (1) (b) Crude oil (2) (c) Crude oil (3)

(d) Coal

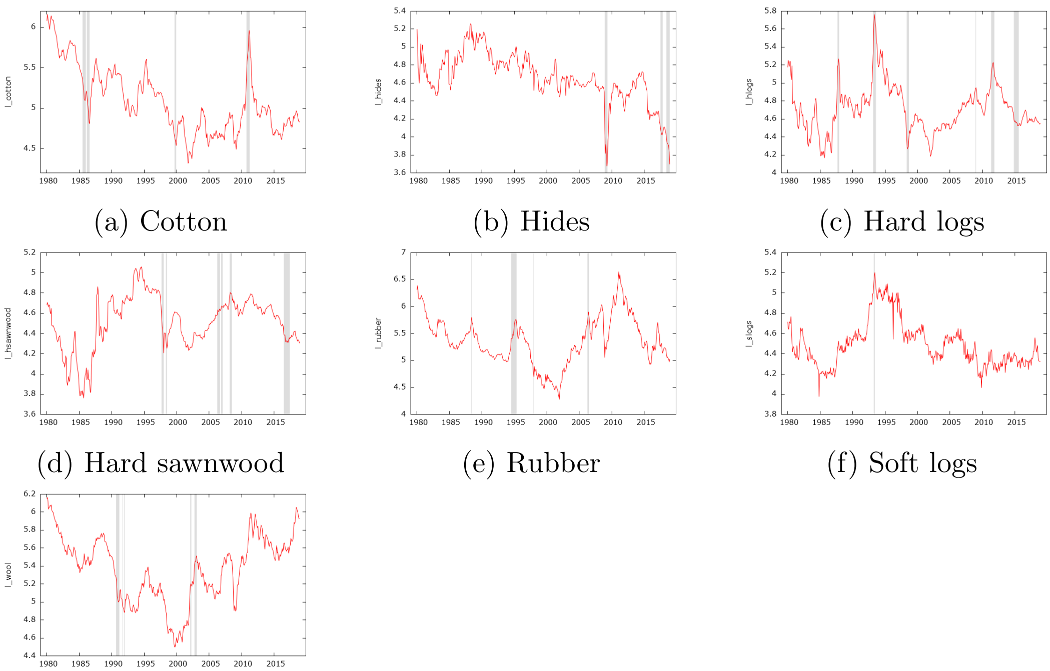
Figure 1.3: Explosive roots multiple bubble tests (Phillips et al., 2015), energy. Episodes of bubbles in shaded areas.



(a) Lamb (b) Poultry (c) Salmon

(d) Shrimps

Figure 1.4: Explosive roots multiple bubble tests (Phillips et al., 2015), livestock products. Episodes of bubbles in shaded areas.

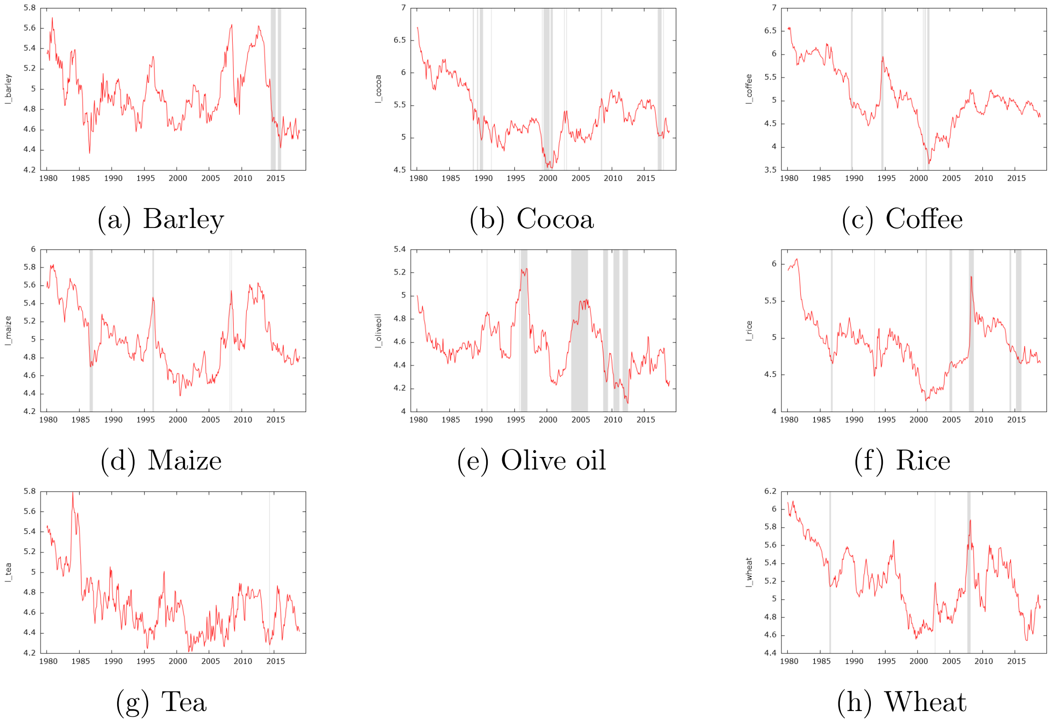


(a) Cotton (b) Hides (c) Hard logs

(d) Hard sawnwood (e) Rubber (f) Soft logs

(g) Wool

Figure 1.5: Explosive roots multiple bubble tests (Phillips et al., 2015), raw materials. Episodes of bubbles in shaded areas.



(a) Barley (b) Cocoa (c) Coffee

(d) Maize (e) Olive oil (f) Rice

(g) Tea (h) Wheat

Figure 1.6: Explosive roots multiple bubble tests (Phillips et al., 2015), food. Episodes of bubbles in shaded areas.

Literature on volatility has reached consensus on the existence of volatility clusters in commodity prices (see Beck (1993); Ramirez and Fadiga (2003)). Beck (2001) underlines the asymmetries of shocks impacting conditional variance: a commodity price increase generates a higher volatility. This phenomenon is explained with the storage model, indeed an ARCH process is found only in storable commodities, but not in non-storable ones. Apart for the GARCH approach, another frontier of research on volatility handles with stochastic volatility models (see, for instance, Brooks and Prokopczuk (2013)). An exhaustive review about volatility in commodity prices is found in Prakash et al. (2011).

1.3 Price determinants

Price formation theory, accounting for mean reversion, starts from the assumption that price movements reflect the underlying demand/supply balance. Market fundamentals establish the equilibrium price and therefore the demand for a particular commodity - given by the sum of demand for direct consumption and the demanded inventories - and the respective supply, are the most important drivers. For this reason theoretical models of commodity prices focus mainly on these factors (see Chapter 2). However, from an empirical point of view, many other factors have been studied as responsible for commodity price movements, and are now considered as more and more crucial in capturing the causes of fluctuations. In this Section, we will focus only on the main drivers which have been proposed - and empirically tested - to be responsible of price movements.

A challenging option would be trying to split a long-run component from a short-run one, in order to understand which are the determinants able to shape the trajectory followed by price series on a long time horizon, and which ones are the causes of temporary deviations from the equilibrium level. The uncommon boom and consequent collapse of all categories of commodity prices occurred during the last decade could be explained with the joint effect of many determinants, but are more the result of exceptional events rather than the epiphany of a new paradigm for a long-term trend. As empirical evidence shows, the fact that this atypical price behaviour interested *all* commodities, and not only some specific groups, suggests that there may be in act some common implications, as price transmission or co-movement, and that many "external" drivers have acquired importance, aside from market fundamentals of a specific market. These "external" drivers include macroeconomic factors and speculation, and have been considered crucial for commodity price formation within the literature, with different perspectives across works.

1.3.1 Market fundamentals

The demand/supply balance is clearly the base upon which prices are formed, if assumed that markets have to clear; commodity markets in particular should be considered as peculiar markets, given the intrinsic features of constrained supply and quite inelastic demand (many of these products are in fact staple items). However, whereas supply-side shocks drive mostly short and medium-term fluctuations in commodity prices, demand shocks and tendencies have an effect which could be persistent in the long-term, if not permanent (Kilian, 2009; Jacks and Stuermer, 2015). Supply shocks are mainly driven by weather and climate conditions for what concerns agricultural commodities, and by minerals mining conditions in the case of energy and metal commodities. For instance, Kaufmann (2011) explains the oil price rise in 2007-08 not with a rising-demand effect, as standard in the literature, but rather with a supply shock, mainly caused by a stagnating growth of non-OPEC countries oil production since 2004. Of course there are also variables able to modify the long-run supply levels, such as the efforts in R&D and investment within the reference sector, thus influencing yields and productivity, or the policies implemented by institutions which could widely affect production of particular groups of commodities[6], but still, the short-term components are the most monitored, since responsible of large part of volatility fluctuation coming from market fundamentals. As highlighted in Kilian (2009), the decomposition of

[6]For instance, particulates emission regulations and biofuel incentives affecting energy commodities production.

real oil price fluctuations reveals that demand shocks, rather than oil supply shocks, are the main drivers.

Commodities demand is affected, globally, by the macroeconomic conditions of the system, namely the GDP level and rate of growth, demographic composition and trajectories, but also the usage of commodities for speculative activities and the implemented US monetary policy, as summarized in Cooke (2009). Some works agree on the attribution of the rapid surge of prices started in the second half of 2000s to the rising world demand, mostly due to emerging countries' performance, and in particular to China's one (as, for instance, stressed in Frankel and Rose (2010)). Also Kilian (2009) states that the surge registered after 2003 was primarily due to the cumulative effects of positive global demand shocks. The subsequent price decrement should be viewed, in this context, as a consequence of a general stagnating demand; global GDP growth indeed, assessed at 2.9 percent in 2016 (OECD and FAO, 2017), the lowest rate since 2009, trained by a slowdown in economic growth of emerging economies. The projections for the forthcoming seasons provided by World Bank (2017) forecast prices of metals and energy to rise moderately in line with the present situation of falling inventories combined with strengthening world demand, whereas agricultural markets are now well-supplied, thanks to favourable weather conditions and high stocks-to-use ratios. The relaxed pressure on supply-demand balance on food sector should overall provide stable prices in the near future.

Another possibility is that some other price determinants have acquired more significance now, with respect to the past, so that market fundamentals remain still important for explaining price fluctuations, but are not predominant, or at least, are not the only drivers. In this view, recent price movements should be analysed with care, taking into account that the observed commodity prices result from the combined effect of market forces and other factors. In line with this, a recent study referring to agricultural prices confirms that market fundamentals appear, in the short-term, to be playing a smaller role than in the past (Baffes and Haniotis, 2010). Specifically, authors attribute price spikes occurred in the past to negative supply shocks, and find no evidence for an effect on world prices due to dietary changes and income growth in middle-income countries. Roache (2010), examining food price volatility, states that the sources of long-run volatility are related mostly to supply-demand fundamentals, but attributes the recent rise of these prices' volatility mostly to other factors, such as US inflation and USD exchange rate's variations.

Within the literature, many contributions try to decree which elements, between demand and supply, is to be considered the most relevant in explaining long-run fluctuations: Jacks and Stuermer (2015) provide evidence on the dynamic effects of global demand shocks, commodity supply shocks and inventory demand shocks on real commodity prices, concluding that only the former and the latter should be considered the chief drivers of fluctuations of commodity prices. Moreover, the effect of global demand shocks is to be considered able to extend across metal, agricultural and soft commodities, capturing thus common patterns, whereas commodity supply shocks do play some role in explaining fluctuations, but only for particular commodities and only with a limited influence in terms of impact and with transitory duration. On the contrary, Borensztein and Reinhart (1994) provide support for the effects of supply-side adjustments, acquiring importance since the 1980s. According to them indeed, the decline of prices registered in the 1980s and 1990s can be largely attributed to supply growth, thanks to technological developments, particularly evident in the agricultural sector. The agricultural policies of industrial countries reinforced the expansion, while at the same time developing economies implemented financial market improvements, exporting-oriented policies and consequent openness to international trade; in addition to this, from the 1990s, the collapse of the Soviet Union improved import of food and export of metals for the related countries.

1.3.2 Macroeconomic and microeconomic drivers

Macroeconomic factors could potentially affect price movements, in both levels and volatility, through many different channels; the reason for the relevance here, is due to the possibility of influencing demand and supply of commodities even over long time horizons. Within the literature, among the proposed factors belonging to this category there are the US dollar exchange rate fluctuations, the real interest rate movements, the trends on inflation (again, of the US), and the dynamics of GDP, or economic activity in general. Inflation and exchange rates of reference are those of the US because almost all international trade of commodities is implemented in US dollars, being dollar the reference currency at which commodities are priced. As stressed in Frankel (1986), ignoring the role of macroeconomic and financial factors in the determination of commodity price is misleading, as not only the exchange rate, but also monetary policy has a high capacity of influencing these prices. While expected future inflation has a positive effect on commodity prices in the present, an increase in the real interest rate has a negative effect on prices, because of investors' shift out of commodities and into bonds. The concept is more extensively expanded in Frankel (2006), in which the negative effect of high interest rates on demand for storable commodities (or, conversely, positive effect on supply), is summarised via three main channels: the increasing incentive for extraction at present rather than in the future, the decrement of firms' willingness to carry inventories and the convenience for speculators to shift out of commodity contracts. All these forces reduce market prices of commodities and empirical evidence of high US interest rates in the 1980s and low ones in the late 2000s coincides with periods of low/high prices. The effect of exchange rates is manifested through channels as the international purchasing power parity and the effect on margins for producers with non-USD costs; in particular, a rise in the value of the USD should result in a fall in commodity prices, expressed in dollars (Gilbert, 1989).

Many empirical tests have been carried out to validate these theories and to quantify the effects on commodity prices: the yet mentioned work of Roache (2010), focused on price volatility, lists as potential factors, other than the above commented market fundamentals, the volatilities of real US interest rates and US dollar and the future markets volumes. The results include a negative relationship between real interest rate levels and food price volatility, but no effect between the latter and the real interest rate volatility, and a positive effect of USD exchange rate volatility. The general conclusion is that US inflation and USD exchange rate volatilities mostly contributed to the recent rise of food prices volatility, rather than other examined factors as speculation, agricultural policies, interconnections with oil prices and global weather patterns. Gilbert (2010) shows, by using Granger causality tests, that GDP growth, monetary expansion and fluctuations of the exchange rate are the main determinants for explaining changes in agricultural commodity prices over a 38-year time span, but attribute the recent price boom mainly to index futures investment, through which monetary and financial activities have influenced food prices. Baffes and Haniotis (2010) relate the recent price boom, apart with the world economic growth, with a fiscal expansion occurred in many countries combined with an easy monetary policy, and a depreciation of US dollar; in addition, low past investment - hitting especially extractive commodity markets - together with financialisation, geopolitical concerns and adverse climate conditions, respectively for energy and agricultural markets, led to lower stocks-to-use ratios during the last period. Indeed, as noted in Frankel and Rose (2010), also factors which could be referred to as "microeconomic" do play a role: the study includes among these determinants inventory levels, uncertainty proxies and spot-forward spread, concluding that most strong and consistent effects come from the micro side, rather than the macro factors. Karali and Power (2013) state that over the period 2006-2009, commodity specific factors dominated the common factors in explaining price volatility, specifying that these common macroeconomic and microeconomic effects are important across a wide range of commodities from 1990 to 2005; long-run volatility is affected mostly by changes in inflation, industrial production, inventories and the long-term/short-term interest rate spread.

Finally, another mechanism of transmission to commodity prices regards the channel through which oil price's dynamics translate into other commodity prices; this effect takes place not only because of the important role of oil in the production and transport phase of many commodities, but also for some substitution effects - consider for instance the increasing demand for biofuels - and interdependencies existing between macroeconomic conditions and crude oil price[7]. By the way this relationship between oil and other commodities will

[7]As reported in Erdem and Ünalmış (2016), while, before the 1990s, business cycles in the US economy appear to follow cycles in oil prices, thus suggesting that oil price fluctuations were mainly driven by supply factors, since then the situation has reversed, being oil prices

be further analysed in a co-movement framework, in Section 1.4.

1.3.3 Financial drivers

Last, but not least important, some contributions investigate the implications of an increasing speculative activity for price formation, but generally, this proposal holds better to explain short-term movements, rather than long-term ones. In theory, speculation should encourage price discovery, thus promoting stability; however, some adverse short-term effects may prevail over the long-term positive effects. Recent empirical analyses provide evidence for a positive linkage between increments in financial activity and last decade observed changes in prices (Cooke, 2009; Bruno et al., 2016). Also Kaufmann (2011) hypothesises that speculative pressures matter, combined with market fundamentals, and states that the former factor's importance has increased over the last decade. On the contrary, Kilian and Lee (2014), working on crude oil prices, find minimal evidence of speculative demand shifts over the period 2007-08, and attributes much of the fluctuations in oil price to market fundamentals, particularly to shifts in global demand for oil. Another possibility is that short-term movements of commodity prices could be explained through bubbles, or, the occurrence of explosive episodes; anyway, according to Etienne et al. (2014), bubbles represent only a small portion of prices' behaviour in the considered 42-years period, and a large share of this price explosiveness occurs during downward price movements. Evidence against the bubble explanation is also found in Irwin et al. (2009), whose purpose is to ascertain the role of speculation in the recent prices' surge; even in this case, fundamentals appear to be the most relevant drivers, rather than an excessive financialisation. Finally, Fattouh et al. (2013) provide a review of the linkage between speculation and prices, focusing on the oil market. Their major finding does not support the positive association between the phenomenon of speculation and oil spot price movements after 2003. Instead, the authors state that it is necessary to distinguish between speculation and *excessive* speculation, that this distinction is not carefully addressed within the empirical works, and that the common movement of spot and future prices is more the result of common economic fundamentals, rather than a financialisation of the oil future market.

1.4 Co-movement

After having discussed the early literature on commodity prices, the corresponding time series properties and the drivers of their dynamics, it is necessary to introduce the real research question of this work, that is: *is there evidence for commodity prices to move together?* If a positive answer exists, then it is necessary to go another step further: is this behaviour detected on a short hori-

more influenced by growing demand for fuel commodities.

zon perspective, or do the series tend to co-move even in the long-run? And more importantly, *why*? From the previous sections some generalities emerge: different commodities appear to share the same price stochastic properties, and many academic contributions have found similarities not only in the general price movements, but also in the underlying explanations. An exiguous number of factors can explain more or less the majority of price fluctuations occurred within different time spans, even if considering a wide variety of commodities; clearly, the hypothesis of co-movement of many commodity prices seems to hold, and indeed, starting from the Nineties, this topic has become predominant, and an increasing strand of literature has examined the tendency of price series to co-move. The way in which this concept of co-movement is implemented and conceptualised varies according to the *a priori* assumptions about the stochastic properties of the series; prices can be indeed assumed to behave as interdependent processes, so that common fluctuations are to be seen as a consequence of reciprocal causation and price transmission across different commodity prices, or they could be thought as independent processes which share the common exogenous drivers. In the first case the co-movement is the result of shock transmission from one series to the other, and the challenge would be detecting the series which pull the movements of all the others. If the hypothesis of common exogenous drivers holds, instead, price series could actually be considered as independent processes, and the co-movement could be detected exploiting the idea of common factors, being them latent or observable; these common factors would be nothing else but the drivers affecting the entire set of commodity prices. The implications of the two representations are not trivial, as they lead to opposite modelling choices and, consequently, to different policy conclusions. Whereas the first strand of literature relied more on the first assumption of price interdependence, being in levels or in variances, and regarding both short and long-term movements, recently empirical works are exploiting the possibility of explaining the co-movements of price series through common long-run trends; this latter assumption involves the idea of cointegration and enfolds two different but complementary specifications: the ECM form and the common trends representation by Stock and Watson (1988). From the common trends representation it is possible to disentangle the short-term movement from the long-term one, as in Gonzalo and Granger (1995). Another tool of particular interest here is given by Dynamic Factor Models, which will be presented in Chapter 3, allowing to estimate some common unobservable factors responsible for the joint dynamics of many series.

An even more drastic hypothesis is that of *excessive* co-movement (from now on, the ECH), introduced by Pindyck and Rotemberg (1990): it corresponds to fluctuations which are not accounted for by business cycle and trend factors, thus remaining unexplained. The topic is nevertheless controversial, as it opens various possibilities such that of defining which is the actual meaning of the word *excessive* and how to interpret the conclusions. Even if there is still not consensus about the validity of the ECH, research is going on examining the intrinsic dynamics which could make commodity prices co-moving (whether this co-movement is excessive or not). Two facts have been crucial for the renewed interest on this topic during recent years: the newly happened upsurge in commodity prices, and the development of new econometric tools for detecting the common movements of different series.

1.4.1 The excessive co-movement hypothesis

Starting from the introduction of the ECH, several works have provided evidence for an excess co-movement of price series, which often remains unexplained. Pindyck and Rotemberg (1990) examine prices of seven commodities (through the correlations of monthly log changes in prices) which should be, theoretically, as unrelated as possible, being nor complements or substitutes in neither consumption or production - and found a persistent tendency for these series to move together, with the impossibility to explain this phenomenon through the effect of inflation or industrial production. The resulting detected co-movement is thus considered as excessive. The authors provide two possible explanations, one regarding a speculative behaviour, the other concerning common price reactions to non-economic factors. With respect to speculation, the authors hypothesise that liquidity constraints may exist: drops in one commodity price could lead speculators to liquidate also activities concerning other commodities; alternatively, the co-movement is the result of herd behaviour in financial markets[8]. The second hypothesis has to due to tandem reactions to the presence of some equilibrium "sunspots", bubbles, or simply changes in market psychology.

Deb et al. (1996), in contrast with Pindyck and Rotemberg (1990), find no or little evidence for excessive co-movement, after having provided some definition of *excess*, using univariate and multivariate GARCH models and distinguishing between long and short-run covariances; in particular, the paper shows that the Pyndick and Rotemberg results are not robust when relaxing the assumption of normality on regression residuals and once accounting for heteroskedasticity, so that the found excessive co-movement could merely be a false artefact. Another result provided by Lescaroux (2009), finds evidence against the ECH, again, once that a distinction between short and long run price variations is implemented. Indeed, if short term fluctuations are removed, there is little room for excess co-movement and the examined commodity prices demonstrate to be rather unrelated, only reflecting respective market fundamentals. In this case, the degree of co-movement is evaluated by decomposing prices at cycle frequencies and measuring how - and to what extent - these cycles are related. Results of Lescaroux (2009) are similar to those obtained by Ai et al. (2006), even if the former work takes into account possible exces-

[8] by this, the authors mean a situation in which traders could alternatively take long or short positions on *all* commodities, without a plausible economic reason.

sive co-movements between oil and metal commodities, while the latter refers to agricultural prices. The authors of the last cited study indeed conclude, as in the case of Pindick and Rotemberg, that macroeconomic indicators alone could not explain all the common dynamics of commodity prices, but if market fundamentals are included in the analysis, little of correlations among prices remains unexplained.

1.4.2 Price transmission and interdependence

Interdependence of price processes may refer to only short-run dynamics, or exclusively to long-run fluctuations, or both of them. Another possibility is that co-movement regards only the volatility (the case of volatility spillovers); more plausibly, the truth may be a combination of this entire set of formalisations.

Whereas the case of latent factors - which will be examined in the next Section - does not necessarily imply a causation phenomenon (in the sense that common movement tends to be attributed to the unobserved factors dynamics, rather than to the other price series), in this specification the key question is to understand which are the commodity prices that lead levels or volatility movements, that spill over different series and links different markets. As noted by Akram (2009), positive spillover effects among different commodity prices are not easily explained by the economic theory, because of the hard task of quantifying, and then splitting, the substitution and the income effects. Nevertheless, in the majority of cases, empirical investigation analyses the interconnections between oil prices and those of other commodities, with the belief that the former anticipates other prices' movements. Energy markets could affect other prices through more than one channel: not only because fuel commodities are directly exploited as production inputs and in transportation phase, but also because oil price reflects the overall economic production, thus influencing commodities' demand. The relationship is further extended for some agricultural commodities, such as maize or sugar, as the biofuel market expansion has contributed to exacerbate competition in both categories; some agricultural commodities, used for the production of biofuels, are now becoming direct substitutes for fuel commodities, but despite this, the dynamics of the emerging biofuel market are heavily shaped by interested countries' internal policies.[9] The impacts of this market in general, and the effects on commodity prices are empirically examined in terms of long-run relationship and causal links (as an example, see Balcombe and Rapsomanikis (2008); Hamelinck and Faaij (2006); Tyner (2010)). Usually, the mainstream point of view sees oil prices as drivers of biofuel-productive commodities.

The contributions on the topic generally exploit cointegration analysis to detect long-run common movements, and in the majority of cases focus on

[9] Mitchel (2008) attributes the recent rise in food prices, for instance, mostly to the increased production of biofuels within the EU and the US; Gilbert (2010), however, find little evidence for this link.

the relationship between energy and other types of commodities, especially agricultural ones, due to the biofuels matter. However, some works interpret the increasing correlation of oil with other commodity prices with through the financialisation hypothesis (Tang and Xiong, 2012).

Chaudhuri (2001) states that real commodity prices and real oil prices are cointegrated. In Bakhat and Würzburg (2013), threshold cointegration is used, and results indicate that: oil price and beverage prices are not cointegrated at all, oil price is linearly cointegrated with natural gas price, threshold cointegrated with aluminium and nickel, and asymmetric threshold cointegrated with food and raw materials. Also Natanelov et al. (2011) test for cointegration and threshold cointegration for oil, gold and some agricultural commodity prices, but find a counter-intuitive result: crude oil price follows the other commodity prices, which, apart from the case of coffee, move first. Zhang et al. (2010), investigating the long-run co-movement between fuel and agricultural commodity prices, fail to detect a cointegrating relationship, but instead find some limited short-run common movements, mainly dragged by sugar price. A similar conclusion is found in Sari et al. (2012), which analyse the role of future prices of energy commodities and grains, in addition to cross-market impacts; whereas for the long-run it is impossible to detect relationships, in the short-run there exists a two-directions feedback between the two markets. The wide differences of conclusions make one thing clear: that further research is needed in order to get closer to the truth.

1.4.3 Latent common factors

Recently, as anticipated, empirical literature is moving towards the possibility of modelling co-movement of prices exploiting latent factors, thanks to the more and more sophisticated econometric techniques; this is accomplished by specifying a DFM (Chapter 3). The commodity price series are treated as observed variables and are represented by the first equation, whereas the drives of common movement enters into the state equation as unobservable components[10]. The techniques range form the parametric approach of the Kalman filter to the non-parametric principal components analysis; what is important here is that the estimated latent factors could be then exploited in second-step estimations (Stock and Watson, 2011).

Following the pioneering work of Bernanke et al. (2005), several studies empirically have indeed started to test for co-movement in commodity prices by extracting latent factors and using then a Factor Augmented VAR (FAVAR) approach to determine which are the causal relations among different variables; the estimated factors are included into a VAR model with some macroeconomic variables, to check the relative importance of the price movement drivers introduced in Section 1.3.

[10]The structure and logic of DFM will be further discussed in Chapter 3.

On the heels of Bernanke et al. (2005), the works on co-movement in this framework have consistently grown, and is still increasing: Vansteenkiste (2009) exploits the extraction of one common factor from a large set of non-energy commodities to attribute price fluctuations to macroeconomic variables, and particularly: the USD exchange rate, the real US interest rate, oil prices and other input costs, some financial variables and demand.
Byrne et al. (2013) extract one common factor from primary commodity prices to capture the effect of fundamental price co-movements, attributing this common movement - at least in the short and medium terms - mainly to real interest rate fluctuations and risk, whereas Lombardi et al. (2012) associates the two extracted factors' fluctuations with real exchange rates and economic activity explanations, failing to find a relevant role for the interest rate and strong presence of spillover effects from oil to non-energy commodities.
Poncela et al. (2014), while identifying the extent of co-movement among commodity prices using DFMs and FAVARs, find that, since December 2003, raw materials are more strongly synchronised, other than exhibiting the same feature of common movement in response to the same shock. For the authors, the most crucial result regards the rising importance of uncertainty to explain non-energy commodity price fluctuations, although supply and demand conditions have a consistent role for sharp movements of series. In a work of Yin and Han (2015), a DFM is implemented to decompose commodity prices into global, sectoral and idiosyncratic components by estimating three types of factors: the single global factor, affecting all the set of commodities, some sectoral factors, responsible for movements of the particular segment of series (i.e. metal commodities, or agricultural), and finally some commodity specific factors. Evidence shows that the importance of the global and the sectoral factors is greater since 2004, signalling an increasing integration among markets.
Finally, also Alam and Gilbert (2017) and Delle Chiaie et al. (2017) highlight the increasing and more pronounced co-movement expressed through one single common factor since the 2000s with respect to the previous decades, but whereas the former work refers to agricultural commodities, the latter includes also oil price and other variables among commodities and indices. Moreover, while the paper of Alam and Gilbert (2017) stresses the relevance of monetary policy and the dynamics of the real interest rate, Delle Chiaie et al. (2017) find that the global factor is closely related to global demand conditions.

Studies of commodity prices co-movement and latent factors are becoming even more frequent because of the implementation of DFMs able to take into account some multivariate stylised facts, as the tendency of many time series to be integrated and often cointegrated. Chapter 3 will focus on DFMs.

1.5 Concluding remarks

This Chapter aims to provide an introduction to the topic of this work: the common movement of commodity prices by use of latent factors. In doing so, the main strands of literature about commodity price dynamics are reviewed, ranging from the stochastic properties of spot prices of the principal commodities - and the hypothesis of a long-term declining trend - to the main drivers of price fluctuations. Finally, empirical evidence for co-movement is analysed. For what concerns the univariate properties of the examined prices, some tests are performed in order to compare the results with those of the literature. The main conclusions are that commodity prices are in general non-stationary, are often affected by explosive behaviour and exhibit autoregressive conditional heteroskedasticity.

Next Chapters will focus on the analysis of co-movement of different commodity prices, taking into account the non-stationary behaviour detected here. In particular, Chapter 2 will incorporate latent factors into a theoretical framework accounting for market fundamentals, whereas Chapter 3 will let commodity price series speak freely, by estimating a new DFM accounting for cointegration structure of data.

Table 1.3: Unit root tests: PP and KPSS, log commodity prices and first differenced series. P-values in parentheses.

	Levels				Differences			
	Phillips-Perron		KPSS		Phillips-Perron		KPSS	
	constant	constant and trend	constant	constant and trend	constant	constant and trend	constant	constant and trend
Aluminium	-3.6192	-3.8397	2.7701	0.35390	-18.129	-18.144	0.10247	0.047757
	(0.0057)	(0.0154)	(<0.01)	(<0.01)	(4.6218e-035)	(3.0759e-041)	(>0.10)	(>0.10)
Barley	-2.7757	-2.7976	0.41864	0.41128	-18.837	-18.815	0.049606	0.049535
	(0.0626)	(0.1989)	(0.0689)	(<0.01)	(9.5407e-036)	(1.1628e-042)	(>0.10)	(>0.10)
Beef	-2.9764	-2.4593	2.9601	1.6657	-16.188	-16.261	0.35543	0.049208
	(0.0379)	(0.3484)	(<0.01)	(<0.01)	(2.3739e-032)	(1.2535e-036)	(0.0967)	(>0.10)
Coal	-2.1795	-2.1948	1.2029	1.1836	-16.038	-16.056	0.14380	0.039017
	(0.2142)	(0.4909)	(<0.01)	(<0.01)	(4.2835e-032)	(4.4749e-036)	(>0.10)	(>0.10)
Cocoa	-3.1326	-2.8355	2.4940	1.4015	-16.999	-17.041	0.25197	0.071679
	(0.0249)	(0.1852)	(<0.01)	(<0.01)	(1.2617e-033)	(1.2168e-038)	(>0.10)	(>0.10)
Coffee	-2.3877	-2.2307	3.6783	1.1251	-16.488	-16.503	0.16729	0.052044
	(0.1458)	(0.4708)	(<0.01)	(<0.01)	(7.5995e-033)	(2.8857e-037)	(>0.10)	(>0.10)
Rapeseed oil	-3.0691	-3.0163	0.79855	0.78605	-23.564	-23.559	0.090518	0.052689
	(0.0296)	(0.1288)	(<0.01)	(<0.01)	(8.1763e-036)	(1.2242e-048)	(>0.10)	(>0.10)
Copper	-2.1642	-2.7652	2.1868	0.85956	-14.957	-14.983	0.15179	0.072051
	(0.2199)	(0.2111)	(<0.01)	(<0.01)	(4.6470e-030)	(4.6867e-033)	(>0.10)	(>0.10)
Cotton	-2.7756	-3.1663	4.3106	0.79200	-13.125	-13.137	0.075921	0.021907
	(0.0626)	(0.0926)	(<0.01)	(<0.01)	(5.7282e-026)	(1.8640e-027)	(>0.10)	(>0.10)
Hides	-2.1326	-3.4564	4.4240	0.45186	-17.022	-17.018	0.062139	0.040189
	(0.2320)	(0.0454)	(<0.01)	(<0.01)	(1.1662e-033)	(1.3866e-038)	(>0.10)	(>0.10)
Lamb	-2.0927	-3.1370	4.7621	0.49735	-16.287	-16.270	0.051279	0.049026
	(0.2478)	(0.0989)	(<0.01)	(<0.01)	(1.6199e-032)	(1.1858e-036)	(>0.10)	(>0.10)
Lead	-2.3201	-3.0984	2.3558	1.1412	-17.407	-17.465	0.27419	0.088271
	(0.1660)	(0.1079)	(<0.01)	(<0.01)	(3.4264e-034)	(1.1215e-039)	(>0.10)	(>0.10)
Soft logs	-2.4452	-2.5084	1.4839	0.95006	-30.414	-30.380	0.054651	0.056701
	(0.1300)	(0.3239)	(<0.01)	(<0.01)	(2.5514e-023)	(2.3546e-043)	(>0.10)	(>0.10)
Hard logs	-3.3474	-3.3385	0.29799	0.29331	-16.514	-16.503	0.046495	0.041226
	(0.0134)	(.0614)	(<0.01)	(<0.01)	(6.8952e-033)	(2.8885e-037)	(>0.10)	(>0.10)
Maize	-2.5609	-2.5740	1.6442	0.97374	-16.235	-16.228	0.061616	0.037406
	(0.1020)	(0.2926)	(<0.01)	(<0.01)	(1.9739e-032)	1.5423e-036	(>0.10)	(>0.10)
Nickel	-2.6995	-2.7492	0.71241	0.40808	-15.215	-15.199	0.050607	0.049578
	(0.0748)	(0.2173)	(0.0142)	(<0.01)	(1.4218e-030)	1.1065e-033	(>0.10)	(>0.10)
Crude oil (1)	-2.3755	-2.6348	1.8614	1.2188	-16.175	-16.168	0.13820	0.070745
	(0.1493)	(0.2650)	(<0.01)	(<0.01)	(2.4922e-032)	(2.2304e-036)	(>0.10)	(>0.10)
Crude oil (2)	-2.3583	-2.6616	1.9883	1.2404	-14.569	-14.561	0.14021	0.067663
	(0.1544)	(0.2533)	(<0.01)	(<0.01)	(2.9515e-029)	(8.1453e-032)	(>0.10)	(>0.10)
Crude oil (3)	-2.4663	-2.6283	1.6414	1.1663	-15.407	-15.393	0.11222	0.069981
	(0.1245)	(0.2679)	(<0.01)	(<0.01)	(6.0650e-031)	(3.0930e-034)	(>0.10)	(>0.10)
Olive oil	-2.7169	-2.8513	1.3961	0.34085	-17.342	-17.325	0.050089	0.050184
	(0.0719)	(0.1797)	(<0.01)	(<0.01)	(4.1836e-034)	(2.4388e-039)	(>0.10)	(>0.10)
Swine	-2.1236	-3.6956	5.5041	1.1438	-19.119	-19.096	0.021389	0.022058
	(0.2356)	(0.0235)	(<0.01)	(<0.01)	(5.6784e-036)	(3.2374e-043)	(>0.10)	(>0.10)
Poultry	-2.9561	-3.9738	3.0472	1.2480	-12.601	-12.601	0.087262	0.012278
	(0.0399)	(0.0101)	(<0.01)	(<0.01)	(1.1197e-024)	(9.1802e-026)	(>0.10)	(>0.10)
Rice	-2.7660	-2.6513	1.8240	0.93348	-14.354	-14.353	0.13023	0.056644
	(0.0640)	(0.2577)	(<0.01)	(<0.01)	(8.5255e-029)	(3.4061e-031)	(>0.10)	(>0.10)
Rubber	-2.5628	-2.5619	0.82683	0.82529	-16.467	-16.467	0.12287	0.087709
	(0.1016)	(0.2982)	(<0.01)	(<0.01)	(8.2220e-033)	(3.5832e-037)	(>0.10)	(>0.10)
Salmon	-2.3965	-2.2646	5.0030	1.5986	-14.950	-14.967	0.16377	0.019771
	(0.1432)	(0.4521)	(<0.01)	(<0.01)	(4.8018e-030)	(5.2094e-033)	(>0.10)	(>0.10)
Hard sawnwood	-2.3839	-2.4672	1.1132	0.54399	-15.648	-15.629	0.062567	0.061269
	(0.1469)	(0.3444)	(<0.01)	(<0.01)	(2.1412e-031)	(6.6914e-035)	(>0.10)	(>0.10)
Soft sawnwood	-4.5984	-4.6489	1.2121	1.0453	-32.797	-32.758	0.030629	0.032311
	(0.0001)	(0.0009)	(<0.01)	(<0.01)	(6.3958e-017)	(5.2279e-038)	(>0.10)	(>0.10)
Shrimps	-2.1831	-3.5401	6.5716	0.37980	-17.676	-17.669	0.060578	0.022917
	(0.2129)	(0.0363)	(<0.01)	(<0.01)	(1.5524e-034)	(3.6915e-040)	(>0.10)	(>0.10)
Sunflower oil	-3.2949	-3.2907	0.85188	0.80449	-14.733	-14.717	0.042436	0.035491
	(0.0157)	(0.0690)	(<0.01)	(<0.01)	(1.3393e-029)	(2.8128e-032)	(>0.10)	(>0.10)
Tea	-3.3667	-3.7902	3.2581	1.0234	-17.141	-17.130	0.049330	0.019479
	(0.0127)	(0.0178)	(<0.01)	(<0.01)	(7.8954e-034)	(7.3239e-039)	(>0.10)	(>0.10)
Tin	-2.3698	-2.1759	1.6798	1.6469	-16.071	-16.163	0.37302	0.065849
	(0.1510)	(0.5014)	(<0.01)	(<0.01)	(3.7599e-032)	(2.2949e-036)	(0.0889)	(>0.10)
Uranium	-2.3935	-2.4157	1.1383	1.1230	-19.374	-19.438	0.38706	0.17630
	(0.1441)	(0.3708)	(<0.01)	(<0.01)	(3.7536e-036)	(7.2905e-044)	(0.0828)	(0.0337)
Wheat	-2.7667	-2.8756	2.7556	0.82033	-16.918	-16.915	0.083519	0.037120
	(0.0639)	(0.1715)	(<0.01)	(<0.01)	(1.6529e-033)	(2.5195e-038)	(>0.10)	(>0.10)
Wool	-2.1483	-2.1673	1.3983	1.3881	-16.512	-16.607	0.32289	0.032917
	(0.2260)	(0.5062)	(<0.01)	(<0.01)	(6.9554e-033)	(1.5472e-037)	(>0.10)	(>0.10)
Zinc	-2.7337	-2.8017	0.66323	0.57495	-16.451	-16.452	0.055000	0.028766
	(0.0691)	(0.1974)	(0.0212)	(<0.01)	(8.7029e-033)	(3.9271e-037)	(>0.10)	(>0.10)
Gold	-2.0093	-2.6409	1.9895	1.6233	-18.418	-18.627	0.78621	0.16259
	(0.2829)	(0.2623)	(<0.01)	(<0.01)	(2.3204e-035)	(2.8275e-042)	(>0.10)	(0.0416)
Silver	-4.0017	-4.1534	1.5359	1.4867	-16.459	-16.600	0.59493	0.17073
	(0.0015)	(0.0056)	(<0.01)	(<0.01)	(8.4594e-033)	(1.6075e-037)	(0.0310)	(0.0369)
Platinum	-2.7086	-3.0883	1.8589	0.99823	-16.628	-16.648	0.22200	0.14970
	(0.0733)	(0.1103)	(<0.01)	(<0.01)	(4.5604e-033)	(1.2076e-037)	(>0.10)	(0.0490)

Table 1.4: Tests for fractional integration. P-values in parentheses.

	Robinson		Phillips (GPH modified)		
	Estimated d	Test statistic (H0: $d = 0$)	Estimated d	Test statistic (H0: $d = 0$)	Test statistic (H0: $d = 1$)
Aluminium	0.82	10.1817	0.68	4.2643	-2.2698
		(0.00)		**(0.00)**	**(0.023)**
Barley	0.98	12.2515	1.20	5.1307	1.4588
		(0.00)		**(0.00)**	**(0.145)**
Beef	1.01	12.6653	0.86	6.2959	-1.0359
		(0.00)		**(0.00)**	**(0.30)**
Coal	1.03	12.8542	0.95	4.6639	-0.3886
		(0.00)		**(0.00)**	**(0.698)**
Cocoa	0.91	11.3802	0.98	5.7380	-0.1325
		(0.00)		**(0.00)**	**(0.895)**
Coffee	0.98	12.2147	1.05	5.8992	-0.3521
		(0.00)		**(0.00)**	**(0.725)**
Rapeseed oil	0.97	12.169	0.72	5.0281	-2.0106
		(0.00)		**(0.00)**	**(0.044)**
Copper	0.94	11.7242	1.01	6.8688	0.0925
		(0.00)		**(0.00)**	**(0.926)**
Cotton	0.87	10.8598	0.74	4.9034	-1.8752
		(0.00)		**(0.01)**	**(0.061)**
Hides	0.74	9.20383	0.79	4.7667	-1.4838
		(0.00)		**(0.00)**	**(0.138)**
Lamb	0.87	10.8386	0.68	4.4530	-2.2811
		(0.00)		**(0.00)**	**(0.023)**
Lead	1.02	12.7302	0.87	6.5707	-0.9483
		(0.00)		**(0.00)**	**(0.343)**
Soft logs	0.86	10.7379	0.90	5.6395	-0.6923
		(0.00)		**(0.00)**	**(0.489)**
Hard logs	0.82	10.2113	0.41	2.0549	-4.2001
		(0.00)		**(0.053)**	**(0.000)**
Maize	0.87	10.8381	0.90	5.1405	-0.7230
		(0.00)		**(0.00)**	**(0.470)**
Nickel	0.94	11.7086	0.96	5.7045	-0.2813
		(0.00)		**(0.00)**	**(0.779)**
Crude oil (1)	0.92	11.4832	0.94	10.2646	-0.4294
		(0.00)		**(0.00)**	**(0.668)**
Crude oil (2)	0.89	11.0739	0.92	9.3813	-0.5456
		(0.00)		**(0.00)**	**(0.585)**
Crude oil (3)	0.87	10.9214	0.86	10.9168	-0.9793
		(0.00)		**(0.00)**	**(0.327)**
Olive oil	0.95	11.8979	0.76	5.1975	-1.7180
		(0.00)		**(0.00)**	**(0.086)**
Swine	0.78	9.7802	0.65	3.6716	-2.5277
		(0.00)		**(0.00)**	**(0.011)**
Poultry	0.62	7.7151	0.69	3.7539	-2.22
		(0.00)		**(0.00)**	**(0.026)**
Rice	0.89	11.1067	1.05	6.2706	0.3265
		(0.00)		**(0.00)**	**(0.744)**
Rubber	0.91	11.3048	0.75	3.6113	-1.8061
		(0.00)		**(0.00)**	**(0.071)**
Salmon	0.86	10.7059	0.79	6.0083	-1.5175
		(0.00)		**(0.00)**	**(0.129)**
Hard sawnwood	1.00	12.5201	0.73	5.0202	-1.9343
		(0.00)		**(0.00)**	**(0.053)**
Soft sawnwood	0.63	7.8529	0.82	4.4328	-1.3087
		(0.00)		**(0.00)**	**(0.191)**
Shrimps	0.83	10.3977	0.67	3.8164	-2.3357
		(0.00)		**(0.00)**	**(0.020)**
Sunflower oil	0.76	9.5307	0.78	4.3901	-1.56
		(0.00)		**(0.00)**	**(0.119)**
Tea	0.68	8.4481	0.56	3.7998	-3.1390
		(0.00)		**(0.00)**	**(0.069)**
Tin	0.93	11.6432	0.82	4.8091	-1.3031
		(0.00)		**(0.00)**	**(0.193)**
Uranium	1.07	13.3625	1.09	7.9628	0.6351
		(0.00)		**(0.01)**	**(0.525)**
Wheat	1.02	12.7206	0.97	7.8164	-0.2298
		(0.00)		**(0.00)**	**(0.818)**
Wool	1.03	12.8517	0.94	4.9489	-0.4403
		(0.00)		**(0.00)**	**(0.660)**
Zinc	1.02	12.7347	0.93	5.8085	-0.5169
		(0.00)		**(0.00)**	**(0.605)**
Gold	1.13	14.0722	1.03	5.5752	0.1986
		(0.00)		**(0.00)**	**(0.843)**
Silver	0.94	11.7045	0.77	4.3956	-1.6591
		(0.00)		**(0.00)**	**(0.097)**
Platinum	0.88	10.9367	0.72	3.4668	-2.0087
		(0.00)		**(0.00)**	**(0.045)**

Table 1.5: ARCH tests (12 lags)

	Mean		Test	
	p	q	Statistic	P-value
Aluminium	5	4	50.4146	0.0000
Barley	2	3	47.876	0.0000
Beef	2	3	38.4912	0.0001
Coal	4	5	60.8587	0.0000
Cocoa	0	1	13.3835	0.3418
Coffee	1	0	38.0405	0.0002
Rapeseed oil	5	4	88.6059	0.0000
Copper	3	2	33.2955	0.0009
Cotton	4	3	125.16	0.0000
Hides	1	2	41.0991	0.0000
Lamb	0	1	36.9926	0.0002
Lead	3	2	80.5701	0.0000
Soft logs	0	1	39.8726	0.0000
Hard logs	4	3	60.1878	0.0000
Maize	1	0	6.54499	0.8862
Nickel	1	0	20.9691	0.0508
Crude oil (1)	3	3	47.0302	0.0000
Crude oil (2)	2	3	30.4658	0.0024
Crude oil (3)	1	0	49.1155	0.0000
Olive oil	1	0	18.3381	0.1058
Swine	5	2	57.7967	0.0000
Poultry	5	5	93.4749	0.0000
Rice	4	3	98.373	0.0000
Rubber	1	0	63.8902	0.0000
Salmon	5	4	28.8653	0.0041
Hard sawnwood	0	1	61.0053	0.0000
Soft sawnwood	4	3	58.6227	0.0000
Shrimps	1	0	54.7222	0.0000
Sunflower oil	1	2	10.0691	0.6099
Tea	1	2	36.2511	0.0003
Tin	2	0	39.2325	0.0000
Uranium	0	5	137.764	0.0000
Wheat	1	0	23.9807	0.0205
Wool	1	0	17.9848	0.1162
Zinc	1	0	20.3161	0.0613
Gold	2	3	54.2141	0.0000
Silver	0	1	53.4596	0.0000
Platinum	0	1	49.5019	0.0000

Chapter 2

Modelling commodity price dynamics: market fundamentals and latent factors

2.1 Introduction

Commodity prices modelling is traditionally challenging and opens several possibilities; literature has been prolific but, at the same time, several issues are not yet fully explored. The way primary commodity prices are modelled mainly depends on the specific question researchers choose to look at. In addition, many theoretical choices may be specific only for some kinds of commodities, making aggregation harder and implying a lack of some generality. The specific features of primary commodities make these markets unique with respect to other assets, and *ad hoc* specifications for price formation are required (i.e. the possibility of commodities to be stored).

The resulting choices for modelling prices should reach a compromise between the necessity of a good specification, founded upon reliable theoretical properties, and the requirement for a certain degree of generalisation: the inclusion of too many variables or assumptions reduces model's tractability (especially in a multi-commodity framework) and the consequent empirical exploitation can be no more taken for granted[1]. This Chapter concerns the development of a model which should be able to capture the joint movement of different commodity prices, movement that could be split in two components: one short- and the other long-term oriented, both reflecting market fundamentals.

This work is inspired mainly by those of Schwartz (1997); Schwartz and Smith (2000), encompassing also the general idea of Gilbert (1995). This latter work - the starting point of the present study - derives a four structural equations model for the aluminium market (respectively, consumption, pro-

[1]For instance, non-negativity constraints imposed in the standard storage model imply non linearities and prevent us to derive analytical solutions.

duction, net imports and storage equations) and then reduces the number of state variables from nine to only two, by aggregating them according to their short or long-term influence on price movements. The price and stockholding reduced equations are then estimated by iterated non-linear three-stage least squares, together with the consumption, production and net imports equations. Finally, thanks to a huge number of inter- and intra-equation restrictions, the structural parameters are recovered by testing the Rational Expectations hypothesis. Similar models have been considered and developed by Perali and Pieroni (2003); Pieroni and Ricciarelli (2005); Boschi and Pieroni (2009); Bonfatti (2012), all exploiting the intuition that thanks to market fundamentals reflecting price disequilibria it is possible to re-write and estimate the storage equation without explicitly considering the expected price at time $t+1$, which of course is a non-observable variable.

The former approach by Schwartz (1997) links instead the stochastic behaviour of commodity prices with a number of factors able to explain the overall dynamics. The particular case of a two-factor model, for instance, allows for a short-term mean reverting variation in prices and uncertainty in the equilibrium level to which prices revert through the long-term component, reflecting fundamental changes that are expected to persist (Schwartz and Smith, 2000). Given that the two factors are not observable, Kalman filtering techniques are required for the estimation. The unobserved factors can be recovered by use of the Kalman smoother.

The model presented here is developed combining these two traditions. A system of equations expressing the market dynamics is introduced, for a multivariate framework; the equilibrium price is found with the assumption that at each time, markets have to clear. This is consistent with the assumption that in competitive markets, price equations should correspond to the law of demand (in case of storable commodities including demand for stocks and for direct consumption) and supply (see Trivedi (1990)). Following Gilbert (1995), the estimated price equation will depend on the two unobservable short- and long-run components, and on some exogenous variables introduced in the three market equations system. The main difference is that the two aforementioned variables will be introduced in a different way and then treated as latent factors, one capturing the mean reverting common variations and the other one the non-stationary part. Specifically, the intuition is that the latent variables are estimated in spite of the unobservable price expectations. The idea that speculators form expectations about the price of a specific commodity by looking at the common dynamics of a larger set of commodity prices reflects the fact that commodity markets are interrelated and common movement plays a crucial role in determining the way expectations are formed. This is consistent with the strand of literature explaining that there is a positive relationship between speculation and co-movement (see Chapter 1), and makes it possible to include this common movement through the channel of speculators' expectations. The final reduced price equations will constitute in fact a Dynamic Factor Model, written in state space form: there will be two equations, one for the vector of commodity prices and the other for the law of motion of the latent factors. In this way the issue of lack of - or unreliable - data on commodity stocks is not present at all, since an equation for stocks is not estimated. The main novelty consists on the multi-commodity framework, which allows to explicitly include the common movement through the way expectations are formed, a task that has never been accomplished in the literature, to our knowledge. Estimation of a model of this kind would require Kalman filtering and smoothing, however this has proven to be rather infeasible, so that it will be done by use of a new procedure presented and developed in Chapter 3.

This Chapter is organised as follows. Section 2.2 introduces the general idea of the model, also providing a brief recap of the underlying literature. Sections 2.2.1, 2.2.2 and 2.2.3 present the consumption, production and storage behavioural equations. Section 2.3 closes the model with the market clearing condition and provides the final estimable specification with the inclusion of latent factors. Section 2.4 report the estimation technique and the related principal results, introducing also the idea and the aim of Chapter 3. Section 2.5, finally, summarises and concludes.

2.2 A three structural equations model for market fundamentals

The theoretical framework about how to capture commodity price dynamics dates back to Gustafson (1958) and the early attempts to build what will be the milestone of this field: the storage model. This model relates speculators' behaviour and their expectations on future price changes. Briefly, the intuition is that the price of a commodity adjusts thanks to the beliefs of stockholders: when the current price is above the expected value of next period, they would not store the commodity, while the opposite occurs when actual price is below the level. The model implies thus a situation of risk neutrality, so that additional stocks are carried as long as there exist a positive return. In the stock-out case, which is the absence of storing incentives, the dynamics of prices are simply supposed to follow the path of the underlying demand/supply balance. The possibility of stock-out splits thus the model in two regimes, in which stocks are either zero or positive.

After the seminal works of Gustafson (1958), Samuelson (1971) and Scheinkman and Schechtman (1983), the standard storage model is considered to be that of Deaton and Laroque (1992) and Williams and Wright (1991)[2]. The standard model confirmed to be able to explain the most relevant features of the dynamics of price *univariate* time series, particularly isolated spikes and conditional high price volatility. However, empirical tests of the model

[2]For a detailed review of the storage model, see Wright (2001) and Cafiero et al. (2011).

(Deaton and Laroque, 1992) failed to reproduce all the commodity price autocorrelation. As suggested in Cafiero et al. (2011), the partial rejection could be possibly due to assumptions such as stationarity and linearity of consumption demand and supply modelled merely as random shocks; the residual autocorrelation, the part which remains unexplained, may be plausibly due to trends in the yields or structural changes in demand or supply functions, which are not entering in the model. The non-linearities (arising from the non-negativity constraints imposed by the presence of storage) in demand prevents the practitioners to derive an analytical solution of the model, which is instead solved via dynamic programming. This, and the impossibility to include many state variables in the supply and demand equations, poses a limit to the exploitation of the storage model for the purpose of this work. As pointed out by Gilbert (1995), a solution for the storage function can be derived in closed form if one is prepared to ignore the non-negativity constraint, switching from a two-regimes situation to a singular linear stockholding equation of the form:

$$S_t = d_0 + d_1(E_t[P_{t+1}] - (1 + r_t)P_t), \tag{2.1}$$

where S_t is the storage demand at time t for a particular commodity, P_t denotes the price for that same commodity, $E_t[P_{t+1}]$ is the expectation at time t of the price at time $t + 1$, r_t is the interest rate and d_1 is assumed as a constant parameter (being A the coefficient of market absolute risk aversion):

$$d_1 = \frac{1}{AE_t[P_{t+1} - E[P_{t+1}]]^2}.$$

Such formulation assuming risk aversion become quite different from those of the traditional group of storage models. Models of this kind, inspired the Muth (1961)'s one of rational expectations, simply ignore the non-negativity constraint because of the huge gain of a linear price model. In any case, this could be justified if there is a large amount of non-speculative inventories able to nullify the possible negative speculative stockholdings (Gilbert, 1990).

The main difficulty in estimating a structural stock equation as in Equation (2.1) - thus regressing the level of stocks on the expected speculative gain - arises from the impossibility to observe the expected price $E_t[P_{t+1}]$. The inclusion of the Rational Expectations Hypothesis (REH) in Muth's model suggests that this issue could be easily handled. The REH implies that

$$P_{t+1} = E_t[P_{t+1}] + \epsilon_{t+1},$$

or in words, the realised price will differ from the expected one by an innovation, with $E_t\epsilon_{t+1} = 0$. Empirically, handling a structural storage equation of this kind will require some kinds of substitution for the unobserved expected price, but estimated models perform poorly (Gilbert, 1990). Suggested approaches imply to substitute a future price, when available, to substitute the fitted values of an ARIMA model or to substitute the actual price and estimate by Instrumental Variables to control for the measurement error. Alternatively, the inversion of the equation may be considered, with price as the dependent variable, but the issue of unreliable or incomplete stock data persists, as stocks series enters in the equation as a covariate. One solution consists in deriving the solved commodity price equation by embodying the stock equation in a system of demand and supply equations, thus estimating the reduced form of commodity price equation, as in Gilbert (1995).

In general, the storage model is still the best approach for capturing the exact behaviour of single commodity prices, or for analysing the relationship between inventories and price. The following model is addressed, instead, in splitting the co-movement of different prices into two main drivers, and the focus is more on the latent factors aspect than in the model parameters and the related policy implications. For this reason, an alternative approach is adopted. The log-linear specification of the three behavioural equations allows for a general application to different categories of commodities and ease the transition to an estimable reduced form price equation. Consumption, production and storage equations are here constructed to be rather simplistic; more accurate specifications require to focus on some particular kind of commodities and within a more specific framework. For instance, supply models for agricultural commodities may include a harvesting equation, other than potential production expressed as investment in trees (an example is given by Wickens and Greenfield (1973)). Nevertheless, to capture the global effect of consumption, production and speculation on different prices in a short- and long-term perspective, a condensed specification is preferred, since the key point is in the interdependences and common factors. More importantly, the model has to be handled empirically, so that the provided equations should be easily led to a plausible estimable specification. The linear tractability gained in the stocks equation by deleting the stock-out possibility also allows us to include more variables in the consumption and supply equations. As a matter of facts, one of the most discussed limits of the storage model is the size of harvest (thus, the production side) merely modelled as a random disturbance, together with a rather simple demand equation. This is the direct consequence of the desire of emphasising *stocks*, but in any case, as suggested in Deaton and Laroque (1996), some of the positive autocorrelation of prices should be explained through supply and demand fundamentals. Throughout the forthcoming sections, the various approaches followed by existing works will be listed together with the adopted specification. Further differences between this kind of models - *à la* Gilbert (1995) and the competitive storage model will be shown when the storage behaviour will be discussed.

2.2.1 The consumption side

As seen in Chapter 1, the effects of demand shocks are more likely to be persistent in the long run. This is due to the fact that demand for primary goods such as commodities is rather sticky and is primarily due to the level of global income and of global population. The global demand for commodities could be split in two parts: the demand for direct consumption and that for the speculative and storing activity of the stockholders. This Section deals with the former, whereas the latter enters in the storage equation (see Section 2.2.3).

Basic microeconomic theory relates the demand of a good to three variables: its own price, the price of the other goods, and income. Assuming the examined goods are ordinary, the demand of a specific commodity should be decreasing if its own price increases, and increasing if price of substitutes and income increase. Tipically, cross-price effects will be in general more relevant among closely related commodities, whereas should be absent for totally different goods (for example, there is no economic reason to assume this effect will exist between a metal commodity and a crop one). Generally, demand equations in empirical models are based on these assumptions, even if not explicitly micro-founded. The consumption of a good is generally assumed to depend on these three variables, even in an aggregated perspective such as the present one, with differences within literature between linear and log-linear forms. Univariate models often do not consider other goods' prices - even if there are some exceptions - even if according to economic theory both substitutes and complements prices should enter in the equations.

In any case, note that instead of a single demand equation, in this work there is a *system* of n demands, one for each commodity. Particularly, this is not a complete demand system, because we are modelling demand for n goods and prices, whereas there exist $n + m$ prices (with $m > 1$) which are not considered in the system. In microeconomic theory, incomplete demand systems arise when the model concerns only a group of commodities which form a subset of total individual budget, but this is not an issue if one is interested to impose integrability conditions and recover the associated utility function's parameters. It is possible to correctly specify the log-linear system and to recover the associated functional form by imposing integrability conditions (Epstein, 1982; LaFrance and Hanemann, 1989; LaFrance, 1985, 1986, 1990; Von Haefen, 2002). These results could be particularly useful as they provide specific restrictions, can add to the modelled demand some micro-founded properties and more importantly, allow to link formulations of these kind to micro-founded utility functions.[3] Integrability conditions could be exploited in

[3]The majority of applied works directly specify a demand function without maximising an utility function subject to a budget constraint (or applying the Roy's identity to the indirect utility function), due to gains in simplicity. The importance of integrability depends upon which is the purpose of the investigation. For instance, in welfare analysis and consumer surplus' theory, integrability is an essential element. In this kind of analyses, checking for integrability of the demand system provide useful parameter restrictions for estimation purposes and could be used as a test of demand modelling choices, by possible rejection of the integrability conditions. Nevertheless, as pointed in LaFrance (1986), if data reject the integrability conditions, this may be due to overly restrictive hypotheses, rather than

further analyses to impose restrictions for recovering the structural parameters of the current model, which is beyond the scope of this work.

Among the choices about how to model consumption of commodities in literature, Gilbert (1995) relates the consumption of aluminium to its own lagged price, a trend-modified industrial production index and a OECD construction index. Consumption demand for cocoa of Bonfatti (2012) is a function of its real current price and real income, given by the weighted GDP of major consuming countries. Pei and Tilton (1999) adopt a similar specification, but exploiting a partial adjustment model. Boschi and Pieroni (2009) slightly modify the Gilbert model for aluminium by inserting in the demand function also the role of monetary policy, measured by the real interest rate. There exist several models adopting specific demand functions according to what is the commodity under analysis. For instance, Shi et al. (2018), also working with aluminium, include in the demand equation the price of copper, the heavy industrial development, proxied by fixed investment and a very specific variable, namely development of the automobile manufacturing industry, among the others. Demand for Norwegian electricity is specified as a linear autoregressive distributed lag function, including temperature and some exogenous variables as the price of alternative fuels, activity level and day-length (Johnsen, 2001). Zink et al. (2016) emphasise the role of secondary markets for metals, including in a structural equations model two different demand linear functions. In both of them, demanded quantity depend on the price and on the differential between the two prices.

Generally, for metal demands, price enters in the consumption equation with lags, whereas in the case of agricultural commodities it is more common to use the current price. This reflects the intuition that consumption of staple goods, such as agricultural commodities, immediately adjusts to changes in prices, whereas it takes some time to modify consumption patterns after price shocks in the case of other commodities. The opposite reasoning applies for the formation of the supply equation. Since this model concerns many different commodities, the general specification including only current price is preferred. In this contest, it is more likely that price appear with some lags in the *production* side, rather than in the consumption equation; this happens because for many commodities, mining or cropping are planned in advance, and so suppliers make their decisions in advance looking at current prices, and then the output will become available only after some periods of time. On the contrary, we assume consumers decide to buy the desired commodities by looking at the observed prices.

A log-linear multivariate specification

In this work, commodities' consumption is the result of the combined effect of world real economic activity and the set of commodity prices; the latter

irrational consumer choices.

is comprehensive of complements and substitutes. Another variable affecting global consumption, as remarked in Chapter 1, is global population. However, it would be more correct to give different weights to demographic dynamics for the consumpion of commodities according to their kind (for instance, population matters more for food commodities, and particularly livestocks). Another issue is finding data on demographic variables in appropriate medium-high frequency; indeed data on world population are not disposable at a frequency higher than yearly. For this reason a specification accounting only on prices and world economic activity is preferred. The reason why a proxy of world income such as GDP is not taken into account in spite of some index of global economic activity has to be found in data availability. World GDP is provided by the World Bank only at an annual frequency. There are quarterly measures of GDP but only taking into consideration some aggregates of countries (i.e. the OECD ones), but some countries which impact is expected to be very huge are not included, as China. One solution could be the construction of an *ad hoc* global weighted GDP starting from single countries data, one other could rely on interpolation of data to an higher frequency. Since this work exploits monthly data, however, quarterly data are still better than annual ones. The issue is very controversial within the field and for this reason some monthly indexes of real economic activity have been developed and then used instead of proxies of world income, in analyses similar to the current one. These indexes are usually based on the changes in prices of some industrial commodities (thus not well suited for the present scope) or are built upon the volume of industrial raw material shipping (Kilian, 2009; Ravazzolo and Vespignani, 2015; Hamilton, 2019). The Kilian index, in particular, consists in a "cyclical variation in global real economic activity based on percentage changes in representative single-voyage ocean shipping freight rates available for various bulk dry cargoes, further differentiated by the size of the vessel and the shipping route". The rates of growth, averaged and adjusted for US inflation, are then de-trended; the obtained index is stationary. The original Kilian (2009) has been corrected as suggested in Hamilton (2019) and the final series is presented in Kilian (2019).

To resume, the global consumption C_{it}, capturing the demanded quantity for a specific commodity i, could thus be expressed as a function of its price, P_{it}, the prices of all the other $n-1$ commodities and world real economic activity x_t. With all these quantities expressed in logs (denoted in lower case), it is possible to interpret all the coefficients as elasticities. The log-linear consumption equation is thus:

$$c_{it} = \alpha_0 + \sum_{j=1}^{n} \alpha_{ij} p_{jt} + \alpha_2 x_t, \tag{2.2}$$

where α_0 is the constant term, $\sum_{j=1}^{n} \alpha_{ij}$ are the commodity own and cross elasticities, p_{jt} is the log of j-th commodity price, α_2 is the real economic activity elasticity and x_t is the Kilian index[4]. This index enters in the system modelled as a AR(1) stationary process with drift:

$$x_t = \mu^x + \rho^x x_{t-1} + \epsilon_t^x. \tag{2.3}$$

Being ϵ_t the innovation term.

2.2.2 The production side

Commodity production modelling is particularly arduous because of the very different approaches which could be adopted; in general terms, the main contributions to the supply specification come from the partial adjustment model and adaptive expectations formulation (Nerlove, 1958a,b) and the production function approach. Other authors have opted for modifications and improvements of the two cited models. In models concerning not only the supply side but encompassing the whole market, as in the present work, the majority of empirical contributions selected linear or log-linear specifications, with the inclusion of different variables in each case. The proposed specification takes inspiration from Nerlovian models, but the final result will be slightly different.

Nerlovian models

Nerlovian models start from the assumption that an economic cause, such as a price change, will produce its effect (i.e. the variation of demanded quantity of a good) only after some lag in time. This is the intuition upon which the so called *distributed lags* are built, which are an extension of the Cobweb theorem result. In such a framework, suppliers take decisions observing the current price, but in the period the output will become available the price will be different. At the same time, suppliers adjust at each period the total production in a given proportion, compared to the long run equilibrium quantity. The basic Nerlovian model can be summarised in three equations, of which the first is the supply of a given crop commodity, the second models the formation of price expectations and the third expresses the adjustment of production:

$$A_t^* = a_0 + a_1 P_t^e + a_2 Z_t + v_t \tag{2.4}$$
$$P_t^e - P_{t-1}^e = b(P_{t-1} - P_{t-1}^e) \tag{2.5}$$
$$A_t - A_{t-1} = c(A_t^* - A_{t-1}), \tag{2.6}$$

in which: A_t^* is the maximum output (or desired, potential); P_t^e and P_{t-1}^e are the expected prices at time t and $t-1$, respectively; Z_t is a vector containing other exogenous variables influencing production; v_t is the error term; A_t and A_{t-1} express total quantity of output produced at time t and $t-1$; b is

[4]Note that the Kilian index is already expressed in logarithms, so it is not necessary to further transform this variable.

the so called expectation coefficient; c is the adjustment coefficient[5]. Models of this kind have been widely exploited for agricultural commodities, and in particular to formalise acreage response, so that A_t expresses the total area under cultivation rather than production, and A_t^* the total amount of disposable land, being land an essentially fixed resource. Askari and Cummings (1977); Rao (1989) provide detailed expositions of Nerlovian models, the contributions from next works and the estimated supply elasticities for many commodities. Among the modifications that have been proposed, there are those concerning variables used by Nerlove (the formation of price expectations, the choice of a proper deflator, output measurements), those regarding which factors should enter in Z_t, and some proposals to better suit particular kinds of commodities, primarily perennial, slow-maturing crops and livestock products. The original Nerlovian model was indeed thought for one-season crop commodities. For what concerns the second point, the most frequently included variables are the prices of the main substitute commodities, weather conditions, technological progress or yields.

The Nerlovian model encompasses adaptive expectations, so that it is possible to relate the entire supply equation to observed terms, i.e. from P_t^e and P_{t-1}^e to P_{t-1} and P_t. In the particular case in which b is equal to 1, equation 2.5 simplifies to

$$P_t^e = P_{t-1} \tag{2.7}$$

which is the same result of the standard Cobweb theorem. This means that suppliers expect price will be the same as observed, but, again, it occurs some time for the production process, so that there is a lag of one period.

Later on, production function approach gained popularity due to the underlying microeconomic properties, i.e. the assumption of profit maximising suppliers. Supply simply derives from the maximisation of a profit function in which a certain functional form is specified for production. The introduction of many functional forms has been crucial in this sense. The main issues arise from the difficulty of selecting a proper production function, or the cost function, and the problem of having the necessary data. Specifying a proper model in this way and for many commodities could become very arduous.

As stressed, agricultural commodity supply is often modelled exploiting models *à la Nerlove*, rather than following the latter approach. Production for metals is often modelled with the distinction of two fundamental equations, one for primary supply and the other for secondary materials, coming from recycling. Production function approach is often used to model primary supply.

The proposed specification

Both approaches have indisputable credit, but there is a general lack of multi-commodities applications. Specifying a functional form and a profit function may be difficult because of the many dissimilarities among different types of

[5] $0 < b \leq 1$ e $0 < c \leq 1$. If the variables are expressed in logs, b and c are elasticities.

commodities. At the same time, the usual way to include other commodity prices in Nerlovian applications is to insert the most important substitute in Z_t, and nothing more. The specification adopted here relies on the belief that supply of commodities is the joint result of two main definitions: total capacity and capacity utilisation, with an inspiration from Nerlovian models, even if not properly equal to them. The idea of a fixed amount of resources entering in equation (2.6) of Nerlovian models perfectly fits the situation of commodities production, since mines and land are finite resources.

Capacity, which could be thought as the A_t^* variable in equation (2.4) and (2.6), expresses the total disposable resources which could be used to produce a particular commodity i. Here it is named W_t^*. Suppliers have to decide how to allocate that fixed amount of resources among different commodities, and they choose according to the relative prices of the various commodities. Taking all quantities in logs, it is expressed as:

$$w_t^* = \sum_{j=1}^{n} \beta_{ij} p_{jt}^e,$$

where w_t^* expresses the total disposable resources at time t which could be used to produce a particular commodity i. All the β_{ij} are in this case the own and cross elasticities. The logarithm of prices are denoted with small letters, p_{jt}, whereas expected prices, which could not be observed, are assumed to be equal to lagged price, as assumed by adaptive expectations hypotheses. However, for the sake of simplicity, equation (2.7) is preferred to equation (2.5). In this way it is possible to substitute $P_{it}^e = P_{i,t-1}$. As a matter of facts, the majority of empirical works include $P_{i,t-1}$ in the supply equation, rather than P_{it}. This intuition reflects the fact that in the production phase of commodities, and especially for agricultural ones, there is a considerable lag of time between suppliers' decision making and output. Capacity utilisation, instead, is the result of the choices made by suppliers among different commodities; once allocated to each product, resources are then exploited for production. The produced quantity for each commodity is so the result of the multiplied amount of used resources w_{it} and the associated productivity y_{it}. The log formalisation is:

$$q_{it} = w_{it} + y_{it}.$$

Again, q_{it} denotes the log of produced quantity Q_{it}. Nerlovian formulations[6] imply that:

$$w_{it} = \gamma w_t^* + (1 - \gamma) w_{i,t-1},$$

with γ being the adjustment elasticity.

[6] See Nerlove (1958a,b) for adaptive expectations and proportional adjustment of production.

Rearranging and substituting, it is possible to write the production equation as:

$$q_{it} = \gamma(\sum_{j=1}^{n} \beta_{ij} p_{j,t-1}) + (1-\gamma) w_{i,t-1} + y_{it}, \tag{2.8}$$

which is the final production equation for the i-th commodity.

2.2.3 The storage equation

Assuming a system without storage, the equilibrium price would be found by the simple equality of the demanded quantity of a given commodity, and the produced one. However, an assumption of this kind would seem rather unrealistic: for most commodities, inventories are carried, and the amount that could be stored could basically be as large as desired. Of course some commodities could suffer some deterioration, with it generally assumed to be proportional to the quantity stored (Williams and Wright, 1991). As mentioned above, storage has been introduced in commodity price modelling by Gustafson (1958) and since then, the majority of literature has taken into account the behaviour of the stockholders. The intuition is that storage in a particular period depends on the price which could be expected in the forthcoming period, whereas at the same time, that price will depend on the aggregate level of stocks at the same period; it clearly emerges that the way expectations are formed plays a crucial role, given that the problem involves uncertainty. Since the seminal work of Muth (1961), rationality has become a fundamental assumption, thus implying a forward-looking behaviour. Other assumptions of the basic storage model include risk neutrality and price-taker stockholders, and more importantly, the non negativity of aggregate storage. The latter has several consequences, ranging from the impossibility to borrow from the future to the non-linearities which prevents to solve the model analytically.

In the standard storage model, profit maximising stockholders of a commodity hold an amount of inventory I_t; given a constant interest rate, r, and a rate δ at which stocks physically deteriorate, inventories are:

$$I_t = 0 \quad \text{if} \quad \frac{1-\delta}{1+r} E_t[p_{t+1}] < p_t$$

$$I_t \geq 0 \quad \text{if} \quad \frac{1-\delta}{1+r} E_t[p_{t+1}] = p_t$$

where the cost of holding inventories is given by $(1-\delta)/(1+r) < 1$. Inventories are demanded if expected profits are positive, while they are zero if stockholders expect losses from holding them (Deaton and Laroque, 1992, 1996).

In order to gain in analytical tractability of the storage equation, which could be treated only by dynamic programming in this formulation, here the non-negativity constraint is ignored as in Gilbert (1995); Pieroni and Ricciarelli

(2005); Boschi and Pieroni (2009). Assuming also the deterioration rate δ as negligible (which is especially true for non-perishable commodities), the storage demand equation for the i-th commodity becomes, with variables in logarithms as before:

$$s_{it} = \eta_0 + \eta_1(E_t[p_{i,t+1}] - p_{it} - r_t), \tag{2.9}$$

which is simply the logarithmic near equivalent of (2.1). η_1 expresses the degree of risk aversion and η_0 is assumed constant in Gilbert (1995), but in this case the hypothesis of $\eta_0 = s_{i,t-1}$ of Boschi and Pieroni (2009) is preferred, in which the intercept represents the initial state of the variable s_{it}. In this way the model predicts that speculators will react to some market imbalances by observing the current price, which may be above or below the expected price and modifying the stockholding behaviour to regress back to the optimal equilibrium.

The interest rate r_t is not assumed as constant, but it is rather modelled as:

$$r_t = \mu^r + \rho^r r_{t-1} + \epsilon_t^r, \tag{2.10}$$

where ϵ_t^r is a white noise. The series is considered as a stationary process[7]. Note that, on the contrary of the consumption and production equations, here the price of the i-th commodity is not influenced by the prices of all the other commodities, at least not directly. This assumption reflects the behaviour of speculators who look at the expected gain on that particular commodity, as standard.

2.3 The market clearing condition

The system determining price behaviour of a particular commodity i is composed by Equations (2.2), (2.8) and (2.9). Two exogenous variables complete the model, namely a proxy of global real economic activity and the real US interest rate. The market clearing identity allows to close the model and is essential to find the equilibrium price for the n commodities. Markets clear when the total availability of each commodity equals the total demand. The former is the sum of production and stocks deriving from previous period storage activities, whereas the latter includes consumption and the current demand for stocks:

$$q_{it} + s_{i,t-1} = c_{it} + s_{it}. \tag{2.11}$$

[7] Literature generally considers interest rate as stationary on the long run and non stationary for brief time spans. Unit root tests performed on real US interest rates (ADF, Phillips-Perron, KPSS) conclude that the hypothesis of stationarity has to be preferred.

implying that:

$$\gamma(\sum_{j=1}^{n}\beta_{ij}p_{j,t-1}) + (1-\gamma)w_{i,t-1} + y_{it} - [\alpha_0 + \sum_{j=1}^{n}\alpha_{ij}p_{jt} + \alpha_2 x_t] = \gamma_1(E_t[p_{i,t+1}] - p_{it} - r_t). \quad (2.12)$$

An equation of the kind of (2.12) could not be estimated, primarily because of the impossibility to observe price expectations. An option to deal with expectations is to solve the model in a rational expectations framework, following Muth (1961). This has already been done (Gilbert, 1995; Pieroni and Ricciarelli, 2005; Boschi and Pieroni, 2009), and estimation would require to jointly consider stock and price as dependent from two state variables summarising the information about short- and long-run dynamics.

Alternatively, if the common movement hypothesis holds, then it is reasonable to assume that stockholders' expectations reflect the overall dynamics of commodity prices, more than only some specific price. If one price is expected to rise and there is tendency of all the series to move together, then also all the other prices will be expected to rise. With this intuition, it is possible to think that price expectations could be substituted with a an expression of some common factors, specifically capturing short-term and long-term dynamics upon which expectations are formed. Note that there is a linkage proposed by the literature between speculation and co-movement, suggesting that it is reasonable to assume speculators' behaviour influence the common movement through expectations as a channel.

In particular, we assume that stockholders do observe or rather perceive these common drivers as insiders of the related markets, which however are not observable by the analyst and practitioner. This is expressed by substituting $E_t[p_{i,t+1}] = \Lambda F_t$, in which F_t contains the dynamics of the short- and long-term common dynamics and Λ relates each commodity price to the underlying common factors. The short-term fluctuations are those expected to disappear after a given time span, whereas the long-term movements will have a permanent effect, determining the overall trend of commodity prices. The short- and long-term decomposition of the market fundamentals has already been proposed in the model of Gilbert (1995), in which the two state variables z_{1t} and z_{2t} summarised information of the whole model. In particular, the z_{2t} variable is constructed to measure the gap between production (net of imports) at $p_t = \bar{p}$ and the consumption trend, thus represented the long-term fundamental, whereas the short term fundamental z_{1t} is interpreted as the market excess supply when at a reference price of $p_t = \bar{p}$. After any manipulations, the Gilbert price equation (to be jointly estimated with the stock equation) is dependent on the two derived state variables, plus a constant, the exogenous variables, and the idiosyncratic component. The current specification derives a different solution, but with f_{1t} and f_{2t} - which have the same economic meaning of the two market fundamentals developed in Gilbert (1995) - inserted into the F_t vector of latent common factors. Note that these latent common factors should capture the co-movement originated by the same common underlying drivers, whereas the interdependencies among commodity prices are yet captured by the consumption and production equations. In particular, complementarity and substitutability in consumption make prices of a given commodity dependent also from other prices; supply modelling introduces this interdependency with the allocation of a common (among the sub-groups) fixed resource, which is the key point of nerlovian formulations. As a consequence, the system of the final price equations comprehend the coexistence of price interdependences and of common latent drivers.

In particular, for what concerns the latter point, assume that

$$E_t[p_{i,t+1}] = \lambda_{i0} + \lambda_{i1} f_{1t} + \lambda_{i2} f_{2t}$$

in which f_{1t} expresses the short-term variations in commodity prices and thus is a stationary process reverting towards 0, and f_{2t} is a $I(1)$ process representing the long-term price common fluctuations. Then it is reasonable to assume that the parameters λ_{i1} and λ_{i2} are positive, as a positive excess of supply (demand), induced by both short- and long-term variations, induces price expectations of *declining* (increasing) prices with respect to the market balance expressed in the constant term λ_0. The two latent factors are represented by the two processes:

$$f_{1t} = \xi_1 f_{t-1} + \epsilon_t^{f1} \tag{2.13}$$

$$f_{2t} = f_{t-1} + \epsilon_t^{f2}. \tag{2.14}$$

The former is a stationary AR(1) process whereas the latter is a random walk[8]; ϵ_t^{f1} and ϵ_t^{f2} are white noises.

Now the price equilibrium can be found as with this representation Equation (2.12) becomes estimable.

2.3.1 The reduced form price equation

The final estimable reduced form price equation will be, in matrix notation and condensing terms:

$$P_t = \alpha + \Pi P_{t-1} + BF_t + HX_t + \varepsilon_t \tag{2.15}$$

$$F_t = A_1 F_{t-1} + \omega_t. \tag{2.16}$$

Instead of a single price equation, we have a state space representation of two equations which has to be jointly estimated. The second equation specifies the law of motion of the latent common factors and is called *state* equation.

[8]The model can be extended to other specifications for the two latent variables, as for example letting the stationary one to be a AR(2); here we present the most simple case.

Equation (2.15) is obtained by setting $\alpha = \frac{\gamma_1\lambda_0 - \alpha_0}{\gamma_1 - \alpha ij}$[9], and grouping the commodity-specific terms into the idiosyncratic term $\varepsilon_t = \frac{(1-\gamma)w_{i,t-1}+y_{it}}{\gamma_1-\alpha ij}$. Matrix Π contains the VAR parameters in reduced form, matrix B is $n \times q$ and links prices to the unobserved factors, thus contains the parameters λ_{i1} and λ_{i2}.[10]. H is the $n \times k$ matrix, with $k = 2$ being the number of exogenous variables, containing the parameters $\alpha_2/(\gamma_1 - \alpha ij)$ and $\gamma_1/(\gamma_1 - \alpha ij)$. Finally, A_1 is a diagonal matrix of dimension 2×2 containing ξ_1 and 1 on the diagonal. P_t and P_{t-1} are $n \times 1$ vectors containing all the log commodity prices at time t and at $t-1$, respectively ; X_t is the 2×1 vector containing the two exogenous variables of the model.

An equation of the kind of (2.15) expresses a VAR model with an unobservable component F_t (a $q \times 1$ vector) and requires to be estimated through the Kalman filter and Maximum Likelihood. It is relevant to stress that whereas Π contains information on the co-movement generated from price interdependencies, B focuses instead on the other kind of common movement. We remind also that this is a non-stationary VAR, as the analysis in Chapter 1 has stated.

The structural parameters of the model cannot be recovered if not by imposing appropriate restrictions. The most suitable solution would be to impose sign restrictions, which however allows only to constraint the *sign* and not the value of the parameters, and to exploit microeconomic theory and the already mentioned integrability conditions. The matter is very controversial and appropriate studies focusing on the topic are for sure needed. However, this is not the scope of the present work, as the focus is on the development of a multi-commodity model encompassing common latent factors and the subsequent estimation of the common movement of commodity prices. Since this is not trivial, as next Section will show, we have chosen to focus on these aspects. The structural model presented here is a starting point of further analyses which may want to try to develop a more sophisticate model. To our knowledge, this is the first attempt trying to build a structural model for *many* commodity prices embodying also latent variables.

2.4 Estimation

Equations (2.15) and (2.16) can be estimated through Kalman filtering techniques, given the fact that F_t is a vector of unobservable variables. There are nowadays several techniques for estimating a system with latent factors (see the literature review about DFMs provided in Chaper 3), but Kalman filtering and smoothing is the only *parametric* one, meaning that it is possible to impose *a priori* a precise structure to the two equations, the number of dynamic factors (and their stochastic properties) and the distribution of the disturbances. More importantly, it allows to provide theoretically related con-

[9] Of course for estimation purposes, λ_{i0} is set as constant λ_0.
[10] q is the number of latent factors, in this case equal to 2.

straints on the parameters. The Kalman filter is a recursive procedure allowing the estimation of a model written in state space form, containing observed and unobserved variables; the unobserved component is recovered using the information contained in the observed at time t and is continuously updated as new information becomes available. The estimation requires an initialisation for both the mean and variance. Model parameters are then estimated by ML; Harvey (1990) provides an excellent review for the functioning of the kalman filter. Extensions for the non-stationary cases are implemented through the diffuse Kalman filter, which provides an arbitrarily large value for the initial state of the variance.

Unfortunately, estimation via Kalman filter becomes infeasible for large systems, as the number of parameters increases (Chapter 3 will deepen the matter) and it becomes more common to encounter convergence problems of the algorithm. Moreover, the present model has not not only a multivariate setting, but also a non-stationary set-up. For this reason, plain estimation with the standard techniques has proven to be impossible. We failed to estimate the model presented above and synthesised in Equations (2.15) and (2.16) even taking into account only three commodity prices and imposing severe restrictions.[11]

This has been the starting point for the development of a suitable technique able to estimate the model presented above, and has paved the way for the work of Chapter 3 of this Thesis. Even if there are no other fully parametric estimators allowing to impose a precise structure to the equations, we have tried to combine the efficiency of Kalman filtering and other less problematic approaches which could allow to impose a certain structure to the model. In this way we have been able to estimate a stationary and a non-stationary factor, responsible for the common movement of commodity prices in short- and long-term perspective, respectively. The developed methodology is explained in Chapter 3, where the theoretical framework is exposed. Therefore, we refer to next Chapter for the explanations; here we provide an example applying this new technique to the model presented above.

2.4.1 Data presentation

The commodity prices included in the analysis are those presented in Chapter 1 and Appendix A. As mentioned, two exogenous variables are included in the model: the index of real economic activity developed in Kilian (2019) and the US real interest rate (calculated with the Fred of St. Louis economic data about the FED funds and CPI). These two series are publicly available and

[11]We tried different estimations with different combinations of prices (changing also the number of included series), different numerical optimisation methods for the computation of the likelihood, and different restrictions, imposing each element of matrix Π or of H equal to zero. Convergence was not achieved in any case. The encountered problems regard dimensionality, the number of parameters, non-stationarity and a not-concave likelihood.

are disposable at a monthly frequency, as the commodity price series. The time span is from January 1980 to December 2018. Unit root tests performed on commodity prices have been shown in Chapter 1, whereas both the Kilian index and the interest rate series are considered as stationary, as confirmed by the ADF, Phillips-Perron and KPSS tests.

For the example provided in this Chapter, we have selected 10 commodity prices,[12] including different categories (food, beverages, raw materials, energy and metals). In particular, the considered series are the log real prices of beef, coal, coffee, copper, maize, crude oil (brent), rubber, wheat, wool and zinc.

2.4.2 Results

The methodology presented in this work (Chapter 3) combines the decomposition of a system of time series as in Gonzalo and Granger (1995), taking into account the cointegration structure of data, and the Dynamic Factor Model estimation. Since the theoretical framework, the algorithm and the general empirical application without imposing any sort of theoretical constraint to data are presented in next Chapter, here we will only resume the principal steps of estimation and explain how we have modified the procedure to impose the exact structure of the theoretical model discussed above.

The proposed procedure consists in two main steps. In the first one, both a cointegration analysis and a Permanent-Transitory decomposition are carried out; the second step deals with proper factor extraction, exploiting the new hybrid techniques combining principal components and Kalman filter and smoother. Specifically, we specify a VECM (we remind that in this case the associated VAR has one lag as in Equation (2.15), and we include in the VECM the two exogenous variables of real economic activity and real interest rate), we perform a cointegration test as in Johansen (1991), and then we split the common trends of the system from the cointegration relations (Gonzalo and Granger, 1995), as explained in next Chapter. In this way we end up with the 10 commodity prices decomposed into a non-stationary part - represented by the common trends - and a stationary one, given by the cointegration relations. Since the trace test assesses a cointegration rank equal to 1, data are decomposed in 9 series corresponding to the common trends and 1 stationary series corresponding to the cointegration relation. The 9 common trends are differenced to achieve stationarity.

At this point, we apply a DFM estimation through the algorithm proposed by Doz et al. (2012) to the decomposed system[13]. We impose a precise structure

[12]Estimation through the Kalman filter would require a small system of variables. Therefore, this example also focuses on a small sub-sample of the whole set of commodity prices.

[13]Note that this estimation procedure is accomplished by subsequent steps, so that whereas in theory it should be possible to test whether a model including the factor variables is better than the same model omitting them, in this case things are more complicated, as the factor structure is imposed to a transformation of the observed series, and after the estimation of the VECM without factors.

to the DFM, and precisely that there is a stationary factor (extracted by the cointegration relation and the 9 differenced common trends of the system) and a non-stationary one that is the cumulation of the stationary one [14], and that factors enter in the state equation with one lag and in the observation equation (the price equation) with no lags, as specified by Equations (2.13) and (2.14).

Figure 2.1 plots the two extracted common factors (corresponding to Equations (2.13) and (2.14)), capturing the short-term co-movement and the long-term dynamics which persist in time.

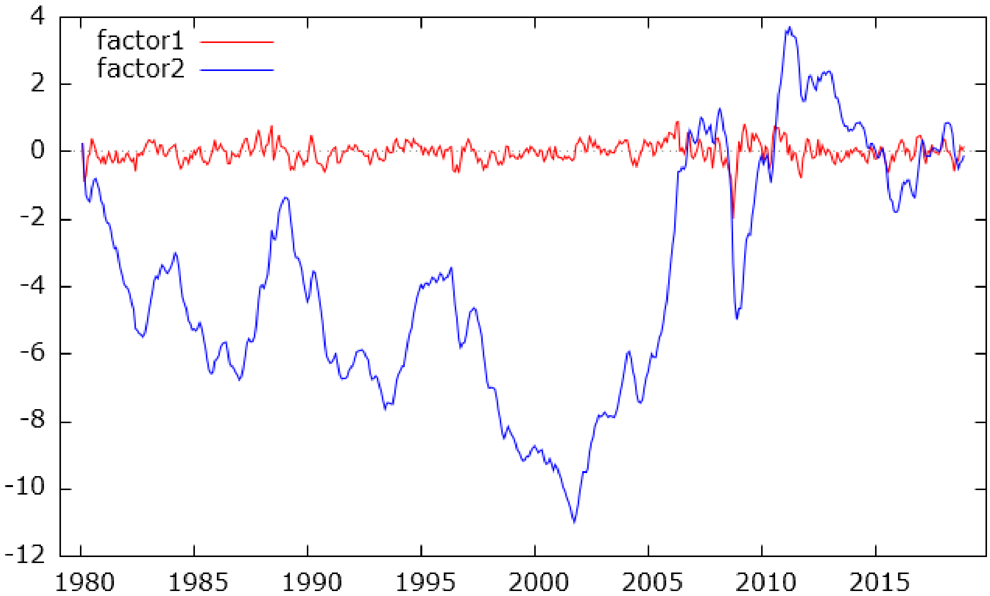


Figure 2.1: The two extracted factors: factor1 is $I(0)$ and factor2 is $I(1)$ and obtained by recumulation

Note that the non-stationary extracted factor share similar characteristics with the IFM all commodities index shown in Chapter 1, as reported in figure 2.2.[15]

The matrix of loadings B can be easily recovered, as shown in Chapter 3. Specifically, the B matrix is decomposed in two matrices, one containing the transitory effects - associated with the stationary factor - and the other the permanent one, containing the loading of the non-stationary factor to each price.[16] Table 2.1 report matrices $B_{transitory}$ and $B_{permanent}$, respectively.

It can be noted that the transitory common movement is marginal for all the series, whereas the non-stationary long-term common component has a greater impact. This is shown also in the following pictures (see Figure 2.4),

[14] seeChapter 3

[15] Obviously, other than the time span difference, the two series have some differences due to the way they are constructed, but also to different ways they are scaled and indexed.

[16] These are the λ_{i1} and λ_{i2} parameters.

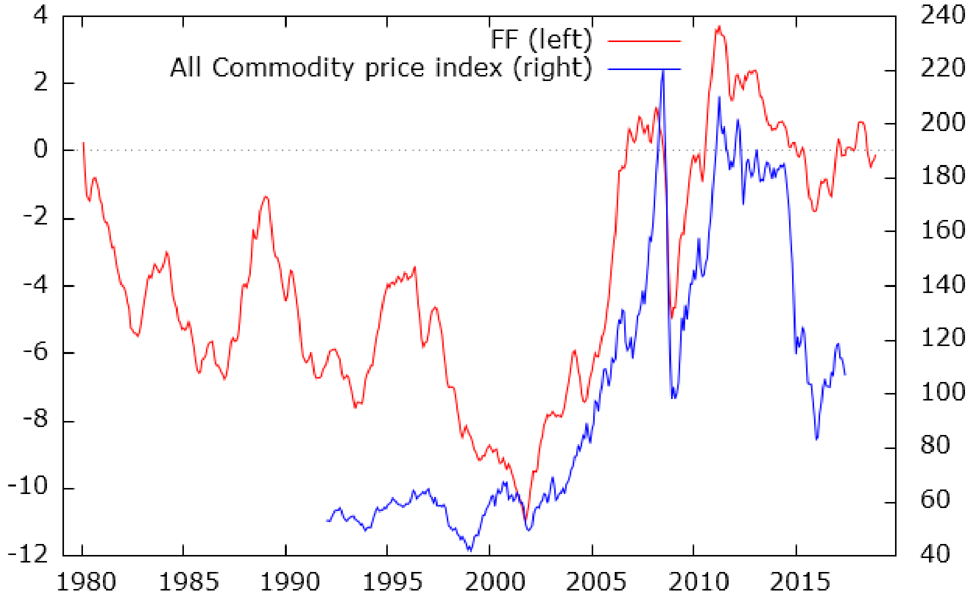


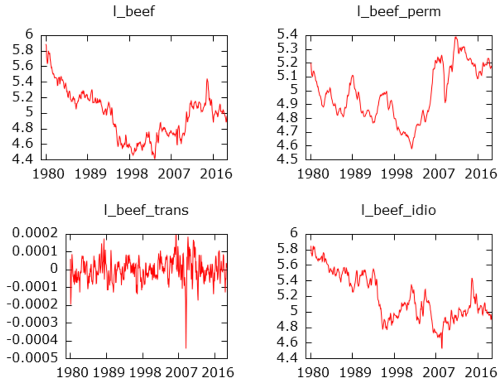
Figure 2.2: The extracted non-stationary factor and the IMF all commodities index

Table 2.1: Matrix of loadings B

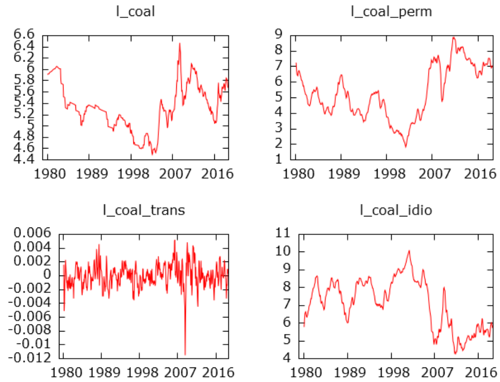
	$B_{transitory}$	$B_{permanent}$
Beef	0.000222940164024085	0.0552792364233032
Coal	0.00579127147332642	0.48120384956755
Coffee	0.00076790389176874	0.432634440231626
Copper	0.00287412357606405	1.01458625791353
Maize	-0.000517435508424343	0.359004216715108
Crude oil	0.00706464745346684	0.679159314932755
Rubber	0.00171054091253928	0.704171419087032
Wheat	0.000565139134804586	0.42754153703902
Wool	0.000748121795880983	0.43346409529183
Zinc	4.96654196449125e-005	0.499322636423137

in which for each price, we report its fitted value (decomposed in transitory and permanent components) plus the idiosyncratic part. Both the permanent and the idiosyncratic component appear dominant in the determination of prices, while the transitory component is negligible for all the ten commodity prices.

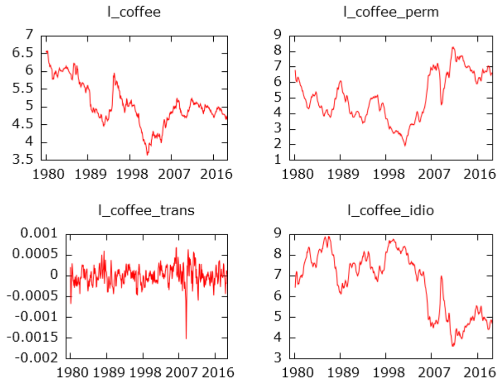
We can therefore conclude that the common movement interest commodity prices on a long-term perspective, and that joint fluctuations of more series are rather marginal. This confirm the hypothesis of macroeconomic factors, such



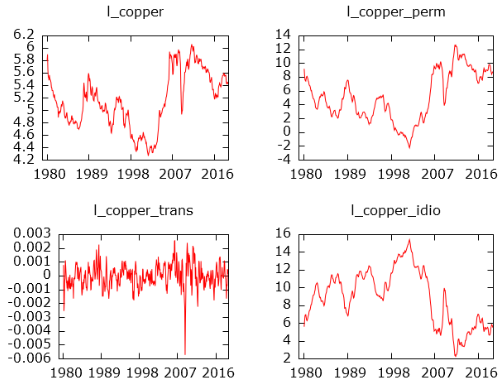
(a) Beef



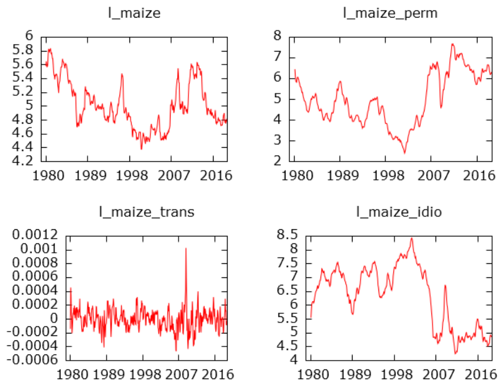
(b) Coal



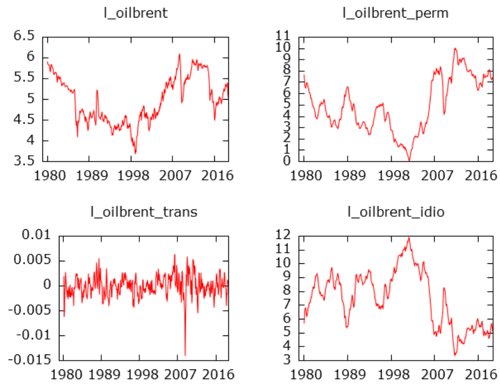
(c) Coffee



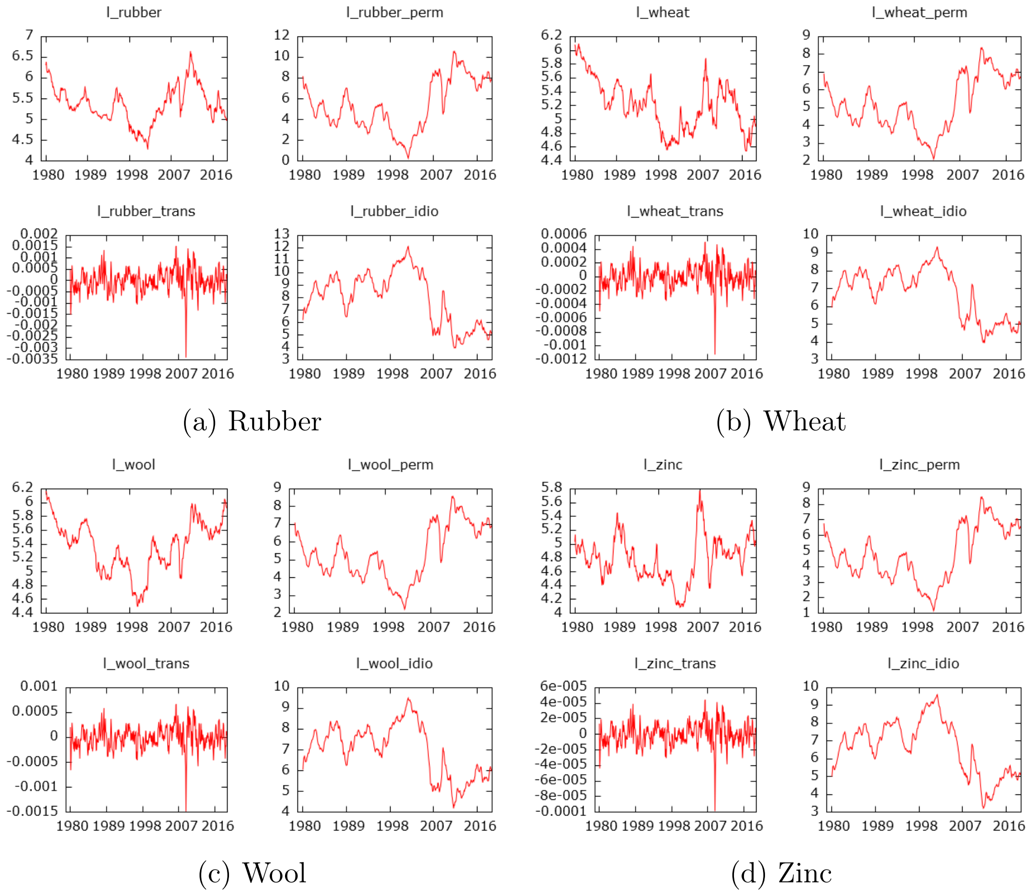
(d) Copper



(e) Maize



(f) Crude oil



(a) Rubber (b) Wheat

(c) Wool (d) Zinc

Figure 2.4: Commodity prices, fitted values (Permanent and Transitory components) and idiosyncratic terms

as the economic growth or other demand-driven causes, are what causes price to move together. The extracted long-run factor is decreasing until the mid-2000s and then it experiences an upsurge, thus suggesting that it is possible that the declining trend hypothesis holds until a change of regime, and that from then on, prices will stay at higher levels. However, it is also possible that this high peak is to be considered as an isolated episode and that prices will be declining again soon. Which of the two hypothesis is going to prevail is a matter of how the relation between the cost-saving technological improvement and pressure on limited resources will change.

For sure further research is needed to improve both the model and its empirical tractability, but in any case this starting point provides a good compromise between the necessity to theoretically derive a model for commodity prices co-movement and its estimation. Moreover, there is still the issue of imposing costraints deriving from theory which are also necessary to recover the structural parameters of the model. Of course this will need effort also in finding a suitable estimation strategy able to incorporate such restrictions.

Given that we stated the impossibility of estimating the model through the standard Kalman filtering techniques, we tried to look at the analysis of co-movement from the opposite perspective: starting from data rather than theoretical modelling. This has been the motivation upon which next Chapter is built.

2.5 Concluding remarks

In this Chapter we have laid the foundations for the development of a theoretical model for the short- and long-term common movement of many commodity prices. The model starts from three structural equations for consumption, production and storage of commodities, and then embodies two latent factors - responsible for the short-term movements and for the long-term dynamics, respectively - into price expectations formation. Two exogenous variables close the model: real economic activity, proxied by the index developed in Kilian (2009, 2019), and US real interest rate. These variables enters into the consumption and storage equations, respectively. From the market clearing condition it is possible to derive a reduced form price equation, which is what has to be estimated together with an equation specifying the law of motion of the two latent factors (the state equation). Estimation of the price and state equations require Kalman filtering techniques, but unfortunately this has proved impossible due to convergence problems, an issue that is well known in the literature. In particular, these convergence problems regard dimensionality, the number of parameters, non-stationarity and a not-concave likelihood. This failure has provided the idea of developing a new procedure for estimating non-stationary DFMs (and this is accomplished in Chapter 3), and by use of this methodology we have been able wo extract the two common factors, capturing short- and lon-run common movement of ten commodity prices.

The price series can be then decomposed in a Trend and a Cycle component, plus an idiosyncratic term, and we find that whereas the Trend and the Idiosyncratic components have a greater weight in explaining each commodity price, the Transitory component is rather marginal. This result implies that long-term drivers, such as global income and other demand-driven changes, are more suitable to explain common movements of different commodity prices, and that there is a substantial amount of price movements that is commodity-specific. The extracted non-stationary factor registers a downward pattern until an upsurge in the mid-2000s, but it is unclear from this time-span to understand whether the PSH or a resource scarcity pressure would prevail in a longer-term horizon.

Next studies should focus on imposing appropriate restrictions to the model for recovering the structural parameters, other than providing other estimation methods, if possible. The three structural equations model here developed should be seen as a starting point of how to model commodity prices with market fundamentals and latent variables to capture the common-movement.

Chapter 3

A cointegration-based Permanent-Transitory decomposition for non-stationary Dynamic Factor Models

3.1 Introduction

In this Chapter we present the main and more substantial contribution of this work: non-stationary Dynamic Factor Models (DFMs) are analysed, in order to propose a new procedure able to capture the co-movement of a large vector of time series split into Permanent and Transitory dynamics, other than taking into account the cointegration relations among variables. The model proposed in Chapter 2 allows estimation, when and if possible, only with a small number of commodity prices. Once assessed the failure in estimating the final reduced form price equation and the state equation by use of the standard techniques, we have started to think about an alternative procedure. With the proposed new methodology, we are able to exploit the whole dataset of 38 prices - presented in Chapter 1 - to estimate stationary and non-stationary factors responsible for common movement of commodity prices, but more importantly, we contribute to fill an important gap in the literature about non-stationary and cointegrated DFMs. The contribution is thus twofold: from the one hand, we present an alternative way for estimating a particular kind of DFMs; non-stationarity is not yet fully studied, and possibility of cointegration has been marginally explored. There are only few works allowing $I(1)$ and cointegrated factor models, and with this procedure it is also possible to disentangle the short-term dynamics from the long-term one. From the other hand, we can exploit the new procedure to answer the research question of this work and determine if there is co-movement among commodity prices, and more importantly we can estimate it.

In particular, the methodology adopted here allows to properly take into account the cointegration relations among variables and to split the common movement in a long term non-stationary part and a short-term stationary component. Whereas with the other existing procedures it is impossible to include some *a priori* economic features of the system, this alternative allows to explicitly take into account for the cointegration structure of the data. Specifically, we include a Permanent-Transitory decomposition of the vector of variables, as proposed by Gonzalo and Granger (1995), and impose a factor structure to the transformed series. At the end, it will be possible to recover the permanent and the transitory component for each variable.

DFMs are a powerful tool for summarising information of a larger number of variables of time series into a small number of factors. In particular, the vector of variables is split into a common component, summarising the joint movement of all the observables, and an idiosyncratic component, which is variable-specific. Although the general framework of DFMs, which assumes stationarity of both the common and the idiosyncratic component - and thus of the observed variables - is now standard within the literature, there are still few works taking into account the possibility of non-stationarity, and more precisely of cointegration. This work aims at contributing to this literature by combining a decomposition of the time series system *à la* Gonzalo and Granger (1995) with a DFM, with the intuition that the non-stationary part of the system can be estimated with the "differencing and re-cumulating" technique as suggested in Bai and Ng (2004). The former procedure consists in splitting a cointegrated system of variables into a stationary part, corresponding to the cointegration relations, and a non-stationary one, corresponding to the common trends; at the end, the system is decomposed in a Transitory and a Permanent component, respectively. The latter procedure instead implies differencing the whole system in order to achieve stationarity, performing the estimation of factors and parameters in differences and then recovering the true factors by simple integration. However, the main issue of this technique regards the possible loss of information if there exists some cointegration relationship which in this way is simply removed.

The intuition of the methodology proposed here consists in as a first step in which a cointegration analysis is carried out, possibly by dividing variables in blocks, and then the system is decomposed following Gonzalo and Granger (1995). After the decomposition, there are in facts r $I(0)$ series corresponding to the cointegration relationships - being r the rank of cointegration - and $n-r$ $I(1)$ series corresponding to the common trends of the system. Anyway, those $I(1)$ series are not cointegrated among themselves by definition. With this crucial intuition, it is possible to extract common factors *after* having decomposed the system and in particular, it is possible to difference the $n-r$ series corresponding to the common trends without loss of information. The DFM estimation is then straightforward. The $I(1)$ factors are then recovered by re-cumulating the $I(0)$ factors. The idiosyncratic components are allowed to be $I(1)$. At the end, the permanent and transitory components can be properly recovered. The Chapter furthermore provides the application of the proposed technique to the 38 commodity real prices of different markets, including energy, metal and agricultural ones. The common movement of different price series is split in short-term and long-term fluctuations. The series are divided in blocks for analysing cointegration, corresponding to different categories of commodities.

The rest of the Chapter is organised as follows: Section 3.2 introduces and summarises the existing literature on DFMs, with particular focus on non-stationary DFMs. Section 3.3 introduces and explains the novelty of this work, that is the P-T decomposed DFM proposed here. Section 3.4 provides an empirical analysis of co-movement of commodity prices using the methodology explained in Section 3.3, and Section 3.5 concludes.

3.2 Dynamic factor models

DFMs were introduced by Geweke (1977); Sargent and Sims (1977) and since then have been widely used in time series analysis, especially within the macroeconomic field. DFMs extract some latent factors from a higher-dimensional vector of time series variables, capturing the common dynamics of the system. In particular, the vector of observables is decomposed in two parts: the common component consists in a linear combination of the factors, and the idiosyncratic component refers to features which are specific to individual series. Since the latent factors are able to capture the common dynamics of the system, it is reasonable to exploit them as a tool for analysing co-movement of different variables. A DFM can be written in state space form as a system of two equations describing the evolution of the variables:

$$Y_t = \Lambda_0 F_t + \Lambda_1 F_{t-1} + ... + \Lambda_s F_{t-s} + \varepsilon_t \tag{3.1}$$

$$F_t = A_1 F_{t-1} + A_2 F_{t-2} + ... + A_p F_{t-p} + u_t, \tag{3.2}$$

with Y_t being the $n \times 1$ vector of observables, F_t the $q \times 1$ vector of unobserved common factors, ε_t is the idiosyncratic component and u_t is the vector of dynamic factor shocks. The idiosyncratic components are assumed to be uncorrelated with the factor shocks at all leads and lags. Matrices $\Lambda_1, ..., \Lambda_s$ and $A_1, ..., A_p$, of dimensions $n \times q$ and $q \times q$, contain the dynamic factor loadings and the factor autoregressive coefficients, respectively. Equation (3.1) is called *observation* or *measurement* equation and Equation (3.2) is known as *state* or *transition* equation. In the static case, $s = 0$ and the factors enter in the observation equation without lags.

Estimation of equations (3.1) and (3.2) requires to recover both the unobserved factors and the parameters, and can be performed exploiting different techniques; Stock and Watson (2011) provide an exhaustive review of the estimators in chronological order of appearance. Once factors have been estimated, it is possible to exploit them for several purposes, such as forecasting (see for instance Stock and Watson (2002); Boivin and Ng (2005)), construction of indicators of economic activity and other indices (Altissimo et al., 2001), estimation of DSGE models (Boivin and Giannoni, 2006), instrumental variable purposes (Kapetanios and Marcellino, 2010) and inclusion in further step factor-augmented vector autoregressions (FAVAR models) (Bernanke et al., 2005).

All DFM literature has been developed assuming stationarity of both processes in Equations (3.1) and (3.2). Next Sections will analyse DFM estimation techniques under the standard framework of stationarity and with some extensions to non-stationary case.

3.2.1 Stationary framework

Following Stock and Watson (2011), DFM estimation can be categorised into three generations. The first one encompasses *small* parametric models which can be estimated via Maximum Likelihood and the Kalman filter and smoother. The second generation overcomes the problem of dimensionality introducing non-parametric estimation trough principal components and other averaging methods, which are feasible for large n. The third generation of estimators combines the optimality of the first generation ones with the great advantage of handling a big number of time series typical of the second generation ones.

The first generation estimators - i.e. the Kalman filter - provide optimal estimates of the latent factors, under the model assumptions and parameters: assuming that the number of factors and the structure of the equations is known, the model can indeed be written in state space form and Kalman filtering and smoothing can extract the unobservable factors. Model parameters are estimated by Maximum Likelihood, although the maximisation entails a non linear optimisation which severely restricts the number of series which could be handled.[1] For this reason, first generation entails only low-dimensional models. Despite the many advantages of specifying and imposing the exact state-space formulation - even though in case of model misspecification, factors extracted with these algorithms can be non-robust - and the efficient estimates of the factors, handling this kind of techniques becomes unfeasible as n increases. Among the advantages of Kalman-filtering techniques, it is possible to incorporate restrictions deriving from the economic theory, since the structure of equations is specified *a priori*; then, they allow to deal with data irregularities such as missing observations or mixed frequencies (see for instance Bańbura and Modugno (2014)). Most importantly, Kalman filter and smoother allow to obtain more efficient factors with respect to principal components procedure, in the case of some specifications including non-stationarity (Poncela and Ruiz, 2016).

[1]For further details, see Harvey (1990); Engle and Watson (1981); Quah and Sargent (1993).

The second generation of DFMs, those for large n, raised in popularity starting from the 2000s, and since then, DFMs have become a popular tool for macroeconomic and time series analysis. This second generation deals with non parametric estimation exploiting cross-sectional averaging methods - usually principal components - and allowing for the inclusion of a great number of time series. In practice, F_t is treated as a parameter of q dimension, estimated using the Y_t vector of data with $n \times 1$ dimension. In this case no restrictions are imposed in the specification of the idiosyncratic noises and in the factors. Consistency of the space spanned by the factors has been proved (see for instance Forni et al. (2000) for the generalised dynamic factor model, Bai et al. (2008); Bai and Wang (2016) for a review of the existing literature. As for the determination of the number of factors, Bai and Ng (2002) have proposed some Information Criteria to determine q in a factor model with large n, and Bai and Ng (2007) have done the same for *dynamic* factors.

Finally, the third generation consists in hybrid methods which combine the advantages of the first and second generation estimators. In particular, Doz et al. (2011) propose a two step estimator that firstly estimates parameters of the model via OLS on PC and then in a second step updates the extracted factors with the Kalman smoother. In Doz et al. (2012), the estimator proposed in Doz et al. (2011) is iterated, so that the Kalman smoother is combined with the EM algorithm. In this way, the dimensionality burden of the Kalman filter and smoother is solved by use of principal components, but the sub-optimality of this methodology is compensated by the exploitation of first generation estimators.

3.2.2 Non-stationary framework

As mentioned, DFM literature has been developed assuming stationarity of both processes in Equations (3.1) and (3.2). However, it is well known that usually many macroeconomic time series are non-stationary and frequently also cointegrated. As proven in Chapter 1, this is also the case for the majority of commodity prices. For this reason, extensions for the non-stationary case, possibly accounting also for cointegration, are needed, even if literature here is more recent and still not complete. One of the most common practices is to difference the whole system and then work within the stationary framework. Nevertheless, when dealing with multivariate time series this should be made with care, because of possibility of failing to detect cointegration. Getting rid of non-stationarity by differencing each individual series implies throwing away important information and may lead to distort results. This is the reason why it is crucial to develop techniques accounting at the same time for non-stationarity and cointegrated factor models.

The connection between cointegration and common factors dates back to the common trend representation of Stock and Watson (1988), even if decompositions of that kind do not exactly coincide with factor extraction.

Non-stationarity has been included in the procedures based on Kalman filter and smoother, in a univariate case, with De Jong et al. (1991); Jong and Chu-Chun-Lin (1994), providing initialisation for non stationary Kalman filtering, such as a diffuse prior.[2] Moving to the multivariate case, Peña and Poncela (2006) have built a non-stationary factor model and estimated it via ML and the EM algorithm. For what concerns PC, first attempts to develop non-stationary DFMs are the methodologies of Bai and Ng (2004); Bai (2004), even if they are referring to panel data structure. The former relates to the so-called method of "differencing and recumulating", thus working with differenced series and then recovering the extracted differenced factors by integration, the latter proposes to perform a slightly modified PC procedure directly to data in levels. In particular, differencing variables rules out the problem of non-stationarity, assuring that the first difference of factors and the model parameters can be estimated consistently (to obtain the factors is it then sufficient to integrate the estimated $\Delta\hat{F}_t$). The great advantage of the methodology of Bai and Ng (2004) is that it can be used regardless stationarity of the idiosyncratic component, that means, all or part of the integrated portion of the system not necessarily is captured by the common component, but may be specific to individual series. Bai and Ng (2004) also provide proofs of the consistency of the PC estimates with "differencing and recumulating". On the contrary, the procedure suggested by Bai (2004) performs the estimation of non-stationary factors - and cointegration among factors is eventually allowed - directly to data in levels, which include $I(1)$ series. However, in this case the idiosyncratic term is allowed only to include $I(0)$ processes, which is equivalent to say that all the non-stationarity has to be captured by the common component. This is a hard assumption, since the case of non-stationary idiosyncratic component is found within several datasets. For this case, a methodology robust to both stationary and non-stationary idiosyncratic components, such as that of Bai and Ng (2004) could be safer. Of course "differencing and recumulating" implies getting rid of non-stationarity *a priori* and thus lead to possible misspecification if in facts there exists some cointegration relationship. Barigozzi et al. (2015, 2016) study non-stationary DFMs in which the idiosyncratic components are allowed to be either $I(0)$ or $I(1)$ and the factors are $I(1)$. Since the factors are cointegrated, they are modelled in ECM representation. Corona et al. (2017) point out that the estimators proposed by Bai and Ng (2004); Barigozzi et al. (2016) are asymptotically equivalent, but with some finite sample differences if deterministic trends are included in the model.

Extensions of hybrid procedures to the non-stationary framework have also been proposed: Corona et al. (2017) update the Doz et al. (2011) estimator to non-stationarity using first differenced data and then recumulating the factors and applying the PC directly in levels in the case of $I(0)$ idiosyncratic compo-

[2]The Kalman filter requires starting values for both the mean and the variance. The diffuse Kalman filter simply provides a very large starting value for the variance.

nents. They show that unless the idiosyncratic component is non-stationary, thus in case both the procedures can be used, extracting factors directly in $I(1)$ series in levels is better than differencing variables and then integrating factors. Also Barigozzi and Luciani (2017) extend the approaches of Doz et al. (2011, 2012) to non-stationarity by developing a Quasi ML estimator based on the EM algorithm combined with the Kalman filter and smoother estimators of the factors. They furthermore prove consistency and provide rates of convergence for both the factors and paramenters.

Finally, the work by Barigozzi and Luciani (2017) is also the only one considering a non-stationary large datatset DFM in which long-run co-movements can be disentangled from short-run co-movements. Their methodology consists in a first disentanglement of common component from the idiosyncratic noise (with the estimator described above) and a second step in which they split common trends from common cycles, applying a non-parametric Trend-Cycle decomposition to the extracted common factors. In particular, the second decomposition consists in identifying the common trends as the linear combinations of the factors and the common cycles as the deviations from long-run equilibria (the cointegration space).

3.3 A cointegration-based P-T decomposition for DFMs

The proposed methodology is similar to the one of Barigozzi and Luciani (2017), meaning that the Transitory part can be split from the Permanent one, but is in a certain way *opposite*, as this decomposition is done before proper factor extraction. The procedure consists in taking into account the cointegration and common trends structure of data and to incorporate it into the DFM. Within this framework, it is possible to split a permanent and a transitory component of the factor structure. The theoretical framework of the procedure is structured as follows.

Assume we have a vector of $I(1)$ variables as in Barigozzi and Luciani (2017), that is

$$Y_t = \mathcal{T}_t + \mathcal{C}_t + \xi_t,$$

in which the common component is the sum of $\mathcal{T}_t + \mathcal{C}_t$, where $\mathcal{T}_t$ represents the trend component and $\mathcal{C}_t$ the cycle component; both are assumed to have a factor structure, that is, to be driven by a small number of shocks q; the idiosyncratic component is given by ξ_t. Note that ξ_t may be $I(1)$.

Assume also that the cointegration rank is $0 \leq r < n$. Then, a Permanent-Transitory decomposition can be computed as follows: the Granger representation theorem ensures that a VECM representation exists

$$\Delta Y_t = \alpha\beta' Y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-i} + \epsilon_t; \tag{3.3}$$

assuming that p is finite, a decomposition as in Gonzalo and Granger (1995) implies the splitting of the vector of series in $Y_t = \mathcal{A}_1 m_t + \mathcal{A}_2 z_t$, in which $m_t = \alpha_\perp Y_t$ and $z_t = \beta' Y_t$. The matrix $\alpha_\perp$ is obtained from by setting $\alpha'_\perp \alpha = 0$ and has dimensions $n \times (n-r)$. Matrices $\mathcal{A}_1$ and $\mathcal{A}_2$ are obtained, respectively, with $\mathcal{A}_1 = \beta_\perp(\alpha'_\perp \beta_\perp)^{-1}$ and $\mathcal{A}_2 = \alpha(\beta'\alpha)^{-1}$. Two important properties that the two terms must satisfy are that the factors m_t must not be cointegrated and that z_t must not cause m_t on the long run.

Thus, for given values of the parameters, this decomposition[3] yields

$$G(L)Y_t = \left[\begin{array}{c} \beta' \\ \alpha'_\perp(1-L) \end{array} \right] Y_t = \left[\begin{array}{c} z_t \\ \Delta m_t \end{array} \right] = W_t \tag{3.4}$$

using the standard notation. Note that $\beta' Y_t$ corresponds to the cointegration relations of the system, whereas $\alpha'_\perp Y_t$ captures the common trends.[4]

By construction, $W_t \sim I(0)$, since the non-stationary part enters with lags. Note that we can define the inverse of the $G(L)$ filter as

$$G(L)^{-1} = \left[\begin{array}{cc} \alpha(\beta'\alpha)^{-1} & \beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\frac{1}{(1-L)} \end{array} \right]$$

where of course $\frac{1}{(1-L)}$ is the cumulation operator, and therefore write

$$Y_t = G(L)^{-1} W_t.$$

Now assume that W_t has an approximate factor structure as standard:

$$W_t = \Lambda^*(L) f_t^* + e_t$$

where $\Lambda^*(L)$ is a matrix polynomial of order s.[5] The equation above can also be written in the so-called "static form"

$$W_t = \Lambda f_t + e_t,$$

where $f_t = \left[f_t^* | f_{t-1}^* | \cdots | f_{t-s}^* \right]$ and the matrix Λ is partitioned accordingly.

By partitioning the loading matrix Λ appropriately,

$$\left[\begin{array}{c} z_t \\ \Delta m_t \end{array} \right] = \left[\begin{array}{c} \Lambda_z \\ \Lambda_\Delta \end{array} \right] f_t + e_t.$$

Putting the above together we have

$$Y_t = \left[\begin{array}{cc} \alpha(\beta'\alpha)^{-1} & \beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\frac{1}{(1-L)} \end{array} \right] \left[\left[\begin{array}{c} \Lambda_z \\ \Lambda_\Delta \end{array} \right] f_t + e_t \right] = \mathcal{T}_t + \mathcal{C}_t + \xi_t$$

[3] Clearly, $\alpha_\perp$ is not the only possibility here. For instance, the decomposition proposed by Kasa (1992) uses $\beta_\perp$ as an alternative. Importantly, using $\beta_\perp$ instead of $\alpha_\perp$ may lead to gain in consistency, given the fact that α is estimanted, whereas β is by constuction super-consistent.

[4] This latter components is lagged because of non-stationarity.

[5] Doz et al. (2012) examine the case of a static factor model, using $s = 0$ whereas here we extend the possibility of having a dynamic structure, so that our set up is more general.

where:

$$\begin{aligned}\mathcal{C}_t &= \alpha(\beta'\alpha)^{-1}\Lambda_z f_t \\ \mathcal{T}_t &= \beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\Lambda_\Delta f_t^c \\ \xi_t &= G(L)^{-1}e_t\end{aligned}$$

and f_t^c is the cumulation of f_t, that is, $f_t = \Delta f_t^c$; note that the idiosyncratic shocks ξ_t will be $I(1)$, as a rule. The $I(1)$ process f_t^c is a q-variate process whose first difference is the vector of $I(0)$ factors.

Note that the usual procedure, followed by practitioners, to estimate DFMs on differenced variables and then re-cumulating the estimated factors, is equivalent to choosing $r = 0$. This procedure, instead, makes it possible to decompose the common component in long- and short-term components in a very natural way.

In practice, a DFM model is applied to data that have been centred and standardised, so that the workflow goes as follows:

1. Estimate the matrix β on the original data Y_t (possibly, by blocks[6]) and compute the z_t series as $z_t = \beta' Y_t$;

2. Estimate α by OLS as
$$\Gamma(L)\Delta Y_t = \mu_t + \alpha z_{t-1} + \epsilon_t;$$

3. compute the GG-decomposed vector W_t as in equation (3.4);

4. compute the vector of standard deviations σ so that $Z_t = \langle\sigma\rangle^{-1}\left[W_t - \bar{W}\right]$, where the notation $\langle x\rangle$ indicates a diagonal matrix that has x on its diagonal; note that $\langle\sigma\rangle$ can be written as
$$\langle\sigma\rangle = \begin{bmatrix}\langle\sigma_z\rangle & 0 \\ 0 & \langle\sigma_\Delta\rangle\end{bmatrix};$$

5. compute the factors in the DFM
$$Z_t = \Lambda f_t + e_t$$
and partition the loading matrix Λ as
$$\Lambda = \begin{bmatrix}\Lambda_z \\ \Lambda_\Delta,\end{bmatrix}$$
where Λ_z has r rows and Λ_Δ has $(n-r)$;

[6] Cointegration is hardly detected in very large systems.

6. recover the Permament and Transitory components of the factor structure as

$$
\begin{aligned}
\mathcal{T}_t &= \beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\langle\sigma_\Delta\rangle\Lambda_\Delta f_t^c && (3.5)\\
\mathcal{C}_t &= \alpha(\beta'\alpha)^{-1}\langle\sigma_z\rangle\Lambda_z f_t && (3.6)
\end{aligned}
$$

   and the idiosyncratic component ξ_t as the difference $Y_t - \mathcal{T}_t - \mathcal{C}_t$.

Factors of the DFM can be computed in several way, as exposed in previous Sections. Here we will exploit the Doz et al. (2012) estimation procedure,[7] since it is the more complete among the alternative (it performs the principal component analysis to initialise the estimation, incorporates the efficiency of the Kalman smoother *and* iterates the procedure to refine the estimates.)

Considering the model written in static form, we have that

$$
\begin{gathered}
W_t = \Lambda f_t + e_t\\
f_t = A f_{t-1} + u_t,
\end{gathered}
$$

and the principal component qk factors are straightforwardly obtained by choosing and storing in a new matrix $\hat{\Lambda}_{PC}$ the eigenvectors associated to the largest qk eigenvalues of the correlation matrix of W_t. At this point, factors are recovered as $\hat{f}_{PC,t} = \hat{\Lambda}'_{PC} W_t$. Once $\hat{\Lambda}_{PC}$ and $\hat{f}_{PC,t}$ are available, the Doz et al. (2011) estimator performs an additional PCA to obtain the initial state of the q factors f_t, which is determined as $\hat{V}'\hat{f}_{PC,t}$, being $\hat{V}$ the matrix of the eigenvectors associated to the q largest eigenvalues of the covariance matrix of the residuals of the regression of $\hat{f}_{PC,t}$ on its lags. Subsequently, matrix of parameters A and the $n \times n$ and $q \times q$ covariance matrices of e_t and u_t are estimated by use of multivariate least squares.

Once all the previous steps are accomplished, the factor estimates $\hat{f}_{PC,t}$ are updated exploiting the Kalman smoothing[8] to produce $\hat{f}_{TS,t}$ given all the previously estimated parameters and the final estimate of the dynamic factors is given by the first sub-vector of $\hat{f}_{TS,t}$ composed by q elements. The procedure is then iterated many times to refine estimates. Doz et al. (2012) perform the iteration using a Quasi-Maximum Likelihood estimator, which is equivalent of refining using the EM algorithm.[9] The EM iterations are performed until a chosen criterion is met and convergence is achieved. At the end the final estimate of the factors $\hat{f}_{ML,t}$ is available.

---

[7] Extended to the dynamic set up.

[8] The initialisation is set at $\hat{f}_{PC,1}$.

[9] We will use the EM algorithm, as also Barigozzi and Luciani (2017) have previously done.

3.4 Empirical Analysis: co-movement of commodity prices

The methodology described above has been used to estimate a DFM with the aim of capturing the common movement of several commodity prices. In particular, by use of this procedure, it is possible to capture the short-run and the long-run co-movement of prices belonging to different categories, thus distinguishing the transitory common fluctuations which are more likely to affect the set of commodities only on a short-time horizon, such as supply and stock-driven unbalances, from the common dynamics which tend to persist and should reflect structural changes in the behaviour of markets (i.e. in the demand side). There is large empirical evidence of a tendency of many commodity prices to move together (see Chapter 1), even if according to economic theory, each price should simply reflect the corresponding demand-supply balance.

The dataset is the one presented in Chapter 1: the 38 real commodity prices, in logarithms, listed in Appendix A. Unit root tests (Chapter 1) assess the non-stationarity of the majority of the series, so that we can conclude that the vector of variables $Y_t \sim I(1)$.

3.4.1 Main results

The starting point has been to analyse cointegration relationships by blocks of commodities, with the assumption that each block should not be cointegrated with another one. Each block is specific for a certain kind of commodities, so that we end up with five blocks: for metals, energy, livestocks, other food commodities and agricultural raw materials, respectively. Analysis by blocks has been necessary, on the one hand, because of the dimensionality issue affecting tests of cointegration such as the one of Johansen (1991), but on the other hand it allows to incorporate some economic theory, assuming there should be no particular need for one kind of commodities belonging to a certain market to be cointegrated with one another of different kind. On the contrary, assuming there may exist cointegration relations among similar commodities makes sense from an economic point of view, because in each group commodity prices are grouped within similar good prices, which may be affected by the same drivers. The sub-group contegration analyses allows to take into consideration what in previous Chapters has been called the co-movement generated as interdependency. One intrinsic characteristic of the new proposed cointegration-based DFM is indeed that for very large DFM it becomes problematic; however, for medium-sized problems one could conceivably assume to split Y_t into blocks, and that cointegration only occurs within blocks. Commodity prices are in this context a valid example. Note that this concept of independent blocks with within cointegration structure but absence of between cointegration corresponds to the definition of complete separation in cointegrated systems of Granger and Haldrup (1996). First block comprehend metal commodity prices, specifically aluminium, copper, lead, nickel, tin, uranium, zinc, gold, silver and platinum prices. Energy block lists the three disposable crude oil prices (brent, WTI and Dubai Fateh) and coal. Beef, lamb, swine, poultry, salmon and shrimps prices are grouped in the livestock block. The group of raw materials is given by prices of cotton, hides, soft logs, hard logs, rubber, hard sawnwood, soft sawnwood and wool. Finally, the last block comprehends food prices, in particular barley, cocoa, coffee, rapeseed oil, maize, olive oil, rice, sunflower oil, tea and wheat.

After specifying a VECM for each block of commodity prices[10], the cointegration rank and block-specific matrix β has been estimated following Johansen (1991).[11] The final matrix of cointegration is block diagonal, obtained by recursively and diagonally adding each block-specific matrix. Matrix α is then estimated by OLS. Table 3.1 summarises the results obtained by rejecting the Johansen (1991) trace test at a threshold γ of 0.01 and 0.05, respectively.[12]

Table 3.1: Cointegration analysis by blocks

		$\gamma = 0.01$		$\gamma = 0.05$	
	VAR length	Cointegration rank	Common trends	Cointegration rank	Common trends
Metals	2	2	8	3	7
Energy	2	2	2	4	0
Livestock	2	2	4	2	4
Raw materials	2	1	7	1	7
Food	2	3	7	9	1

As it is shown, the results change significantly, as the total cointegration rank is $r = 10$ in the case of $\gamma = 0.01$ (with an association of $n-r = 28$ common trends) and $r = 19$ with $\gamma = 0.05$. For what concerns the food block, the two situations are opposite, as in one case 7 common trends are detected within the group, and only 1 in the other. Energy group switch from a situation of no cointegration - which mean the system is stationary - to a cointegrated block with 2 common trends. This suggests that the choice of the cointegration rank has to be made with great care.

Here we will focus on the analysis performed by choosing the situation of $\gamma = 0.01$. This implies we end up with z_t of dimensions 10×1 and Δm_t of dimensions 28×1 (see Equation (3.4)). At this point a proper factor extraction is performed, by use of the Doz et al. (2012) algorithm.

The number of factors, which selection is as always in the hands of the practitioner, can be determined following Information Criteria of Bai and Ng (2007). In this case, the determined number of dynamic factors q is 3. Figures 3.1 and 3.2 report the three extracted factors and their respective cumulation (these correspond to f_t and f_t^c, respectvely). These series express the joint

[10]The Hannan-Quinn criterion has been used for determining the proper lag specification.

[11]Note that it may be possible to extend this methodology by using other cointegration tests.

[12]Deterministic component: unrestricted constant.

common movement of the 38 commodity prices, and specifically capture the $I(0)$ short-run common dynamics and the $I(1)$ long-run co-movement. Note that each of the three extracted non-stationary factors tells a different story about the general tendencies of commodity prices over a long-term horizon. The first factor captures a substantial upsurge of prices starting from the 1990s; the second factor has not a clear direction, but highlights a great positive peak and the consequent fall until the rapid increase of next decade; the third one is slightly decreasing. Clearly, it is impossible to assess if commodity prices will enter in a scarcity era or are still in a declining-trend path, by looking at the results of the common movement analysis.

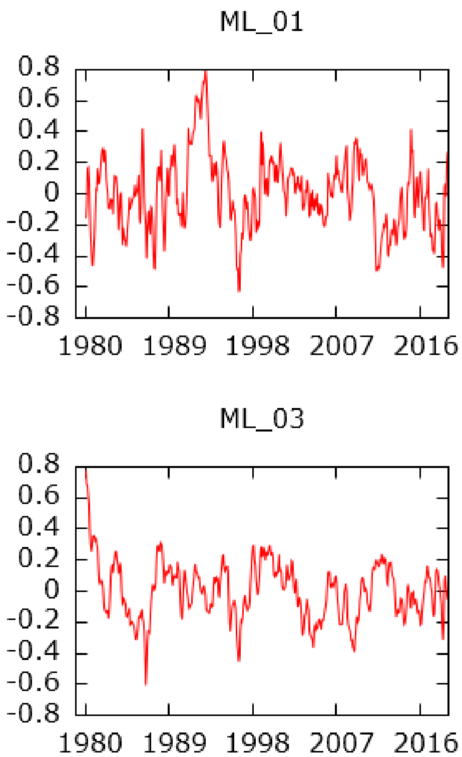


Figure 3.1: Extracted factors, Doz et al. (2012) estimator

Now both matrices $\alpha(\beta'\alpha)^{-1}\Lambda_z$ and $\beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\frac{1}{1-L}\Lambda_\Delta$ are recovered; these represent the loadings of the common component on each series.[13] Note that, whereas matrices $\alpha(\beta'\alpha)^{-1}$ and $\beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\frac{1}{1-L}$ capture the effects of the sub-group common movement, matrices Λ_z and Λ_Δ contain the effects of the general co-movement driven by the three external latent factors. Clearly, $\alpha(\beta'\alpha)^{-1}\Lambda_z$ and $\beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\frac{1}{1-L}\Lambda_\Delta$, which constitute the final Permanent and Transitory weights, are a combination of both. Finally, each series can be split in both $\mathcal{T}_t$ and $\mathcal{C}_t$, plus the idiosyncratic noise ξ_t. This allows to see which - among

[13]See Appendix C for the loading matrices.

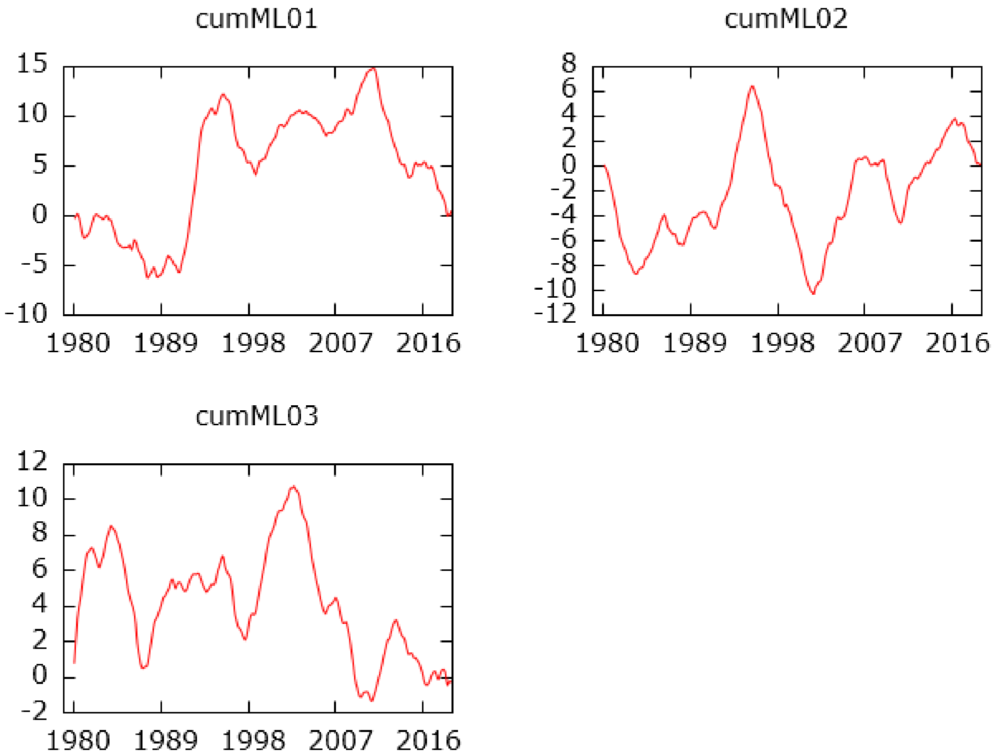


Figure 3.2: Cumulation of extracted factors, Doz et al. (2012) estimator

Table 3.2: Factors VAR parameters, Doz et al. (2012) estimator, 3 factors and two lags

	F_{t-1}			F_{t-2}		
factor 1	1.050	-0.030	0.178	-0.169	0.027	-0.182
factor 2	0.068	1.025	0.250	-0.045	-0.120	-0.270
factor 3	-0.075	0.051	1.239	0.043	-0.075	-0.327

the three determinants - is the most significant in determining the dynamics of the considered price, and to see how the different Permanent components (the long-run trends) differ from some series to others.

Commodity prices of the food blocks are well filtered by the P-T decomposition with dynamic factors; Figures 3.3 and 3.5 shows some examples, specifically for wheat and cocoa prices. The mentioned Figures picture the original series and its decomposition in the common component (Permanent and Transitory) and the idiosyncratic part. Figure 3.4 plots the original series of log real wheat price against its long-term common component, which is obtained by combining common latent factors and food common trends. Figures 3.5, 3.6, 3.7, 3.8, 3.9, 3.10, 3.11 and 3.12 show some other examples. Good results are obtained also within the raw material prices and livestocks, and quite good on energy group. The only block which results not well-filtered is metals one, with the exceptions of lead and nickel prices. This may be due to the absence of proper trends common to more commodity prices, given that the block groups heterogeneous metals, from precious to non-precious ones. However, the same analysis performed by excluding the precious commodities (platinum, gold and silver) does not improve the results, and further investigation is of course needed to understand the causes.

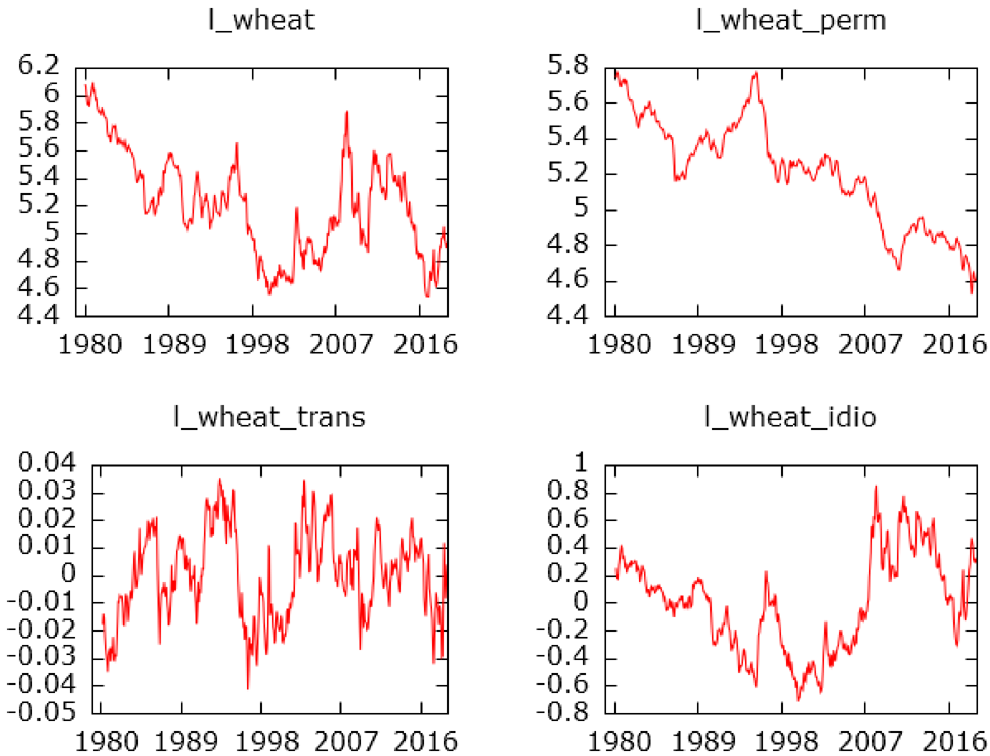


Figure 3.3: P-T decomposition with dynamic factors, wheat log real price

As it can be seen from the reported Figures (and from the loading matrices reported in Appendix C) the Transitory component is not much relevant to explain the commodity prices co-movement, as it accounts always for a very small part of the original series. This is confirmed for *all* the 38 series. The Permanent component is much more significant to explain the original log real price, and the idiosyncratic component is always relevant, meaning that there is a large part of prices composition which is commodity-specific. If the Permanent component is more relevant than the Transitory component, than it means that the causes of the common movement of commodity prices is driven by long-run factors, more than short-term ones. These include demand-driven

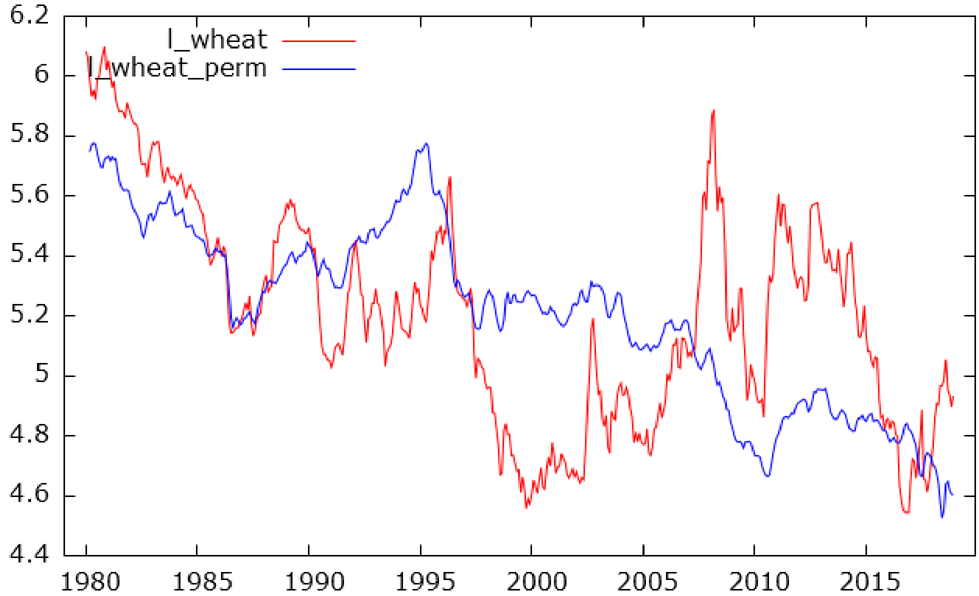


Figure 3.4: Wheat log real price and its Permanent component

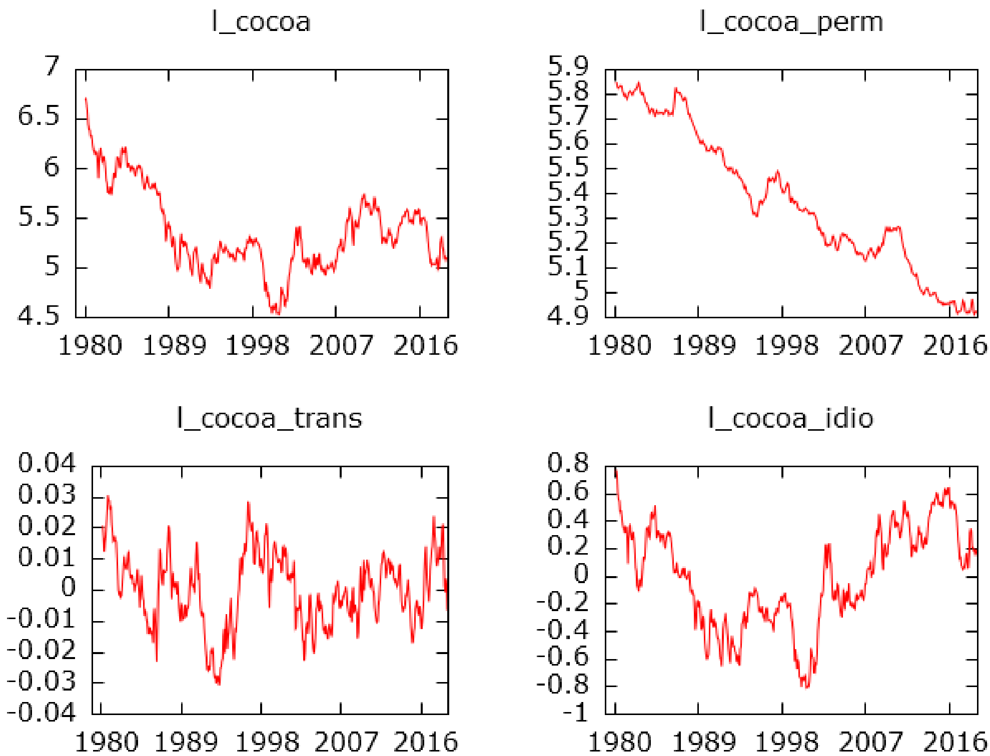


Figure 3.5: P-T decomposition with dynamic factors, cocoa log real price

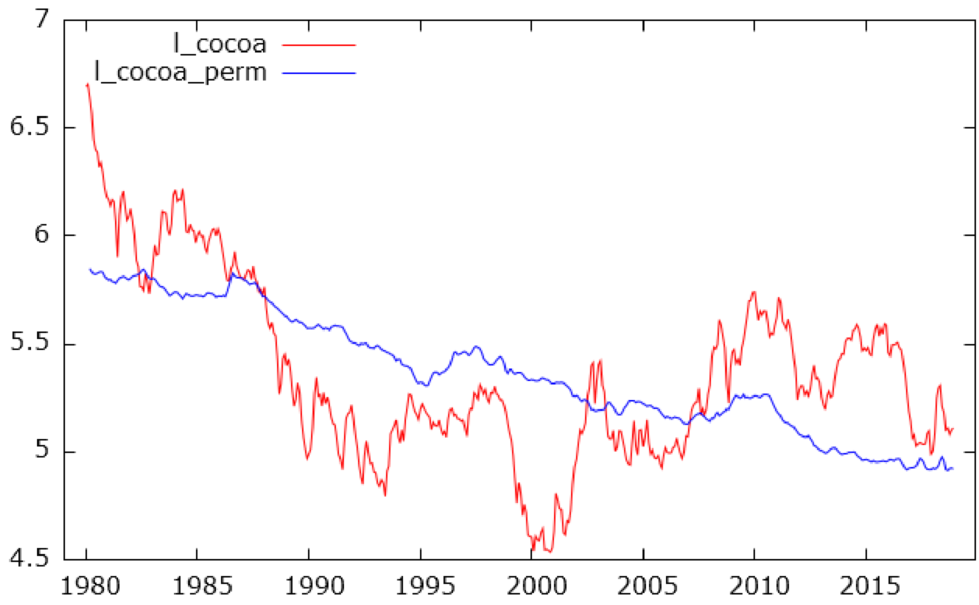


Figure 3.6: Cocoa log real price and its Permanent component

explanations (see Chapter 1) as demographic and economic growth and taste shifts, and some supply-oriented long-term factors as the level of investment.

3.4.2 Other extensions

The empirical analysis has been repeated with the inclusion of the two exogenous variables presented in Chapter 2, specifically the Kilian (2019) index and the US real interest rate. The inclusion of exogenous variables enters within the VECM estimation phase. The results of the new cointegration and common trends analysis is summarised in Table 3.3.

Table 3.3: Cointegration analysis by blocks, with two exogenous variables

		$\gamma = 0.01$		$\gamma = 0.05$	
	VAR length	Cointegration rank	Common trends	Cointegration rank	Common trends
Metals	1	1	9	2	8
Energy	2	4	0	4	0
Livestock	2	3	3	3	3
Raw materials	2	1	7	1	7
Food	2	5	5	8	2

Focusing, as it has been done before, on the case of $\gamma = 0.01$, we end up with less common trends and a greater number of cointegration relations. With the Kilian index and the real interest rate included, the number of dynamic factors detected by use of the IC of Bai and Ng (2007) is equal to 2. Figure 3.13 plots them and their cumulated counterparts.

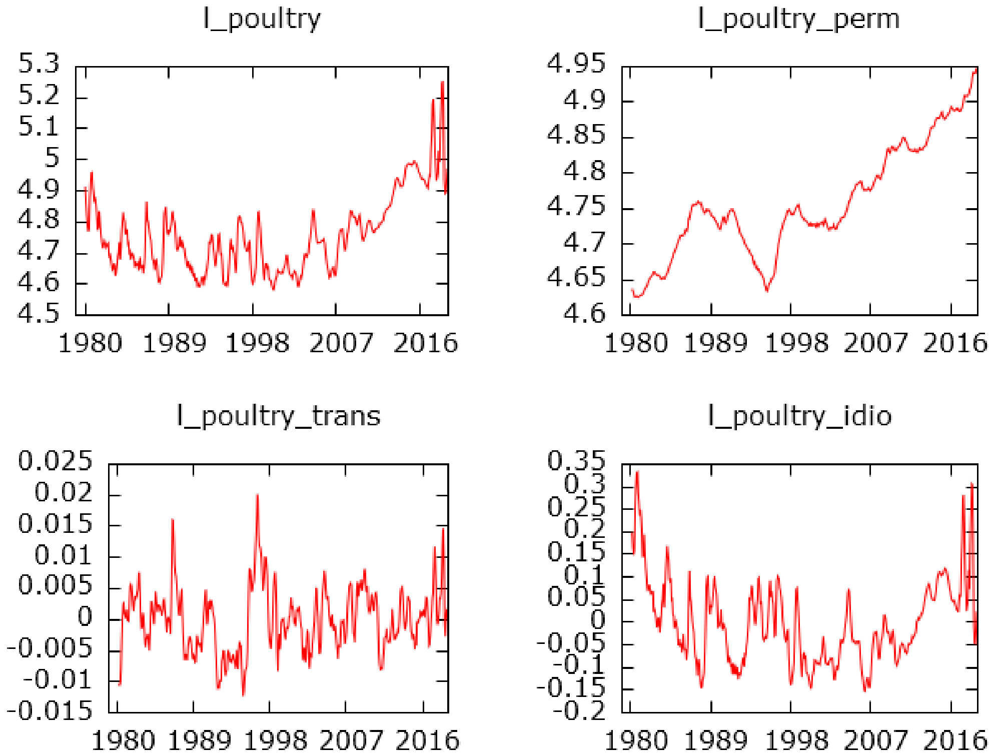


Figure 3.7: P-T decomposition with dynamic factors, poultry log real price

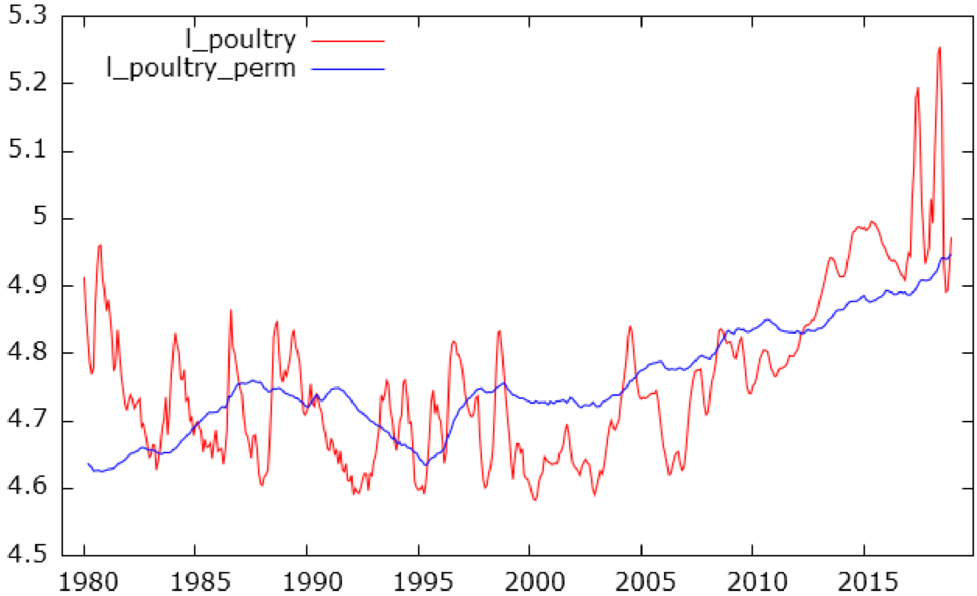


Figure 3.8: poultry log real price and its Permanent component

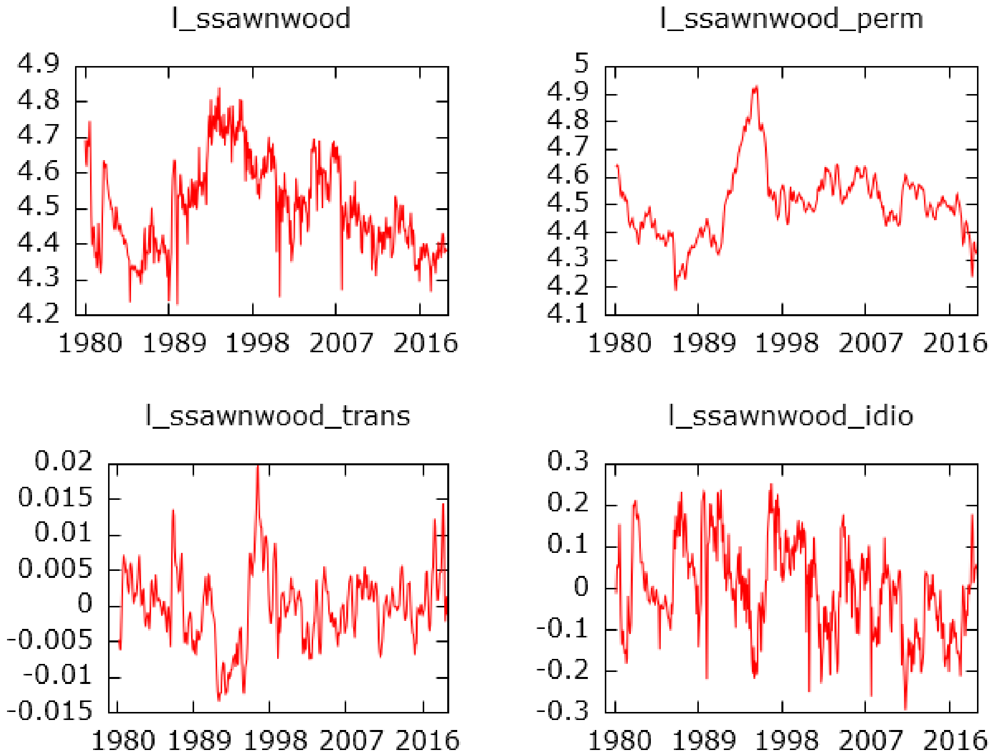


Figure 3.9: P-T decomposition with dynamic factors, soft sawnwood log real price

Even if some very marginal improvements occur for the blocks of food, livestock and raw materials, in this case the Permanent component of the energy commodities is totally absent, given that the performed cointegration tests consider the subsystem as stationary. The situation for the metals group does not change nor improve. Here we report Figures 3.14 and 3.15 as an example for the wheat price, in order to compare the situation with the decomposed series without the inclusion of the exogenous variables. The Permanent component results to be more smoothed with the inclusion of the two exogenous variables and a clear declining trend is found; the idiosyncratic component seem to be less pervasive; the Transitory component, instead, is quite irrelevant in both cases.

Finally, a comparison of the current application has been made with the case of a standard DFM estimated as in Bai and Ng (2004). Note that this case correspond to the very same analysis with the assumption of a cointegration rank equal to 0, thus assuming that there are no cointegration relations, and the whole system is differenced to proceed to factor extraction. In this case of course it is not possible to decompose the series and to split the common component in Permanent and Transitory parts, so the only comparison can be performed with the extracted factors. Getting rid completely of cointegration

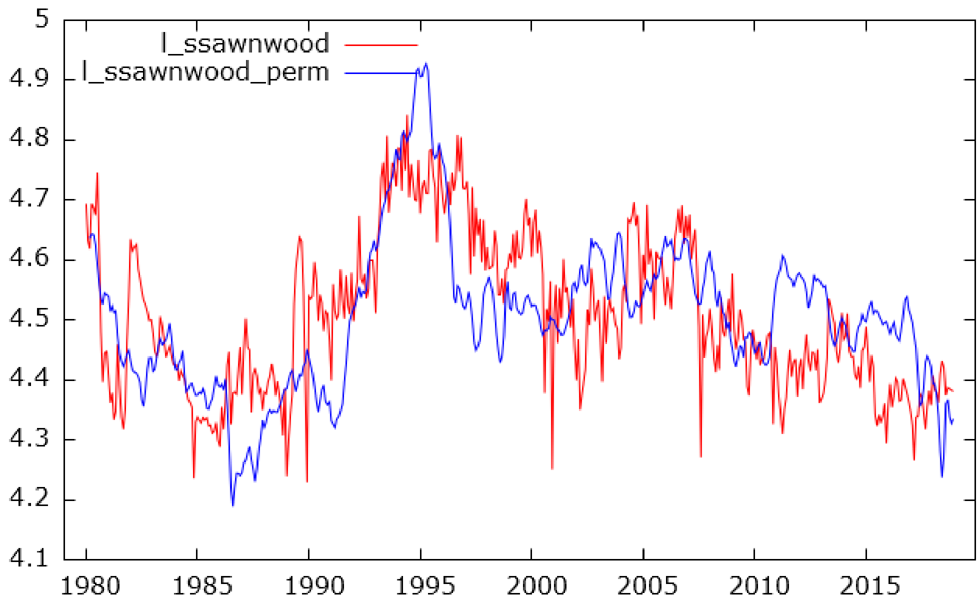


Figure 3.10: Soft sawnwood log real price and its Permanent component

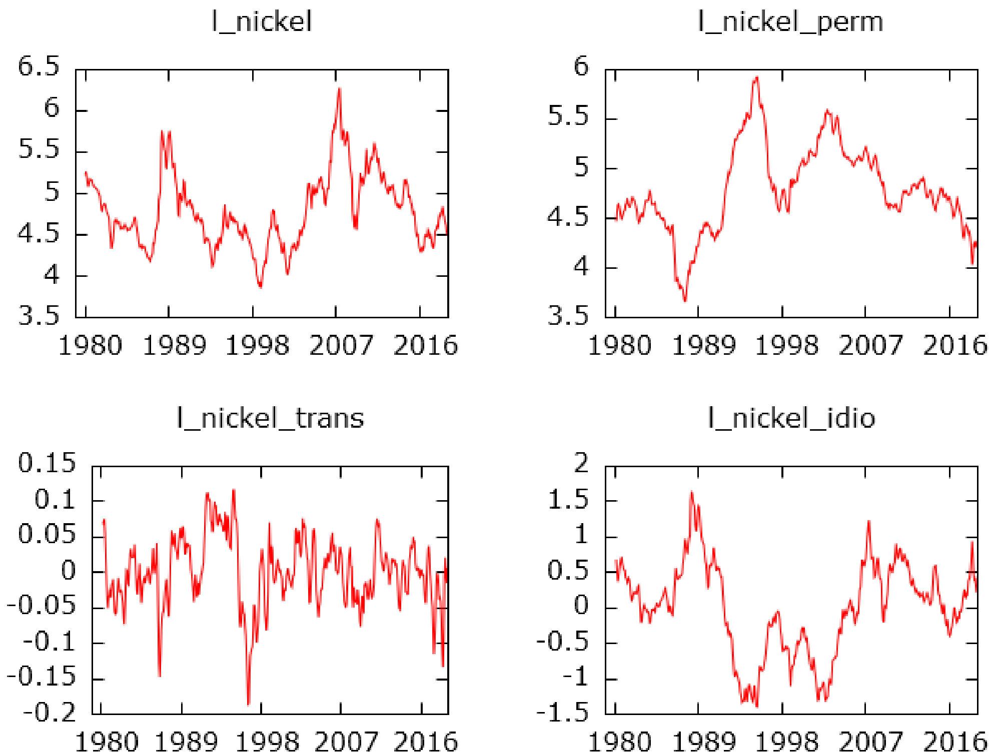


Figure 3.11: P-T decomposition with dynamic factors, nickel log real price

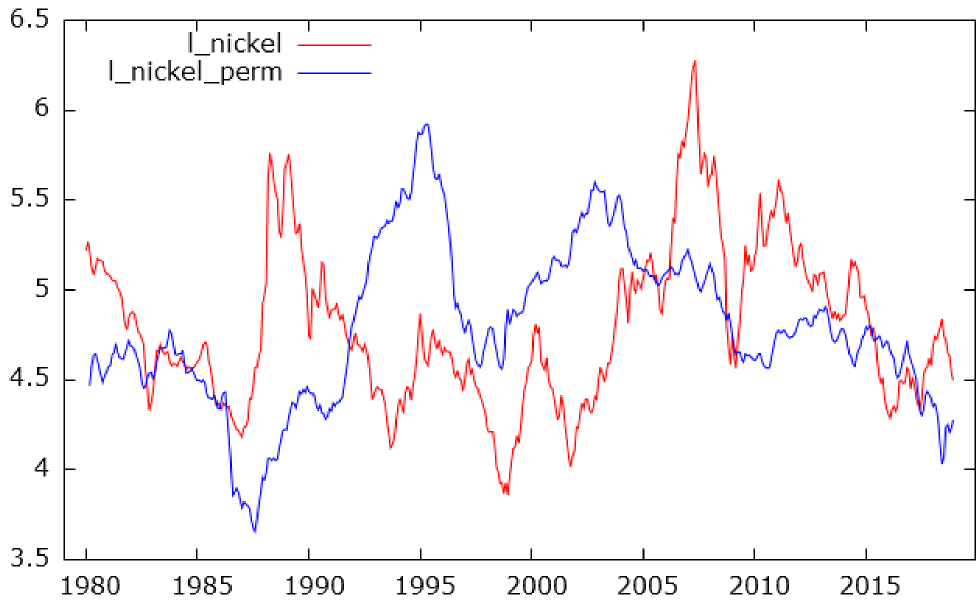


Figure 3.12: Nickel log real price and its Permanent component

lead to a proper determination of number of dynamic factors equal to 1, using the aforementioned Information Criteria. However, we extracted three common factors as in the proposed P-T based DFM, in order to compare them. Figure shows the three factors, estimated with the Doz et al. (2012) algorithm as before, and their cumulation. Table 3.4 reports the VAR parameters of the extracted factors. The P-T decomposed DFM here proposed allows to obtain smoother estimates of the factors, whereas those extracted with the simple "differencing and recumulating" technique exhibit more noise and show less persistence. However, the majority of the common movement of the 38 series is captured by the first factor, whereas the contribution of the other two is marginal. By looking at this factor, it is clear that the overall joint dynamics of the 38 commodity prices are captured by a series exhibiting a slightly decreasing trend till the beginning of 2000s and then an upsurge since then. It is interesting to note that there are some slight similarities among the three non-stationary factors extracted in this context and with the proposed methodology: the second one does not exhibit a clear direction in both cases, whereas the third is decreasing. However, the three factors extracted with the proposed methodology are smoother, whereas those of the case of $r = 0$ have more noise. The factors obtained with the proposed cointegration-based DFM are nevertheless of more difficult interpretation since they are extracted from common trends and cointegration relations and not directly from prices (besides, they are *three* and not only one, and are very different from one another, as one captures an upsurge since the 1990s, one is stable but with great variability and the other is slightly declining.)

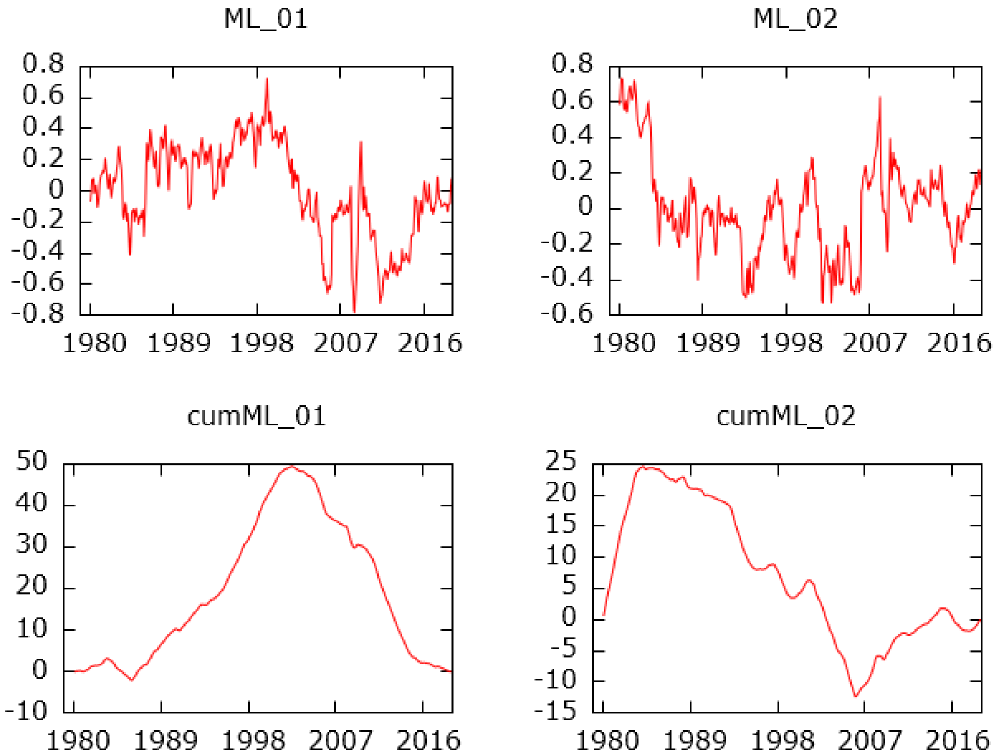


Figure 3.13: Extracted factors and their cumulation, analysis with exogenous variables.

The three factors extracted with the new procedure are able to synthesise the joint movements of commodity prices, and could be included and exploited for further analyses, as mentioned in Section 3.2. Importantly, we are able to split the Permanent from the Transitory common component, and we have stated that the short-run common movement is rather marginal in contributing to the formation of commodity prices, whereas the Permanent and idiosyncratic components have more weight.

Table 3.4: Factors VAR parameters, $r = 0$, Doz et al. (2012) estimator in first differenced variables, 3 factors and two lags

	F_{t-1}			F_{t-2}		
factor 1	0.708	-0.318	0.073	-0.057	-0.003	-0.022
factor 2	-0.140	0.256	-0.218	0.145	-0.149	0.183
factor 3	-0.028	0.064	0.287	0.121	-0.049	-0.089

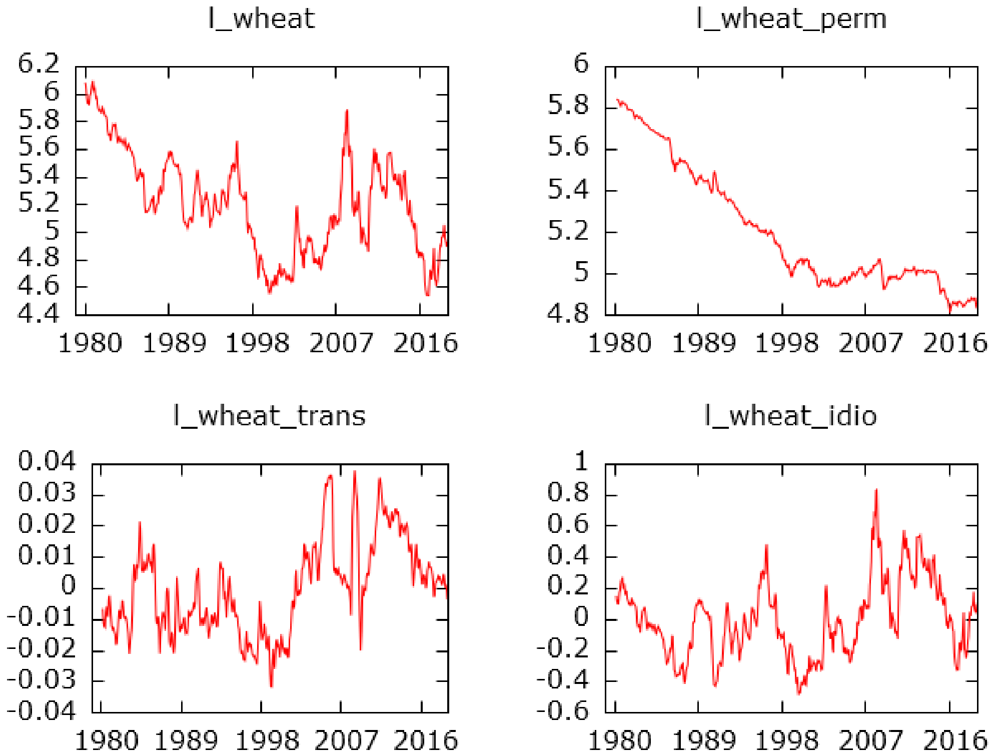


Figure 3.14: P-T decomposition with dynamic factors, wheat log real price, analysis with exogenous variables

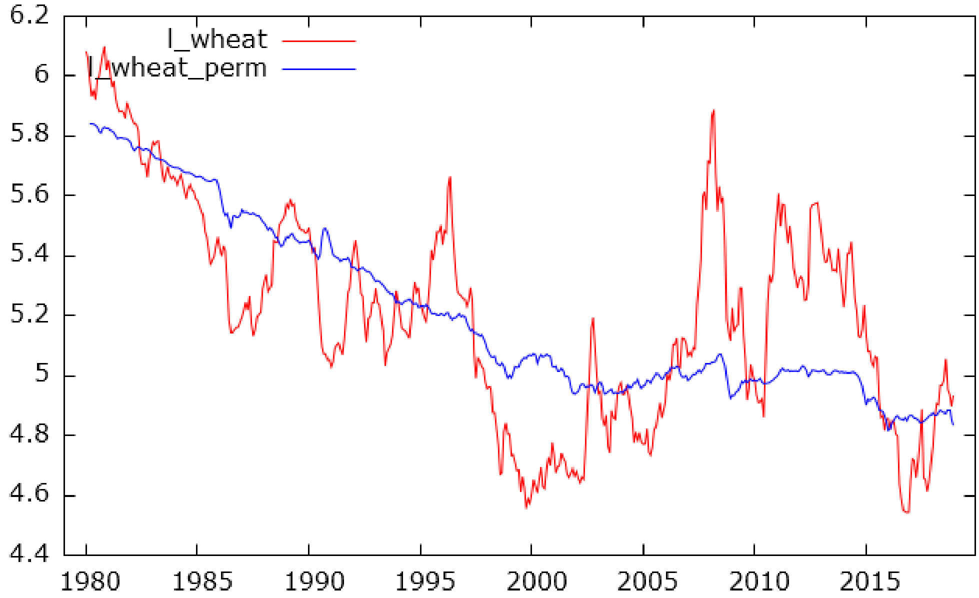


Figure 3.15: Wheat log real price and its Permanent component, analysis with exogenous variables

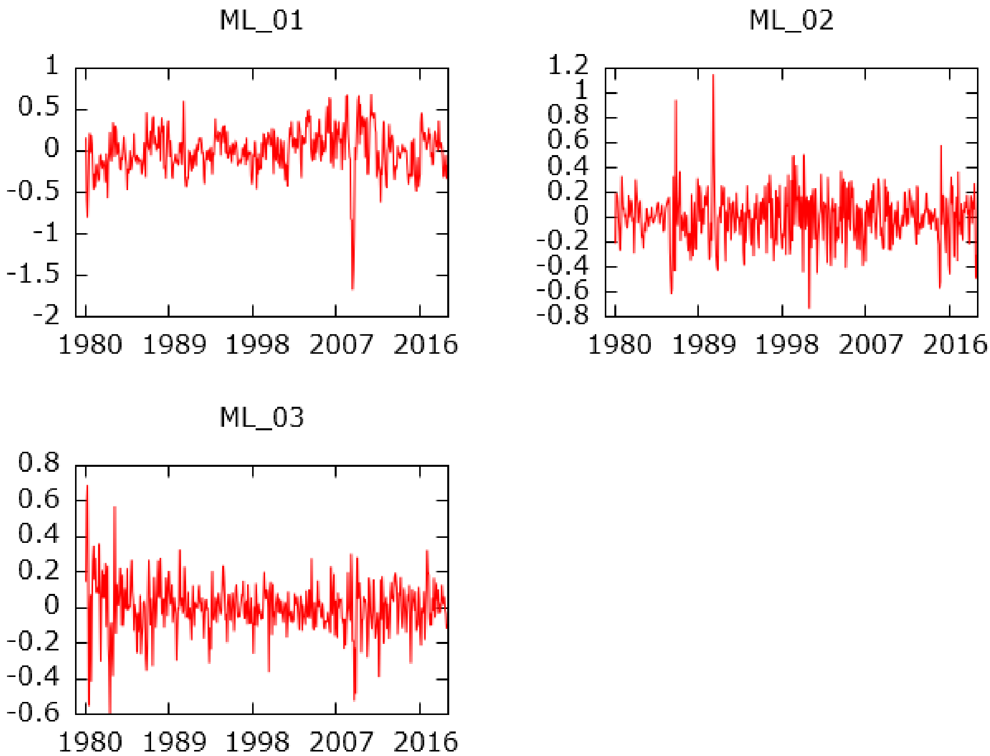


Figure 3.16: Latent factors, $r = 0$, Doz et al. (2012) estimator

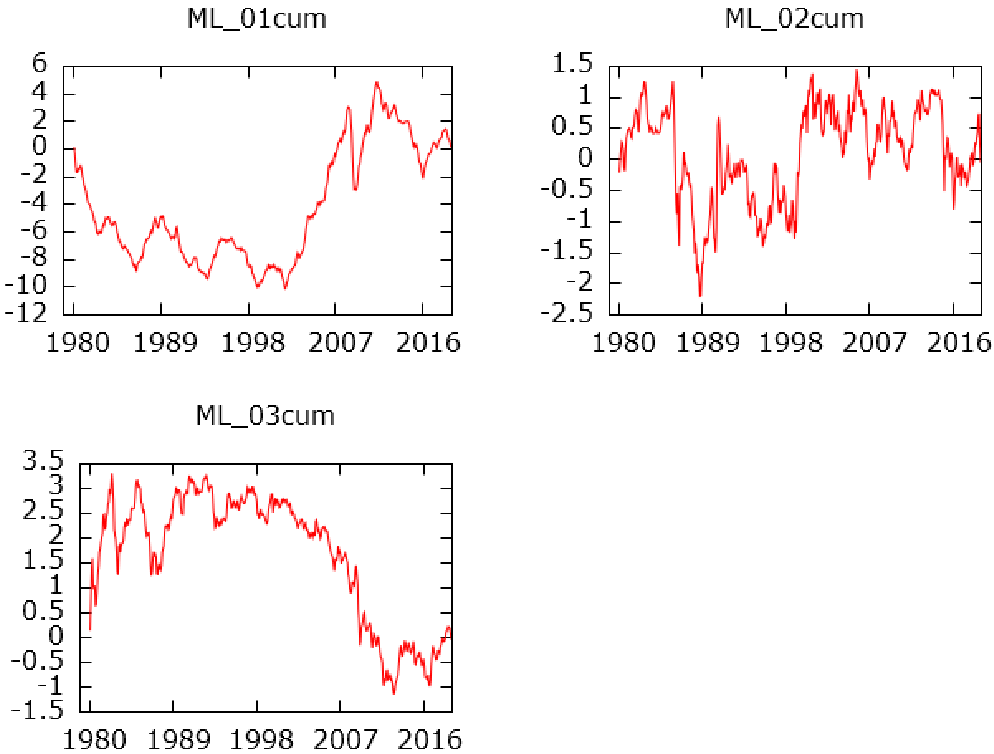


Figure 3.17: Cumulation of latent factors, $r = 0$, Doz et al. (2012) estimator

3.5 Concluding remarks

This Chapter proposes a new methodology to estimate a Dynamic Factor Model by taking in consideration the cointegration structure of data, and allowing to decompose the system in a long-term Permanent component, a short-term Transitory component and a idiosyncratic component that is specific to each individual series. This methodology consists in a first step of cointegration analysis, a decomposition as in Gonzalo and Granger (1995) in common trends and cointegration relations and a consequent factor extraction. The latent factors capture the common component of the system, here split in short-run and long-run dynamics. Variables can be finally decomposed using the loading matrices.

The proposed procedure is applied to 38 real commodity prices belonging to different categories, in order to analyse their common movement by taking into consideration the cointegration structure of data. Commodity prices are split in five blocks - for metals, energy, livestock, raw materials and food prices - and Johansen (1991) cointegration tests performed to each block conclude that the system has 10 cointegration relations and 28 common trends. At this point, a DFM is estimated; there are several alternatives for the estimation of a large n DFM; we exploit the Doz et al. (2012) proposal because it combines the efficiency of the Kalman filter and smoother with the many computational advantages of principal components, plus the iteration of the procedure for refinements.

Three common factors are able to summarise the common movement of the 38 commodity prices, and specifically, the Permanent common component is much more important in explaining the price dynamics. On the other hand, the Transitory component seems not so crucial in explaining the commodity prices co-movement. It is also important to stress that the idiosyncratic component is not marginal, meaning there is an important share of price series which is commodity-specific. The decomposition proposed here leads to good results for all the series with the exception of metal commodity prices; further investigation is needed to deepen the matter and to improve the results.

Extensions of the current study should also compare the presented methodology with existing ones, such other Trend-Cycle decompositions, or check if results are robust to other cointegration tests.

Conclusions

This work examines the common movement of commodity prices splitting the short-term variations from the long-run common fluctuations. Commodity prices co-movement is an important aspect monitored by researchers because of its implications in price formation and future movements. According to economic theory, the price of each commodity should reflect the related market balance. If many commodity prices share the same dynamics, then it means that there are *other* important drivers which are responsible for price movements. Whereas part of this common movement can be explained by price interdependencies (such as complementarity or substitutability of a set of goods), there is still part of co-movement which interest very different markets which have no reason to be thought as interrelated. Detecting this common movement is not trivial as it opens several challenges both from a theoretical and from an empirical perspectives.

Here we tried to fill part of this gap by proposing an analysis of co-movement developed in three Chapters. Chapter 1 reviews other studies on commodity prices general dynamics and provide the first univariate results, by exploiting a set of 38 commodity spot monthly prices available from the IMF primary commodity database. Chapter 2 proposes a first attempt of modelling commodity markets by including latent factors responsible for co-movement. The model consists in three structural equations determining consumption, production and storage on a multi-commodity framework, plus a market clearing condition which allows to find the equilibrium price. Latent factors are introduced through stockholders, which act as the insiders of the markets and form price expectations by looking at the factors which are unobservable for the practitioner. Chapter 3, which is the main core of this work, contributes both to propose a new estimation procedure for non-stationary and cointegrated Dynamic Factor Models and to exploit this methodology to empirically assess the co-movement of the 38 considered commodity prices.

After having confirmed the non-stationarity hypothesis for the majority of the series and having built the model, we have extracted latent factors from the dataset, using different examples and specifications. We have been able to detect the short- and long-term prices common movement, but we have found that long-term one is more important in determining price movements, whereas common short-term fluctuations are marginal. This suggests that long-term common drivers are more crucial, such as the increasing global demand for commodities with respect to supply. From the extracted factors, no one clear global trend emerges, thus it is difficult to understand if there is a shift from a declining to an increasing trend era. More specifically, some series, especially belonging to food category, exhibit more evidently a declining trend followed by a change of direction starting from the mid-2000s. This suggests that for some commodities the worry of an increasing demand/supply pressure is indeed relevant and in act (for instance, meat prices, for which the most plausible explanation seems the increasing global demand).

Further research is needed to fully explore the topic, which offers several possibilities and is far from being completely examined. In particular, future analyses should focus on the improvement of the presented model, which only wants to represent a starting point of modelling latent factors with market fundamentals, on a way to derive and impose parameter restrictions and on other estimation methods development.

Bibliography

Ai, C., Chatrath, A., and Song, F. (2006). On the comovement of commodity prices. *American Journal of Agricultural Economics*, 88(3):574–588.

Akram, Q. F. (2009). Commodity prices, interest rates and the dollar. *Energy economics*, 31(6):838–851.

Alam, M. R. and Gilbert, S. (2017). Monetary policy shocks and the dynamics of agricultural commodity prices: evidence from structural and factor-augmented var analyses. *Agricultural Economics*, 48(1):15–27.

Altissimo, F., Bassanetti, A., Cristadoro, R., Forni, M., Hallin, M., Lippi, M., Reichlin, L., and Veronese, G. (2001). Eurocoin: a real time coincident indicator of the euro area business cycle.

Ardeni, P. G. (1989). Does the law of one price really hold for commodity prices? *American Journal of Agricultural Economics*, 71(3):661–669.

Arezki, R., Hadri, K., Loungani, P., and Rao, Y. (2014). Testing the prebisch–singer hypothesis since 1650: Evidence from panel techniques that allow for multiple breaks. *Journal of International Money and Finance*, 42:208–223.

Askari, H. and Cummings, J. T. (1977). Estimating agricultural supply response with the nerlove model: a survey. *International economic review*, pages 257–292.

Baffes, J. and Etienne, X. (2016). Analysing food price trends in the context of engel's law and the prebisch-singer hypothesis. *Oxford Economic Papers*, 68(3):688–713.

Baffes, J. and Haniotis, T. (2010). Placing the 2006/08 commodity price boom into perspective.

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics*, 122(1):137–183.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2004). A panic attack on unit roots and cointegration. *Econometrica*, 72(4):1127–1177.

Bai, J. and Ng, S. (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics*, 25(1):52–60.

Bai, J., Ng, S., et al. (2008). Large dimensional factor analysis. *Foundations and Trends® in Econometrics*, 3(2):89–163.

Bai, J. and Wang, P. (2016). Econometric analysis of large factor models. *Annual Review of Economics*, 8:53–80.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of econometrics*, 73(1):5–59.

Bakhat, M. and Würzburg, K. (2013). Co-integration of oil and commodity prices: A comprehensive approach. *Economics for Energy Working Paper WP FA05*.

Balcombe, K. and Rapsomanikis, G. (2008). Bayesian estimation and selection of nonlinear vector error correction models: the case of the sugar-ethanol-oil nexus in brazil. *American Journal of Agricultural Economics*, 90(3):658–668.

Bańbura, M. and Modugno, M. (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, 29(1):133–160.

Barigozzi, M., Lippi, M., and Luciani, M. (2015). Dynamic factor models, cointegration, and error correction mechanisms. *arXiv preprint arXiv:1510.02399*.

Barigozzi, M., Lippi, M., and Luciani, M. (2016). Non-stationary dynamic factor models for large datasets. *arXiv preprint arXiv:1602.02398*.

Barigozzi, M. and Luciani, M. (2017). Common factors, trends, and cycles in large datasets. *arXiv preprint arXiv:1709.01445*.

Beck, S. (2001). Autoregressive conditional heteroscedasticity in commodity spot prices. *Journal of Applied Econometrics*, 16(2):115–132.

Beck, S. E. (1993). A rational expectations model of time varying risk premia in commodities futures markets: theory and evidence. *International Economic Review*, pages 149–168.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (favar) approach. *The Quarterly journal of economics*, 120(1):387–422.

Bleaney, M. and Greenaway, D. (1993). Long-run trends in the relative price of primary commodities and in the terms of trade of developing countries. *Oxford Economic Papers*, pages 349–363.

Boivin, J. and Giannoni, M. (2006). Dsge models in a data-rich environment. Technical report, National Bureau of Economic Research.

Boivin, J. and Ng, S. (2005). Understanding and comparing factor-based forecasts. Technical report, National Bureau of Economic Research.

Bonfatti, A. (2012). Commodity market analysis under rational expectations: the case of the world cocoa market.

Borensztein, E. and Reinhart, C. M. (1994). The macroeconomic determinants of commodity prices. *Staff Papers*, 41(2):236–261.

Boschi, M. and Pieroni, L. (2009). Aluminium market and the macroeconomy. *Journal of Policy Modeling*, 31(2):189–207.

Brooks, C. and Prokopczuk, M. (2013). The dynamics of commodity prices. *Quantitative Finance*, 13(4):527–542.

Bruno, V. G., Büyükşahin, B., and Robe, M. A. (2016). The financialization of food? *American Journal of Agricultural Economics*, 99(1):243–264.

Byrne, J. P., Fazio, G., and Fiess, N. (2013). Primary commodity prices: Co-movements, common factors and fundamentals. *Journal of Development Economics*, 101:16–26.

Cafiero, C., Bobenrieth, E., and Bobenrieth, J. (2011). Storage arbitrage and commodity price volatility. *Safeguarding food security in volatile global markets*, pages 301–326.

Cashin, P. and McDermott, C. J. (2002). The long-run behavior of commodity prices: small trends and big variability. *IMF staff Papers*, 49(2):175–199.

Chaudhuri, K. (2001). Long-run prices of primary commodities and oil prices. *Applied Economics*, 33(4):531–538.

Cooke, B. (2009). *Recent food prices movements: A time series analysis*, volume 942. Intl Food Policy Res Inst.

Corona, F., Poncela, P., and Ruiz, E. (2017). Estimating non-stationary common factors: implications for risk sharing. *Computational Economics*, pages 1–24.

Cuddington, J. T. (1992). Long-run trends in 26 primary commodity prices: A disaggregated look at the prebisch-singer hypothesis. *Journal of Development Economics*, 39(2):207–227.

Cuddington, J. T., Ludema, R., and Jayasuriya, S. A. (2002). Reassessing the prebisch-singer hypothesis: Long-run trends with possible structural breaks at unknown dates.

Cuddington, J. T. and Nülle, G. (2014). Variable long-term trends in mineral prices: The ongoing tug-of-war between exploration, depletion, and technological change. *Journal of International Money and Finance*, 42:224–252.

Cuddington, J. T. and Urzúa, C. M. (1989). Trends and cycles in the net barter terms of trade: a new approach. *The Economic Journal*, 99(396):426–442.

De Jong, P. et al. (1991). The diffuse kalman filter. *The Annals of Statistics*, 19(2):1073–1083.

Deaton, A. and Laroque, G. (1992). On the behaviour of commodity prices. *The Review of Economic Studies*, 59(1):1–23.

Deaton, A. and Laroque, G. (1996). Competitive storage and commodity price dynamics. *Journal of Political Economy*, 104(5):896–923.

Deaton, A. and Laroque, G. (2003). A model of commodity prices after sir arthur lewis. *Journal of Development Economics*, 71(2):289–310.

Deb, P., Trivedi, P. K., and Varangis, P. (1996). The excess co-movement of commodity prices reconsidered. *Journal of Applied Econometrics*, pages 275–291.

Delle Chiaie, S., Ferrara, L., and Giannone, D. (2017). Common factors of commodity prices. *La Défense-EconomiX*.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431.

Doz, C., Giannone, D., and Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on kalman filtering. *Journal of Econometrics*, 164(1):188–205.

Doz, C., Giannone, D., and Reichlin, L. (2012). A quasi–maximum likelihood approach for large, approximate dynamic factor models. *Review of economics and statistics*, 94(4):1014–1024.

Du, X., Cindy, L. Y., and Hayes, D. J. (2011). Speculation and volatility spillover in the crude oil and agricultural commodity markets: A bayesian analysis. *Energy Economics*, 33(3):497–503.

Engle, R. and Watson, M. (1981). A one-factor multivariate time series model of metropolitan wage rates. *Journal of the american statistical association*, 76(376):774–781.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.

Epstein, L. G. (1982). Integrability of incomplete systems of demand functions. *The Review of Economic Studies*, 49(3):411–425.

Erdem, F. P. and Ünalmış, İ. (2016). Revisiting super-cycles in commodity prices. *Central Bank Review*, 16(4):137–142.

Etienne, X. L., Irwin, S. H., and Garcia, P. (2014). Price explosiveness, speculation, and grain futures prices. *American Journal of Agricultural Economics*, 97(1):65–87.

FAO (2017). Food outlook. biannual report on global food markets.

Farooki, M. and Kaplinsky, R. (2013). *The impact of China on global commodity prices: The global reshaping of the resource sector*. Routledge.

Fattouh, B., Kilian, L., and Mahadeva, L. (2013). The role of speculation in oil markets: What have we learned so far? *The Energy Journal*, pages 7–33.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics*, 82(4):540–554.

Frankel, J. A. (1986). Expectations and commodity price dynamics: The overshooting model. *American Journal of Agricultural Economics*, 68(2):344–348.

Frankel, J. A. (2006). The effect of monetary policy on real commodity prices. Technical report, National Bureau of Economic Research.

Frankel, J. A. and Rose, A. K. (2010). Determinants of agricultural and mineral commodity prices.

Geweke, J. (1977). The dynamic factor analysis of economic time series. *Latent variables in socio-economic models.*

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of time series analysis*, 4(4):221–238.

Gilbert, C. L. (1989). The impact of exchange rates and developing country debt on commodity prices. *The Economic Journal*, 99(397):773–784.

Gilbert, C. L. (1990). The rational expectations hypothesis in models of primary commodity prices. Technical report, The World Bank.

Gilbert, C. L. (1995). Modelling market fundamentals: a model of the aluminium market. *Journal of Applied Econometrics*, 10(4):385–410.

Gilbert, C. L. (2010). How to understand high food prices. *Journal of agricultural economics*, 61(2):398–425.

Gonzalo, J. and Granger, C. (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business & Economic Statistics*, 13(1):27–35.

Granger, C. W. and Haldrup, N. (1996). Separation in cointegrated systems, long memory components and common stochastic trends. *Long Memory Components and Common Stochastic Trends*.

Grantham, J. (2011). Time to wake up: days of abundant resources and falling prices are over forever. *GMO Quarterly Letter*.

Grilli, E. R. and Yang, M. C. (1988). Primary commodity prices, manufactured goods prices, and the terms of trade of developing countries: what the long run shows. *The World Bank Economic Review*, 2(1):1–47.

Gustafson, R. L. (1958). *Carryover levels for grains: a method for determining amounts that are optimal under specified conditions*. Number 1178. US Department of Agriculture.

Hall, S. G., Psaradakis, Z., and Sola, M. (1999). Detecting periodically collapsing bubbles: a markov-switching unit root test. *Journal of Applied Econometrics*, pages 143–154.

Hamelinck, C. N. and Faaij, A. P. (2006). Outlook for advanced biofuels. *Energy Policy*, 34(17):3268–3283.

Hamilton, J. D. (2019). Measuring global economic activity. Technical report, National Bureau of Economic Research.

Harri, A., Nalley, L., and Hudson, D. (2009). The relationship between oil, exchange rates, and commodity prices. *Journal of Agricultural and Applied Economics*, 41(2):501–510.

Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge university press.

Harvey, D. I., Kellard, N. M., Madsen, J. B., and Wohar, M. E. (2017). Long-run commodity prices, economic growth, and interest rates: 17th century to the present day. *World Development*, 89:57–70.

Hotelling, H. (1931). The economics of exhaustible resources. *Journal of political Economy*, 39(2):137–175.

IMF (2016). Commodity special feature from world economic outlook.

Irwin, S. H., Sanders, D. R., and Merrin, R. P. (2009). Devil or angel? the role of speculation in the recent commodity price boom (and bust). *Journal of Agricultural and Applied Economics*, 41(2):377–391.

Jacks, D. and Stuermer, M. (2015). What drives commodity prices in the long run? Technical report.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica: journal of the Econometric Society*, pages 1551–1580.

Johnsen, T. A. (2001). Demand, generation and price in the norwegian market for electric power. *Energy Economics*, 23(3):227–251.

Jong, P. d. and Chu-Chun-Lin, S. (1994). Stationary and non-stationary state space models. *Journal of Time Series Analysis*, 15(2):151–166.

Kapetanios, G. and Marcellino, M. (2010). Factor-gmm estimation with large sets of possibly weak instruments. *Computational Statistics & Data Analysis*, 54(11):2655–2675.

Karali, B. and Power, G. J. (2013). Short-and long-run determinants of commodity price volatility. *American Journal of Agricultural Economics*, 95(3):724–738.

Kasa, K. (1992). Common stochastic trends in international stock markets. *Journal of monetary Economics*, 29(1):95–124.

Kaufmann, R. K. (2011). The role of market fundamentals and speculation in recent price changes for crude oil. *Energy Policy*, 39(1):105–115.

Kellard, N., Wohar, M. E., et al. (2002). Trends and persistence in primary commodity prices. *Royal Economic Society*.

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99(3):1053–69.

Kilian, L. (2019). Measuring global real economic activity: Do recent critiques hold up to scrutiny? *Economics Letters*, 178:106–110.

Kilian, L. and Lee, T. K. (2014). Quantifying the speculative component in the real price of oil: The role of global oil inventories. *Journal of International Money and Finance*, 42:71–87.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1-3):159–178.

LaFrance, J. T. (1985). Linear demand functions in theory and practice. *Journal of Economic Theory*, 37(1):147–166.

LaFrance, J. T. (1986). The structure of constant elasticity demand models. *American Journal of Agricultural Economics*, 68(3):543–552.

LaFrance, J. T. (1990). Incomplete demand systems and semilogarithmic demand models. *Australian Journal of Agricultural Economics*, 34(2):118–131.

LaFrance, J. T. and Hanemann, W. M. (1989). The dual structure of incomplete demand systems. *American Journal of Agricultural Economics*, 71(2):262–274.

Lescaroux, F. (2009). On the excess co-movement of commodity prices—a note about the role of fundamental factors in short-run dynamics. *Energy Policy*, 37(10):3906–3913.

Lipsey, R. E. (1994). Quality change and other influences on measures of export prices of manufactured goods and the terms of trade between primary products and manufacture. Technical report, National Bureau of Economic Research.

Lombardi, M. J., Osbat, C., and Schnatz, B. (2012). Global commodity cycles and linkages: a favar approach. *Empirical Economics*, pages 1–20.

Maples, J. G. and Brorsen, B. (2017). Time series modeling of cash and futures commodity prices. In *NCCC-134: Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management, St. Louis, MO, April*, pages 24–25.

Mensi, W., Beljid, M., Boubaker, A., and Managi, S. (2013). Correlations and volatility spillovers across commodity and stock markets: Linking energies, food, and gold. *Economic Modelling*, 32:15–22.

Mitchel, D. (2008). A note on rising food prices. Technical report.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica: Journal of the Econometric Society*, pages 315–335.

Myers, R. J. (1994). Time series econometrics and commodity price analysis: A review. *Review of Marketing and Agricultural Economics*, 62(2).

Natanelov, V., Alam, M. J., McKenzie, A. M., and Van Huylenbroeck, G. (2011). Is there co-movement of agricultural commodities futures prices and crude oil? *Energy Policy*, 39(9):4971–4984.

Nelson, C. R. and Plosser, C. R. (1982). Trends and random walks in macroeconmic time series: some evidence and implications. *Journal of monetary economics*, 10(2):139–162.

Nerlove, M. (1958a). Adaptive expectations and cobweb phenomena. *The Quarterly Journal of Economics*, 72(2):227–240.

Nerlove, M. (1958b). Distributed lags and estimation of long-run supply and demand elasticities: Theoretical considerations. *Journal of Farm Economics*, 40(2):301–311.

OECD and FAO (2017). Agricultural outlook 2017-2026.

Pei, F. and Tilton, J. E. (1999). Consumer preferences, technological change, and the short-run income elasticity of metal demand. *Resources Policy*, 25(2):87–109.

Peña, D. and Poncela, P. (2006). Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference*, 136(4):1237–1257.

Perali, F. and Pieroni, L. (2003). Fundamental market analysis with rational expectations: a model for primary commodities.

Peterson, H. H., Tomek, W. G., et al. (2000). Implications of deflating commodity prices for time-series analysis. In *2000 Conference, April 17-18 2000, Chicago, Illinois*, number 18944. NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management.

Phillips, P. C. et al. (1999). *Unit root log periodogram regression.* Citeseer.

Phillips, P. C. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346.

Phillips, P. C., Shi, S., and Yu, J. (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the s&p 500. *International Economic Review*, 56(4):1043–1078.

Phillips, P. C., Wu, Y., and Yu, J. (2011). Explosive behavior in the 1990s nasdaq: When did exuberance escalate asset values? *International economic review*, 52(1):201–226.

Pieroni, L. and Ricciarelli, M. (2005). Testing rational expectations in primary commodity markets. *Applied Economics*, 37(15):1705–1718.

Pindyck, R. S. and Rotemberg, J. J. (1990). The excess co-movement of commodity prices. *The Economic Journal*, 100(403):1173–1189.

Poncela, P. and Ruiz, E. (2016). Small-versus big-data factor extraction in dynamic factor models: An empirical assessment. In *Dynamic Factor Models*, pages 401–434. Emerald Group Publishing Limited.

Poncela, P., Senra, E., and Sierra, L. P. (2014). Common dynamics of nonenergy commodity prices and their relation to uncertainty. *Applied Economics*, 46(30):3724–3735.

Prakash, A. et al. (2011). *Safeguarding food security in volatile global markets.* Food and Agriculture Organization of the United Nations.

Prebisch, R. (1962). The economic development of latin america and its principal problems. *Economic Bulletin for Latin America.*

Quah, D. and Sargent, T. J. (1993). A dynamic index model for large cross sections. In *Business cycles, indicators and forecasting*, pages 285–310. University of Chicago Press.

Ramirez, O. A. and Fadiga, M. (2003). Forecasting agricultural commodity prices with asymmetric-error garch models. *Journal of Agricultural and Resource Economics*, pages 71–85.

Rao, J. M. (1989). Agricultural supply response: A survey. *Agricultural economics*, 3(1):1–22.

Ravazzolo, F. and Vespignani, J. (2015). A new monthly indicator of global real economic activity.

Reinhart, C. M. and Wickham, P. (1994). Commodity prices: cyclical weakness or secular decline? *Staff Papers*, 41(2):175–213.

Roache, S. K. (2010). *What explains the rise in food price volatility?* Number 10-129. International Monetary Fund.

Robinson, P. M. (1995). Log-periodogram regression of time series with long range dependence. *The annals of Statistics*, pages 1048–1072.

Samuelson, P. A. (1971). Stochastic speculative price. *Proceedings of the National Academy of Sciences*, 68(2):335–337.

Sargent, T. J. and Sims, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. *New methods in business cycle research*, 1:145–168.

Sari, R., Hammoudeh, S., Chang, C.-L., and McAleer, M. (2012). Causality between market liquidity and depth for energy and grains. *Energy Economics*, 34(5):1683–1692.

Scheinkman, J. A. and Schechtman, J. (1983). A simple competitive model with production and storage. *The Review of Economic Studies*, 50(3):427–441.

Schroeder, T. C. and Goodwin, B. K. (1991). Price discovery and cointegration for live hogs. *Journal of Futures Markets*, 11(6):685–696.

Schwartz, E. and Smith, J. E. (2000). Short-term variations and long-term dynamics in commodity prices. *Management Science*, 46(7):893–911.

Schwartz, E. S. (1997). The stochastic behavior of commodity prices: Implications for valuation and hedging. *The Journal of Finance*, 52(3):923–973.

Shi, W., Wang, G., Zhao, X., Feng, X., and Wu, J. (2018). Price determination in the electrolytic aluminum industry: The role of electricity prices. *Resources Policy*.

Singer, H. W. (1950). The distribution of gains between investing and borrowing countries. *The American Economic Review*, 40(2):473–485.

Slade, M. E. (1982). Trends in natural-resource commodity prices: an analysis of the time domain. *Journal of Environmental Economics and Management*, 9(2):122–137.

Stock, J. H. and Watson, M. (2011). Dynamic factor models. *Oxford handbook on economic forecasting*.

Stock, J. H. and Watson, M. W. (1988). Testing for common trends. *Journal of the American statistical Association*, 83(404):1097–1107.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Tang, K. and Xiong, W. (2012). Index investment and the financialization of commodities. *Financial Analysts Journal*, 68(5):54–74.

Trivedi, P. K. (1990). The prices of perennial crops: the role of rational expectations and commodity stocks. *Primary commodity prices: economic models and policy.*, pages 72–102.

Tyner, W. E. (2010). The integration of energy and agricultural markets. *Agricultural Economics*, 41(s1):193–201.

Vansteenkiste, I. (2009). How important are common factors in driving non-fuel commodity prices? a dynamic factor analysis. Technical report, European Central Bank.

Von Haefen, R. H. (2002). A complete characterization of the linear, log-linear, and semi-log incomplete demand system models. *Journal of Agricultural and Resource Economics*, pages 281–319.

Wickens, M. R. and Greenfield, J. (1973). The econometrics of agricultural supply: an application to the world coffee market. *The Review of Economics and Statistics*, pages 433–440.

Williams, J. C. and Wright, B. D. (1991). *Storage and Commodity Markets*.

World Bank (2017). Commodity markets outlook.

Wright, B. (2001). Storage and price stabilization. *Handbook of Agricultural Economics*, 1:817–861.

Yamada, H. and Yoon, G. (2014). When grilli and yang meet prebisch and singer: Piecewise linear trends in primary commodity prices. *Journal of International Money and Finance*, 42:193–207.

Yin, L. and Han, L. (2015). Co-movements in commodity prices: Global, sectoral and commodity-specific factors. *Economics Letters*, 126:96–100.

Zhang, Z., Lohr, L., Escalante, C., and Wetzstein, M. (2010). Food versus fuel: What do prices tell us? *Energy policy*, 38(1):445–451.

Zink, T., Geyer, R., and Startz, R. (2016). A market-based framework for quantifying displaced production from recycling or reuse. *Journal of Industrial Ecology*, 20(4):719–729.

Appendix A

Dataset presentation

Table A.1: Data description, IMF database of primary commodity prices

	Commodity price description
Aluminium	Aluminum, 99.5% minimum purity, LME spot price, CIF UK ports, US$ per metric ton
Barley	Barley, Canadian no.1 Western Barley, spot price, US$ per metric ton
Beef	Beef, Australian and New Zealand 85% lean fores, CIF U.S. import price, US cents per pound
Coal	Coal, Australian thermal coal, 12,000- btu/pound, less than 1% sulfur, 14% ash, FOB Newcastle/Port Kembla, US$ per metric ton
Cocoa	Cocoa beans, International Cocoa Organization cash price, CIF US and European ports, US$ per metric ton
Coffee	Coffee, Robusta, International Coffee Organization New York cash price, ex-dock New York, US cents per pound
Rapeseed oil	Rapeseed oil, crude, fob Rotterdam, US$ per metric ton
Copper	Copper, grade A cathode, LME spot price, CIF European ports, US$ per metric ton
Cotton	Cotton, Cotton Outlook 'A Index', Middling 1-3/32 inch staple, CIF Liverpool, US cents per pound
Hides	Hides, Heavy native steers, over 53 pounds, wholesale dealer's price, US, Chicago, fob Shipping Point, US cents per pound
Lamb	Lamb, frozen carcass Smithfield London, US cents per pound
Lead	Lead, 99.97% pure, LME spot price, CIF European Ports, US$ per metric ton
Soft Logs	Soft Logs, Average Export price from the U.S. for Douglas Fir, US$ per cubic meter
Hard Logs	Hard Logs, Best quality Malaysian meranti, import price Japan, US$ per cubic meter
Maize	Maize (corn), U.S. No.2 Yellow, FOB Gulf of Mexico, U.S. price, US$ per metric ton
Nickel	Nickel, melting grade, LME spot price, CIF European ports, US$ per metric ton
Crude oil	1) Crude Oil (petroleum), Dated Brent, light blend 38 API, fob U.K., US$ per barrel 2) Crude Oil (petroleum), Dubai Fateh Fateh 32 API, US$ per barrel 3) Crude Oil (petroleum), West Texas Intermediate 40 API, Midland Texas, US$ per barrel
Olive oil	Olive Oil, extra virgin less than 1% free fatty acid, ex-tanker price U.K., US$ per metric ton
Swine	Swine (pork), 51-52% lean Hogs, U.S. price, US cents per pound
Poultry	Poultry (chicken), Whole bird spot price, Ready-to-cook, whole, iced, Georgia docks, US cents per pound
Rice	Rice, 5 percent broken milled white rice,
Rubber	Rubber, Singapore Commodity Exchange, No. 3 Rubber Smoked Sheets, 1st contract, US cents per pound
Salmon	Fish (salmon), Farm Bred Norwegian Salmon, export price, US$ per kilogram
Hard Sawnwood	Hard Sawnwood, Dark Red Meranti, select and better quality, C&F U.K port, US$ per cubic meter
Soft Sawnwood	Soft Sawnwood, average export price of Douglas Fir, U.S. Price, US$ per cubic meter
Shrimps	Thailand Whiteleg Shrimp 70 Shrimps/Kg Spot Price
Sunflower oil	Sunflower Oil, US export price from Gulf of Mexico, US$ per metric ton
Tea	Tea, Mombasa, Kenya, Auction Price, US cents per kilogram, From July 1998,Kenya auctions, Best Pekoe Fannings. Prior, London auctions, c.i.f. U.K. warehouses
Tin	Tin, standard grade, LME spot price, US$ per metric ton
Uranium	Uranium, NUEXCO, Restricted Price, Nuexco exchange spot, US$ per pound
Wheat	Wheat, No.1 Hard Red Winter, ordinary protein, Kansas City, US$ per metric ton
Wool	Wool, coarse, 23 micron, Australian Wool Exchange spot quote, US cents per kilogram
Zinc	Zinc, high grade 98% pure, US$ per metric ton
Gold	Gold, Fixing Committee of the London Bullion Market Association, London 3 PM fixed price, US$ per troy ounce
Silver	Silver, London Bullion Market Association, USD/troy ounce
Platinum	Platinum, LME spot price, USD/troy ounce

Table A.2: Summary Statistics, observations 1980:01 to 2018:12

Variable	Mean	Median	S.D.	Min	Max
aluminium	110.	98.7	38.4	64.1	310.
barley	149.	135.	46.6	78.7	302.
beef	153.	152.	49.7	82.4	361.
coal	225.	202.	96.2	89.7	646.
cocoa	236.	194.	115.	93.2	815.
coffee	195.	144.	132.	37.9	727.
rapoil	194.	179.	63.8	84.3	410.
copper	192.	171.	86.3	71.5	430.
cotton	178.	150.	81.0	75.1	474.
hides	109.	105.	28.4	39.4	192.
iron	275.	153.	254.	95.1	1.21e+003
lamb	107.	106.	34.2	48.9	253.
lead	224.	187.	112.	83.0	667.
slogs	91.8	85.6	23.3	53.3	182.
hlogs	116.	109.	35.0	64.5	319.
maize	160.	142.	58.8	79.8	343.
nickel	134.	112.	70.7	47.5	532.
oilbrent	170.	140.	93.7	39.7	441.
oilDF	175.	143.	99.8	44.3	470.
oilWTI	156.	126.	78.5	42.4	409.
oliveoil	99.9	94.0	24.5	58.6	188.
swine	183.	130.	119.	37.8	638.
poultry	117.	114.	15.4	97.8	191.
rice	150.	135.	70.1	63.4	437.
rubber	238.	208.	115.	72.1	772.
salmon	163.	131.	91.1	57.5	483.
hsawnwood	94.3	96.9	23.3	43.0	157.
ssawnwood	91.3	89.0	12.1	68.7	127.
shrimps	95.8	94.2	36.5	43.3	210.
sunoil	188.	175.	60.0	76.0	447.
tea	114.	104.	41.3	67.8	329.
tin	208.	172.	126.	59.6	672.
uranium	271.	213.	188.	72.5	1.22e+003
wheat	196.	181.	74.7	94.0	445.
wool	229.	230.	80.7	90.1	475.
zinc	126.	120.	43.0	59.6	321.
gold	226.	197.	105.	88.8	555.
silver	219.	184.	171.	75.7	1.77e+003
platinum	171.	156.	71.8	79.9	463.

Appendix B

Loading matrices

Table B.1: Matrix of transitory loadings: $\alpha(\beta'\alpha)^{-1}\Delta_z$

aluminum	0.26144	1.7349	1.2188	-0.21430	-1.4963	-0.88540
copper	0.12196	0.78486	0.24015	-0.053364	-0.70523	-0.12195
lead	-0.085597	0.14861	-0.033272	0.092387	-0.13655	0.035151
nickel	0.24394	0.31339	0.21035	-0.19635	-0.26486	-0.10801
tin	0.43009	0.40965	0.31178	-0.35324	-0.34750	-0.20096
uranium	-0.44219	4.2647	0.16474	0.69906	-3.8782	-0.29899
zinc	-1.6014	4.0742	6.0180	1.9434	-3.6511	-4.5219
gold	-0.16423	2.7907	0.52654	0.33646	-2.5623	-0.55981
silver	4.7981	7.3890	2.6779	-3.3051	-6.6820	-2.4134
platinum	1.3831	1.3803	3.2129	-0.53516	-1.2357	-2.7997
oilbrent	-0.44527	-0.10752	0.71622	0.44245	0.10448	-0.66354
oilDF	-0.53658	-0.13286	0.85735	0.53263	0.12868	-0.79527
oilWTI	-0.46641	-0.11338	0.74889	0.46333	0.11008	-0.69404
coal	-0.021079	0.057697	0.14274	0.031294	-0.047922	-0.11363
beef	-0.022862	-0.026306	0.044337	0.026254	0.018173	-0.044375
lamb	-0.0088295	0.055415	0.045959	0.015570	-0.041595	-0.028432
swine	-0.095840	-0.094061	0.27507	0.13211	0.10411	-0.22602
poultry	0.0082306	0.010966	-0.038369	-0.012830	-0.014982	0.029227
salmon	-0.023006	0.043486	-0.00015540	0.031150	-0.040070	0.0012571
shrimps	0.0048265	0.059609	-0.054866	0.0052259	-0.037644	0.062793
cotton	0.021840	0.0042398	-0.018542	-0.010904	0.0012063	0.023545
hides	-0.019851	0.096580	0.24678	0.037826	-0.081598	-0.20217
slogs	0.053421	0.29499	0.30689	-0.027301	-0.25675	-0.28316
hlogs	-0.026734	0.19943	0.062711	0.046520	-0.17105	-0.059575
rubber	0.0073007	0.29466	0.45342	0.033440	-0.24810	-0.40081
hsawnwood	-0.094667	0.17074	0.068021	0.099851	-0.14795	-0.051686
ssawnwood	0.022105	0.16846	0.21845	-0.0073711	-0.14650	-0.20008
wool	0.069942	0.085184	0.065078	-0.058107	-0.072025	-0.053037
barley	0.079998	0.0044369	0.48622	-0.060061	0.032229	-0.39606
cocoa	-0.0075528	-0.088009	-0.086676	0.010050	0.075453	0.064662
coffee	-0.14485	0.57111	0.20917	0.14311	-0.50574	-0.19011
rapoil	-0.049470	0.16083	0.26022	0.056373	-0.12624	-0.21409
maize	0.035594	0.085683	0.30830	-0.022855	-0.055537	-0.25366
oliveoil	-0.12283	0.21997	0.0049474	0.10212	-0.19809	-0.015551
rice	-0.022439	0.070411	0.15541	0.019898	-0.051375	-0.13099
sunoil	0.0031361	-0.050093	0.078309	-0.0033706	0.052931	-0.064146
tea	0.0085833	-0.012476	0.17755	-0.0071417	0.027198	-0.15295
wheat	0.020961	0.13146	0.26164	-0.015945	-0.10102	-0.21902

Table B.2: Matrix of permanent loadings: $\beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\Lambda_\Delta$

aluminum	0.031616	0.30241	0.27429	0.12096	0.031745	0.069663
copper	0.11116	0.13133	-0.00036261	-0.031002	-0.0012945	0.024353
lead	0.034900	0.0016670	0.0058440	-0.017117	-0.0074333	0.048340
nickel	0.023810	0.15232	0.14493	0.060098	0.016767	-0.014046
tin	0.16350	0.18746	0.084907	0.014153	0.00088685	-0.013318
uranium	-0.051110	-0.41170	0.21065	-0.33141	-0.012121	-0.058552
zinc	0.80030	-0.30997	0.062259	-0.49208	-0.18710	0.31499
gold	-0.086994	-0.077913	0.72509	-0.33759	-0.020314	0.47830
silver	-1.4189	0.12187	4.6997	-1.8229	0.27324	2.9777
platinum	0.41710	0.38803	0.41630	-0.32090	-0.039634	0.62620
oilbrent	0.023965	0.066593	0.052984	0.035684	0.0091972	-0.022144
oilDF	0.056457	0.048299	0.041709	0.011859	-0.0018410	-0.0073534
oilWTI	0.044619	0.054006	-0.049149	0.053037	-0.0092991	0.056895
coal	0.044946	0.020570	0.029367	-0.013354	-0.0040365	-0.0074682
beef	0.016033	0.028500	0.060534	0.00085085	0.0019922	0.0016513
lamb	0.019260	0.021467	0.031358	0.012767	-0.0038805	0.024161
swine	-0.10081	-0.055755	-0.084554	-0.037132	0.0030047	-0.024933
poultry	-0.0020163	-0.013161	-0.023308	-0.0051560	-0.0034284	0.0053284
salmon	-0.023108	-0.055463	-0.062240	-0.0082283	-0.0030053	0.016760
shrimps	0.022646	0.028330	0.031882	0.0022947	-0.0046198	0.019885
cotton	-0.0083898	0.027361	0.016149	-0.0065770	0.0078353	-0.021532
hides	-0.0064120	-0.00080270	0.031606	-0.0049652	-0.0017674	0.0098219
slogs	0.0016305	-0.014448	-0.018100	-0.011446	-0.0098683	0.020118
hlogs	0.021268	0.014667	0.044775	-0.013247	-0.0039347	0.0054213
rubber	-0.0020173	0.0074912	-0.0080789	-0.018325	0.00064892	-0.020511
hsawnwood	0.0095228	-0.020687	-0.0026056	-0.019929	-0.0050374	-0.0025018
ssawnwood	-0.0033890	-0.012363	-0.014453	-0.0085954	-0.0052315	0.0040693
wool	-0.0086291	-0.0040075	-0.047800	0.018339	-0.0028286	0.031861
barley	0.041324	0.075364	0.023726	0.031863	0.0020405	-0.0094405
cocoa	-0.012522	-0.042102	0.0064642	-0.015707	0.00037979	0.012995
coffee	0.013669	-0.10658	-0.036533	-0.037571	-0.016957	0.036389
rapoil	0.017567	-0.0025878	-0.028092	-0.0077273	-0.0085036	0.026914
maize	0.013765	0.074465	-0.0039160	0.0044415	0.0042815	-0.012972
oliveoil	-0.0090721	-0.024122	-0.024820	-0.0078092	-0.0019876	0.0054885
rice	0.018640	-0.021383	-0.00030458	-0.016797	-0.0052490	0.0084712
sunoil	-0.00012512	0.00013969	-0.0075743	-0.0092384	0.0050798	0.0070919
tea	-0.014171	-0.0041929	0.0013692	0.010893	-0.00064232	-0.0032034
wheat	0.023198	0.069499	0.010597	-0.0074825	0.0011423	-0.025381