Dynamic modeling of commodity futures prices

Paul Karapanagiotidis*

Thesis Draft
June 22, 2014

Abstract

Theory suggests that physical commodity prices may exhibit nonlinear features such as bubbles and various types of asymmetries. This paper investigates these claims empirically by introducing a new time series model apt to capture such features. The data set is composed of 25 individual, continuous contract, commodity futures price series, representative of a number of industry sectors including softs, precious metals, energy, and livestock. It is shown that the linear causal ARMA model with Gaussian innovations is unable to adequately account for the features of the data. In the purely descriptive time series literature, often a threshold autoregression (TAR) is employed to model cycles or asymmetries. Rather than take this approach, we suggest a novel process which is able to accommodate both bubbles and asymmetries in a flexible way. This process is composed of both causal and noncausal components and is formalized as the mixed causal/noncausal autoregressive model of order (r, s). Estimating the mixed causal/noncausal model with leptokurtic errors, by an approximated maximum likelihood method, results in dramatically improved model fit according to the Akaike information criterion. Comparisons of the estimated unconditional distributions of both the purely causal and mixed models also suggest that the mixed causal/noncausal model is more representative of the data according to the Kullback-Leibler measure. Moreover, these estimation results demonstrate that allowing for such leptokurtic errors permits identification of various types of asymmetries. Finally, a strategy for computing the multiple steps ahead forecast of the conditional distribution is discussed.
Keywords: commodity futures, mixed causal/noncausal model, nonlinear dynamic models, commodity futures, speculative bubble.
JEL: C22, C51, C52, C58

*University of Toronto, Department of Economics, p.karapanagiotidis@utoronto.ca

1 Introduction

Financial theory has proposed general approaches for pricing financial assets and their derivatives, based on arbitrage pricing theory [Ross (1976)], or equilibrium models: for example the Capital Asset Pricing Model [Sharpe (1964)] or Consumption-Based Capital Asset Pricing Model [Breeden (1979)]. Traders have also relied on technical analysis for insight into price movements [see e.g. Frost (1986)].

These approaches are generally applied separately on the different segments of the market, each segment including a set of basic assets plus the derivatives written on these basic assets. These segments are used for different purposes and can have very different characteristics. A standard example is the stock market, where the basic assets are the stocks and the derivatives are both options written on the market index and futures written on the index of implied volatility, called the VIX. These derivatives have been introduced to hedge and trade against volatility risk. A large part of the theoretical and applied literature analyzes this stochastic volatility feature.

Another segment also largely studied is the bond market, including the sovereign bonds, but also the bonds issued by corporations and the mortgage backed securities; the associated derivatives in this case are insurance contracts on the default of the borrowers, such as Credit Default Swaps (CDS) or Collateralized Debt Obligations (CDO). These derivatives have been introduced to manage the counterparty risks existing in the bond market.

This paper will focus on another segment, that is the segment of commodities. This segment includes the spot markets, derivatives such as the commodity futures with and without delivery, and derivatives such as options, puts and calls, written on these futures.

This segment has special features compared to other segments, such as the stock

market for instance. At least three features make the commodity markets rather unique:

i) The basic assets are physical assets. There is a physical demand and a physical supply for these commodities and by matching their demand and supply, we may define a "fundamental price" for each commodity. It is known that the analysis of these fundamental prices can be rather complex even if it concerns the real economy only. This is mainly a consequence of both shifts in demand and supply and of various interventions to control the fundamental price of commodities. What follows are examples of such effects which differ according to the commodity.

   Cycles are often observed on commodity prices. They can be a consequence of costly, irreversible investment, made to profit from high prices. For instance, farmers producing corn can substitute into producing cattle, when grain prices are low. The production of milk (or meat) will increase and jointly the production of grain will diminish. As a consequence the prices of milk (or meat) will decline, whereas the price of grain will increase. This creates an incentive to substitute grain to cattle in the future and so forth, which introduces cycles in the price evolution of both corn and cattle. Other substitutions between commodities can also create a change of trend in prices. For example, the development of alternative fuel derived from soy created a significant movement in soy prices.

   These complicated movements can also be affected by different interventions to sustain and/or stabilize the prices. The interventions can be done by governments (e.g. U.S., or European nations) for agricultural commodities, as well as by (monopolistic or oligopolistic) producers such as the Organization of Petroleum Exporting Countries (OPEC) for petroleum production or the De Beers company for diamonds. The real demand and supply will affect the spot prices and futures con-

tracts with delivery.

ii) Recently the commodity markets have also experienced additional demand and supply pressures by financial intermediaries. These intermediaries are not interested in taking delivery of the underlying products upon maturity and are only interested in cashing in on favourable price changes in the futures contracts. This behaviour betrays the original purpose of the futures markets which was to enable both producers and consumers to hedge against the risk of future price fluctuations of the underlying commodity.

    To try to separate the market for the physical commodity from simply gambling on their prices, pure intangible assets have been introduced that are the commodity futures without delivery. Thus the market for commodity derivatives has been enlarged. As usual, the speculative effect is proportional to the magnitude and importance of the derivative market. This speculative effect is rather similar to what might be seen in the markets for CDS or on the implied volatility index (VIX).

iii) The different spot and futures markets for commodities are not very organized and can involve a small number of players and very often feature a lack of liquidity.

The economic literature mainly focuses on two features of commodity prices, that are their cross-sectional and serial heterogeneity, respectively. Below, I will discuss the literature specific to each. The cross-sectional analysis tries to understand how the prices of futures contracts with delivery are related with the spot prices, or to explain the difference between the prices of futures with and without delivery. The analysis of the serial heterogeneity of prices focuses on the nonlinear dynamic features due to either the cycles and rationing effects coming from the real part of the market, or the speculative bubbles created by the behaviour of financial arbitrageurs.

The questions above can be considered from either a structural, or a descriptive point of view. A "structural" approach attempts to construct a theoretical model involving the relevant economic variables of interest which may be important in explaining relationships which drive commodity spot and futures prices. The descriptive approach does not explain "why" these series exhibit particular features, but rather provides a framework to estimate the relationships between the prices, make forecasts, and price the derivatives.

What follows is a discussion on how these two approaches above have been addressed in the literature.

i) Cross-sectional heterogeneity

The study of cross-sectional heterogeneity of commodity futures prices has its roots in both the *theory of normal backwardation* and the *theory of storage*. The Keynesian theory of normal backwardation implies a greater expected future spot price than the current futures contract price, assuming that producers are on net hedgers and that speculators, in order to take on the risk offered by producers, must be offered a positive risk premium.

Of the two theories, the theory of storage has probably had the greater influence. Instead of focusing on the net balance of trader's positions as in the theory of normal backwardation, the theory of storage focuses on how the levels of inventory, that is the "stocks," of the underlying commodities affect the decisions of market participants. Inventories play an important role since it is known that both the consumption and supply of many commodities are inelastic to price changes. For example, it is known that gasoline and petroleum products are everyday necessities and both their consumption and production adjust slowly to price changes. Moreover, given real supply and demand shocks the inelastic nature of these markets can lead to wild price fluctuations. There-

fore, the role of inventories is important in buffering market participants from price fluctuations, by avoiding disruptions in the flow of the underlying commodities, and by allowing them to shift their consumption or production intertemporally.

The cost of storage is essentially a "no arbitrage" result. Let the difference of the current futures price and the spot price be known as the *basis*. If the basis is positive, it must necessarily equal the cost of holding an inventory into the future, known as the *cost of carry,* since otherwise a trader could purchase the good on the spot market, enter into a futures contract for later delivery, and make a sure profit (or loss). From the reverse point of view, the basis could never be negative since holders of inventories could always sell the good at the spot price, and enter a futures contract to buy at the lower price, with no cost of carry.

However, empirical examination of the basis reveals that it is often negative. Kaldor (1939) was the first to suggest a solution to this problem known as the *convenience yield.* The convenience yield measures the benefit of owning physical inventories, rather than owning a futures contract written on them. When a good is in abundance, an investor gains little by owning physical inventories. However, when the good is scarce, it is preferable to hold inventories. Therefore, in equilibrium the basis should be equal to the difference between the cost of carry and the convenience yield, permitting the basis to be negative when inventories are scarce.

Working (1933,1948,1949) used the theory of storage to describe the relationship between the price of storage and inventories for the wheat market, called the "Working curve" or the storage function. The Working curve is positively sloped and for some positive threshold storage level, relates inventories to the costs of storing them; however, below this positive threshold of inventories, the function takes on negative values, illustrating that positive inventories can be held even when the returns from storage are

negative, thereby incorporating the notion of Kaldor's convenience yield into the storage function.

Later work generalized these results in considering motivations for both storage behaviour and the convenience yield. For example, Brennan (1958) considered storage from the speculative point of view, suggesting that on the supply side, in addition to cost of storage, we expand the notion of the convenience yield to include a risk premium to holders of inventories who may speculate upon, and benefit from, a possible rise of demand on short notice.

Modern structural models distinguish between what is the fundamental price connected with the underlying physical supply and demand, from the cost of storage and any speculation. For example, in looking at oil price speculation, Knittel and Pindyck (2013) address what is meant by the notion of "oil price speculation" and how it relates to investment in oil reserves, inventories, or derivatives such as futures contracts. Although the price of storage is not directly observed, it can be determined from the spread between futures and spot prices. In their model there are two interrelated markets for a commodity: the *cash market* for immediate or "spot" purchase/sale, and the "storage market" for inventories. The model attempts to distinguish between the physical supply and demand market and the effect of speculators on both the futures and spot prices.

Other structural work on the basis has employed the CAPM model. For example Black (1976) studied the nature of futures contracts on commodities, suggesting that the capital asset model of Sharpe (1964) could be employed to study the expected price change of the futures contract. Dusak (1973) also studied the behaviour of futures prices within a model of capital market equilibrium and found no risk premium for U.S. corn, soybeans, and wheat futures between 1952 and 1967. Breeden (1979) developed the consumption CAPM model which allowed us to consider the futures price as composed

of both an expected risk premium and a forecast of the future spot price.

Econometrically, Fama and French (1987) found evidence that the response of futures prices to storage-cost variables was easier to detect than evidence that futures prices contain premiums or power to forecast spot prices.

Other econometric work has been purely descriptive in attempting to model the basis process itself. For example, Gibson and Schwartz (1990) model the convenience yield as a mean reverting continuous time stochastic process, where the unconditional mean represents the state of inventories which satisfy industry under normal conditions.

The cost of storage also imposes a natural constraint on inventories in that they cannot be negative; this has effects which show up empirically. For example, inventory levels and the basis tend to share a positive relationship as the theory of storage and convenience yield would suggest. Brooks et al. (2011) employ actual physical inventory levels data on 20 different commodities between 1993-2009 and show that inventory levels are informative about the basis, so that when inventories are low the basis is possibly negative (and vice versa). They also find that futures price level volatility is a decreasing linear function of inventories so that when the basis is negative, price volatility is higher. Empirical evidence also suggests that the basis behaves differently when it is positive versus when it is negative. For example, Brennan (1991) expanded the work of Gibson and Schwartz (1990) by incorporating the non-negativity constraint of inventories and so the convenience yield is downward limited.

Finally, there is econometric evidence that corroborates Brennan (1958) above. Sigl-Grub and Schiereck (2010), employ commitment of traders information on 19 commodity futures contracts between 1986 and 2007 (using the commitment of traders information as a proxy for speculation) and find that the autoregressive persistence of futures returns processes tend to increase with speculation.

ii) Price dynamics

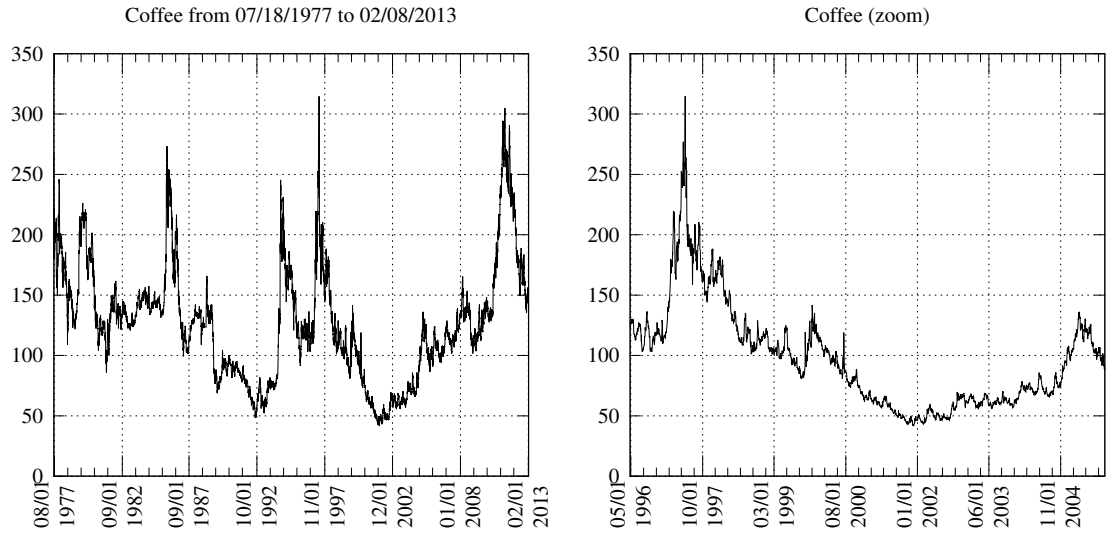
Another part of the literature tries to understand the nonlinear dynamic patterns observed in futures prices that can manifest as either cycles or speculative bubbles. Generally, we observe more or less frequent successive peaks and troughs in the evolution of prices. These peaks and troughs have non standard patterns which can be classified according to the terminology in Ramsey and Rothman (1996) where they distinguish the concepts of "longitudinal" and "transversal" asymmetry. The notion of longitudinal asymmetry employed in Ramsey and Rothman (1996) builds upon other previous work, for example the study of business cycle asymmetry from Neftci (1984).

Longitudinal asymmetry refers to asymmetry where the process behaves differently when traveling in direct time versus in reverse time. For example, longitudinal asymmetry may manifest as a process where the peaks rise faster than then they decline (and behaves in the opposite way in reverse). **Figure 1** provides a plot which illustrates these features for the coffee price level, continuous futures contract without delivery. In the right panel (which provides a zoom) we can see how the peaks tend to rise quickly, but take a long time to decline into the trough.

Transversal asymmetry is characterized by different process dynamics above and below some horizontal plane in the time direction; that is, in the vertical displacement of the series from its mean value. For example, the coffee process also exhibits transversal asymmetry in that the peaks in the positive direction are very sharp and prominent, while the troughs are very drawn out and shallow (again see **Figure 1** right panel). So, a series can be both longitudinally and transversely asymmetric.

The theoretical literature has been able to derive price evolutions with such patterns as a consequence of self-fulfilling prophecies. The initial rational expectation (RE) models were linear: the demand is a linear function of the current expected future prices

Figure 1: Plots of daily continuous contract futures price level series, Coffee with zoom



and exogenous shocks on demand, and the supply is a linear function of the current price and of supply shocks. In this way we can consider the path of equilibrium prices. Muth (1961) was the first to employ such a framework which incorporated expectations formation directly into the model.

Since the equilibrium in RE models is both with respect to prices and information, these models have an infinite number of solutions, even if the exogenous shocks have only linear dynamic features. Some of these solutions have nonlinear dynamic features which are similar to the asymmetric bubble patterns described above. Among these solutions featuring bubbles, some can exhibit isolated bubbles and others can demonstrate a sequence of repeating bubbles. For example, Blanchard (1979) and Blanchard and Watson (1982) derived RE bubble models for the stock market which presumed the price process is composed of both the fundamental competitive market solution for price

[1] plus a non-stationary martingale component that admits a rational expectation representation [Gourieroux, Laffont, and Monfort (1982)], but exhibits bubble like increases or decreases in price. Blancard and Watson (1982) described a possible piecewise linear model for the martingale bubble component which spurred later authors to test statistically for the presence of this component. Later, Evans (1991) suggested that such econometric tests may be limited in their ability to detect a certain important class of rational bubbles which exhibit repeating explosive periods.

Generally these basic modeling attempts were focused on the stock market and it is not clear what analog there is (if any) of the "fundamental" price of the futures contract without delivery. Moreover, they take into account only the expected prices, not the level of volatility and they incorporate linear functions for the price, and so the solution may not be unique.

More recent RE models have exhibited features consistent with the asymmetries discussed above with regards to both Ramsey and Rothman (1996) and the cost of storage models and the natural asymmetry which occurs since inventories cannot be negative. For example, Deaton and Laroque (1996) construct a RE model of commodity spot prices, in which they generate a "harvest" process[2] which drives a competitive price in agricultural markets composed of both final consumers and risk-neutral speculators. From an intertemporal equilibrium perspective, when the price today is high (relative to tomorrow) nothing will be stored so there will be little speculation; however, when the price tomorrow is high (relative to today), speculation will take place and storage will be positive. Because inventories cannot be negative, the market price process under storage will follow a piecewise linear dynamic stochastic process.

[1] That is, where price is the linear present value of future dividends.

[2] The process may possibly be serially correlated. The authors discuss at least the major differences that occur in the model dynamics when harvests are i.i.d. versus serially correlated.

Moreover, both theory and evidence suggests that RE models might take the form of a noncausal process. For example, Hansen and Sargent (1991) showed that if agents in the commodity futures market can be described by a linear RE model, and have access to an information set strictly larger than that available to the econometrician modeling them, then the true shocks of the moving average representation that describe the RE equilibrium process will not represent the shocks the econometrician estimates given a purely causal linear model. In fact, the shocks of the model will have a non-fundamental representation and we say that the model is at least partly "noncausal." Of course, modeling a process as partly noncausal does not imply that agents somehow "know the future." Rather, it simply represents another equivalent linear representation.

Through simulation studies, Lof (2011) also showed that if we simulate the market asset price from both an RE model with homogenous agents and that from a model with boundedly rational agents with heterogenous beliefs [based on the model by Brock and Hommes (1998)], and then estimate both a purely causal model and a model with a noncausal component on this data (given that the econometrician has full information) we find that on average the rational expectations model is better fit by the causal model, while the heterogenous agents model is better fit by a noncausal model.

Given these features, the time series literature has rapidly realized that the standard linear dynamic models, that is, the autoregressive moving average (ARMA) processes with Gaussian shocks, are not appropriate for representing the evolution of either commodity spot or futures prices. Indeed, they are not able to capture the nonlinear dynamic features due to asymmetric cycles and price bubbles described above. For describing the cycles created through the dynamics of investment between two substitutable commodities among producers (see the discussion of the example of cattle vs. grain above), it is rather natural to consider an autoregressive model with a threshold, that is, the thresh-

old autoregressive model (TAR) introduced by Tong and Lim (1980) in the time series literature. Indeed, the cycles associated with substitutable products are in some ways analogous to the predator-prey cycle for which the TAR model was initially introduced. The TAR model has been applied on commodity prices to study the integration between corn and soybean markets in North Carolina by Goodwin and Piggotts (2001) and U.S. soybeans and Brazilian coffee by Ramirez (2009) to compare the asymmetry of such cycles.

Contribution of the paper

Our paper contributes to the empirical literature on commodity futures prices by implementing nonlinear dynamic models apt to reproduce the patterns of speculative bubbles observed on the commodity price data. To focus on speculative bubbles and not on the underlying cycles of the fundamental spot price, we consider the continuous contract futures price series available from Bloomberg on which it is believed that the speculative effects will be more pronounced. We propose to analyze such series by means of the mixed causal/noncausal models where the underlying noise defining the process has fat tails. Indeed, it has been shown in Gourieroux and Zakoian (2012) that such models can be used to mimic speculative bubbles, or more generally peaks and troughs with either longitudinal or transversal asymmetry. The estimation of such mixed models will be performed on 25 different physical commodities, across five different industrial sectors, to check for the robustness of this modeling.

The rest of the paper is as follows. Section 2 discusses the details of the futures contracts including the underlying commodities, the markets they are traded in, and the features of the data series themselves including summary statistics. Section 3 shows that the linear causal ARMA models with Gaussian innovations are unable to adequately capture the structure of this commodity data. Section 4 introduces the theory

of mixed causal/noncausal processes, and discusses the special case of the noncausal Cauchy autoregressive process of order 1. This section also demonstrates how the mixed causal/noncausal process can accommodate both asymmetries and bubble type features. Section 5 then introduces the mixed causal/noncausal autoregressive model of order (r, s) and discusses its estimation by approximated maximum likelihood. Section 5.2 then details the results of estimating the mixed causal/noncausal autoregressive model to the commodity futures price level data. Section 6 then compares the estimated unconditional distributions of both the purely causal and mixed models according to the Kullback-Leibler measure. Section 7 then considers the appropriate method for forecasting the mixed causal/noncausal model given data on the past values of the process and applies this method to forecast the futures data. Finally, the technical proofs and the other material related to the data series are gathered in the appendices.

2 Description of the asset and data

2.1 The forward contract

A forward contract on a commodity is a contract to trade, at a future date, a given quantity of the underlying good at a price fixed in advance. Such a forward contract will stipulate:

- The names of those entering into the contract, i.e. the buyers and sellers.
- The date at which the contract is entered into at some time t.
- The date at which the contract matures at some future time $t + h$.

- The forward delivery price $f_{t,t+h}$, negotiated and set in the contract at time t to be paid at the future time $t + h$.
- The monetary denomination of the contract.
- The characteristics and quality of the underlying good, often categorized by pre-specified "grades."
- The amount and units of the underlying good; typically commodity contracts will stipulate a number of predefined base units e.g. 40,000 lbs of lean hogs.
- Whether the good is to be delivered to the buyers upon maturity at time $t + h$ (otherwise the buyer will have to pick up the good themselves).
- It will also specify the location of delivery if applicable and the condition in which the good should be received.

Historically, such forward contracts were introduced to serve an economic need for producers or consumers to be able to hedge against the risk of price fluctuations in which they sell or purchase their products. For example, a producer of wheat might be subject to future supply and demand conditions that are unpredictable. As such a risk adverse producer would enter into a forward contract which would ensure a stable price at a certain date in the future for their products. Therefore, despite whether the price of their product rises or falls they can be certain of receiving the forward price. As another example, consider the consumer's side of the problem, where an airline company wishes to guarantee a stable future price for inputs, e.g. jet fuel, in order to provide customers with relatively unchanging prices of their outputs i.e. airline tickets.

Such traditional forward contracts still exist as bilateral agreements between two parties, sold on so called "over the counter" (OTC) markets. These contracts still fulfill

an important role for certain groups, for example large organizations such as national governments since the parties involved are unlikely to default on their end of the contract. However, if the investor is not sure of the financial integrity of the opposite party, such a forward contract is by construction subject to counterparty risk. Therefore, as opposed to nations which have the power to recover from counterparty loses and are self insured, contracts catering to other types of investors must somehow incorporate an insurance scheme into the contract itself to accommodate counterparty risk.

Counterparty risk presents itself as the forward contract approaches maturity since if the forward price is below (resp. above) the spot price, $f_{t,t+h} < (\text{resp. } >) p_{t+h}$, then the contract is profitable only to the buyer (resp. seller), except if the seller (resp. buyer) defaults.

2.2 The futures contract

A futures contract on a commodity is a forward contract, but with an underlying insurance in place against possible counterparty risk. The insurance is paid by means of insurance premia, called "margin" on the futures markets. There is an initial premium or initial margin, and intermediary premia, or "margin calls."

Therefore a futures contract with delivery contains the same information and contractual stipulations as the forward contract. It still represents an agreement to either buy or sell some underlying good at a future date, given a predetermined "futures price" $F_{t,t+h}$ set at time t today. However, in addition it will also specify

- A margin call scheme which:
  - Stipulates the *initial margin*; that is the amount the trader must first put up as collateral to enter into futures contracts.

◇ Implements a mechanism whereby the margin account balance is maintained a certain level sufficient to cover potential losses. If the margin account balance drops below a threshold amount, the trader is obliged to put up more collateral, known as the *margin call*.

Generally, the price of a futures contract with delivery, $F_{t,t+h}$, differs from the price of a similar forward contract $f_{t,t+h}$, since it must account for the price of the underlying insurance against counterparty risks.

A futures contract requires the presence of an "insurance provider" usually either a broker, or a clearing house. This provider will fix the margin rules for both the buyer and seller and manage a reserve account to be able to hedge the counterparty risks in case of default of either party unable to fulfill margin calls.[3]

Of course, the clearing house plays a second very important role: namely that of "clearing the market" by trying to match demand and supply between buyers and sellers of contracts. As a consequence, the clearing house facilitates the formation of futures prices $F_{t,t+h}$ as equilibrium prices. Therefore, we must distinguish between brokers themselves who act as intermediaries, and the clearing house and brokering platforms which also serve a more central purpose.

Finally, if the date and magnitude of the margin calls were known at the date of the futures' contract issue, the contract with delivery would simply reflect a portfolio (or sequence) of forward contracts which are renewed each day [Black (1976)]. However, the margin calls are fixed by the brokers or the clearing house according to the evolution of the risk, i.e. to the observed evolution of the spot prices, but also to the margin rules

[3]There also exists a counterparty risk of the insurance provider itself. For instance, in 1987 the clearing house for commodity futures in Hong Kong defaulted. This "double default" counterparty risk is not considered in our analysis.

followed by their competitors and so the interpretation as a portfolio of forwards is no longer valid.

2.3 The futures contract without delivery

In the market for futures with delivery, historically some intermediaries or investors have demonstrated that they are not on the market simply to buy or sell physical goods for future delivery and that they do not actually take delivery of the underlying physical good. Rather these investors are on the market simply to speculate on the future price of the contract.

Given this trend, futures contracts without delivery have been introduced where instead of taking delivery of the commodity they receive cash. When you do not have delivery of a physical good, the derivative product becomes a purely "financial" asset. Therefore there has been an attempt to separate these two types of instruments: a financial market designed purely for speculative purposes and a "real" market that provides a mechanism for both producers and consumers to hedge against the risk of price fluctuations.

This trend towards differentiation of futures with and without delivery was designed to suppress the effect that speculation may have on the spot price of the underlying good. For example, traders who are in a loss position may be unable to offset their positions rapidly enough as maturity of the futures contract with delivery approaches. Given this situation they are forced to purchase or sell the underlying good in the spot markets in order to meet their contractual obligation. If many traders are in this situation simultaneously and on the same side of the market, the effect could have a dramatic impact on the spot price.

2.4 Organization of the markets

In recent years, the futures commodity markets have become more organized. There is standardization of the financial products and the margin rules. For example the Standard Portfolio Analysis of Risk (SPAN) system has become common place as an instrument to determine the margin levels (both the clearing houses associated with the Chicago Mercantile Exchange (CME) and Intercontinental Exchange (ICE) have adopted its use). The system represents a computational algorithm which determines each trading day the risk for each commodity future by scanning over sixteen different possible price and volatility scenarios given the time to maturity of the contract. The sixteen scenarios consider various possible gains or losses for each futures contract, with each gain or loss classification representing a certain fraction of the margin ratio.[4] The results of these tests are used to define the appropriate margin call requirements for the different participants. Even if the SPAN methodology is a standard one, the choice of the risk scenarios depends on the clearing house. Finally, the SPAN system is not perfect and is likely to be modified in the near future. See for example, the "CoMargin" framework discussed in Cruz Lopez et al. (2013).

Interestingly, the OTC forward markets are slowly becoming more organized like the futures markets. For example the European Market Infrastructure Regulation (EMIR) that entered into force on August 16, 2012, was designed to promote the trading of standardized forward contracts on exchanges or electronic trading platforms which are cleared by central counterparties and non-centrally cleared contracts should be subject to higher capital requirements. Generally there is concern that the clearing houses need to play a larger role in their function of mitigating counterparty risk, especially as it

[4] See `https://www.theice.com/publicdocs/clear_us/SPAN_Explanation.pdf` available on the ICE exchange website.

pertains to large valued contracts which could effect the economic base if they were left to default.[5]

2.5 Example of a futures contract

Figure 2 provides an example of a set of futures contracts with delivery written on coffee and traded on the ICE exchange.[6] There are different contracts available for different maturities, which are listed on the far left column. Coffee production generally occurs in both the northern and southern hemispheres – there is a northern harvest taking place between October and January and a southern harvest between May and September. Given these differing harvests, coffee futures mature every two months from March to September and every three months onward until the following March. Furthermore, there exist contracts currently available for purchase that mature quite far into the future. For example, the coffee future contract currently with the longest time to maturity is the contract for March 2016 delivery.

The date this chart was accessed is also given as September 19th, 2013. Therefore, when we speak of the futures price $F_{t,t+h}$, within the context of our model with daily data (see the data section below) the time t would be the current date given above, and the period h would represent the number of trading days until the contract matures. Such contracts with delivery stipulate a *last trading day* which is typically the last business day prior to the 15th day of the given contract's maturity month. For instance, given the December 2013 contract, the last business day before December 15th will fall on Friday

[5]However, having the clearing house play a more predominant role also raises concerns over systemic risk – that is, could clearing houses themselves become "too big to fail" institutions? See the H. Plumridge (December 2nd, 2011) , "What if a clearing house failed?," *Wall Street Journal,* accessed Sept. 20, 2013 at `http://online.wsj.com/article/SB10001424052970204397704577074023939710652.html`.

[6]The chart is provided by TradingCharts.com at `http://tfc-charts.w2d.com/marketquotes/KC.html`.

December 13th, 2013 (resp. Friday March 14th, 2014; Thursday May 15th, 2014; etc; for the subsequent contracts).

The "open," "high," "low," and "last," describe the intraday trading activity of the current trading session; that is, the opening price, the highest and lowest prices, and the last price paid, respectively. The table also displays the last change in price, the current volume of trades, and the set price and open interest from the last trading session of the prior day. "Open interest" (also known as open contracts or open commitments) refers to the total number of contracts that have not yet been settled (or "liquidated") in the immediately previous time period, either by an offsetting contractual transaction or by delivery. Therefore, a larger open interest can complement the volume measure in interpreting the level of liquidity in the market. As contracts approach maturity, both the volume and open interest levels tend to rise; contracts with very distant times to maturity are not very liquid.

Figure 2: Coffee futures contracts, ICE exchange

Commodity Futures Price Quotes For
Coffee (ICE Futures)
(Price quotes for ICE Futures Coffee delayed at least 10 minutes as per exchange requirements)

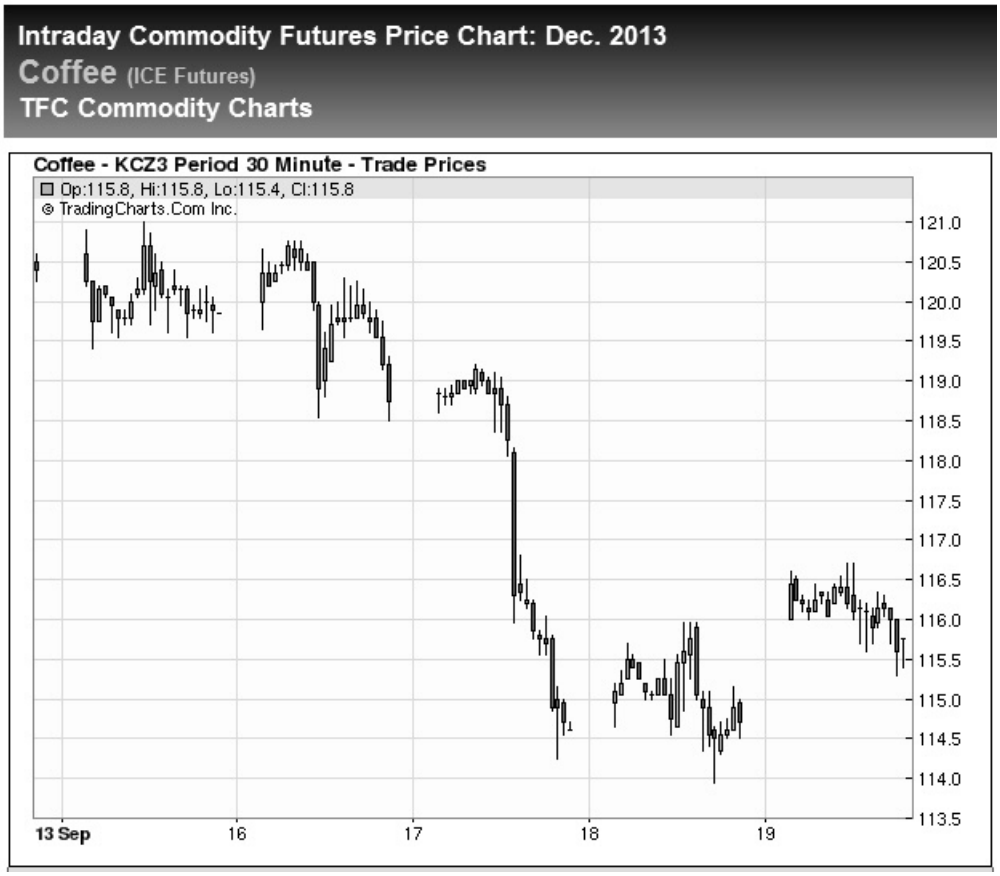
REFRESH DATA

Click for Chart	Current Session								Prior Day		Opt's
	Open	High	Low	Last	Time	Set	Chg	Vol	Set	Op Int	
Dec'13	116.00	116.70	115.60	115.90	12:06 Sep 19	-	1.00	11128	114.90	103661	Call Put
Mar'14	119.00	119.70	118.70	119.05	12:06 Sep 19	-	1.10	3136	117.95	27679	Call Put
May'14	120.80	121.75	120.70	121.20	12:06 Sep 19	-	1.20	1731	120.00	11070	Call Put
Jul'14	123.10	123.70	122.75	123.00	12:06 Sep 19	-	1.00	597	122.00	6804	Call Put
Sep'14	124.95	125.55	124.75	125.20	12:06 Sep 19	-	1.25	238	123.95	2908	Call Put
Dec'14	127.55	128.00	127.30	127.65	12:06 Sep 19	-	1.05	77	126.60	3311	Call Put
Mar'15	130.00	130.40	129.70	130.05	12:06 Sep 19	-	0.95	17	129.10	599	Call Put
May'15	132.10	132.10	131.25	131.55	12:04 Sep 19	-	0.70	7	130.85	108	Call Put
Jul'15	133.80	133.80	132.90	133.15	12:04 Sep 19	-	0.60	5	132.55	92	Call Put
Sep'15	135.35	135.35	134.35	134.60	12:04 Sep 19	-	0.45	18	134.15	173	Call Put
Dec'15	137.15	137.15	135.95	136.25	12:04 Sep 19	-	0.20	18	136.05	471	Call Put
Mar'16	-	-	-	137.70 *	12:04 Sep 19	-	-	-	137.70	-	Call Put

Figure 3 provides a candlestick plot of the typical intraday trading activity between September 13th, 2013, and September 19th, 2013, for the coffee future contract with delivery in December 2013. Note that trading does not occur 24 hours a day (rather the trading day takes place between 8:30AM-7:00 PM BST[7]) and so there are discontinuities in the price series. The thin top and bottom sections of the candlestick, called the *shadows*, represent the high and low prices, and the thick section called the *real body*, denotes the opening and closing prices. Each candlestick describes trading activity over a 30 minute period.[8]

[7]British Summer Time as the ICE exchange is located in London, England.
[8]There are 21 candlesticks each day, representing the 10.5 opening hours.

Figure 3: Coffee futures with delivery in December 2013, ICE exchange, intraday price $ US



2.6 Data on the commodity futures contracts

2.6.1 The continuous contract

The discussion above illustrates some of the difficulties in analyzing price data for derivative products. For example, many of the products are very thinly traded with low liquidity. Moreover, some products may only be available on one trading platform and not another. For example, many futures contracts with delivery are available mutually exclusively either on CME, or the ICE, and their associated clearing houses do

not necessarily follow identical margin schemes. Also, OTC product data may only be available through certain brokers proprietary trading platforms.

Perhaps the most consequential problem we face in attempting to analyze futures contracts data is that the individual contracts of various maturities will eventually expire and so we need a method whereby we can "extend" the futures price series indefinitely. However, even in accomplishing this task we must consider that the contracts of various maturities, while written on the same underlying good are not quite the same "asset" and so the asset itself is changing over time. Therefore, we need some method to, not only extend the series, but to standardize the price measurements across time and maturity, and ensure that when we construct the series we are taking prices which are relevant, e.g. with sufficient liquidity to be appropriately representative, deriving in essence a new asset that no longer matures. In doing so we would also like to be able to bring together information on prices available from different trading platforms in one place.

The Bloomberg console offers a solution to this problem by amalgamating futures data for delivery from both the ICE and CME exchanges into one system. Bloomberg also offers what is called called a *continuous contract* which mimics the behaviour of a typical trader who is said to "roll over" the futures contract as it approaches maturity. "Rolling over" refers to the situation where a trader would close out, or "zero," their account balance upon the approach of a futures contract's maturity, if they do not intend on taking delivery, by first purchasing an offsetting futures contract and then simultaneously reinvesting in another future with a further expiration month. In this way, an artificial asset is created which tracks this representative trader's futures account holdings across time indefinitely. Details on how this is accomplished, as well as other methods that can be employed, are outlined in Appendix 10. Users of the Bloomberg console can customize criteria which define the rollover strategy, e.g. volume of trades

or open interest; in this paper I choose to employ the continuous contract that mimics the rolling over of the futures contract with the shortest time to maturity known as the "front month" contract.

2.6.2 Industry sectors

I will consider a number of physical commodity futures contracts for a broad range of products. The commodities are divided into various industry sectors that are expected to behave similarly to each other. The industry sectors are given in **Table 1**.

Table 1: Commodity sectors

Energy	Metals	Softs	Soy	Livestock
Brent crude oil	Copper	Corn	Soybeans	Lean hogs
Light crude oil	Gold	Rice	Soybean meal	Live cattle
Heating oil	Palladium	Wheat	Soybean oil	
Natural gas	Platinum	Sugar		
Gas oil	Silver	Orange juice		
Gasoline RBOB		Cocoa		
		Coffee		
		Cotton		
		Lumber		

Within each futures contract itself there are specified a number of different product grades. At the exchange level it is determined that any products which match pre-specified grade criteria are considered part of the same futures contract. This is to promote standardization of contracts and volume of trades. For example, the coffee future discussed above is specified on the ICE exchange as the "Coffee C" future with exchange code KC. This future allows a number of grades and a "Notice of Certification" is issued based on testing the grade of the beans and by cup testing for flavor. The Exchange uses certain coffees to establish the "basis". Coffees judged better are at

a premium; those judged inferior are at a discount. Moreover, these grades are established within a framework of deliverable products, for example from the ICE product guide for this KC commodity future we have that "Mexico, Salvador, Guatemala, Costa Rica, Nicaragua, Kenya, New Guinea, Panama, Tanzania, Uganda, Honduras, and Peru all at par, Colombia at 200 point premium, Burundi, Venezuela and India at 100 point discount, Rwanda at 300 point discount, and Dominican Republic and Ecuador at 400 point discount. Effective with the March 2013 delivery, the discount for Rwanda will become 100 points, and Brazil will be deliverable at a discount of 900 points."

2.6.3 Energy

Brent crude oil is a class of sweet light crude oil (a "sweet" crude is classified as containing less than 0.42% sulfur, otherwise it is known as "sour"). The term "light" crude oil characterizes how light or heavy a petroleum liquid is compared to water. The standard measure of "lightness" is the American Petroleum Institute's API gravity measure. The New York Mercantile Exchange (NYMEX) defines U.S. light crude oil as having an API measure between 37 (840 kg/m^3) and 42 (816 kg/m^3) and foreign as having between 32 (865 kg/m^3) and 42 API.

Therefore, various grades are defined in the standardized contract. Both foreign and domestic light crude oil products are required to admit various characteristics based on sulfur levels, API gravity, viscosity, Reid vapor pressure, pour point, and basic sediments or impurities. Exact grade specifications are available in the CME Group handbook, Chapter 200, 200101.A and B.

The price of Brent crude is used as a benchmark for most Atlantic basin crude oils, although Brent itself derives from North Sea offshore production. Other important benchmarks also include North America's West Texas Intermediate and the middle

east UAE Dubai Crude which together track the world's internationally traded crude oil supplies. The representative light crude oil future employed in this paper is written on West Texas Intermediate and exchanged by the CME Group. The delivery point for (WTI) light crude oil is Cushing, Oklahoma, U.S., which is also accessible to the international spot markets via pipelines. Likewise, the Brent crude oil future is exchanged by ICE and admits delivery at Sullom Voe, an island north of Scotland.

Heating oil is a low viscosity, liquid petroleum product used as a fuel for furnaces or boilers in both residential and commercial buildings. Heating oil contracts take delivery in New York Harbor. Just as in crude oil contracts, very detailed stipulations exist regarding product quality grades; see the CME handbook, Chapter 150, 150101. Natural gas is a hydrocarbon gas mixture consisting primarily of methane, used as an important energy source in generating both heating and electricity. It is also used as a fuel for vehicles and is employed in both the production of plastics and other organic chemicals. Natural gas admits delivery at the Henry Hub, a distribution hub on the natural gas pipeline system in Erath, Louisiana, U.S. Contract details are available in the CME handbook, Chapter 220, 220101. Gas oil (as it is known in Northern Europe) is Diesel fuel. Diesel fuel is very similar in its physical properties to heating oil, although it has commonly been associated with combustion in Diesel engines. Gas oil admits delivery in the Amsterdam-Rotterdam-Antwerp (ARA) area of the Netherlands and Belgium. Contract grade specifications are available from the exchange, ICE.

The Gasoline RBOB classification stands for Reformulated Blendstock for Oxygenate Blending. RBOB is the base gasoline mixture produced by refiners or blenders that is shipped to terminals, where ethanol is then added to create the finished ethanol-blended reformulated gasoline (RFG). Gasoline RBOB admits delivery in New York Harbor and quality grade details are outlined in the CME handbook, Chapter 191,

191101.

2.6.4 Metals

Gold and silver, have both traditionally been highly sought after precious metals for use in coinage, jewelry, and other applications since before the beginning of recorded history. Both also have important applications in electronics engineering and medicine. The CME exchange licenses storage facilities located within a 150 mile radius of New York city, in which gold or silver may be stored for delivery on exchange contracts. The quality grades for gold and silver are defined in the CME handbook, Chapters 113 and 112, respectively.

Platinum, while also considered a precious metal, also plays an important role, along with the metal Palladium in the construction of catalytic converters. Catalytic converters are used in the exhaust systems of combustion engines to render output gases less harmful to the environment. Palladium also plays a key role in the construction of hydrogen fuel cells. Finally, copper is a common element used extensively in electrical cabling given its good conductivity properties. Platinum, Palladium, and Copper offer a number of delivery options, including delivery to warehouses in Zurich, Switzerland. See the CME handbook Chapters 105, 106 and 111 respectively.

2.6.5 Softs and Livestock

"Soft goods" are typically considered those that are either perishable or grown in an organic manner as opposed to "hard goods" like metals which are extracted from the earth through mining techniques.

In the grains category we have corn, rice, and wheat which are all considered "cereal grains"; that is, they represent grasses from which the seeds can be harvested as food.

Sugar, derived from sugarcane, is also a grass but the sugar is derived not from the seeds but from inside the stalks. Corn, rice, and wheat all admit a number of standardized delivery points within the U.S. See the CME handbook chapters 10, 14, and 17 for grade specifications and delivery options. Sugar delivery point options and grade details are available online from ICE, under the Sugar No.11 contract specification.

Orange juice is derived from oranges which grow as the fruit of citrus tree, typically flourishing in tropical to subtropical climates. The juice traded is in frozen concentrated form. Orange juice is deliverable to a number of points in the U.S., including California, Delaware, Florida, and New Jersey warehouses. See the ICE FCOJ Rulebook available online for further information and quality grade details. Coffee is derived from the seeds of the coffea plant, referred to commonly as coffee "beans." Cocoa represents the dried and fully fermented fatty seeds contained in the fruit of the cocoa tree. Finally, cotton is a fluffy fibre that grows around the seeds of the cotton plant. Delivery point information and quality grade details for Coffee, Cocoa, and Cotton are also available via the ICE Rulebook chapters available online.

In the soy category we have soybeans, a species of legume widely grown for its edible beans; soybean meal which represents a fat-free, cheap source of protein for animal feed and many other pre-packaged meals; and finally, soybean oil is derived from the seeds of the soy plant and represents one of the most widely consumed cooking oils. All three soybean products admit a number of standardized delivery points within the U.S. See the CME handbook chapters 11, 12, and 13 for grade specifications and delivery options.

Lean hogs refers to a common type of pork hog carcass used typically for consumption. A lean hog is considered to be 51-52% lean, with 0.80-0.99 inches of back fat at the last rib, with a 170-191 lbs. dressed weight (both "barrow" and "gilt" carcasses).

Live cattle are considered 55% choice, 45% select, yield grade 3 live steers (a castrated male cow). Finally, lumber is traded as random length 2×4's between 8-20 feet long. Lean hogs futures are not delivered but are cash settled based on the CME Lean Hog Index price. Cattle is to be delivered to the buyer's holding pen. Lumber shall be delivered on rail track to the buyer's producing mill. See CME handbook Chapters 152, 101, and 201, respectively for details.

2.6.6 Data sources

The following **Table 2** outlines the dates for which there exists data for each commodity futures price series, the time to maturity, currency denomination, commodity exchange and code, and basic unit/characteristics of the product traded.

Table 2: Commodity specifications

Commodity	Start date	CEM	Currency unit	Exchange	Code	Basic unit
Soybean meal	7/18/1977	FHKNZ	U.S.$/st	CME	ZM/SM	100 st's
Soybean oil	7/18/1977	FHKNZ	U.S.$/100lbs	CME	ZL/BO	60,000 lbs
Soybeans	7/18/1977	FHKNX	U.S.$/100bushel	CME	ZS/S	5,000 bushels
Orange juice	7/18/1977	FHKNUX	U.S.$/100lbs	ICE	OJ	15,000 lbs
Sugar	7/18/1977	HKNV	U.S.$/100lbs	ICE	SB	112,000 lbs
Wheat	7/18/1977	HKNUZ	U.S.$/100bushel	CME	ZW/W	5,000 bushels
Cocoa	7/18/1977	HKNUZ	U.S.$/MT	ICE	CC	10 MT
Coffee	7/18/1977	HKNUZ	U.S.$/100lbs	ICE	KC	37,500 lbs
Corn	7/18/1977	HKNUZ	U.S.$/100bushel	CME	CZ/C	5,000 bushels
Cotton	7/18/1977	HKNZ	U.S.$/100lbs	ICE	CT	50,000 lbs
Rice	12/6/1988	FHKNUX	U.S.$/100hw	CME	ZR/RR	2,000 hw
Lumber	4/7/1986	FHKNUX	U.S.$/mbf	CME	LBS/LB	110 mbf
Gold	7/18/1977	GMQZ	U.S.$/oz	CME	GC	100 troy oz
Silver	7/18/1977	HKNUZ	U.S.$/100oz	CME	SI	5,000 troy oz
Platinum	4/1/1986	FJNV	U.S.$/oz	CME	PL	50 troy oz
Palladium	4/1/1986	HMUZ	U.S.$/oz	CME	PA	100 troy oz
Copper	12/6/1988	HKNUZ	U.S.$/100lbs	CME	HG	25,000 lbs
Light crude oil	3/30/1983	All	U.S.$/barrel	CME	CL	1,000 barrels
Heating oil	7/1/1986	All	U.S.$/gallon	CME	HO	42,000 gallons
Brent crude oil	6/23/1988	All	U.S.$/barrel	ICE	CO	1,000 barrels
Gas oil	7/3/1989	All	U.S.$/MT	ICE	QS?	100 MT
Natural gas	4/3/1990	All	U.S.$/mmBtu	CME	NG	10,000 mmBtu
Gasoline RBOB	10/4/2005	All	U.S.$/gallon	ICE	HO	42,000 gallons
Live cattle	7/18/1977	GJMQVZ	U.S.$/100lbs	CME	LE/LC	40,000 lbs
Lean hogs	4/1/1986	GJMQVZ	U.S.$/100lbs	CME	HE/LH	40,000 lbs

The units are described as follows. A barrel is considered to be 42 U.S. gallons. An mmBtu is one million British Thermal Units, a traditional unit of energy equal to about 1055 joules per Btu. An MT is one metric tonne, which is a unit of mass approximately equal to 1,000 kilograms. Lbs and oz are the abbreviations for pounds and ounces, respectively. A "Troy oz" is a slightly modified system whereby one troy oz is equal to approximately 1.09714 standard oz. A bushel is a customary unit of dry volume, equivalent to 8 gallons. An mbf is a specialized unit of measure for the volume of lumber in the U.S, called a "board-foot." A board-foot (or "bf") is the volume of a one-foot length of a wooden board, one foot wide and one inch thick. Therefore an mbf is one million such board-feet. Finally, an "st" or short tonne is a unit of mass smaller than the metric tonne, equivalent to approximately 907 kilograms.

The column CEM represents the range of "contract ending months" that each futures contract may be specified for. The month codes are as follows: F - January, G - February, H - March, J - April, K - May, M - June, N - July, Q - August, U - September, V - October, X - November, and Z - December. These are the standard codes employed by the exchanges.

All series end on February 8th, 2013, and represent daily closing prices for those days the commodities are traded on the exchange. In June 2007 the CBOT (Chicago Board of Trade) which acted as the exchange for soy products, wheat corn, and rice, merged with the CME (Chicago Mercantile Exchange) to form the CME Group. Moreover, most of the energy futures were originally traded on the NYMEX (New York Mercantile Exchange) and the metals were traded on the COMEX (Commodity Exchange; a division of the NYMEX). However, on August 18, 2008, the NYMEX (along with the COMEX) also merged with the CME Group. Gas oil was originally traded on the IPE (International Petroleum Exchange) which was acquired by ICE (IntercontinentalEx-

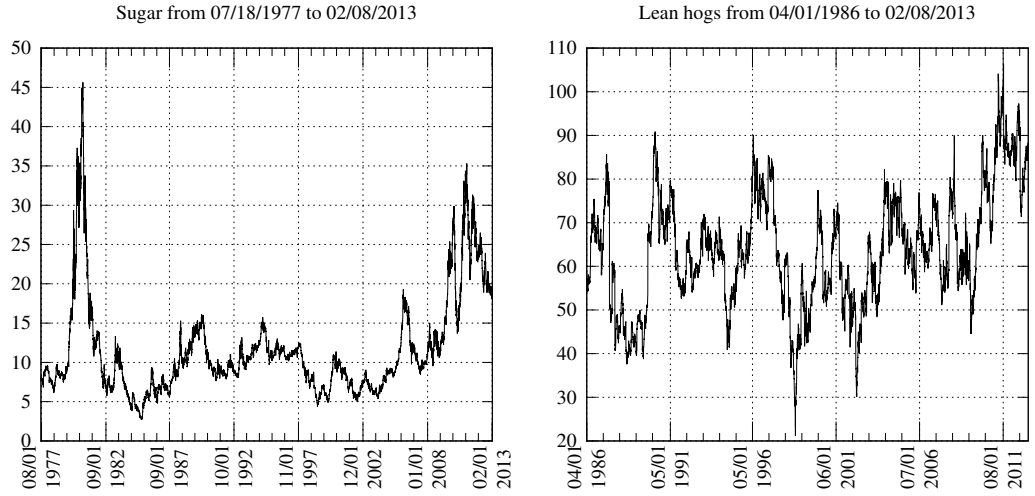
change) in 2001. Therefore, care must be taken in interpreting the various exchange codes which have changed over time.

For most CME contracts, the last trading day is typically the 15th business day before the first day of the contract month. The delivery date is then freely chosen as any day during the contract month.

2.7 Features of the price level series

When dealing with financial data we typically consider the continuously compounded returns series, $r_t = ln(P_t/P_{t-1})$, since the price level process is nonstationary and so we are obliged to transform the initial price data. However, in the case of futures price data without delivery, an examination of the time evolution of the price level processes does not necessarily suggest the presence of trends, either of the stochastic type (i.e. random walk), or due to a deterministic increase or decrease.

Figure 4: Plots of daily continuous contract futures price level series, Sugar and Lean hogs



For example, let us consider the two plots in **Figure 4**, that display the time evo-

lution of the futures prices of sugar and lean hogs. Both series do not exhibit obvious deterministic time trends and their dramatic bubbles (especially in sugar) suggest that they cannot have been generated by a random walk. Interestingly, lean hogs exhibits the well known "pork cycle," or cyclical patterns related to pork production.

The price level series all exhibit a very high level of linear persistence in the sense that their estimated autocorrelation function, $\hat{\rho}(s)$, are all $\hat{\rho}(1) \approx 1$ with small, but significant, $\hat{\rho}(s)$ for some $s > 1$ (see **Table 3** for the autocorrelation at lag 1). Moreover, their normalized spectral densities exhibit extremely sharp peaks at the zero frequency and are near zero elsewhere in the spectrum. Of course, this is suggestive of a unit root process, however, augmented Dickey-Fuller unit root tests of the series are inconclusive in rejecting the null of a unit root (including a constant, but no time trend).[9]

This is unsurprising given what we know about the properties of some exotic parametric processes which are able to elude detection by traditional unit root testing (see for example the causal representation of the noncausal AR(1) model with i.i.d. Cauchy innovations discussed later in Section 4.2). A linear unit root test is not of much use if the causal representation of the process may be nonlinear and strictly stationary, with moments that do not exist. Finally, linear unit root tests have been shown to have low power in the presence of nonlinearity (such as multiple regimes, for example).

Since all continuous contract futures series are constructed through the "rolling over" mechanism, they reflect the price of a reconstituted futures contract in which the time to maturity, h, remains fixed throughout the time evolution of the price level, despite the fact that the reconstitution is generated from individual contracts of different maturities each representing daily closing prices for those days these futures contracts are traded on the exchange. The different starting dates for each of the series are given

[9]The estimated spectral density and Dickey-Fuller test results are available upon request.

in **Table 2** and all the continuous contract series end on end on February 8th, 2013.

Summary statistics for the price levels series are given in **Table 3** and plots and histograms of all the price level series are available in Appendix 14 (**Figures 10.i to 11.iv**).

Note some of the salient features from the summary statistics in **Table 3**. If we are to interpret the series as strictly stationary, the sample moments suggest highly leptokurtic unconditional distributions for most of the series. Exceptions to this exist, however, in orange juice, lumber, platinum, copper, gasoline RBOB, and lean hogs. Perhaps more importantly we should consider that most of the series are also positively skewed, again with a few exceptions in gasoline RBOB and lean hogs (and possibly orange juice). Visual examination of the histograms in Appendix 14 corroborate these statistics. Moreover, some of the histograms indicate a bimodal structure, especially among those series that are highly skewed, suggesting the possibility of a mixture between low price and high price regimes. A good example of this is the copper series.

Table 3: Summary statistics - commodity futures price level series

Levels / Series	Quantiles 10%	50%	90%	Mean	Stnd. Dev.	Skewness	Kurtosis	ACF(1)	Sample size
Soybean meal	149.600	185.800	314.200	210.347	70.151	1.729	6.190	0.998	9280
Soybean oil	16.640	23.750	39.993	26.399	10.449	1.709	5.516	0.999	9280
Soybeans	503.750	629.000	1057.600	716.563	249.577	1.755	5.735	0.998	9280
Orange juice	79.250	115.125	170.350	118.926	33.531	0.592	2.663	0.998	9280
Sugar	6.040	9.830	20.503	11.586	6.343	1.946	7.283	0.998	9280
Wheat	267.250	357.500	622.750	401.672	151.036	1.878	6.656	0.998	9280
Cocoa	991.000	1621.000	2971.100	1835.268	744.051	0.926	3.466	0.997	9280
Coffee	64.700	124.450	192.000	126.325	48.051	0.699	3.495	0.997	9280
Corn	203.750	258.250	435.000	298.578	126.933	2.097	7.126	0.998	9280
Cotton	49.059	65.150	85.720	67.665	19.798	2.688	16.481	0.997	9280
Rice	5.360	8.440	14.601	9.243	3.557	0.844	3.503	0.999	6309
Lumber	181.700	261.700	366.920	267.773	70.562	0.463	2.458	0.996	7005
Gold	277.700	385.400	964.230	510.664	351.245	2.202	7.139	0.999	9280
Silver	4.400	6.037	18.050	9.406	7.680	2.272	7.910	0.998	9280
Platinum	367.200	534.000	1555.420	755.715	463.352	1.169	3.096	0.999	7009
Palladium	111.000	206.150	645.140	286.657	203.778	1.303	3.935	0.999	7009
Copper	74.000	115.400	358.860	168.275	111.428	1.060	2.562	0.999	6309
Light crude oil	16.400	26.740	85.712	38.103	27.475	1.371	3.827	0.999	7793
Heating oil	45.733	67.655	264.865	112.316	86.145	1.292	3.484	0.999	6944
Brent crude oil	15.796	25.410	100.128	41.547	32.501	1.205	3.199	0.999	6427
Gas oil	147.000	226.500	894.875	375.818	281.273	1.161	3.180	0.999	6160
Natural gas	1.631	3.142	7.366	3.987	2.478	1.370	4.950	0.998	5964
Gasoline RBOB	153.220	223.895	304.360	227.116	57.877	0.023	2.309	0.995	1920
Live cattle	60.500	71.488	95.100	75.023	15.871	1.219	4.915	0.998	9280
Lean hogs	46.550	63.345	81.380	63.726	13.133	0.165	2.830	0.995	7009

* Note that ACF(1) represents the autocorrelation function at lag 1 and T is the sample size. Also the kurtosis measure employed here is *not* the excess kurtosis.

3 The linear causal ARMA model

In this section we show that the causal linear ARMA model, with Gaussian innovations, is unable to adequately capture the features of the futures price level data.

In order to assess the ARMA model's ability to fit the price level data, I estimate a number of different ARMA(p, q) specifications and choose among the best fitting according to the Akaike information criteria (AIC). The software used to estimate the ARMA model is the popular "R project for statistical computing" available for download at `http://www.r-project.org/`. In order to facilitate the (p, q) parameter search we employ the *auto.arima()* function in the R *forecast* package due to Hyndman and Khandakar (2008). Given computational constraints, maximum orders of $p + q = 13$, $p \leq 10$ and $q \leq 3$ are chosen. AIC's are specified not to be approximated and the "stepwise" selection procedure is avoided to make sure all possible model combinations are tested.

The *arima()* routine called by *auto.arima()* obtains reasonable starting parameter values by conditional sum of squares and then the parameter space is more thoroughly searched via a Nelder and Mead (1965) type algorithm. The pseudo-likelihood function is computed via a state-space representation of the ARIMA process, and the innovations and their variance found by a Kalman filter. Since the assumption of Gaussian shocks may be misspecified, robust sandwich estimator standard errors are employed of the type introduced by White (1980).

If the ARMA model captures the nonlinear features of the data, the residuals (e_t) should be approximately representative of a strong white noise series. Therefore, we test for this feature in two ways: 1) we employ the Ljung-Box test with the null of weak white noise residuals [Ljung and Box (1978)] and 2) the BDS test with the null of

independent residuals [Brock, Dechert and Scheinkman, and LeBaron (1996)].

3.1 Test specifications

The Ljung-box test statistic is given as

$$LB(S) = T \sum_{s=1}^{S} \frac{T+2}{T-s} \hat{\varrho}(s)^2, \tag{1}$$

where $\varrho(s)$ is the estimated autocorrelation function of the ARMA model residuals. The null hypothesis is that the autocorrelation function of the ARMA residuals is jointly 0 up to the Sth lag. Finally, $LB(S) \sim \chi^2(S)$, if the residuals are representative of the true theoretical (ϵ_t) which is a strong white noise (and neglecting the fact that $\hat{\varrho}(s)$ is an estimated quantity itself).

The BDS test was designed to be employed on the residuals of a best fitting linear model in order to look for deterministic chaos in the residual nonlinear structure. This test involves the correlation dimension technique originally developed by Grassberger and Procaccia (1983) to detect the presence of chaotic structure by embedding overlapping subsequences of the data in *k*-space. Given a k-dimensional time series vector $\boldsymbol{x}_{t,k} = (x_t, x_{t+1}, \ldots, x_{t+k-1})'$ called the k-history, the BDS test treats this k-history as a point in a k-dimensional space. The BDS test statistic, called the *correlation integral* is given as

$$C_k(\epsilon, T) = \frac{2}{T_k(T_k-1)} \sum_{i<j} I_\epsilon(\boldsymbol{x}_{i,k}, \boldsymbol{x}_{j,k}), \quad \text{where} \quad T_k = T - k + 1, \tag{2}$$

and where $I_\epsilon(u, v)$ is an indicator variable that equals one if $\|u - v\| < \epsilon$ and zero otherwise, where $\|\cdot\|$ is the supnorm. The correlation integral estimates the fraction of

data pairs of $\boldsymbol{x}_{t,k}$ that are within ϵ distance from each other in k-space. Despite the original purpose of the test, it is effectively a test for independence since if we can reject the null hypothesis of correlation of $(x_t)_{t=1}^{T_k}$ in every k-dimensional embedding space this is equivalent to being i.i.d. That is, if the k-histories show no pattern in k-dimensional space, then we should have that $C_k(\epsilon, T) \approx C_1(\epsilon, T)^k$.

It is shown that the BDS statistic $\sqrt{T}\left[C_k(\epsilon, T) - C_1(\epsilon, T)^k\right]$ is asymptotically Normal with mean zero and finite variance under the null hypothesis [see Tsay (2010), Ch.4.2.1]. If we cannot reject the null hypothesis the alternative is quite broad since, depending on the correlation structure in the k-dimensional spaces, the nonlinearity could have come about due to either deterministic nonlinearity, i.e. chaos [see Blank (1991), Decoster et al. (1992), and Yang and Brorsen (1993)], or stochastic nonlinearity.

For the Ljung-Box test we specify the number of lags S as $S = ln(T)$ rounded to the nearest integer, where T is the sample size given in **Table 3**. According to Tsay (2010), Ch.2.2, pg.33, simulation studies suggest that this choice maximizes the power of the test. For the BDS test we consider embedding dimensions k up to $k = 15$, which trades off number of dimensions for computational efficiency.

3.2 Results

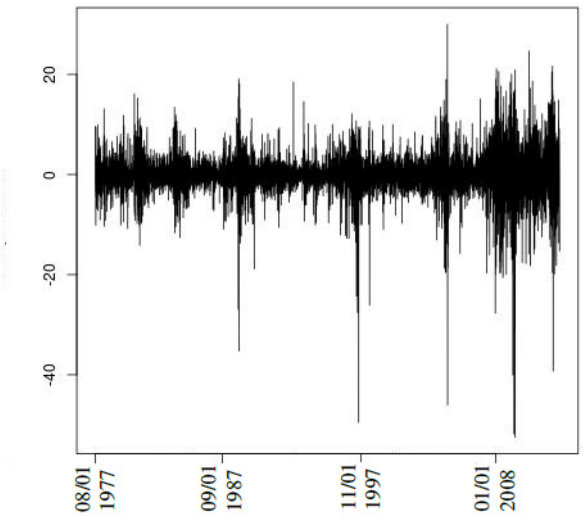
Table 4 presents estimation results for the ARMA model. Generally, for all the series, the best fitting linear ARMA model residuals reject the BDS null hypothesis of i.i.d. shocks at the 1% test significance level (in fact all of the test statistic p-values are extremely close to 0). There is one exception in the *lean hogs* price levels series, where for $\epsilon = 2.6$ (the parameter that defines "near points" in the k-dimensional space, i.e. $\|u - v\| < \epsilon$), we are not able to reject the null hypothesis of i.i.d. residuals (however,

we are able to reject for smaller $\epsilon = 1.95$). The p-values in this case decline monotonically from 0.731 at $k = 2$ down to 0.165 at $k = 15$.

Plots of all the residuals series also suggest ARCH effects (see **Figure 5** for an example). Interestingly, except in the case of coffee, the noises are still weak white according to the Ljung-Box test as we are unable to reject the null hypothesis at the 10% level, although we are able to reject platinum at the 13% level and soybean meal at the 15% level.

Interestingly, the ARMA estimation software is unable to fit an autoregressive model to the gold series, and so we skip testing its residuals for whiteness.

Figure 5: Soybean meal residuals from ARMA model



Clearly, the causal linear ARMA model is not able to fully capture the structure of the data as the residuals are weak white noise, but not i.i.d. Therefore, the evidence presented in this section suggests that we need a better model if we are to adequately capture the nonlinear dynamic features of the futures price level data.

Table 4: ARMA estimation results

Series	p	q[a]	AIC	Log-likelihood	Ljung-Box Pval[b]
Soybean meal	6	2	52395.00	-26188.50	0.15
Soybean oil	8	3	11859.05	-5917.52	0.92
Soybeans	9	2	73548.21	-36762.11	0.52
Orange juice	4	3	42121.61	-21052.80	0.40
Sugar	10	2	7842.39	-3908.20	1.00
Wheat	7	2	67069.47	-33524.74	1.00
Cocoa	8	3	94368.76	-47172.38	0.72
Coffee	4	2	48866.80	-24426.40	0.06
Corn	7	3	59385.84	-29681.92	0.63
Cotton	10	0	32760.78	-16369.39	1.00
Rice	10	3	-4799.02	2413.51	0.96
Lumber	8	3	44027.92	-22001.96	1.00
Gold	0	3	102914.50	-51453.27	n/a
Silver	9	3	7424.04	-3699.02	0.94
Platinum	8	2	55936.82	-27957.41	0.13
Palladium	9	3	48209.69	-24091.84	0.99
Copper	10	0	34719.50	-17348.75	1.00
Light crude oil	7	2	22244.11	-11112.06	0.95
Heating oil	9	2	34465.28	-17220.64	1.00
Brent crude oil	7	2	18807.92	-9393.96	0.90
Gas oil	5	3	44142.24	-22062.12	0.92
Natural gas	3	2	-4178.27	2095.13	0.23
Gasoline RBOB	5	3	11715.32	-5848.66	0.99
Live cattle	6	1	22771.40	-11377.70	0.99
Lean hogs	3	2	23567.63	-11777.81	0.70

[a] The orders of the ARMA(p, q) model are given in the first and second columns.

[b] The column denoted "Ljung-Box Pval" indicates the p-value statistic for this test – therefore we reject the null hypothesis at $x\%$ probability of committing a type I error if $Pval < x$.

4 The linear mixed causal/noncausal process

The linear mixed causal/noncausal process takes the form of a two sided infinite moving average representation,

$$Y_t = \sum_{i=-\infty}^{\infty} a_i \epsilon_{t-i}, \tag{3}$$

where (ϵ_t) is a strong white noise, that is a sequence of independently and identically distributed (i.i.d.) variables, that doesn't necessarily admit finite moments. The mixed causal/noncausal process is composed of both a purely causal component that depends only on past shocks, that is the sum of $a_i \epsilon_{t-i}$ for all $i > 0$, and a purely noncausal component that depends only on future shocks, that is the sum of $a_i \epsilon_{t-i}$ for all $i < 0$. We have a unique representation for (3), up to a scale factor on ϵ_t, except in the case where the white noise (ϵ_t) is Gaussian [see e.g. Findley (1986) and Cheng (1992)]. For Gaussian white noise, there exists an equivalent purely causal linear representation where (ϵ_t^*) is another Gaussian white noise. This implies that for non-Gaussian (ϵ_t), a mixed linear process including a noncausal component (i.e. $\exists i < 0$, $a_i \neq 0$) will necessarily admit a nonlinear causal dynamic. For more details see Appendix 11.1.

4.1 The asymmetries

As an example, let us consider the effect of shocks (ϵ_t) on the model above in (3). Let ϵ_t be distributed Cauchy which admits no first and second-order finite moments. Moreover, let $a_i = \rho_1^i$ for $i \geq 0$, $a_i = \rho_2^i$ for $i \leq 0$, and $|\rho_k| < 1$ for $k = 1, 2$, where we are free to choose ρ_1 and ρ_2 as such.

In choosing various values for ρ_k, $k = 1, 2$, we will see how the general linear causal/noncausal model is able to exhibit bubble like phenomenon with asymme-

tries of the type discussed in Ramsey and Rothman (1996) (see the Introduction, Section 1, (ii), Price dynamics). Consider the simulated sample path of the linear mixed causal/noncausal model with standard Cauchy shocks as depicted in **Figure 6**, where we have zoomed in on a bubble episode to focus on the dynamics.

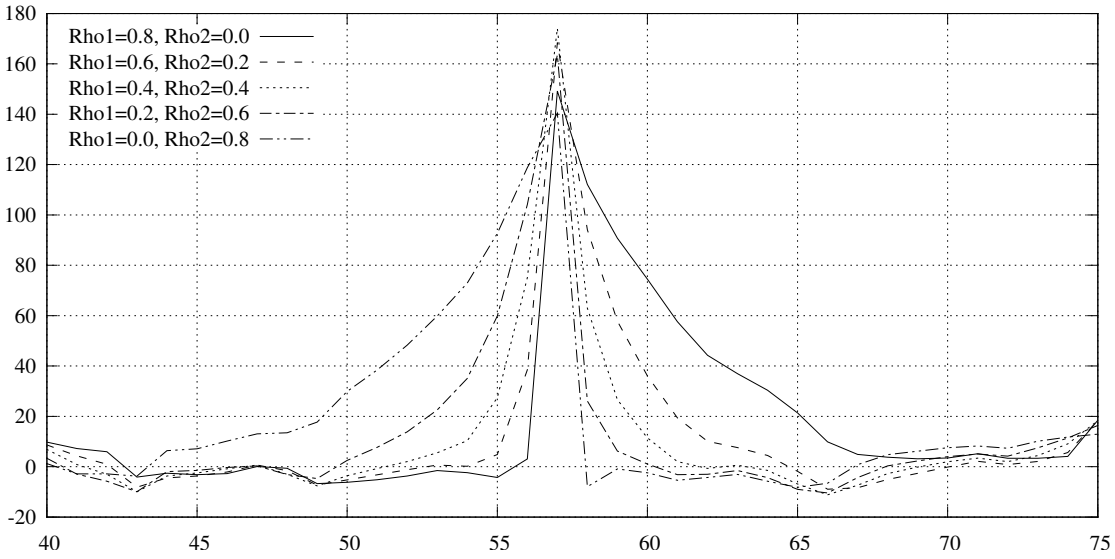
Within **Figure 6** we have an example of a positive shock $\epsilon_t > 0$ around time $t = 57$. Depending on the values chosen for ρ_1 and ρ_2, the bubble's build up and subsequent crash exhibits different rates of ascent and descent. For example, consider the parameter combination $(\rho_1 = 0.8, \rho_2 = 0)$. This represents the purely causal case where the shock occurs at time $t = 57$ and its effect dies off slowly, and so we have a quick rise and a subsequently slow decline. Also consider the opposite case where $(\rho_1 = 0, \rho_2 = 0.8)$. This is the purely noncausal case where the bubble builds up slowly until time $t = 57$ and then quickly declines. The other cases represent mixed causal/noncausal models where the bubble rises and falls at rates which depend on the ratio of $\rho_1/\rho_2 = \alpha$. If $\alpha > 1$ the bubble rises quicker than it declines; if $\alpha < 1$ then it rises slower than it declines, and if $\alpha = 1$ then it behaves symmetrically around time $t = 57$. These asymmetries can be classified within the framework of Ramsey and Rothman (1996) as being *longitudinally* asymmetric in that the probabilistic behaviour of the process is not the same in direct and reverse time.

Of course, for a negative shock $\epsilon_t < 0$ the behaviour would be duplicated, but instead we would see a crash instead of a bubble. This suggests that the mixed causal/noncausal process can also exhibit *transversal* asymmetries, that is asymmetries in the vertical plane, by modifying the distribution of the shocks. For example, if we were to only accept positive Cauchy shocks, $\epsilon_t > 0$, this would induce a process that only exhibited positive bubbles which would represent a transversally asymmetric process.

Therefore, by managing both the moving average coefficients, a_i, and the distribu-

tion of the shocks ϵ_t in (3), the mixed causal/noncausal model can exhibit both longitudinal and transversal asymmetries of the type discussed by Ramsey and Rothman (1996).

Figure 6: The mixed causal/noncausal model with Cauchy shocks



4.2 The purely causal representation

As discussed above, in general we have a unique linear representation as (3) except when the white noise process is Gaussian. This implies that, for fat tailed distributions, such as the t-distribution or Cauchy distribution, the purely causal strong form representation will necessarily admit a nonlinear representation [Rosenblatt (2000)].

4.2.1 Example: The noncausal autoregressive process with Cauchy shocks

Consider the noncausal autoregressive process of order 1 with Cauchy shocks,

$$x_t = \rho x_{t+1} + \epsilon_t, \tag{4}$$

where $|\rho| < 1$ and $\epsilon_t/\sigma_\epsilon$ follows a standard i.i.d. Cauchy distribution. The shocks can be interpreted as backward innovations, defined as $\epsilon_t = x_t - median(x_t|x_{t+1})$, since, strictly speaking, the moments of the Cauchy distribution do not exist.

This process admits both a strong purely causal representation which is necessarily nonlinear with i.i.d. shocks, and a weak form purely causal representation which is linear, but where the shocks are weak white noise and not i.i.d.

More precisely, the noncausal process (x_t) is a Markov process in direct time with a causal transition p.d.f. given as [Gourieroux and Zakoian (2012), Proposition 2, and Appendix 11.3 of this paper]:

$$f_{t+1|t}(x_{t+1}|x_t) = \frac{1}{\sigma_\epsilon \pi} \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + z_t^2} \frac{\sigma_\epsilon^2 + (1-|\rho|)^2 x_t^2}{\sigma_\epsilon^2 + (1-|\rho|)^2 x_{t+1}^2}. \tag{5}$$

In particular the causal conditional moments associated with the equation above exist up to order three, whereas the noncausal conditional moments associated with the forward autoregression in (4), and the unconditional moments, do not exist.

i) The causal strong autoregressive representation

In order to represent (4) as a causal, direct time, process in strong form, we must appeal to the nonlinear (or generalized) innovations of the process [see Rosenblatt (2000), Corollary 5.4.2. or Gourieroux and Jasiak (2005), Section 2.1].

Intuitively, a nonlinear error term, (η_t), of the causal process (x_t) is a strong white

noise where we can write the current value of the process x_t as a nonlinear function of its own past value x_{t-1} and η_t, say,

$$x_t = G(x_{t-1}, \eta_t), \quad \eta_t \sim i.i.d., \tag{6}$$

where x_t and η_t satisfy a continuous one-to-one relationship given any x_{t-1}. For more details see Appendix 11.4.

ii) The causal weak autoregressive representation

Only the Gaussian autoregressive processes possess both causal and noncausal strong form linear autoregressive representations. The noncausal AR(1) Cauchy model therefore admits only a weak form linear representation given as [Gourieroux and Zakoian (2012), Section 2.3]:

$$x_t = E_{t|t-1}[x_t|x_{t-1}] + \eta_t^* \sqrt{Var_{t|t-1}[x_t|x_{t-1}]}. \tag{7}$$

The representation is weak since (η_t^*) is a weak white noise (not i.i.d.) and $\eta_t^* \sqrt{Var_{t|t-1}[x_t|x_{t-1}]} = \epsilon_t^*$ is conditionally heteroskedastic. That is, the weak innovations also display GARCH type effects.

The conditional moments of x_t are given as:

$$E_{t|t-1}[x_t|x_{t-1}] = sign(\rho)x_{t-1} \quad \text{and} \tag{8a}$$

$$E_{t|t-1}[x_t^2|x_{t-1}] = \frac{1}{|\rho|}x_{t-1}^2 + \frac{\sigma_\epsilon^2}{|\rho|(1-|\rho|)}. \tag{8b}$$

Interestingly, from equation (8a), we see that for $\rho > 0$ the process exhibits a unit root (this is the martingale property), but is still stationary; this unit root is expected

since the unconditional moments of x_t do not exist. Usually when we consider the properties of a unit root model this is within the context of models with a nonstationary stochastic trend. However, in the example above the causal process (x_t) has a unit root when being strongly stationary. So the unit root does not generate a stochastic trend, but can generate bubbles due to the martingale interpretation [see Gourieroux and Zakoian (2012), and the discussion in Section 4.3].

4.3 Other bubble like processes

As described in Gourieroux and Zakoian (2012), several other examples of martingale processes with bubbles have been introduced in the literature. However, none of these processes are as easy to introduce into a general dynamic framework as the set of mixed causal/noncausal processes.

Interestingly, these previous bubble processes are piecewise linear, but still maintain the martingale property. For example, the bubble process introduced in Blanchard and Watson (1982) is given by:

$$x_{t+1} = \begin{cases} \frac{1}{\pi} x_t + \epsilon_{t+1}, & \text{with probability} \quad \pi, \\ \epsilon_{t+1}, & \text{with probability} \quad (1-\pi), \end{cases} \tag{9a}$$

where ϵ_t is a Gaussian error term and $\pi \in (0, 1)$. This is a martingale process, with a piecewise linear dynamic in that given the latent state, the parameter on the autoregression switches between zero and $1/\pi$.

Evans (1991) proposes to model the explosive rate parameter, (θ_t), say, as a Bernoulli random variable, $\mathcal{B}(1, \pi)$. Again, this process represents one that is piecewise linear, but in this case is also a multiplicative error term model, with (u_t) representing an i.i.d. pro-

cess with $u_t \geq 0$, $E_t[u_{t+1}] = 1$, and with parameters $0 < \delta < (1+r)\alpha$ where $r > 0$, $\pi \in (0, 1]$, and

$$x_{t+1} = \begin{cases} \left(\delta + \frac{1}{\pi}(1+r)\theta_{t+1}\left(x_t - \frac{\delta}{(1+r)}\right)\right) u_{t+1} & \text{if} \quad x_t > \alpha \\ (1+r)x_t u_{t+1} & \text{if} \quad x_t \leq \alpha. \end{cases} \tag{10a}$$

In this case the regime is not latent, but is a function of the observable x_t. In this way, the process is an extension of the self-exciting threshold autoregression of Tong and Lim (1980).

For illustration I have simulated sample paths from the two bubble processes above along with the causal AR(1) Cauchy process (see **Figure 7**). The Blanchard and Watson process is simulated by choosing $\pi = 0.8$ and $\epsilon_t \sim IIN(0, 1)$. The Evans process is simulated in accordance to the parameters chosen in simulating bubbles for Table 1, on page 925, of their paper; that is, we have $\alpha = 1$, $\delta = 0.5$, $1 + r = 1.05$, $\pi = 0.75$ and a sample path of length $T = 100$ is generated. Moreover, u_t is log-normally distributed, where $u_t = \exp[y_t - \tau^2/2]$ and $y_t \sim IIN(0, 0.05^2)$. Finally, the causal AR(1) Cauchy is simulated by choosing $\rho = 0.8$ in equation (4) and $\sigma = 0.1$ as the scale parameter of the Cauchy distribution.

The bubble processes above were constructed for very specific theoretical reasons. The Blanchard and Watson (1982) process is given as an example of a bubble consistent with the rational expectation hypothesis and the Evans (1991) process is given as an example of a stationary process with periodically collapsing bubbles that defies standard linear unit root testing. Alone, and without further modification, neither process should be considered a serious candidate to model bubbles in commodity futures price levels. On the other hand, unlike these previous bubble processes, the AR(1) Cauchy model is

easily introduced in a mixed causal/noncausal framework.

5 Estimation of the mixed causal/noncausal process

In this section we introduce the mixed causal/noncausal autoregressive model which will be estimated in an attempt to model the asymmetric bubble features exhibited by the commodity futures price level data. The model is a linear parameterization of the general mixed causal/noncausal model in (3) and represents the mixed causal/noncausal analog of the causal autoregressive model. The model is discussed in the next Section 5.1 and estimation of the model via maximum likelihood is discussed in Section 5.2.

5.1 The mixed causal/noncausal autoregressive model of order (r, s)

Definition 5.1. *The mixed causal/noncausal autoregressive process of order* (r, s)

Let (x_t) be a univariate stochastic process generated by a linear autoregressive mixed causal/noncausal model with order (r, s). The process is defined by

$$\alpha(L)x_t = \epsilon_t^*, \quad \epsilon_t^* \sim i.i.d., \tag{11a}$$

$$\text{where} \quad \alpha(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \ldots - \alpha_p L^p, \tag{11b}$$

such that L is the lag operator (i.e. $Lx_t = x_{t-1}$ and $L^{-1}x_t = x_{t+1}$), $p = r + s$, and the operator $\alpha(z)$ can be factorized as $\alpha(z) = \phi(z)\varphi^*(z)$. We have that $\phi(z)$ (of order r) contains all its roots strictly outside the complex unit circle and $\varphi^*(z)$ (of order s) contains all its roots strictly inside the unit circle.[10]

Therefore, $\phi(z)$ represents the purely causal autoregressive component and $\varphi^*(z)$ rep-

[10]To ensure the existence of a stationary solution, we assume that all roots have a modulus strictly different from 1.

resents the purely noncausal autoregressive component [Breidt et al. (1991)].

5.1.1 Moving average representation of the stationary solution

If $\alpha(L)$ has no roots on the unit circle, and ϵ_t belongs to a L^ν-space with $\nu > 0$ (that is $E[|\epsilon_t|^\nu] < \infty$), then a unique stationary solution to the difference equation defined in (11b) exists [see Appendix 11.1]. We can write:

$$x_t = \alpha(L)^{-1}\epsilon_t^* = \sum_{l=-\infty}^{\infty} \gamma_l \epsilon_{t-l}^*, \tag{12}$$

where the series of moving average coefficients is absolutely summable, $\sum_{l=-\infty}^{\infty}|\gamma_l| < \infty$.

The strong stationary representation is derived as follows. Let us factorize $\phi(L)$ and $\varphi^*(L)$ as

$$\phi(L) = \prod_{j=1}^{r}(1 - \lambda_{1,j}L), \quad \text{where} \quad |\lambda_{1,j}| < 1, \tag{13a}$$

$$\text{and} \quad \varphi^*(L) = \prod_{k=1}^{s}(1 - \frac{1}{\lambda_{2,k}}L), \quad \text{where} \quad |\lambda_{2,k}| < 1. \tag{13b}$$

The noncausal component can also be written as

$$\varphi^*(L) = \frac{(-1)^s L^s}{\prod_{k=1}^{s}\lambda_{2,k}} \prod_{k=1}^{s}(1 - \lambda_{2,k}L^{-1}). \tag{14}$$

We get the Taylor series expansions

$$(1-\lambda_{1,j}L)^{-1} = \sum_{l=-\infty}^{\infty} \lambda_{1,j}^{l} L^{l}, \tag{15a}$$

$$\text{and} \quad (1-\lambda_{2,k}L^{-1})^{-1} = \sum_{l=-\infty}^{\infty} \lambda_{2,k}^{l} L^{-l}, \tag{15b}$$

which are valid because the roots are such that $|\lambda_{1,j}| < 1, \forall j$ and $|\lambda_{2,k}| < 1, \forall k$. Thus we get

$$\begin{aligned} x_t = \phi(L)^{-1}\varphi^*(L)^{-1}\epsilon_t^* &= \frac{\prod_{k=1}^{s}\lambda_{2,k}}{(-1)^s L^s} \frac{1}{\prod_{j=1}^{r}(1-\lambda_{1,j}L)\prod_{k=1}^{s}(1-\lambda_{2,k}L^{-1})}\epsilon_t^* \\ &= \frac{\prod_{k=1}^{s}\lambda_{2,k}}{(-1)^s L^s} \prod_{j=1}^{r}\left(\sum_{l=-\infty}^{\infty}\lambda_{1,j}^{l}L^{l}\right)\prod_{k=1}^{s}\left(\sum_{l=-\infty}^{\infty}\lambda_{2,k}^{l}L^{-l}\right)\epsilon_t^* = \sum_{l=-\infty}^{\infty}\gamma_l\epsilon_{t-l}^*. \end{aligned} \tag{16}$$

5.1.2 An alternative representation

Since such a representation in (11a) is defined up to a scale factor on ϵ_t^*, another equivalent representation is given as

$$\Phi(L)x_t = \phi(L)\varphi(L^{-1})x_t = \epsilon_t, \tag{17a}$$

$$\text{where} \quad \varphi(L^{-1}) = 1 - \varphi_1 L^{-1} - \varphi_2 L^{-2} - \ldots - \varphi_s L^{-s}, \tag{17b}$$

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_r L^r, \tag{17c}$$

and (ϵ_t) is the sequence of i.i.d. random variables defined as $\epsilon_t = -(1/\varphi_s^* L^s)\epsilon_t^* = -(1/\varphi_s^*)\epsilon_{t+s}^*$.

We can always map the parameters from model (17) to (11) since we have $-(1/\varphi_s^* L^s)\varphi^*(L) = \varphi(L^{-1})$, where the coefficients of $\varphi(L^{-1})$ are given as $\varphi_i = -\varphi_{s-i}^*/\varphi_s^*$ for $i = 1, \ldots, s-1$, and $\varphi_s = 1/\varphi_s^*$ for $i = s$, and the roots of $\varphi^*(L)$

and $\varphi(L^{-1})$ are inverses (in the sense that $\varphi^*(z) = \varphi(1/z) = 0$ for some complex z where $|z| < 1$).

From the original representation in equation (11) we have that

$$\alpha(L)x_t = \epsilon_t^* \Leftrightarrow x_t - \alpha_1 x_{t-1} - \cdots - \alpha_p x_{t-p} = \epsilon_t^*, \tag{18}$$

and so under this standardization, the autoregressive coefficient associated with the current time period, x_t, is normalized to one (i.e. $\alpha_0 = 1$).

However, given the alternative representation we have that

$$\frac{1}{\varphi_s^*}x_{t+s} + \frac{\varphi_1^*}{\varphi_s^*}x_{t+(s-1)} + \cdots + x_t + \cdots + \phi_{r-1}x_{t-(r-1)} + \phi_r x_{t-r} =$$

$$\varphi_s x_{t+s} + \varphi_{s-1}x_{t+(s-1)} + \cdots + x_t + \cdots + \phi_{r-1}x_{t-(r-1)} + \phi_r x_{t-r} = \epsilon_t, \tag{19}$$

and so under this alternative standardization, the autoregressive coefficient chosen as equal to 1 does not coincide with the most recent time period, $t + s$; rather the standardization applies the autoregressive coefficient equal to 1 to the "intermediate" value x_t, where the autoregression depends also on s future lags and r past lags.

5.2 ML estimation of the mixed causal/noncausal autoregressive model

In estimating the parameters of the mixed causal/noncausal autoregressive process in Section 5.1, we can apply the usual maximum likelihood estimation (MLE) approach. The likelihood function represents the distribution of the sample data, conditional on the parameters of the model. The maximum likelihood method estimates the parameters of

the model as the values of the parameters which maximize this likelihood function. Let θ represent the vector of parameters, including the vectors of causal and noncausal autoregressive coefficients, $\boldsymbol{\phi}$ and $\boldsymbol{\varphi}$, and the parameters characterizing the fat tailed, t-distributed, error term,[11] that are its degree of freedom parameter λ and scale σ. The maximum likelihood estimator is given as

$$\hat{\theta}_{mle} = \underset{\theta}{\operatorname{argmax}}\, f(\boldsymbol{x_T}|\theta), \tag{20}$$

where

$$\begin{aligned} &f(\boldsymbol{x_T}|\theta) = f(x_T, x_{T-1}, x_{T-2}, \ldots, x_1|\theta) \\ &= f(x_T|x_{T-1}, \ldots, x_1; \theta) f(x_{T-1}|x_{T-2}, \ldots, x_1; \theta) \ldots f(x_3|x_2, x_1; \theta) f(x_2|x_1; \theta) f(x_1|\theta), \end{aligned} \tag{21}$$

and $\boldsymbol{x_T} = \{x_T, x_{T-1}, x_{T-2}, \ldots, x_1\}$ is the joint vector of sample data.

i) Approximation of the likelihood in the causal autoregressive model

In causal time series analysis of autoregressive models, say for the autoregressive model of order p, we know that the likelihood function can be approximated by neglecting the effect of starting values. For example, the causal AR(p) model's likelihood:

$$\begin{aligned} &f(\boldsymbol{x_T}|\theta) = f(x_T, x_{T-1}, x_{T-2}, \ldots, x_1|\theta) \\ &= f(x_T|x_{T-1}, \ldots, x_{T-p}; \theta) f(x_{T-1}|x_{T-2}, \ldots, x_{T-p-1}; \theta) \ldots f(x_3|x_2, x_1; \theta) f(x_2|x_1; \theta) f(x_1|\theta), \end{aligned} \tag{22}$$

[11] We employ either a t-distributed or skew t-distributed error term in order to identify the mixed causal/noncausal model. See Appendix 11.5.

can be approximated by neglecting the conditional densities of the initial values x_t for all $t \leq p$. For large sample size, T, this approximation error becomes negligible and the estimator obtained by maximizing the approximated likelihood is still asymptotically efficient.

ii) Approximation of the likelihood in the mixed causal/noncausal autoregressive model

The maximum likelihood approach can also be used in the general framework of the mixed causal/noncausal processes, that is the parameters estimated by:

$$\hat{\theta}_{mle} = \underset{\theta}{\operatorname{argmax}} f(\boldsymbol{x_T}|\theta), . \tag{20}$$

Under standard regularity conditions, including the strong stationarity of the process and appropriate mixing conditions, the ML estimator is consistent and its asymptotic properties, that is its speed of convergence and asymptotic distribution, are easily derived [see Breidt et al. (1991)].

However, in practice the closed form expression of the likelihood, $f(\boldsymbol{x_T}|\theta)$, is difficult to derive and the likelihood function has to be approximated, without loosening the asymptotic properties of the ML estimator. Two approaches are typically suggested:

i) Take the autoregressive expression $\alpha(L)x_t = \epsilon_t$, and approximate the likelihood by:

$$\prod_{t=p+1}^{T} f_\epsilon(\alpha(L)x_t|\beta), \tag{23}$$

where β are the parameters characterizing the distribution of the error. Such an approximation is wrong and leads in general to an inconsistent estimator. The

reason is as follows. Since the approximation is based on the autoregression:

$$x_t - \alpha_1 x_{t-1} - \alpha_2 x_{t-2} - \cdots - \alpha_p x_{t-p} = \epsilon_t, \tag{24}$$

the approximation above is valid if ϵ_t is independent of the explanatory variables, $x_{t-1}, \ldots, x_{t-p}$. But in a mixed model with a noncausal component, ϵ_t appears in the moving average representation of $x_{t-1}, \ldots,$ and x_{t-p}, which creates dependence. This is the well known error-in-variables model encountered in linear models and usually solved by introducing instrumental variables, with, in general, a loss of efficiency.

ii) Consider the moving average expression of $x_t = \sum_{l=-\infty}^{\infty} a_i \epsilon_{t-i}$ with the identification restriction $a_0 = 1$. Set to zero the values of the noise corresponding to the indices outside the observation period, $\{1, 2, \ldots, T\}$, that is, $\epsilon_t = 0$, if $t \leq 0$ and if $t \geq T + 1$. Thus we truncate the moving average representation into:

$$x_t \approx \sum_{i=-T+t}^{T-1} a_i(\alpha)\epsilon_{t-i} = \sum_{\tau=1}^{T} a_{t-\tau}(\alpha)\epsilon_\tau, \quad \text{for} \quad t = 1, \ldots, T, \tag{25}$$

where the dependence on the autoregressive parameters is explicitly indicated. We get a linear system of equations, which relates the observations $\{x_1, \ldots, x_t\}$ and the errors $\{\epsilon_1, \ldots, \epsilon_T\}$ in a one-to-one relationship. Therefore, the joint distribution of $\{x_1, \ldots, x_t\}$ can be deduced from the joint distribution of $\{\epsilon_1, \ldots, \epsilon_T\}$, which has a closed form by applying the change of variables Jacobian formula. However, this approach is difficult to implement numerically, since the matrix of the transformation, $\boldsymbol{A}(\boldsymbol{\alpha})$, with generic elements $a_{t-\tau}(\alpha)$, $t, \tau = 1, \ldots, T$ has a large $T \times T$ dimension. This makes difficult, first the inversion of this matrix, and second the

numerical computation of its determinant.

This explains why a methodology has been introduced to circumvent this numerical difficulty, which explains how to approximately invert this matrix and compute the determinant, by using appropriately both the causal and noncausal components [see Breidt et al. (1991), Lanne and Saikkonen (2008), and Appendix 12 of this paper]. This approximated likelihood is used in our application to commodity futures prices. The approximation requires knowledge of the causal and noncausal orders (r, s) respectively. If they are unknown, the approach is applied to all pairs of orders (r, s) such that $r + s = p$ as given. The selected orders are the ones which minimize the AIC criterion, based on the log-likelihood value.

More precisely, Lanne and Saikkonen (2008) note that the matrix $\boldsymbol{A(\alpha)}$ can be approximately written as

$$\boldsymbol{A(\alpha)} \approx \boldsymbol{A_c(\phi) A_{nc}(\varphi^*)}, \tag{26}$$

where $\boldsymbol{A_c(\phi)}$ (resp. $\boldsymbol{A_{nc}(\varphi^*)}$) depends on the causal (resp. noncausal) autoregressive coefficients only, and is lower (resp. upper) triangular with only 1's on the diagonal. Therefore, the Jacobian

$$|det\left(\boldsymbol{A(\alpha)}\right)| \approx |det\left(\boldsymbol{A_{nc}(\phi)}\right)||det\left(\boldsymbol{A_{nc}(\varphi^*)}\right)| = 1 \tag{27}$$

and can be neglected. Therefore, the likelihood function can be approximated by

$$\prod_{t=r+1}^{T-s} f_\epsilon(\varphi(L^{-1})\phi(L)x_t; \lambda, \sigma), \tag{28}$$

where $\theta = \{\boldsymbol{\phi}, \boldsymbol{\varphi}, \lambda, \sigma\}$ represents the parameters of the model, that is, the vectors of causal and noncausal autoregressive coefficients respectively, and the t-distribution

degree of freedom and scale parameter assumed on (ϵ_t).

They show that the only autoregressive representation which leads to consistent estimators is the representation with the autoregressive coefficient equal to 1 for x_t with r lagged values before, and s lagged values afterwards, as given above in the autoregressive equation (19).

Example 1: Causal AR(1) process

Let us consider the causal AR(1) process

$$x_t = \alpha_1 x_{t-1} + \epsilon_t, \quad \text{where} \quad |\alpha_1| < 1. \tag{29}$$

This is the usual case and so we can employ the MLE to estimate α_1 by maximizing the approximate likelihood function $\prod_{t=2}^{T} f_\epsilon(x_t - \alpha_1 x_{t-1})$. This case does not present a problem since we already have the coefficient in front of x_t equal to 1.

Example 2: Noncausal AR(1) process

However, given the noncausal AR(1) process

$$x_t = \alpha_1 x_{t-1} + \epsilon_t, \quad \text{where} \quad |\alpha_1| > 1, \tag{30}$$

the estimator which maximizes the approximate likelihood function $\prod_{t=2}^{T} f_\epsilon(x_t - \alpha_1 x_{t-1})$ is now biased. Indeed, since x_t can be written as the noncausal moving average $x_t = \sum_{j=0}^{\infty} \left(\frac{1}{\alpha_1}\right)^j \epsilon^*_{t+1+j}$, there now exists a dependence between x_t and x_{t-1}.

The methodology leading to consistent estimation consists in the case of regressing x_{t-1} on x_t, instead of regressing x_t on x_{t-1}. We can rewrite the noncausal regression

above as

$$x_t = \frac{1}{\alpha_1} x_{t+1} - \frac{1}{\alpha_1} \epsilon_{t+1} = \frac{1}{\alpha_1} x_{t+1} + \epsilon^*_{t+1} \quad \text{where} \quad |\alpha_1| > 1, \tag{31}$$

which now restores the independence between the regressand and the regressor, and so the MLE which maximizes $\prod_{t=2}^{T} f_{\epsilon^*}(x_t - \frac{1}{\alpha_1} x_{t+1})$ is asymptotically unbiased.

5.3 Estimation results

In this Section I will evaluate estimation results from the mixed autoregressive model of order (r, s) as applied to the 25 commodity futures price level series. Estimation of the model parameters numerically optimizes the approximated likelihood function discussed in the last section. As in Lanne and Saikkonen (2008) and Lanne, Luoto, and Saikkonen (2012), we assume the regularity conditions of Andrews et al. (2006) are satisfied, which require the likelihood to be twice differentiable with respect to both $\boldsymbol{x_T}$ and θ. The approximated likelihood algorithm is computed in Fortran and the optimization of the likelihood function is performed using a set of Fortran optimization subroutines called the PORT library, designed by David M. Gay at Bell Laboratories [Gay (1990)].

As in Section 3 where the linear causal ARMA model with Gaussian innovations was shown to inadequately capture the features of the price level data, I will again employ the AIC criterion as a measure of model fit, along with Ljung-Box statistics testing the hypothesis that the innovations exhibit no linear autocorrelation. In this way, I will consider the best fitting linear causal ARMA model, with Gaussian innovations, from Section 3 as a benchmark model.

Table 5.i presents the results of maximum likelihood estimation. The mixed AR

model orders, (r, s), were selected via AIC among a possible set of (r, s) values such that $r \leq 10$ and $s \leq 10$. The first row of the results for each series represents the benchmark ARMA model, with Gaussian innovations, from Section 3, while the second and third rows represent the mixed AR(r, s) model with both t-distributed and skew t-distributed errors, respectively. Recall that the mixed causal/noncausal model is only identified for non-Gaussian error terms. The lags column represents the number of lags included in the Ljung-Box statistic, where p-values are provided in their respective columns. Finally, an 'x' marks the model with the lowest normalized AIC.

The estimation results suggest that the mixed causal/noncausal model improves model fit over the baseline causal ARMA model, with Gaussian innovations. When the models are nested, I employ likelihood ratio (LR) tests. In every case the mixed causal/noncausal model improves model fit significantly at the 1% significance level.

In comparing the skewed t-distributed error term mixed causal/noncausal model to the standard t-distribution error term model, the results vary by series. In most cases the skewed t-distribution improves model fit and passes a LR test at the 1% level. Moreover, orange juice, lumber, silver, copper, light crude oil, and gas oil also pass at the 5% level and coffee passes at the 10% level. Series that do not pass LR tests at the 10% level are soybean meal and oil, sugar, corn, cotton, rice, gold, palladium, natural gas, and lean cattle, suggesting that there is little gain in employing a skewed t-distribution on the innovations of these mixed models.

Interestingly, the estimated t-distribution degree of freedom parameter, λ, for the mixed causal/noncausal model error terms range between near 1 (i.e. Cauchy distributed) to around 3 or so in most cases, which suggests bubble like behaviour as discussed in Gourieroux and Zakoian (2012). The only exceptions to this are found in lumber ($\lambda \approx 3.88$), gasoline RBOB ($\lambda \approx 4.93$), and live cattle ($\lambda \approx 3.39$).

Moreover, an examination of the roots of the lag polynomials implied by the estimated parameters also confirms the partly noncausal nature of the series. If we accept only the statistically significant estimated parameters [12] and solve for the roots of the implied causal and noncausal lag polynomials, $\phi(L)$ and $\varphi(L^{-1})$ (from (17a)), we find that the roots of both appropriately lie outside the unit circle.[13] Of course, if the data generating process was purely causal, none of the lags of the noncausal polynomial, $\varphi(L^{-1})$, should be statistically significant.

Moreover, if we fit the best (according to the AIC criterion) purely causal ARMA model, with t-distributed error terms instead of Gaussian ones, we often find that the estimated roots of the causal lag polynomial lie inside the unit circle. This suggests misspecification of the noncausal component, as well as the fact that the noncausality is not identified in the purely causal ARMA model with Gaussian innovations.

For reference I have constructed tables with all of the roots of the lag polynomials of both the causal ARMA models of order (p, q), with t-distributed innovations, and the mixed causal/noncausal AR models of order (r, s) (see **Tables 7.i to 7.iii** within Appendix 14).

For example, estimating purely causal ARMA models with t-distributed innovations suggest the following results: wheat, coffee, rice, gold, platinum, all the energy series except natural gas, and lean hogs all share at least one root with absolute value less than one in their $\frac{\alpha(L)}{\beta(L)} = \delta(L)$ lag polynomial (that is $\frac{\alpha(L)}{\beta(L)} = \delta(L)$ in the ARMA model $\delta(L)x_t = \epsilon_t$, where $\alpha(L)$ and $\beta(L)$ are the AR and MA lag polynomials respectively), suggesting that this polynomial could be factorized and then estimated as

[12] Tested at the 5% level, assuming Normally distributed parameters and employing the inverse of the observed Hessian matrix at the MLE estimated value as the parameter covariance matrix.

[13] Which implies that (11a), $\alpha(L) = \phi(L)\varphi^*(L)$, is such that the roots of $\phi(L)$ lie strictly outside the complex unit circle while those of $\varphi^*(L)$ lie strictly inside.

a mixed causal/noncausal model (instead of the traditional differencing technique employed).

Furthermore, the very large valued roots of the causal polynomial for light crude oil, gas oil, and heating oil, suggest that these series may be better represented as purely noncausal since these large causal roots have little effect on the causal impulse response. This result is confirmed by looking at the mixed causal/noncausal model roots of light crude oil, but not for gas or heating oil which have causal polynomial roots relatively close to 1. Finally, the mixed causal/noncausal representation for soybeans suggests that the process may be better modeled as purely causal, while the results for cotton, live cattle, and lean hogs suggest they may be purely noncausal.

To summarize, our results suggest that most of the futures price series exhibit much better in-sample model fit, according to the AIC criterion, when modeled by a mixed causal/noncausal autoregressive specification that takes into account their possible noncausal components. Moreover, this noncausality is unidentified in the purely causal ARMA model with Gaussian innovations. Finally, estimation of purely causal ARMA models with fat tailed, t-distributed, innovations reinforces the series' noncausal nature, as often the causal lag polynomial roots lie inside the complex unit circle.

Table 5.i: Estimation results of mixed causal/noncausal AR(r, s) models

Series		**p/r**	**q/s**	**AIC**	**Log-likelihood**	**Ljung-Box**	λ
Soybean meal		6	2	52395.000	-26188.500	0.152	∞
	x	10	10	48208.261	-24081.130	0.007	2.070
		10	10	48210.118	-24081.059	0.007	2.072
Soybean oil		8	3	11859.050	-5917.523	0.919	∞
	x	10	10	9211.876	-4582.938	0.135	2.455
		10	10	9213.688	-4582.844	0.138	2.455
Soybeans		9	2	73548.210	-36762.110	0.521	∞
		10	10	69444.844	-34699.422	0.000	2.073
	x	10	10	69438.354	-34695.177	0.000	2.086
Orange juice		4	3	42121.610	-21052.800	0.395	∞
		10	10	38686.959	-19320.480	0.378	2.326
	x	10	10	38683.919	-19317.960	0.389	2.331
Sugar		10	2	7842.392	-3908.196	0.999	∞
	x	2	2	1549.499	-767.750	0.000	1.702
		2	2	1551.289	-767.645	0.000	1.702
Wheat		7	2	67069.470	-33524.740	0.998	∞
		5	5	61896.849	-30935.424	0.000	2.028
	x	5	5	61880.290	-30926.145	0.000	2.047
Cocoa		8	3	94368.760	-47172.380	0.716	∞
		2	1	91804.882	-45896.441	0.000	2.558
	x	10	10	91586.110	-45769.055	0.003	2.584
Coffee		4	2	48866.800	-24426.400	0.064	∞
		10	10	43731.886	-21842.943	0.014	1.923
	x	10	10	43730.300	-21841.150	0.012	1.925
Corn		7	3	59385.840	-29681.920	0.625	∞
	x	2	3	53647.827	-26815.913	0.776	1.811
		2	3	53649.243	-26815.622	0.783	1.811
Cotton		10	0	32760.780	-16369.390	1.000	∞
	x	1	3	27005.831	-13495.916	0.000	2.455
		1	3	27007.812	-13495.906	0.000	2.455

Table 5.ii: Estimation results of mixed causal/noncausal AR(r, s) models

Series		**p/r**	**q/s**	**AIC**	**Log-likelihood**	**Ljung-Box**	λ
Platinum		8	2	55936.820	-27957.410	0.129	∞
		10	10	51667.822	-25810.911	0.000	1.572
	x	10	10	51644.800	-25798.400	0.000	1.585
Rice		10	3	-4799.022	2413.511	0.958	∞
	x	1	3	-7173.685	3593.842	0.013	2.076
		1	3	-7172.345	3594.173	0.013	2.075
Lumber		8	3	44027.920	-22001.960	1.000	∞
		10	10	42939.948	-21446.974	0.562	3.874
	x	10	10	42937.244	-21444.622	0.546	3.876
Gold		0	3	102914.500	-51453.270	n/a	∞
	x	10	10	56917.739	-28435.869	0.000	1.317
		10	10	56919.621	-28435.811	0.000	1.318
Silver		9	3	7424.036	-3699.018	0.935	∞
		10	10	-7052.297	3549.149	0.000	1.063
	x	10	10	-7056.283	3552.141	0.000	1.066
Palladium		9	3	48209.690	-24091.840	0.992	∞
	x	8	8	42569.544	-21265.772	0.000	1.225
		8	8	42571.492	-21265.746	0.000	1.225
Copper		10	0	34719.500	-17348.750	1.000	∞
		10	10	30533.482	-15243.741	0.000	1.349
	x	10	10	30529.777	-15240.889	0.000	1.354
Light crude oil		7	2	22244.110	-11112.060	0.949	∞
		1	3	17297.702	-8641.851	0.015	1.409
	x	1	3	17295.206	-8639.603	0.014	1.415
Heating oil		9	2	34465.280	-17220.640	0.998	∞
	x	10	10	30808.001	-15381.000	0.042	1.535
		8	8	30841.794	-15400.897	0.000	1.538
Brent crude oil		7	2	18807.920	-9393.960	0.901	∞
		10	10	15081.643	-7517.822	0.000	1.458
	x	10	10	15073.528	-7512.764	0.000	1.462

Table 5.iii: Estimation results of mixed causal/noncausal AR(r, s) models

Series		**p/r**	**q/s**	**AIC**	**Log-likelihood**	**Ljung-Box**[a]	λ[f]
[c] Platinum		8	2	55936.820	-27957.410	0.129	∞
[d]		10	10	51667.822	-25810.911	0.000	1.572
[e]	x[b]	10	10	51644.800	-25798.400	0.000	1.585
Gas oil		5	3	44142.240	-22062.120	0.922	∞
		10	10	41116.045	-20535.023	0.259	1.566
	x	10	10	41112.456	-20532.228	0.261	1.574
Natural gas		3	2	-4178.268	2095.134	0.226	∞
	x	1	1	-7772.315	3891.158	0.017	1.666
		1	1	-7771.454	3891.727	0.018	1.666
Gasoline RBOB		5	3	11715.320	-5848.658	0.988	∞
		2	1	11535.858	-5761.929	0.050	4.662
	x	2	1	11526.267	-5756.133	0.056	4.925
Live cattle		6	1	22771.400	-11377.700	0.986	∞
		10	10	20427.885	-10190.943	0.915	3.331
	x	8	8	20426.530	-10193.265	0.873	3.392
Lean hogs		3	2	23567.630	-11777.810	0.704	∞
		0	2	18929.149	-9459.574	0.572	2.728
	x	0	2	18922.375	-9455.188	0.570	2.737

[a] The Ljung-Box statistics are given as p-values, where the lag parameter chosen is the log sample size, $ln(T)$.

[b] The 'x' row for each series denotes the model with the lowest AIC.

[c] The first row in each series is the causal ARMA(p, q) model with Gaussian innovations estimated in Section 3.

[d] The second row is the mixed causal/noncausal AR(r, s) with t-distributed errors.

[e] The third row is the same model but with skew t-distributed errors.

[f] The λ column indicates the estimated degree of freedom parameter for the error term distribution–in the skewed t-distributed case this value represents the sum of the two skew parameters. See Appendix 11.5.

6 Comparison of the estimated unconditional distributions

Another way to evaluate the mixed causal/noncausal autoregressive model is by comparing its model based unconditional distribution by sample histogram. Histograms are

estimated for both the purely causal ARMA and mixed causal/noncausal autoregressive models, both employing t-distributed error terms, by simulating long sample paths of length $T = 200000$, given the model parameters estimated by MLE in Section 5.3.

The mixed causal/noncausal autoregressive model seeks to capture both the asymmetries and bubble features present in commodity futures prices. The transversal asymmetry and bubble features present in the series can be examined visually by considering the sample histograms of the price series presented in **Figures 11.i to 11.iv** in Appendix 14. Note the long, positively-skewed, tails many of the series exhibit, illustrating how these price series tend to spend most of the time in the shallow troughs, occasionally interrupted by brief, but dramatic, positive bubbles.

The metric employed in comparing the estimated unconditional distributions is the Kullback-Leibler divergence measure, which is a non-symmetric measure of the difference between two probability distributions P and Q. Specifically, the Kullback-Leibler measure, from continuous distributions Q to P, denoted $KL(Q,P) = \int_{-\infty}^{\infty} \ln\left(\frac{p(x)}{q(x)}\right) p(x)dx$, is the measure of the information lost when we use Q to approximate P. [14] Since the Kullback-Leibler measure is "information monotonic", as an ordinal measure of making comparisons it is invariant to the choice of histogram bin size. **Table 6** reports the Kullback-Leibler measures of the sample histogram densities for both $KL(P,Q)$ and $KL(Q,P)$ where $p(x)$ denotes the estimated p.d.f. of the sample data and the $q(x)$'s are model based estimates from the simulated sample paths of the purely causal and mixed causal/noncausal autoregressions.

Table 6 is broken into two sections: the two left columns report the Kullback-Leibler measure where the estimated models are used to approximate the sample data. In this

[14] In employing estimated sample histograms I use the discretized version of the Kullback-Leibler formula where areas of zero support are padded with 1^{-315}.

Table 6: Kullback-Leibler divergence measures

Series	KL(Q,P)[a] ARMA	MIXED	KL(P,Q) ARMA	MIXED
Soybean meal	n.s.[b]	00.329	n.s.	97.216
Soybean oil	01.965	00.316	495.751	55.752
Soybeans	n.s.	00.310	n.s.	49.584
Orange juice	00.976	00.216	351.966	60.033
Sugar	01.768	00.500	326.343	168.821
Wheat	00.535	00.427	44.699	32.956
Cocoa	00.625	01.247	230.260	37.961
Coffee	**04.519**	00.216	703.097	81.218
Corn	01.526	00.549	185.980	144.244
Cotton	00.808	**12.710**	114.104	25.918
Rice	00.429	00.311	59.220	123.030
Lumber	00.149	00.136	07.610	08.477
Gold	n.s.	uns.[c]	n.s.	uns.
Silver	n.s.	uns.	n.s.	uns.
Platinum	n.s.	00.662	n.s.	96.789
Palladium	n.s.	01.368	n.s.	440.585
Copper	n.s.	00.832	n.s.	173.295
Light crude oil	n.s.	00.813	n.s.	202.916
Heating oil	n.s.	01.043	n.s.	326.858
Brent crude oil	n.s.	00.759	n.s.	118.503
Gas oil	n.s.	00.709	n.s.	132.528
Natural gas	00.906	00.753	303.694	325.575
Gasoline RBOB	01.429	00.261	483.674	08.649
Live cattle	00.562	**18.227**	31.469	76.491
Lean hogs	02.649	00.032	640.295	03.308
average	**01.346**	01.858	284.154	**121.335**
selective average[d]	01.206	**00.650**		

[a] P represents the sample data.

[b] "n.s." stands for non-stationary, i.e. the simulations from the causal linear model were explosive.

[c] "uns." within the context of the mixed causal/noncausal models implies that the simulated sample paths were, for a lack of better words, "unstable": highly erratic with extremely long tails and extremely irregular, almost "chaotic" type behaviour. In general, while stationary, models with "uns." listed represented poor candidates as having come from the data's DGP.

[d] The selective average omits the extreme outlying cases highlighted in bold.

case if the sample path density has zero support for some region in its domain, it does not punish the prospective model density for allocating too much (resp. too little) probability to this region since this component of the Kullback-Leibler sum is zero. The two right columns report the opposite case where the sample path density is used to approximate the estimated models; in this case, if the model density has zero support in some region of its domain then the sample path density isn't penalized for allocated too much (resp. too little) probability to this region. Finally, smaller values indicate less information lost by the approximation and are preferred.

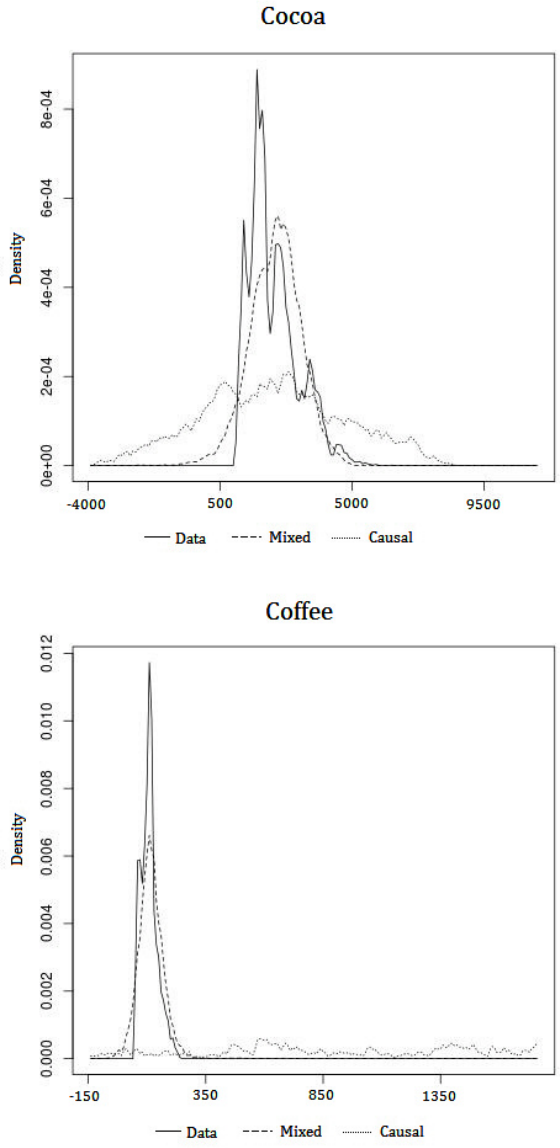
The results of these comparisons suggest the following. First, the Kullback-Leibler measures show that the unconditional distributions generated by the causal ARMA models represent a poor fit to the sample data. The ARMA model seems unable to produce the sharp bubble like behaviour we see in most of the series and the shape of its unconditional density is often much too uniform. It does not exhibit long, positively skewed, tails as are present in many of the estimated histograms of the commodity futures prices as provided in **Figures 11.i to 11.iv** in Appendix 14. Moreover, we often find that the sample paths from the causal ARMA models are explosive, due to the noncausal root in their estimated causal lag polynomials.

The results from the left hand columns of **Table 6** suggest a few distinct outlying Kullback-Leibler measures. For example, cotton and live cattle are extremely large compared to the other series' measures in the case of the mixed causal/noncausal autoregressive model, and coffee represents an outlier in the case of the causal ARMA models. Given the presence of these outliers, I calculate both the average Kullback-Leibler measure across all series and the average omitting these outliers. Given this selective average, we find that, in both the left and right columns (i.e. in the case of both $KL(Q, P)$ and $KL(P, Q)$, respectively), the mixed causal/noncausal model represents

a better fit to the sample data than the purely causal ARMA model.

Finally, **Figure 8** provides an example of the estimated unconditional densities for cocoa and coffee, respectively.

Figure 8: Estimated unconditional densities, Cocoa and Coffee



7 Forecasting the mixed causal/noncausal model

This section will first consider the problem of computing the predictive conditional density of the mixed causal/noncausal model, when the information set includes only the past values of the time series data up to some time t, say $\mathcal{F}_t = \{x_t, x_{t-1}, \ldots, x_1\}$. We then evaluate the ability of the mixed causal/noncausal model to not only fit the training sample, but also its ability to forecast out of sample.

7.1 The predictive distribution

Let us consider the general stochastic process:

$$x_t = h(\ldots, \epsilon_{t-1}, \epsilon_t, \epsilon_{t+1}, \ldots), \quad \text{where} \quad \epsilon_t \sim i.i.d. \tag{32}$$

Moreover, let $\mathcal{F}_t = \{x_t, x_{t-1}, \ldots, x_1\}$ represent the information set generated by the stochastic process up to and including time t.

The best nonlinear forecasts, at date t, for a given horizon h, can be deduced from the conditional distribution of x_{t+h}, given $\mathcal{F}_t$. More precisely, if $a(x_{t+h})$ is a square integrable transformation of x_{t+h}, then its best predictor is simply $E[a(x_{t+h})|\mathcal{F}_t] = \int a(x_{t+h}) f_{t+h|t}(x_{t+h}|\mathcal{F}_t) dx_{t+h}$.

In our framework, the standard moments may not exist and so we cannot choose to predict $a(x_{t+h}) = x_{t+h}$ for example. An alternative approach can be to compute the prediction intervals by considering the quantiles of the predictive distribution. This is the solution adopted here.

Lanne, Luoto, and Saikkonen (2012) suggest a means whereby we can simulate these quantiles. Their numerical algorithm is discussed in Appendix 13. However,

this method is computationally demanding and not necessarily the most straightforward method. Therefore, we begin a discussion below that considers the problem from first principles.

7.2 Equivalence of information sets

Consider the general mixed causal/noncausal model from (17), with causal order r and noncausal order s. It is clear that given the information set $\mathcal{F}_t = \{x_1, \ldots, x_t\}$, this is equivalent to knowing,

$$\mathcal{F}_t \equiv \{x_1, \ldots, x_r, u_{r+1}, \ldots, u_t\}, \tag{33}$$

since $u_t = \phi(L)x_t$ [see the Appendix 12]. Note that u_t represents a shock to the process x_t, which is an autoregressive function of x_t, since $u_t = \phi(L)x_t$, but where the u_t's are not i.i.d. Rather u_t is noncausal autoregressive since we have that $\varphi(L^{-1})u_t = \varphi(L^{-1})\phi(L)x_t = \epsilon_t$, where ϵ_t is i.i.d.

Knowing the latter information set in (33) is also equivalent to knowing,

$$\mathcal{F}_t \equiv \{x_1, \ldots, x_r, \epsilon_{r+1}, \ldots, \epsilon_{t-s}, u_{t-s+1}, \ldots, u_t\}, \tag{34}$$

since $\epsilon_t = \varphi(L^{-1})u_t = \varphi(L^{-1})\phi(L)x_t$. Moreover, this information is also equivalent to,

$$\mathcal{F}_t \equiv \{v_1, \ldots, v_r, \epsilon_{r+1}, \ldots, \epsilon_{t-s}, u_{t-s+1}, \ldots, u_t\}, \tag{35}$$

where $\varphi(L^{-1})x_t = v_t$. Therefore, for the process u_t that is noncausal of order s, predicting u_{t+1} based on the information set $\mathcal{F}_t$, is equivalent to predicting it based on the information subset $\{u_{t-s+1}, \ldots, u_t\}$, since the $\{v_1, \ldots, v_r, \epsilon_{r+1}, \ldots, \epsilon_{t-s}\}$ elements are

independent of the future values $\{u_{t+1}, \ldots, u_{t+h}\}$, for some forecast horizon h.

Therefore, in establishing the predictive density of the mixed causal/noncausal process x_t, we can focus our attention on the problem of predicting the noncausal component u_{t+1} conditional on the past information set $\mathcal{F}_t = \{u_{t-s+1}, \ldots, u_t\}$, since there is a direct relationship between the predictive distributions of u_{t+1} and x_{t+1} in the sense that,

$$f_{x_{t+1|t}}(x_{t+1} - \mu_t | x_t, \ldots, x_1) = f_{x_{t+1|t}}(\phi(L)x_{t+1} | x_t, \ldots, x_1) \tag{36a}$$

$$= f_{u_{t+1|t}}(u_{t+1} | x_t, \ldots, x_1) \tag{36b}$$

$$= f_{u_{t+1|t}}(u_{t+1} | u_t, \ldots, u_{t-s+1}, \epsilon_{t-s}, \ldots, \epsilon_{r+1}, x_r, \ldots, x_1) \tag{36c}$$

$$= f_{u_{t+1|t}}(u_{t+1} | u_t, \ldots, u_{t-s+1}), \tag{36d}$$

$$\text{where} \quad \mu_t = \phi_1 x_t + \phi_2 x_{t-1} + \cdots + \phi_r x_{t-(r-1)}. \tag{36e}$$

Since the change of variables implies a Jacobian determinant of 1, the conditional density of x_{t+1} is just a relocation of the conditional density of u_{t+1}. Here, μ_t represents a location parameter and so $u_{t+1} = x_{t+1} - \mu_t = \phi(L)x_{t+1}$. Therefore, by simulating the quantiles of $f_{u_{t+h|t}}(u_{t+h}|\mathcal{F}_t)$, we are able to generate prediction intervals for x_{t+h}.

The prediction problem of the noncausal process $\varphi(L^{-1})u_{t+1} = \epsilon_{t+1}$, based on past information set $\mathcal{F}_t$, must be considered with some care. In this way we first consider some simple examples. Note that while u_t is a noncausal autoregressive process, we desire the *causal* predictor which is based on the past information set $\mathcal{F}_t$, and this predictor is generally nonlinear for non-Gaussian ϵ_t.

7.3 Examples: the causal prediction problem of the noncausal process

Example 1: AR$(0,1)$ case

Let us consider the prediction problem for the purely noncausal model of order $s = 1$. We get, $x_{t+1} = u_{t+1} = \varphi_1 u_{t+2} + \epsilon_{t+1}$, and $\epsilon_{t+1} \sim i.i.d.$ In this case we desire the predictive density $f_{x_{t+1|t}}(x_{t+1}|\mathcal{F}_t)$, based on the *past* values of the process $\mathcal{F}_t = \{x_t, \ldots, x_1\}$, but where the process (x_t) is noncausal.

Since $x_t = u_t$ and $s = 1$, the predictive density $f_{x_{t+1|t}}(x_{t+1}|x_t)$ depends only on the past information set $\{x_t = u_t\}$, and by Bayes Theorem we get

$$f_{x_{t+1|t}}(x_{t+1}|x_t) = f_{x_{t|t+1}}(x_t|x_{t+1}) f_x(x_{t+1}) / f_x(x_t), \tag{37}$$

where $f_x(\cdot)$ denotes the stationary distribution of the process (x_t). We already know the noncausal transition density $f_{x_{t|t+1}}(x_t|x_{t+1})$, since it is defined by our linear model and our assumption on the shocks ϵ_t, since $u_t = x_t$, and $\varphi(L^{-1})u_t = \epsilon_t$, the conditional density of u_t given u_{t+1} is the same as the density of ϵ_t, up to a location parameter. However, what is not clear is how to deal with the stationary distribution $f_x(\cdot)$ since its analytical expression is unknown in the general case [although it has been derived where $\epsilon_t \sim Cauchy(0, \sigma^2)$ in Gourieroux and Zakoian (2012)]. Lanne, Luoto, and Saikkonen (2012) suggest a means whereby we can circumvent this problem by "enlarging the space" of random variables (see the Appendix 13). This very computationally intensive approach is not the most direct, as we shall show below.

One alternative that works quite well when the order of the noncausal polynomial is low, is to simply approximate the stationary distribution $f_x(x_{t+1})$ in (37) by means of a kernel smoother. For example, given the stationary nature of the data, $x_t = u_t$, a

consistent estimator of $f_x(\cdot)$ is given by the kernel density estimator:

$$f_x(x_t) \approx \frac{1}{Th} \sum_{\tau=1}^{T} K\left(\frac{x_t - x_\tau}{h}\right), \tag{38}$$

where $h > 0$ is an appropriately chosen smoothing parameter defining the *bandwidth* and $K(\cdot)$ is a kernel function, for instance a symmetric function that integrates to one. The Epanechnikov Kernel $K(x) = \frac{3}{4}(1 - x^2)\mathbf{1}_{|x|\leq 1}$, can be shown to be efficient in the mean squared error sense [see e.g. Epanechnikov (1969)].

Example 2: AR$(0, s)$ with $s > 1$

Let us now consider a larger noncausal autoregressive order, where we still face the purely noncausal prediction problem $u_{t+1} = x_{t+1}$. Let the noncausal lag polynomial be of order s: $\varphi(L^{-1}) = 1 - \varphi_1 L^{-1} - \cdots - \varphi_s L^{-s}$.

Again, let us express the predictive density in terms of Bayes theorem, where, since $u_t = x_t$ is a noncausal autoregressive process of arbitrary order s, the prediction depends only on the subset of information given by $\mathcal{F}_t = \{x_t, \ldots, x_{t-s+1}\}$,

$$\begin{aligned} &f_{x_{t+1|t,\ldots,t-s+1}}(x_{t+1}|x_t, \ldots, x_{t-s+1}) \\ &\qquad = \frac{f_{x_{t|t+1,\ldots,t+s}}(x_{t-s+1}|x_{t-s+2}, \ldots, x_t, x_{t+1}) f_{\bar{x}}(x_{t+1}, x_t, \ldots, x_{t-s+2})}{f_{\bar{x}}(x_t, x_{t-1}, \ldots, x_{t-s+1})} \end{aligned} \tag{39}$$

Again, $f_{\bar{x}}(x_{t-s+1}|x_{t-s+2}, \ldots, x_t, x_{t+1})$ is known from our linear noncausal autoregressive model of order s. However, it remains unclear how to deal with the joint stationary density, $f_{\bar{x}}(\cdot)$, of a sequence of s successive values of the process, especially for a larger dimension of s.

Indeed, the kernel estimator will prove problematic for large noncausal orders, s, since we now face a multidimensional smoothing problem. Indeed, as the dimension of

the smoothing problem increases, much more data is required in order to get a reliable estimate of this joint density.

7.4 A Look-Ahead estimator of the predictive distribution

Gourieroux and Jasiak (2013) suggest a direct solution to the problem of computing the predictive density $f_{x_{t+1|t}}(\cdot)$ of the noncausal process when the dimension s is relatively large. The method relies on the "Look-Ahead" estimator of the stationary density $f_{\bar{x}}(\cdot)$ [see Glynn and Henderson (1998) for the introduction of this estimator and Garibotti (2004) for an application]. First we describe the estimator in the univariate framework where the order of the noncausal polynomial is $s = 1$, and then provide an analog for the case where $s > 1$.

7.4.1 Markov process

The Look-Ahead estimator, introduced by Glynn and Henderson (1998), is a relatively simple method which allows us to estimate the stationary distribution of a Markov process, if it exists. Take for example, the Markov process, (x_t), discussed in **Example 1** above, with unique invariant density $f_x(\cdot)$, and transition density $f_{x_{t|t+1}}(\cdot)$ as expressed in (37). This Markov transition density satisfies the Kolmolgorov equation,

$$f_x(x_t^*) = \int f_{x_{t|t+1}}(x_t^*|x_{t+1}) f_x(x_{t+1}) dx_{t+1}, \forall x_t^*, \tag{40}$$

where x_t^* denotes the generic argument of the stationary density. Therefore, given a finite sample from the stationary process, $(x_\tau)_{\tau=1}^t$, we can approximate the stationary

density by

$$\hat{f}_x(x_t^*) = \sum_{\tau=0}^{t-1} f_{x_{t|t+1}}(x_t^*|x_{\tau+1}), \forall x_t^*, \tag{41}$$

where $f_{x_{t|t+1}}(x_t^*|x_{t+1})$ is known explicitly from our linear noncausal autoregressive model.

7.4.2 Markov process of order s

For larger noncausal order $s > 1$, the result is analogous. The two stationary distributions in the numerator and denominator, $f_{\bar{x}}(\cdot)$, can be estimated by the Look-Ahead estimator as

$$\hat{f}_{\bar{x}}(\boldsymbol{x}_t^*) = \sum_{\tau=0}^{t-s} l_{x_{t|t+1}}(\boldsymbol{x}_t^*|\boldsymbol{x}_{\tau+s}), \tag{42}$$

where $\boldsymbol{x}_t = \{x_t, x_{t-1}, \ldots, x_{t-s+1}\}$. The density above is more easily understood as the factorization of the joint noncausal transition density,

$$l_{x_{t|t+1}}(x_t, \ldots, x_{t-s+1}|x_{t+1}, \ldots, x_{t+s}) = \prod_{j=0}^{s-1} f_{x_{t|t+1,\ldots,t+s}}(x_{t-j}|x_{t+1-j}, \ldots, x_{t+s-j}), \tag{43}$$

whose terms are known for all j, given the linear noncausal autoregressive model (they are equal to the density of ϵ_t, up to a location parameter).

7.5 Drawing from the predictive distribution by SIR method

Given the approximate expression for the stationary density functions, $f_{\bar{x}}(\cdot)$, of both the numerator and denominator in (39), provided by the Look-Ahead estimator, we are now free to draw samples from the entire predictive density $f_{x_{t+1|t}}(\cdot)$ directly. One way this can be accomplished is by means of the (SIR) Sampling Importance Resampling

technique [see Rubin (1988), and Smith and Gelfand (1992)].

The SIR method is essentially a reweighted bootstrap simulation. Suppose we have access to some drawings from the continuous probability density $f(x)$, say $\{x_1, \ldots, x_N\}$, but we are unable to draw samples ourselves. The bootstrap procedure directs us to re-sample from the set $\{x_1, \ldots, x_N\}$, each draw having probability $1/N$. The resulting resampled set is then an approximation to draws from $f(x)$, with the approximation error approaching zero as $N \to \infty$. Indeed, for any resampled draw $\hat{x}$ we have,

$$Pr(\hat{x} \leq a) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{x_i \leq a} \to_{n \to \infty} E_f[\mathbf{1}_{x \leq a}] = \int_{-\infty}^{a} f(x)dx. \tag{44}$$

Of course the bootstrap is limited in that if our initial sample from $f(x)$ is small, repeatedly resampling from this limited sample will provide a poor approximation. The SIR allows us to circumvent this problem by allowing us to draw our initial sample from some instrumental density $g(x)$. Then by resampling from this sample, according to the weights $f(x)/g(x)$, we are able to approximate a sample from $f(x)$, rather than a sample from $g(x)$. To show this note that

$$Pr(\hat{x} \leq a) = \sum_{i=1}^{N} \left(\frac{f(x_i)}{g(x_i)} \right) \mathbf{1}_{x_i \leq a} \to_{n \to \infty} E_g[\left(\frac{f(x)}{g(x)} \right) \mathbf{1}_{x \leq a}] = \int_{-\infty}^{a} f(x)dx. \tag{45}$$

The closer is the target $f(x)$ to the instrumental density $g(x)$, the faster the rate of convergence.

Within the context of generating draws from the predictive density of the noncausal process, $f_{x_{t+1|t}}(\cdot)$, we should therefore generate draws from some proposal $g(\cdot)$ which closely approximates the target. Indeed, we have an analytic approximate expression

for $f_{x_{t+1|t}}(\cdot)$ in terms of the product of the noncausal conditional density and the Look-Ahead estimators of the stationary densities (see equation (39)), but we are unable to draw from this density directly.

The SIR method is especially appealing since it can be easily parallelized with reduced computational costs. That is, we can draw N samples from the predictive density in parallel as opposed to say a Metropolis Hastings algorithm, which is inherently sequential in nature.

Moreover, the Basel III voluntary regulation standard on bank capital levels, stress testing, and market liquidity risk, was agreed upon by the members of the Basel Committee on Banking Supervision between 2010 to 2011 and is scheduled to be introduced in 2018. Part of this regulation is the requirement that econometric models employed by financial institutions must include the possibility to simulate future sample paths for asset prices. Of course, this is a prerequisite for performing stress tests. In this respect, the proposed methodology of Lanne, Luoto, and Saikkonen (2012) would be rejected by regulators.

Forecasts up to some horizon 'h'

Given the joint predictive density, conditional on $\mathcal{F}_t$, but out to some horizon $h > 1$, we can use the same SIR method to draw samples as in the case where $h = 1$, since we can factorize the joint density as the product of the expressions given in equation (39) as,

$$g(x_{t+h}, \ldots, x_{t+1}|\mathcal{F}_t) = \prod_{j=1}^{h} f_{x_{t|t,\ldots,t-s+1}}(x_{t+j}|x_{t+j-1}, \ldots, x_{t+j-s+1}) \tag{46a}$$

$$= \prod_{j=1}^{h} \left(f_{x_{t|t+1,\ldots,t+s}}(x_{t-s+j}|x_{t+j-s+1}, \ldots, x_{t+j}) \right) \tag{46b}$$

$$\cdot \frac{f_{\bar{x}}(x_{t+h}, x_{t+h-1}, \ldots, x_{t+h-s+1})}{f_{\bar{x}}(x_t, x_{t-1}, \ldots, x_{t-s+1})}.$$

Therefore, since terms in the product cancel, as h gets large we need only estimate one term in both the numerator and denominator by the Look-Ahead method. Of course, for the SIR simulation with horizon h, we require an h-dimension proposal density $g(\cdot)$.

7.6 Application to commodity futures data

While the method described above is computationally intensive, it is clear that it is ripe for parallelization since we can potentially draw each of the N samples from the h-dimensional predictive density, $g(x_{t+h}, \ldots, x_{t+1}|\mathcal{F}_t)$, at the same time. In this sense, I have implemented the algorithm in parallel using the CUDA development libraries designed and freely available from Nvidia at `http://www.nvidia.ca/object/cuda\_home\_new.html`. All that is required is a Nvidia GPU (graphics processing unit) and knowledge of the C programming language.

In order to evaluate forecasts, I have set aside an additional 107 sample data points beyond the most recent date available within-sample, which is February 8th, 2013. Therefore, this out of sample period extends between February 11th to July 15th, 2013. [15]

As an example I now employ the Look-Ahead estimator of the stationary density, and the SIR method, to generate draws from the predictive density of the mixed causal/noncausal model for the coffee futures series. The parameters of the model are those estimated in section 5.3, where the shock is skew t-distributed.

In the implementation of the SIR approach, the instrumental distribution, that is the importance function has to be chosen close to the conditional distribution used to simulate the future asset price paths, that is, the predictive distribution outlined above. We select as the instrumental distribution a multivariate Gaussian distribution. Such a Gaus-

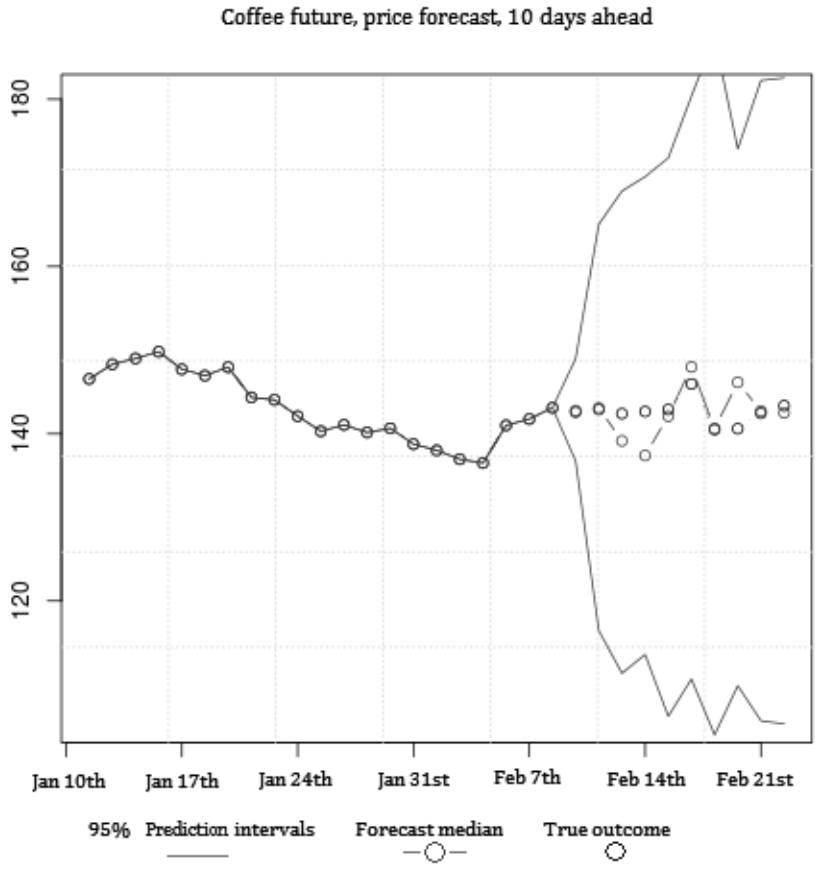
[15] Feburary 9th and 10th fall on a weekend.

sian distribution is parametrized by the vector of means and by the variance-covariance matrix. However the first and second order moments of the conditional distribution do not necessarily exist.

Therefore, the matching of the two distributions has to be based on other existing moments. Among the possible alternatives are calibrations based on the joint characteristic function, or calibration based on the first and second order moments of the square root of the absolute values of future prices, which exist. We have followed the second calibration, which has the advantage of leading to a number of moment restrictions equal to the number of parameters to be matched. Finally note that both the square root marginal and cross moments of the conditional distribution of interest, and of the Gaussian approximation, have no closed form expression and have to be computed numerically; for instance by reapplying the modified Look-ahead estimator for the conditional distribution.

The following plot in **Figure 9** provides the forecasted conditional median, and 95% prediction intervals.

Figure 9: Forecast predictive density for Coffee futures price series



8 Conclusion

The mixed causal/noncausal autoregressive model is able to capture asymmetries and bubble features present in the data on commodity futures prices. It improves model fit over the causal ARMA model with Gaussian innovations, according to the AIC criterion, since the mixed causal/noncausal autoregressive specification takes into account possible noncausality. This noncausality is unidentified in the traditional time series model, that is the purely causal ARMA model with Gaussian innovations. Estimation

of the purely causal ARMA models with fat tailed, t-distributed, innovations emphasizes the noncausal nature of most series, where often the causal lag polynomial roots lie inside the complex unit circle.

Moreover, inspection of the causal and noncausal lag polynomial roots of the mixed causal/noncausal autoregressive models suggest that longitudinal asymmetries can be accounted for by varying the causal and noncausal coefficient weights. Moreover, allowing for a low degree of freedom in the fat tailed t-distribution of the error term can account for bubble like phenomenon and these bubbles can induce transversal asymmetries if the model's shock, ϵ_t, admits a skewed distribution. In this way the model can account for both the longitudinal and transversal asymmetries described in Ramsey and Rothman (1996).

Furthermore, a comparison of the unconditional distributions, by sample histogram and Kullback-Leibler measure, suggest that the mixed causal/noncausal model with t-distributed shocks is a much closer approximation to the data than the equivalent purely causal ARMA model.

Finally, taking into account noncausal components is especially important when producing forecasts. Indeed, the standard Gaussian causal model will provide smooth term structure of linear forecasts with some long run equilibria. These forecasts are misleading in the presence of a noncausal component. Moreover, in many cases, including the energy and metals sectors, the causal polynomial admits explosive roots and so the forecasts do not exist. Employing a mixed causal/noncausal model therefore permits us to forecast the occurrence of future bubbles, including when they begin their build-up, when they crash, and what will be their magnitude.

9 References

ANDREWS, B., R.A. DAVIS AND F.J. BREIDT (2006): "Maximum Likelihood Estimation for All-Pass Time Series Models," *Journal of Multivariate Analysis,* 97, 1638-1659.

AZZALINI, A., AND A. CAPITANIO (2003): "Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t-Distribution," *Journal of the Royal Statistical Society, Series B,* 65, 2, 367-389.

BLACK, F. (1976): "The Pricing of Commodity Contracts," *The Journal of Financial Economics,* 3, 167-179.

BLANCHARD, O.J. (1979): "Speculative Bubbles, Crashes, and Rational Expectations," *Economic Letters*, 3, 387-389.

BLANCHARD, O.J., AND M. WATSON (1982): "Bubbles, Rational Expectations and Financial Markets," *National Bureau of Economic Research,* Working Paper No. 945.

BLANK, S.C. (1991): "Chaos in Futures Markets? A Nonlinear Dynamical Analysis," *The Journal of Futures Markets,* 11, 6, 711-728.

BLOOMBERG L.P. (2013) "Futures Price Data for Various Continuous Contracts," Bloomberg database, University of Toronto, Mississauga, Li Koon Chun Finance Learning Center.

BREEDEN, D. (1979): "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7, 3, 265-296.

BREIDT, J., R. DAVIS, K. LII, AND M. ROSENBLATT (1991): "Maximum Likelihood Estimation for Noncausal Autoregressive Processes," *Journal of Multivariate Analysis*, 36, 175-198.

BRENNAN, M.J. (1958): "The Supply of Storage," *American Economic Review,* 47, 50-72.

————————- (1991): "The Price of Convenience and the Valuation of Commodity Contingent Claims," in *Stochastic Models and Options Values,* ed. by D. Land, and B. Oksendal, Elsevier Science Publishers.

BROCK, W.A., W.D. DECHERT, J. SCHEINKMAN, AND B. LEBARON (1996): "A Test for Independence Based on the Correlation Dimension," *Econometric Reviews,* 15, 3, 197-235.

BROCK, W.A., AND C.H. HOMMES (1998): "Heterogenous Beliefs and Routes to Chaos in a Simple Asset Pricing Model," *Journal of Economic Dynamics and Control,* 22, 8-9, 1235-1274.

BROOKS, C., E. LAZAR, M. PROKOPCZUK, AND L. SYMEONIDIS (2011): "Futures Basis, Scarcity and Commodity Price Volatility: An Empirical Analysis," *International Capital Markets Association Center,* Working Paper, University of Reading.

CHENG, Q. (1992): "On the Unique Representation of Non Gaussian Linear Processes," *Annals of Statistics*, 20, 1143-1145.

CRUZ LOPEZ, J., J.H. HARRIS, C. HURLIN, AND C. PERIGNON (2013): "CoMargin," *Bank of Canada,* Working Paper.

DEATON, A., AND G. LAROQUE (1996): "Competitive Storage and Commodity Price Dynamics," *Journal of Political Economy,* 104, 5, 896-923.

DECOSTER, G.P., W.C. LABYS, AND D.W. MITCHELL (1992): "Evidence of Chaos in Commodity Futures Prices," *The Journal of Futures Markets,* 12, 3, 291-305.

DUSAK, K. (1973): "Futures Trading and Investor Returns: An Investigation of Com-

modity Market Risk Premiums," *Journal of Political Economy,* 81, 1387-1406.

EPANECHNIKOV, V.A. (1969): "Non-Parametric Estimation of a Multivariate Probability Density," *Theory of Probability and its Applications*, 14, 153158.

EVANS, G. (1991): "Pitfalls in Testing for Explosive Bubbles in Asset Prices," *The American Economic Review*, 81, 4, 922-930.

FAMA, E.F., AND K.R. FRENCH (1987): "Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage," *The Journal of Business,* 60, 1, 55-73.

FINDLEY, D.F. (1986): "The Uniqueness of Moving Average Representations with Independent and Identically Distributed Random Variables for Non-Gaussian Stationary Time Series," *Biometrika,* 73, 2, 520-521.

FROST, R. (1986): *Trading Tactics: A Livestock Futures Anthology*, ed. by Todd Lofton, Chicago Mercantile Exchange.

FULKS, B. (2000): "Back-Adjusting Futures Contracts," *Trading Recipes DB,* `http://www.trade2win.com/boards/attachments/commodities/90556d1283158105-rolling-futures-contracts-cntcontr.pdf`.

GARIBOTTI, G. (2013): "Estimation of the Stationary Distribution of Markov Chains," *PhD Dissertation*, University of Massachusetts, Amherst, Department of Mathematics and Statistics.

GAY, D.M. (1990): "Usage Summary for Selected Optimization Routines," *Computing Science Technical Report,* No. 153, `https://r-forge.r-project.org/scm/viewvc.php/*checkout*/pkg/Rnlminb2/inst/doc/PORT.pdf?revision=4506&root=rmetrics`, New Jersey: AT&T Bell Labs.

GIBSON, R., AND E.S. SCHWARTZ (1990): "Stochastic Convenience Yield and the

Pricing of Oil Contingent Claims," *The Journal of Finance,* 45, 3, 959-976.

GLYNN, P., AND S. HENDERSON (1998): *Estimation of Stationary Densities for Markov Chains*, Winter Simulation Conference, ed. by D. Medeiros, E. Watson, J. Carson and M. Manivannan, Piscataway, NJ: Institute of Electrical and Electronics Engineers.

GOODWIN, B.K., AND N.E. PIGGOTT (2001): "Spatial Market Integration in the Presence of Threshold Effects," *The American Journal of Agricultural Economics,* 83, 2, 302-317.

GOURIEROUX, C., AND J. JASIAK (2005): "Nonlinear Innovations and Impulse Responses with Application to VaR Sensitivity," *Annals of Economics and Statistics,* 1-31.

————————- (2013), "Filtering, Prediction, and Estimation of Noncausal Processes," *CREST,* DP.

GOURIEROUX, C., J.J. LAFFONT, AND A. MONFORT (1982): "Rational Expectations in Dynamic Linear Models: Analysis of the Solutions," *Econometrica,* 50, 2, 409-425.

GOURIEROUX, C., AND J.M ZAKOIAN (2012): "Explosive Bubble Modelling by Noncausal Cauchy Autoregressive Process," *Center for Research in Economics and Statistics,* Working Paper.

GRASSBERGER, P., AND I. PROCACCIA (1983): "Measuring the Strangeness of Strange Attractors," *Physica D: Nonlinear Phenomena,* 9, 1, 189-208.

HALLIN, M., C. LEFEVRE, AND M. PURI (1988): "On Time-Reversibility and the Uniqueness of Moving Average Representations for Non-Gaussian Stationary Time Series," *Biometrika*, 71, 1, 170-171.

HANSEN, L.P, AND T.J. SARGENT (1991): "Two Difficulties in Interpreting Vector

Autogressions," in *Rational Expectations Econometrics*, ed. by L.P. Hansen and T.J. Sargent, Boulder, CO: Westview Press Inc., 77-119.

HYNDMAN, R.J. AND Y. KHANDAKAR (2008): "Automatic Time Series Forecasting: The Forecast Package for R", *Journal of Statistical Software*, 27, 3.

JONES, M.C. (2001): "A Skew-t Distribution," in *Probability and Statistical Models with Applications,* ed. by A. Charalambides, M.V. Koutras, and N. Balakrishnan, Chapman & Hall/CRC Press.

KALDOR, N. (1939): "Speculation and Economic Stability," *Review of Economic Studies*, October, 7, 1-27.

KNITTEL, C.R., AND R.S. PINDYCK (2013): "The Simple Economics of Commodity Price Speculation," *National Bureau of Economic Research,* Working Paper No. 18951.

LANNE, M., J. LUOTO, AND P. SAIKKONEN (2012): "Optimal Forecasting of Noncausal Autoregressive Time Series," *International Journal of Forecasting,* 28, 3, 623-631.

LANNE, M., H. NYBERG, AND E. SAARINEN (2011): "Forecasting U.S. Macroeconomic and Financial Time Series with Noncausal and Causal Autoregressive Models: a Comparison," *Helsinki Center of Economic Research,* Discussion Paper No. 319.

LANNE, M., AND P. SAIKKONEN (2008): "Modeling Expectations with Noncausal Autoregressions," *Helsinki Center of Economic Research,* Discussion Paper No. 212.

LJUNG, G., AND E.P. BOX (1978): "On a Measure of a Lack of Fit in Time Series Models," *Biometrika,* 65, 2, 297-303.

LOF, M. (2011): "Noncausality and Asset Pricing," *Helsinki Center of Economic Research,* Discussion Paper No. 323.

MASTEIKA, S., A.V. RUTKAUSKAS, J.A. ALEXANDER (2012): ""Continuous Futures Data Series for Back Testing and Technical Analysis," *International Conference on Economics, Business and Marketing Management,* 29, Singapore: IACSIT Press.

MUTH, J. (1961): "Rational Expectations and the Theory of Price Movements," *Econometrica*, 29, 315-335.

NEFTCI, S.N. (1984): "Are Economic Time Series Asymmetric Over the Business Cycle," *Journal of Political Economy,* 92, 307-328.

NELDER, J.A., AND R. MEAD (1965): "A Simplex Method for Function Minimization," *The Computer Journal,* 7, 4, 308-313.

NOLAN, J. (2009) "Stable Distributions: Models for Heavy Tailed Data," `http://academic2.american.edu/~jpnolan/stable/chap1.pdf`, American University.

RAMIREZ, O.A. (2009): "The Asymmetric Cycling of U.S. Soybeans and Brazilian Coffee Prices: An Opportunity for Improved Forecasting and Understanding of Price Behavior," *Journal of Agricultural and Applied Economics,* 41, 1, 253-270.

RAMSEY, J., AND P. ROTHMAN (1996): "Time Irreversibility and Business Cycle Asymmetry," *Journal of Money and Banking,* 28, 1-21.

ROSENBLATT, M. (2000): *Gaussian and Non-Gaussian Linear Time Series and Random Fields,* New York: Springer Verlag.

ROSS, S. (1976): "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory,* 13, 3, 341-360.

RUBIN, D.B. (1988): "Using the SIR Algorithm to Simulate Posterior Distributions," *Bayesian Statistics*, 3, ed. by J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, Cambridge, MA: Oxford University Press, 395-402.

SHARPE, W.F. (1964): “Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk,” *Journal of Finance*, 19, 3, 425-442.

SIGL-GRUB, C., AND D. SCHIERECK (2010): “Speculation and Nonlinear Price Dynamics in Commodity Futures Markets,” *Investment Management and Financial Innovations,* 7, 1, 62-76.

SMITH, A.F.M, AND A.E. GELFAND (1992): “Bayesian Statistics Without Tears: A Sampling-Resampling Perspective,” *The American Statistician*, 46, 2, 84-88.

TERASVIRTA, T. (1994): “Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models,” *Journal of the American Statistical Association,* 89, 425, 208-218.

TONG, H., AND K.S. LIM (1980): “Threshold Autoregression, Limit Cycles, and Cyclical Data,” *Journal of the Royal Statistical Society, Series B,* 42, 3, 245-292.

TSAY, R.S (2010): *Analysis of Financial Time Series,* 3rd ed., New Jersey: Wiley Press.

WHITE, H. (1980): ”A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity”, *Econometrica*, 48, 4, 817-838.

WORKING, H. (1933): “Price Relations Between July and September Wheat Futures at Chicago Since 1885,” *Wheat Studies of the Food Research Institute,* 9, 6, 187-238.

————————- (1948): “Theory of the Inverse Carrying Charge in Futures Markets.” *Journal of Farm Economics,* 30, 1, 1-28.

————————- (1949): “The Theory of the Price of Storage,” *American Economic Review,* 39, 1254-1262.

YANG, S.R., AND W. BRORSEN (1993): “Nonlinear Dynamics of Daily Futures

Prices: Conditional Heteroskedasticity or Chaos?," *The Journal of Futures Markets,* 13, 2, 175-191.

10 Appendix: Rolling over the futures contract

Consider first, the "fair price" of the futures contract implied by the spot-futures parity theorem. The theorem implies that, given the assumption of well functioning competitive markets, a constant, annual, risk-free rate of interest r_f and a cost of carry c, no arbitrage should ensure that the following relationship between the futures and spot price of the underlying commodity holds at time t:

$$F_{t,t+k} = S_t(1 + \frac{k}{365}(r_f + c)), \tag{47}$$

where $c \in [0, 1]$. That is, given the exploitation of arbitrage opportunities, we should have that the cost of purchasing the underlying good at price S_t today and holding it until $t + k$ (given opportunity cost of capital and cost of carry) should be equal to the current futures price $F_{t,t+k}$. Of course, this relationship implies that as the maturity date approaches (i.e. as $k \to 0$) we have that $F_{t,t} = S_t$.

This relationship is an approximate one and will not hold exactly in reality: indeed, the risk-free rate and the cost of carry vary in time and are uncertain, and some goods are perishable and cannot be stored indefinitely. Nevertheless this relationship is useful for considering the rolling over of futures contracts, since if we keep a given futures contract in a portfolio, its residual maturity will decrease. The formula in (47) demonstrates this effect and the need to adjust the futures price series level if we want it to maintain the same residual maturity.

Upon the approach of the futures' maturity, we also wish to extend the price series and obtain price data for each date. In order to do so we would have to close out our current position and then open a new position in the futures contract of the next maturity. For example, suppose we are holding a futures contract that expires at time $t + k$ and k is approaching 0. We could sell this futures contract and purchase a new contract on the same underlying good but that expires at time $t + k + j$. However in doing so we would clearly incur a loss since we have that:

$$1 + \frac{k}{365}(r_f + c) < 1 + \frac{k + j}{365}(r_f + c) \tag{48}$$

by the spot-futures parity theorem. This is known as *rollover risk* and the difference in

the two prices is called the *calendar spread*.

However, this loss for the trader should not be considered as part of the overall price series historical data we use for forecasting since it represents a predictable discontinuity in the series. Therefore typically futures price series are also adjusted for this calendar spread by the data provider. There are a few ways to go about doing this, each with their pros and cons: [16]

1. Just append together prices without any adjustment. This will distort the series, by including spurious autocorrelation.

2. Directly adjust the prices up or down according to either the new or old contract at the rollover time period. This can be done by simply subtracting the difference between the two price series, or multiplying one of the price series by ratio of the two (i.e. absolute difference or relative difference, respectively). This method works, but it causes either the newer or older contract prices to diverge further and further from their original values as we append additional contracts. Moreover, it leaves the choice of adjustment a rather arbitrary one.

3. Continuously adjust the price series over time. This method melds together the futures contract prices of both the "front month" contract (i.e. the contract with the shortest time-to-maturity) with the contracts of longer times-to-maturity (called the "back month" contracts) in a continuous manner. This allows us the potential to create a continuous contract price which reflects an "unobserved" futures contract which maintains a fixed time-to-maturity as time progresses. Ultimately, we are free to choose a model whereby we can reconstitute the unobserved futures contract price by employing information in the prices of observed contracts of different maturities.

   Example: Smooth transition model

   Consider two futures contracts on the same underlying commodity, one with time-to-maturity k, the other with time-to-maturity $k + j$, where we assume that their prices, $F_{t,t+k}$ and $F_{t,t+k+j}$, approximately satisfy the no arbitrage condition of the spot-futures parity theorem. Moreover, let $\epsilon_{i,t}$ for $i = 1, 2$ be error terms

[16] See Fulks (2000), a widely disseminated PDF document available on the world wide web. Alternatively, Masteika et al. (2012) provides a more recent treatment of the relevant issues.

satisfying the standard assumptions of a regression model. The price variables $F_{t,t+k}$, $F_{t,t+k+j}$, and S_t are observable, as is the current risk free rate $r_{f,t}$. The cost of carry c_t, is unobservable since it includes a convenience yield, and so we must estimate it. Either way, we can then write down the model:

$$F_{t,t+k} = S_t(1 + \frac{k}{365}(r_{f,t} + c_t)) + \epsilon_{1,t} \tag{49a}$$

$$F_{t,t+k+j} = S_t(1 + \frac{k+j}{365}(r_{f,t} + c_t)) + \epsilon_{2,t} \tag{49b}$$

$$P_t = \alpha F_{t,t+k} + (1 - \alpha)F_{t,t+k+j} \tag{49c}$$

where $\epsilon_{i,t}$ represents a residual deviation away from the spot-futures parity fair value, $\alpha = \frac{k}{K}$, where K is an upper bound on $k + j$ (that is it represents the time to maturity when the future is first issued) and j is sufficiently large so that the difference in futures prices aren't negligible (typically $j \geq 30$ since futures contracts of different maturities are indexed by month).

P_t, therefore, represents our estimate of the unobserved contract which incorporates the information in the front and back month contracts. Since the spot-futures parity doesn't hold exactly, P_t reflects not just the spot price S_t, the risk free rate $r_{f,t}$, and the cost of carry c_t; but also some residual error factors $\epsilon_{i,t}$ for $i = 1, 2$.

The Bloomberg console allows the user to specify various criteria which modify how the continuous contract price series is constructed from the front and back month contracts. Any of the 3 methods above are available for use. In constructing the price series data employed in this paper I use a method similar to (3) above but simpler in its weighting. The continuous contract futures price P_t is equal to the front month contract price $F_{t,t+k}$ until the contract has 30 days left to maturity, so that $k = 30$. At that point, the continuous contract reflects the weighted average between the front month and the next back month contract, with the weights reflecting the number of days left until maturity of the front month contract. That is,

$$P_t = \left(\frac{k}{d}\right) F_{t,t+k} + \left(\frac{k-x}{d}\right) F_{t,t+k+j} \tag{50}$$

where $d = 30$ represents the total number of days in the month and k is the number of days remaining in the month. Once $k = 0$, the price is then $P_t = F_{t,t+j}$, until this new

front month contract again has 30 days left until maturity, or $j = 30$. If the difference in time-to-maturity for all contracts is fixed at 30 days (i.e. a different contract matures every month), then this scheme represents the reconstitution of an unobserved futures contract with a fixed time to maturity of 30 days, as time progresses forward indefinitely.

11 Appendix: Mixed causal/noncausal process

In this appendix we provide the definitions of mixed causal/noncausal processes and review several of their properties employed in the main part of the text.

11.1 Strong moving average

The infinite moving average $Y_t = \sum_{i=-\infty}^{\infty} a_i \epsilon_{t-i}$, where (ϵ_t) is a sequence of i.i.d. variables, that is a strong white noise, can be defined for a white noise without first and/or second order moments.

Let us consider the Banach space L^p of the real random variables such that $\|Y\|_p = \sqrt{E[|Y|^p]}$ exists, for a given p. For expository purposes we consider the Banach space which requires $p \geq 1$. However, the existence of the process can also be proved for $0 < p \leq 1$. If $\|\epsilon_t\|_p = \sqrt{E[|\epsilon_t|^p]}$ exists and if the set of moving average coefficients is absolutely convergent, $\sum_{i=-\infty}^{\infty} |a_i| < \infty$, then the series with elements $a_i \epsilon_{t-i}$ is such that $\sum_{i=-\infty}^{\infty} \|a_i \epsilon_{t-i}\|_p = \sum_{i=-\infty}^{\infty} |a_i| \|\epsilon_{t-i}\|_p = \left(\sum_{i=-\infty}^{\infty} |a_i|\right) \|\epsilon_{t-i}\|_p < \infty$, since $\|\epsilon_t\|_p$ is independent of date t. Thus the series with elements $a_i \epsilon_{t-i}$ is normally convergent. In particular the variable $Y_t = \sum_{i=-\infty}^{\infty} a_i \epsilon_{t-i}$ has a meaning for the $\|\cdot\|_p$ convergence in the sense that

$$Y_t = \lim_{n \to \infty, m \to \infty} \sum_{i=-m}^{n} a_i \epsilon_{t-i}, \tag{51}$$

where the limit is with respect to the L^p-norm. Moreover, the limit Y_t has a finite L^p-norm, such that $\|Y_t\|_p \leq \left(\sum_{i=-\infty}^{\infty} |a_i|\right) \|\epsilon_{t-i}\|_p < \infty$.

The L^p convergence implies the convergence in distribution. The distribution of the process (ϵ_t) is invariant with respect to the lag of time, that is to the operator L which transforms the process (ϵ_t) into $L(\epsilon_t) = (\epsilon_{t-1})$. Since the process (Y_t) is derived from the white noise (ϵ_t) by a time invariant function, we deduce that the distribution of (Y_t) is the same as the distribution of $L(Y_t) = (Y_{t-1})$, that is (Y_t) is a strong stationary

process.

Similar arguments apply to any moving average transformation of a strongly stationary process existing in L^p, that is to:

$$X_t = \sum_{i=-\infty}^{\infty} b_j Y_{t-j}, \tag{52}$$

whenever $\sum_{j=-\infty}^{\infty} |b_j| < \infty$, since $\|Y_t\|_p$ is finite and time independent. In particular, we can as usual compound moving averages. From the equations:

$$Y_t = \sum_{i=-\infty}^{\infty} a_i \epsilon_{t-i} = a(L)\epsilon_t, \quad \text{with} \quad a(L) = \sum_{i=-\infty}^{\infty} a_i L^i, \tag{53a}$$

$$X_t = \sum_{j=-\infty}^{\infty} b_j Y_{t-j} = b(L)Y_t, \quad \text{with} \quad b(L) = \sum_{j=-\infty}^{\infty} b_j L^j, \tag{53b}$$

we can deduce

$$X_t = b(L)a(L)\epsilon_t, \tag{54}$$

that is, the moving average representation of process (X_t) in terms of the underlying strong white noise (ϵ_t). The new moving average operator

$$c(L) = b(L)a(L) = \sum_{k=-\infty}^{\infty} c_k L^k \tag{55}$$

admits moving average coefficients given by

$$c_k = \sum_{i=-\infty}^{\infty} a_i b_{k-i} = \sum_{j=-\infty}^{\infty} a_{k-j} b_j, \quad \forall k. \tag{56}$$

11.2 Identification of a strong moving average representation

The question of the identification of a strong moving average representation is as follows. Let us consider a strong moving average process in L^p, $Y_t = \sum_{i=-\infty}^{\infty} a_i \epsilon_{t-i}$. Is it possible to also write this process as $Y_t = \sum_{i=-\infty}^{\infty} a_i^* \epsilon_{t-i}^*$, that is, with different noise and moving average coefficients? Of course the white noise is defined up to a

multiplicative positive scalar c, since

$$Y_t = \sum_{i=-\infty}^{\infty} a_i^* \epsilon_{t-i}^*, \quad \text{with} \quad a_i^* = a_i/c, \epsilon_t^* = c\epsilon_t. \tag{57}$$

The identification conditions below have been derived previously in Findley (1986), Hallin, Lefevre, and Puri (1988), and Cheng (1992).

Identification condition

i) The moving average representation is identifiable up to a multiplicative positive scalar and to a drift of the time index for the noise process, if and only if the distribution of the white noise is not Gaussian.

ii) If the white noise is Gaussian, the process always admits a causal Gaussian representation,

$$Y_t = \sum_{i=0}^{\infty} a_i^* \epsilon_{t-i}^*, \quad \text{with} \quad \epsilon_t^* \sim IIN(0,1). \tag{58}$$

As a consequence the general linear process which is not purely causal, that is which depends on at least one future shock (i.e. $a_i \neq 0$ for at least one negative time index i) cannot admit a linear causal representation. Equivalently, its causal representation will automatically feature nonlinear dynamic features.

11.3 Probability distribution functions of the stationary strong form noncausal representation

It can be shown that the unconditional distribution of the process in equation (4) is given as

$$f_t(x_t) = \frac{1-|\rho|}{\sigma_\epsilon \pi} \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + (1-|\rho|)^2 x_t^2}. \tag{59}$$

[Gourieroux and Zakoian (2012), Proposition 1] This unconditional distribution is independent of date t by the strong stationary property.

Moreover, the Markov transition distribution (conditional density) of the forward-

looking process is given as

$$f_{t|t+1}(x_t|x_{t+1}) = \frac{1}{\sigma_\epsilon \pi} \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + z_t^2}, \quad \text{where} \quad z_t = \frac{x_t - \rho x_{t+1}}{\sigma_\epsilon}, \tag{60}$$

which follows from the definition of the standard Cauchy distribution.

Therefore, from Bayes theorem along with equations (59) and (60), we have that

$$f_{t+1|t}(x_{t+1}|x_t) = f_{t|t+1}(x_t|x_{t+1}) f_{t+1}(x_{t+1}) / f_t(x_t) \tag{61a}$$

$$= \frac{1}{\sigma_\epsilon \pi} \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + z_t^2} \frac{\frac{1-|\rho|}{\sigma_\epsilon \pi} \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + (1-|\rho|)^2 x_{t+1}^2}}{\frac{1-|\rho|}{\sigma_\epsilon \pi} \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + (1-|\rho|)^2 x_t^2}} \tag{61b}$$

$$= \frac{1}{\sigma_\epsilon \pi} \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + z_t^2} \frac{\sigma_\epsilon^2 + (1-|\rho|)^2 x_t^2}{\sigma_\epsilon^2 + (1-|\rho|)^2 x_{t+1}^2}, \tag{61c}$$

which provides the causal transition density of the process [Gourieroux and Zakoian (2012), Proposition 2].

11.4 The causal strong autoregressive representation

A nonlinear causal innovation, (η_t), of the process (x_t) is a strong white noise such that we can write the current value of the process x_t as a nonlinear function of its own past value x_{t-1} and η_t: $x_t = G(x_{t-1}, \eta_t)$, say, where x_t and η_t are in a continuous one-to-one relationship given any x_{t-1} [Rosenblatt (2000)].

Moreover, since the conditional cumulative distribution function of $x_t|x_{t-1}$ is strictly monotone increasing and continuous, it has an inverse. We can write

$$x_t = F^{-1}(\Phi(\eta_t)|x_{t-1}) \quad \text{where} \quad \eta_t \sim IIN(0,1) \tag{62a}$$

$$\Leftrightarrow \eta_t = \Phi^{-1}[F(x_t|x_{t-1})], \tag{62b}$$

and $F(\cdot|x_{t-1})$ is the c.c.d.f. of x_t while $\Phi(\cdot)$ is the c.d.f. of the standard Normal distribution. Therefore, by choosing $G(x_{t-1}, \eta_t) = F^{-1}(\Phi(\eta_t)|x_{t-1})$, we can select a Gaussian causal innovation. The choice of a Gaussian causal innovation is purely conventional.

11.5 Distributions with fat tails

Different distributions with fat tails can be used as the distribution of the baseline shocks (ϵ_t) to construct mixed causal/noncausal linear processes. Below we provide three examples of fat tailed distributions that are employed in this paper, that are the student t-distribution, the skewed student t-distribution [see Jones (2001)], and the "stable" distributions [see Nolan (2009)], respectively.

i) **Student t-distribution:**
This is a distribution on $(-\infty, +\infty)$ with probability density function given as:

$$f(x) = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \tag{63}$$

where $\nu > 0$ is the real degree of freedom parameter and $\Gamma(\cdot)$ is the Gamma function defined as $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$, if $z > 0$.

The p.d.f. is symmetric; it bears the same "bell" shape as the Normal distribution except that the t-distribution exhibits fat tails. As the number of degrees of freedom, ν goes to 1 the t-distribution approaches the Cauchy distribution and as the degree of freedom approaches ∞, the t-distribution approaches the Normal distribution.

Its tail behaviour is such that $E[|x|^p] < \infty$, if $\nu > p$.

ii) **Skewed t-distribution:** [Jones (2001), Section 17.2]
This is a distribution on $(-\infty, +\infty)$ with probability density function given as:

$$f(x) = \frac{1}{2^{\nu-1}\beta(a,b)\sqrt{\nu}} \left(1 + \frac{x}{\sqrt{\nu + x^2}}\right)^{a+1/2} \left(1 - \frac{x}{\sqrt{\nu + x^2}}\right)^{b+1/2}, \tag{64}$$

where $\nu = a+b$, a and b are two positive real valued degrees of freedom parameters and $\beta(a, b)$ represents the Beta function defined as $\beta(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$. If $a > b$ the distribution is positively skewed, negatively skewed if $a < b$, and identical to the t-distribution above if $a = b$. This distribution allows for different magnitudes for the left and right fat tails, respectively.

Another skewed t-distribution has been proposed in the literature as a generalization of the skewed Normal distribution. This alternative skewed t-distribution is

parameterized by only one skewness parameter instead of two as in Jones (2001) [see Azzalini and Capitanio (2003), Section 4, for more details].

iii) **Stable distribution:**

A random variable, x, is said to be "stable," or to have a "stable distribution," if a linear combination of two independent copies of x has the same distribution as x, up to location and scale parameters. That is, if x_1 and x_2 are independently drawn from the distribution of x, then x is stable if if for any constants $a > 0$ and $b > 0$ the random variable $z = ax_1 + bx_2$ has the same distribution as $cx + d$ for some constants $c > 0$ and d. The distribution is said to be strictly stable if $d = 0$.

Generally, we cannot express the p.d.f. of the stable random variable x in an analytical form. However, the p.d.f is always expressable as the Fourier transform of the characteristic function, $\varphi(t)$, which always exists, that is, $f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\varphi(t)e^{-ixt}dt$. The characteristic function is given as:

$$\varphi(t) = \exp\left[it\mu - |ct|^{\alpha}\left(1 - i\beta sign(t)\tan\left(\pi\alpha/2\right)\right)\right] \tag{65}$$

Therefore the distribution is parameterized by $\{\alpha, \beta, c, \mu\}$ where $\alpha \in (0, 2]$ is the stability parameter, $\beta \in [-1, 1]$ is a skewness parameter, $c \in (0, \infty)$ is the scale parameter, and $\mu \in (-\infty, \infty)$ is the location parameter.

The Normal, Cauchy, and Levy distributions are all stable continuous distributions. If $\alpha = 2$ the stable distribution reduces to the Normal distribution. If $\alpha = 1/2$ and $\beta = 1$, it corresponds to the Levy distribution. Finally, if $\alpha = 1$ and $\beta = 0$ the distribution is Cauchy and the p.d.f. is given analytically as:

$$f(x) = \frac{1}{\pi\left(1 + x^2\right)}. \tag{66}$$

Even if the p.d.f. of a stable distribution has no explicit expression, its asymptotic behaviour is known. We have [see Nolan (2009), Th 1.12]:

$$f(x) \sim c^{\alpha}\frac{(1 + sign(x)\beta)sin(\pi\alpha/2)\Gamma(\alpha + 1)/\pi}{|x|^{1+\alpha}}, \quad \text{for large} \quad x. \tag{67}$$

Therefore, $E[|x|^p] < \infty$, if $\alpha > p$. In particular the mean does not exist if $\alpha \leq 1$.

12 Appendix: Approximation of the mixed causal/noncausal AR(r, s) likelihood

This section describes the nature of the matrix transformations which ensure that the MLE estimator is consistent, by regressing both forward (noncausal) and backward (causal) lags on x_t.

It will first be useful to define the following processes u_t and v_t. From (17a), let u_t be defined as

$$u_t = \phi(L)x_t = \varphi(L^{-1})^{-1}\epsilon_t = \sum_{j=0}^{\infty} \varphi_j^* \epsilon_{t+j}, \tag{68}$$

where $\varphi_0^* = 1$ and the right hand side series of moving average coefficients are absolutely summable. We call (68) the *forward looking* moving average representation of x_t.

Moreover, also from (17a) let v_t be defined as

$$v_t = \varphi(L^{-1})x_t = \phi(L)^{-1}\epsilon_t = \sum_{j=0}^{\infty} \phi_j^* \epsilon_{t-j}, \tag{69}$$

where $\phi_0^* = 1$ and the right hand side series of moving average coefficients are absolutely summable. We call (69) the *backward looking* moving average representation of x_t.

The changes of variables above can also be written in matrix form. Consider the time series x_t for $t = 1, \ldots, T$. From (68) and (69), we have $u_t = \phi(L)x_t$ and $v_t = \varphi(L^{-1})x_t$. Therefore, let us introduce the following matrices, $\mathbf{\Phi_c}$ and $\mathbf{\Phi_{nc}}$:

$$\mathbf{\Phi_c} = \begin{pmatrix} \boldsymbol{I}_{r\times r} & & & & & & & & \mathbf{0}_{r\times(T-r)} & \\ \hdashline -\phi_r & -\phi_{r-1} & \ldots & -\phi_1 & 1 & 0 & \ldots & \ldots & \ldots & \ldots \\ 0 & -\phi_r & -\phi_{r-1} & \ldots & -\phi_1 & 1 & 0 & \ldots & \ldots & \ldots \\ & & & & & & \ddots & & & \\ \ldots & \ldots & 0 & -\phi_r & -\phi_{r-1} & \ldots & -\phi_1 & 1 & 0 & \ldots \\ \hdashline & \mathbf{0}_{s\times(T-s)} & & & & & & & \boldsymbol{I}_{s\times s} & \end{pmatrix} \tag{70}$$

and

$$\boldsymbol{\Phi_{nc}} = \left(\begin{array}{ccccccccccc}
1 & -\varphi_1 & \dots & -\varphi_{s-1} & -\varphi_s & 0 & \dots & \dots & \dots & \dots \\
0 & 1 & -\varphi_1 & \dots & -\varphi_{s-1} & -\varphi_s & 0 & \dots & \dots & \dots \\
 & & \ddots & & & & & & & \\
 & & & \ddots & & & & & & \\
 & & & & \ddots & & & & & \\
\dots & \dots & \dots & \dots & 0 & 1 & -\varphi_1 & \dots & -\varphi_{s-1} & -\varphi_s \\
\hdashline
\dots & 0 & -\phi_r & -\phi_{r-1} & \dots & -\phi_1 & 1 & 0 & \dots & 0 \\
 & & & & & & & \ddots & & \\
\dots & \dots & \dots & 0 & -\phi_r & -\phi_{r-1} & \dots & -\phi_1 & 1 & 0 \\
\dots & \dots & \dots & \dots & 0 & -\phi_r & -\phi_{r-1} & \dots & -\phi_1 & 1
\end{array}\right) \tag{71}$$

where the lower partition of $\boldsymbol{\Phi_{nc}}$ has s rows. Therefore, $\boldsymbol{\Phi_c}$ will represent the causal transformation and $\boldsymbol{\Phi_{nc}}$ the noncausal transformation, respectively. Both matrices are of size $T \times T$.

Applying the noncausal transformation to the vector of data, $\boldsymbol{x}$, we have:

$$\begin{bmatrix} v_1 \\ \vdots \\ \vdots \\ v_{T-s} \\ u_{T-s+1} \\ \vdots \\ u_T \end{bmatrix} = \begin{bmatrix} x_1 - \varphi_1 x_2 - \cdots - \varphi_s x_{1+s} \\ \vdots \\ \vdots \\ x_{T-s} - \varphi_1 x_{T-s+1} - \cdots - \varphi_s x_T \\ x_{T-s+1} - \phi_1 x_{T-s} - \cdots - \phi_r x_{T-s+1-r} \\ \vdots \\ x_T - \phi_1 x_{T-1} - \cdots - \phi_r x_{T-r} \end{bmatrix} = \boldsymbol{\Phi_{nc}} \begin{bmatrix} x_1 \\ \vdots \\ \vdots \\ x_{T-s} \\ x_{T-s+1} \\ \vdots \\ x_T \end{bmatrix} \tag{72}$$

Moreover, from $\epsilon_t = \phi(L)\varphi(L^{-1})x_t = \phi(L)v_t$, we have:

$$\boldsymbol{e} = \begin{bmatrix} v_1 \\ \vdots \\ v_r \\ \epsilon_{r+1} \\ \vdots \\ \epsilon_{T-s} \\ u_{T-s+1} \\ \vdots \\ u_T \end{bmatrix} = \begin{bmatrix} v_1 \\ \vdots \\ v_r \\ v_{r+1} - \phi_1 v_r - \cdots - \phi_r v_1 \\ \vdots \\ v_{T-s} - \phi_1 v_{T-s-1} - \cdots - \phi_r v_{T-s-r} \\ u_{T-s+1} \\ \vdots \\ u_T \end{bmatrix} = \boldsymbol{\Phi_c} \begin{bmatrix} v_1 \\ \vdots \\ v_{T-s} \\ u_{T-s+1} \\ \vdots \\ u_T \end{bmatrix} \tag{73}$$

So we have the transformation $\boldsymbol{e} = \boldsymbol{\Phi_c}\boldsymbol{\Phi_{nc}}\boldsymbol{x}$.

Thus the elements of $\boldsymbol{e}$ are mutually independent and the joint density of $\boldsymbol{e}$ is given as:

$$f_e(\boldsymbol{e}|\theta) = f_v(v_1, \ldots, v_r) \left(\prod_{t=r+1}^{T-s} f_\epsilon(\epsilon_t; \lambda, \sigma) \right) f_u(u_{T-s+1}, \ldots, u_T), \tag{74}$$

where $\theta = \{\boldsymbol{\phi}, \boldsymbol{\varphi}, \lambda, \sigma\}$ represents the parameters of the model.

The $\boldsymbol{\Phi_c}$ matrix is lower triangular and its determinant is equal to 1. Therefore, using the change of variables Jacobian formula, we can express the joint density in terms of $\boldsymbol{x}$ as:

$$f_x(\boldsymbol{x}|\theta) = f_v(\varphi(L^{-1})x_1, \ldots, \varphi(L^{-1})x_r) \left(\prod_{t=r+1}^{T-s} f_\epsilon(\varphi(L^{-1})\phi(L)x_t; \lambda, \sigma) \right) f_u(\phi(L)x_{T-s+1}, \ldots, \phi(L)x_T)|det(\boldsymbol{\Phi_{nc}})|. \tag{75}$$

Since the determinant of $\boldsymbol{\Phi_{nc}}$ is independent of sample size,[17] we can approximate

[17] To show this we can employ the partitioned matrix determinant formula: $det\left(\begin{bmatrix} \boldsymbol{A_{11}} & \boldsymbol{A_{12}} \\ \boldsymbol{A_{21}} & \boldsymbol{A_{22}} \end{bmatrix}\right) = det\left(\boldsymbol{A_{11}}\right) det\left(\boldsymbol{A_{22}} - \boldsymbol{A_{21}}{\boldsymbol{A_{11}}}^{-1}\boldsymbol{A_{12}}\right)$, where it can be shown that $\boldsymbol{A_{11}}$ is $(T-s) \times (T-s)$ with determinant 1, and so the second term in the factorization represents the determinant of an $s \times s$ matrix, for all T.

asymptotically the likelihood by using the second factor in the above expression, that is,

$$\prod_{t=r+1}^{T-s} f_\epsilon(\varphi(L^{-1})\phi(L)x_t;\lambda,\sigma). \tag{76}$$

For large samples, T will dwarf $r+s=p$ and so the approximation will be consistent.

Asymptotic properties of the approximated maximum likelihood estimators are discussed in section 3.2 and consistent estimation of the standard errors is detailed in section 3.3, both of Lanne et al. (2008).

13 Appendix: Numerical algorithm for mixed causal/noncausal AR(r,s) forecasts

Solution proposed by Lanne, Luoto, and Saikkonen (2012)

Lanne, Luoto, and Saikkonen (2012) propose to circumvent the problem presented by our ignorance of the stationary distribution $f_x(\cdot)$ by enlarging the space of random variables. They first rewrite (37) as:

$$f_{x_{t+1|t}}(x_{t+1}|x_t) = f_{x_t,x_{t+1}}(x_t,x_{t+1})/f_x(x_t). \tag{77}$$

Then by using the fact that $x_t = u_t = \varphi(L^{-1})\epsilon_t = \sum_{j=0}^{\infty}\varphi_1^j\epsilon_{t+j}$, they choose to employ the mapping $(x_t,x_{t+1},x_{t+2},\dots) \to (\epsilon_t,\epsilon_{t+1},\epsilon_{t+2},\dots)$. This suggests a linear relationship which, by approximating $x_t = u_t \approx \sum_{j=0}^{M}\varphi_1^j\epsilon_{t+j}$ given a sufficiently large truncation lag M, we are able to invert, providing an approximate expression for ϵ_t as a linear function of both x_t and future $\epsilon_{t+1},\epsilon_{t+2},\dots,\epsilon_{t+M}$. For example in this case where the noncausal polynomial is of order 1, we have that $\hat{\epsilon}_t \approx x_t - \sum_{j=1}^{M}\varphi_1^j\epsilon_{t+j}$.

Since, by assumption, the distribution of the shocks ϵ_t is known, the authors are able to compute the probability of these approximated $\hat{\epsilon}_t$'s, and relying upon Monte-Carlo simulation methods, are able to approximate the conditional C.D.F. function of $x_{t+1} = u_{t+1}$. The conditional C.D.F. function at a given value $\alpha \in R$ can be computed from (77) above by means of approximating the following integral by Monte-Carlo simulation, where we average across draws of sufficiently long future paths of $\boldsymbol{\epsilon}_{t+1}^{+} =$

$\{\epsilon_{t+1}, \ldots, \epsilon_{t+M}\}$:

$$F_{x_{t+1|t}}(\alpha|x_t) = \int \mathbf{1}_{\alpha > x_{t+1}} f_{x_{t+1|t}}(x_{t+1}|x_t) dx_{t+1} \tag{78a}$$

$$\approx \int \frac{1}{f_\epsilon(\hat{\epsilon}_t)} \mathbf{1}_{\alpha > x_{t+1}} \left(\sum_{j=0}^{M-1} \varphi_1^j \epsilon_{t+1+j} \right) f_\epsilon(\hat{\epsilon}_t) \prod_{j=1}^{M} f_\epsilon(\epsilon_{t+j}) d\boldsymbol{\epsilon}_{t+1}^{+} \tag{78b}$$

This method has two drawbacks: first, we approximate the above integral by Monte-Carlo simulation of the long future paths of $\boldsymbol{\epsilon}_{t+1}^{+}$. Second, M has to be sufficiently large so that the approximation does not miss the effect of far future shocks. The value of M required to obtain an accurate approximation will grow as the roots of the noncausal polynomial approach 1, and so will the computational requirements of the algorithm.

The numerical method proposed by Lanne, Luoto, and Saikkonen (2012) also works in the more general case where $s > 1$. However, now that the noncausal order is greater than 1 enlarging the space from $(x_{t-s+1}, \ldots, x_t, x_{t+1}, \ldots) \to (\epsilon_{t-s+1}, \ldots, \epsilon_t, \epsilon_{t+1}, \epsilon_{t+2}, \ldots)$ requires us to invert a system of equations. Therefore, we may employ a matrix transformation between the two spaces and this matrix is inverted to provide an approximation to $\epsilon_t, \ldots, \epsilon_{t-s+1}$ in terms of both $x_t, \ldots, x_{t-s+1}$ and future $\epsilon_{t+1}, \ldots, \epsilon_{t+M}$. It is noted in their paper (and in the Appendix here) that the Jacobian determinant of this transformation is always 1. However, while this matrix is sparse, for large s and M it is computationally costly.

Below, we describe their method for the approximate simulation of the conditional c.d.f.,

$$F_{u_{t+h|t}}(\alpha|\mathcal{F}_t) = \int_{-\infty}^{\alpha} f_{u_{t+h|t}}(u_{t+h}|\mathcal{F}_t) du_{t+h} \tag{79}$$

for $h = 1$, when $s > 1$ (which they also generalized to the case where $h > 1$ in their paper). The method is broken down into a number of discussion points as follows:

1. We require the density of $\boldsymbol{\epsilon}_{t+1}^{+} = \{\epsilon_{t+1}, \epsilon_{t+2}, \ldots\}$, conditional on the data $\boldsymbol{x}_t = \{x_t, x_{t-1}, \ldots\ x_1\}$.

2. Since from (68), we have that $u_{t+1} = \sum_{j=0}^{\infty} \varphi_j^* \epsilon_{t+1+j}$, from equation (75) it can be shown that:

$$\frac{f_{x,\epsilon^+}(\boldsymbol{x}_t, \boldsymbol{\epsilon}_{t+1}^{+}|\theta)}{f_x(\boldsymbol{x}_t|\theta)} = p(\boldsymbol{\epsilon}_{t+1}^{+}|\boldsymbol{x}_t; \theta) = \frac{f_{u^-,\epsilon^+}(\boldsymbol{u}_t^{-}(\boldsymbol{\phi}), \boldsymbol{\epsilon}_{t+1}^{+})}{f_{u^-}(\boldsymbol{u}_t^{-}(\boldsymbol{\phi}))}, \tag{80}$$

where θ represents the parameters of the mixed causal/noncausal AR(r,s) model and
$\boldsymbol{u}_t^-(\boldsymbol{\phi}) = \{\phi(L)x_{t-s+1},\dots,\phi(L)x_t\} = \{u_{t-s+1},\dots,u_t\}$.

3. Then, we can use Monte-Carlo simulations to approximate both the numerator and denominator of (80) in order to approximate the desired conditional c.d.f. as:

$$F_{u_{t+1|t}}(\alpha|\mathcal{F}_t) \approx \frac{1}{f_u(\boldsymbol{u}_t^-(\boldsymbol{\phi}))}\int \mathbf{1}_{\alpha>u_{t+1}}\left(\sum_{j=0}^{M-1}\varphi_j^*\epsilon_{t+1+j}\right) f_{u,\epsilon^+}(\boldsymbol{u}_t^-(\boldsymbol{\phi}),\boldsymbol{\epsilon}_{t+1}^+)d\boldsymbol{\epsilon}_{t+1}^+, \tag{81}$$

where under the assumption of some finite M (such that as $M\to\infty$, $(\varphi_j^*)\to 0$), we can approximate u_{t+1} as $u_{t+1}\approx\sum_{j=0}^{M-1}\varphi_j^*\epsilon_{t+1+j}$.

4. In order to do this, however, we need to accomplish a change of variables between $(\boldsymbol{u}_t^-(\boldsymbol{\phi}),\boldsymbol{\epsilon}_{t+1}^+)$ and $(\{\epsilon_{t-s+1},\dots,\epsilon_t\},\boldsymbol{\epsilon}_{t+1}^+)$. Given (68), the approximate mapping between these two sets of variables is given as:

$$\begin{bmatrix} 1 & \varphi_1^* & \dots & \dots & \dots & \dots & \varphi_{M+s-1}^* \\ 0 & \ddots & \ddots & & & & \vdots \\ \vdots & \ddots & 1 & \varphi_1^* & \dots & \dots & \varphi_M^* \\ \vdots & & \ddots & 1 & 0 & \dots & 0 \\ \vdots & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & \dots & \dots & 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{t-s+1} \\ \vdots \\ \epsilon_t \\ \epsilon_{t+1} \\ \vdots \\ \epsilon_{t+M} \end{bmatrix} \approx \begin{bmatrix} u_{t-s+1} \\ \vdots \\ u_t \\ \epsilon_{t+1} \\ \vdots \\ \epsilon_{t+M} \end{bmatrix} \tag{82}$$

which can be written as $\boldsymbol{Ce}\approx\boldsymbol{w}$. Therefore, by inverting $\boldsymbol{C}$ and noting that its determinant is 1, we can write the numerator in (80) as:

$$f_{u^-,\epsilon^+}(\boldsymbol{u}_t^-(\boldsymbol{\phi}),\boldsymbol{\epsilon}_{t+1}^+) \approx \prod_{j=1}^{s} f_\epsilon(\epsilon_{t-s+j}(\boldsymbol{u}_t^-(\boldsymbol{\phi}),\boldsymbol{\epsilon}_{t+1}^+)) \prod_{\tau=t+1}^{t+M} f_\epsilon(\epsilon_\tau), \tag{83}$$

where I have written the elements $\epsilon_{t-s+j}(\boldsymbol{u}_t^-(\boldsymbol{\phi}),\boldsymbol{\epsilon}_{t+1}^+)$ as such to indicate that they are functions of both $\boldsymbol{u}_t^-(\boldsymbol{\phi})$ and $\boldsymbol{\epsilon}_{t+1}^+$.

5. Therefore, if we simulate N i.i.d. draws of the M length vector $\boldsymbol{\epsilon}_{t+1,i}^+$ (i.e. for

$i = 1, \ldots, N$) according to $f_\epsilon(\cdot)$, an approximation to the desired conditional c.d.f. in (81) is given as:

$$F_{u_{t+1|t}}(\alpha|\mathcal{F}_t) \approx \frac{N^{-1}\sum_{i=1}^{N} \mathbf{1}_{\alpha>u_{t+1}} \left(\sum_{j=0}^{M-1} \varphi_j^* \epsilon_{t+1+j,i}\right) \prod_{j=1}^{s} f_\epsilon(\epsilon_{t-s+j}(\boldsymbol{u}_t^-(\boldsymbol{\phi}), \boldsymbol{\epsilon}_{t+1,i}^+))}{N^{-1}\sum_{i=1}^{N} \prod_{j=1}^{s} f_\epsilon(\epsilon_{t-s+j}(\boldsymbol{u}_t^-(\boldsymbol{\phi}), \boldsymbol{\epsilon}_{t+1,i}^+))}. \tag{84}$$

Then, given an appropriately chosen grid of α_i's, we can generate an approximation to the shape of the c.d.f. across its support.

14 Appendix: Tables and Figures

Figure 7: Plots of simulated bubble processes

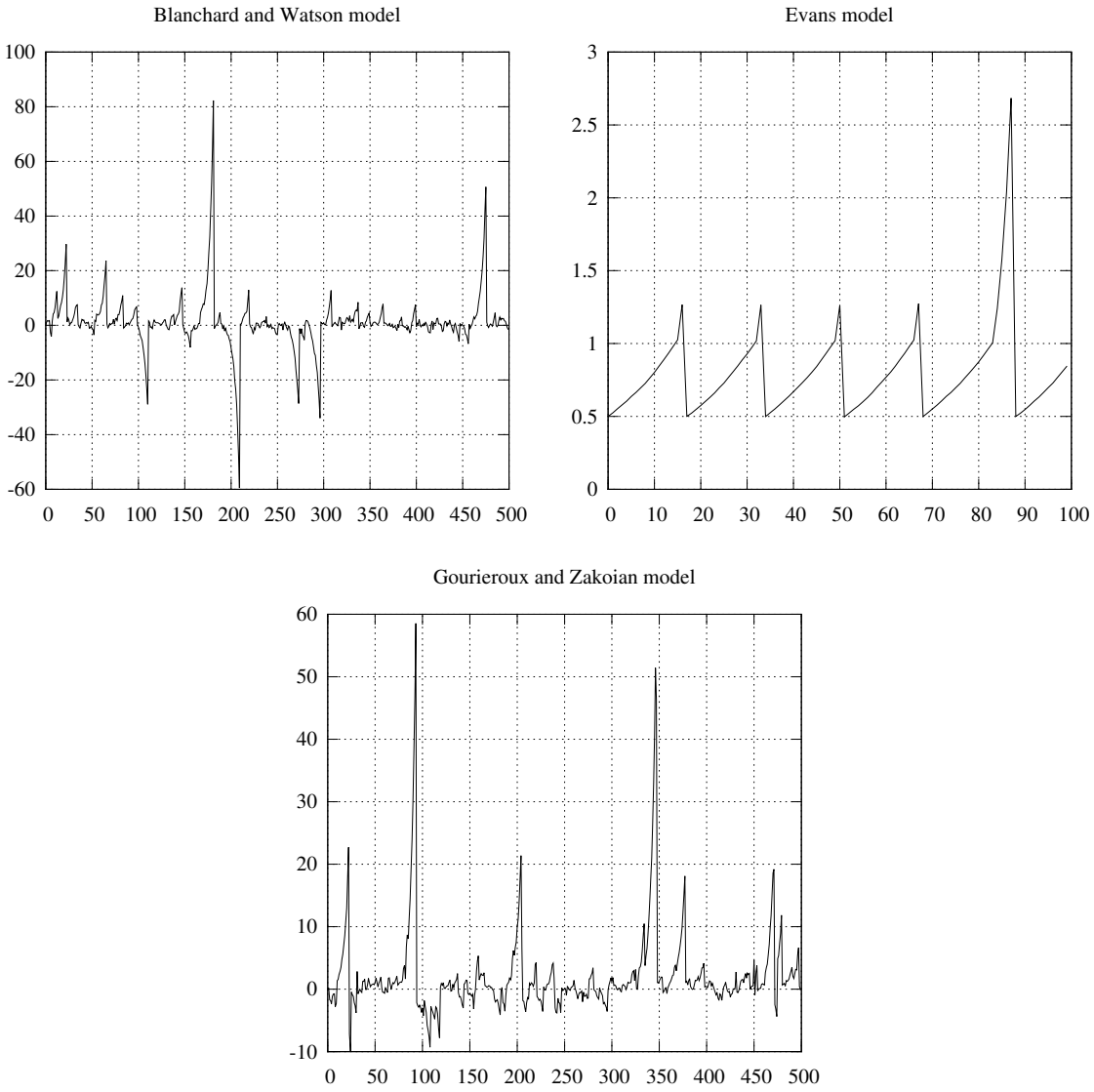


Table 7.i: Lag polynomial roots of the mixed and benchmark models

	Model	p/r,q/s	Sig.p/r	Sig.q/s	cR	cMC	ncR	ncMC	#CC
Soybean meal	skew-t arma	10,0	1,8,9		1.010	1.571			4
						1.581			
						1.582			
						1.583			
	t-dist mixed	10,10	1,3,5,7,9,10	1,2,3,4,6,9	1.385	1.354	-1.716	1.091	4/4
					-2.532	1.414		1.530	
						1.474		1.530	
						1.500		1.561	
Soybean oil	skew-t arma	10,0	1,10		1.033	1.478			4
					1.306	1.558			
						1.600			
						1.619			
	t-dist mixed	10,10	1,2,4,9,10	1,2,3,4,8	1.373	1.341	1.009	1.666	4/3
					-1.797	1.359	1.285	1.669	
						1.390		1.474	
						1.510			
Soybeans	skew-t arma	10,0	1,2,5,8,9		1.028	1.514			4
						1.551			
						1.556			
						1.582			
	skew-t mixed	10,10	1,2,5,8,10	1	-1.559	1.358	0.944		4/0
					1.749	1.464			
						1.477			
						1.558			
Orange juice	skew-t arma	10,0	1,2,3,10		1.033	1.505			4
						1.572			
						1.623			
						1.660			
	skew-t mixed	10,10	1,2,5,9	1,2,5	1.556	1.518	1.060	2.460	4/1
						1.542	1.843		
						1.555	-2.750		
						1.608			
Sugar	skew-t arma	1,2	1	1,2	1.000	4.590			3
						4.756			
						5.010			
						5.487			
	t-dist mixed	2,2	1,2	1,2		4.373	1.002		1/0
							14.637		
Wheat	skew-t arma	5,0	1,5		0.992	2.350			2
						2.655			
	skew-t mixed	5,5	1,2,3,5	1,3,4	1.006	1.814	1.789	2.046	2/1
						2.071	-2.434		
Cocoa	skew-t arma	10,0	1		1.022				0
	skew-t mixed	10,10	1,6,9	1,2,4,9,10	1.436	1.417	-1.435	1.202	4/4
						1.486	1.740	1.408	
						1.499		1.414	
						1.508		1.426	

Table 7.ii: Lag polynomial roots of the mixed and benchmark models

	Model	**p/r,q/s**[a]	**Sig.p/r**[b]	**Sig.q/s**[b]	**cR**[c]	**cMC**[c]	**ncR**[d]	**ncMC**[d]	**#CC**[e]
Coffee	t-dist arma	10,0	1,3		0.995	4.740			1
	skew-t mixed	10,10	1,2,5,6,10	1,2,5,6,7		1.375	1.027	1.684	5/2
						1.403	1.571	1.762	
						1.428	-1.645		
						1.430			
						1.446			
Corn	skew-t arma	2,0	1,2		1.000				0
					51.190				
	t-dist mixed	2,3	1	1,2,3	-32.542		1.002	5.484	0/1
Cotton	skew-t arma	10,0	1,2,6,7		1.007	1.738			3
						1.707			
						1.615			
	t-dist mixed	1,3	0	1,2,3			1.003	5.317	0/1
Rice	skew-t arma	2,2	1,2	1,2	0.997	3.099			3
					2.917	3.332			
					-3.552	3.493			
	t-dist mixed	1,3	1	1,2,3	-15.328		1.001	5.003	0/1
Lumber	skew-t arma	1,1	1	1	1.005	13.181			4
						13.237			
						13.314			
						13.375			
	skew-t mixed	10,10	1,2,4-10	1,5	1.015	1.235	-1.862	1.218	4/2
					-1.454	1.247		1.752	
						1.336			
						1.900			
Gold	t-dist arma	3,0	1,2,3		0.999	5.618			1
	t-dist mixed	10,10	1,2,6,10	1	-1.450	1.395	0.974		4/0
					1.489	1.416			
						1.431			
						1.434			
Silver	skew-t arma	10,0	1,2,4,8		1.003	1.606			3
					-1.874	1.715			
						1.751			
	skew-t mixed	10,10	1,3-6,9,10	1,4,5,7	1.479	1.424	0.996		4
					-1.533	1.424	1.600	1.721	4/2
						1.451	-2.070	1.643	
						1.327			
Platinum	skew-t arma	10,0	1,4,7,8,9		0.957	1.493			4
						1.528			
						1.572			
						1.582			
	skew-t mixed	10,10	1,2,3,5-9	1,2,6-8,10	-1.786	1.355	0.974	1.304	4/4
						1.376	1.257	1.328	
						1.385		1.401	
						1.860		1.594	

[a] (p,q) or (r,s) pairs for ARMA(p,q) and mixed causal/noncausal AR(r, s) models respectively.
[b] Significant lags at the 5% level assuming Normal distributed parameters.
[c] Causal lag polynomial; real roots and modulus of complex roots respectively.
[d] Noncausal lag polynomial; real roots and modulus of complex roots respectively.
[e] Number of complex conjugate roots with the same modulus (causal/noncausal).

Table 7.iii: Lag polynomial roots of the mixed and benchmark models

	Model	p/r,q/s	Sig.p/r	Sig.q/s	cR	cMC	ncR	ncMC	#CC
Palladium	skew-t arma	5,0	1,2,4,5		1.006	2.431			1
					-2.434				
					3.525				
	t-dist mixed	8,8	1,2-7	1,2,3,7,8	-1.618	1.621	0.989	1.547	3/3
						1.632	1.536	1.574	
						1.884		1.619	
Copper	skew-t arma	10,0	1,2,6		1.055	2.020			2
					1.696	2.101			
	skew-t mixed	10,10	1,2,3,6	1,6,7,8		1.728	0.952	1.352	3/3
						1.737	-1.323	1.482	
						1.831		1.751	
Light crude oil	t-dist arma	2,0	1,2		0.999				0
					-23.729				
	skew-t mixed	1,3	1	1,2,3	-14.222		1.002	6.144	0/1
Heating oil	t-dist arma	2,0	1,2		0.999				0
					-27.213				
	t-dist mixed	10,10	1-4,7,9,10	1-6,9,10	1.245	1.279	1.032	1.259	4/4
					-2.553	1.307	-1.505	1.303	
						1.349		1.315	
						1.368		1.372	
Brent crude oil	t-dist arma	2,2	1,2	1,2	0.989	2.466			3
					2.255	2.621			
					-2.716	2.695			
	skew-t mixed	10,10	1,4,9,10	1,2,5,6,9	1.261	1.292	1.068	1.276	4/3
					-1.527	1.331	1.101	1.388	
						1.336	-1.723	1.540	
						1.500			
Gas oil	skew-t arma	1,0	1		0.998				0
	skew-t mixed	10,10	3,7,9,10	1,4,7-10	1.230	1.324	0.925	1.346	4/4
					-2.140	1.328	-1.264	1.483	
						1.341		1.542	
						1.508		1.563	
Natural gas	t-dist arma	1,2	1	1	1.001	34.697			4
						34.765			
						34.839			
						34.886			
	t-dist mixed	1,1	1	1	-31.650		1.001		0/0
Gasoline RBOB	skew-t arma	3,0	1,3		0.972	4.452			1
	skew-t mixed	2,1	2	1		4.390	1.005		1/0
Live cattle	skew-t arma	10,0	1,5		1.019	2.408			1
					1.973				
					-2.543				
	t-dist mixed	10,10	1	3,4,6	0.994			1.896	0/3
								1.728	
								1.891	
Lean hogs	skew-t arma	5,0	1,4,5		0.984	2.555			1
					-2.525				
					2.744				
	skew-t mixed	0,2					1.004		0/0
							55.339		

Figure 10.i: Plots of daily continuous contract futures price level series

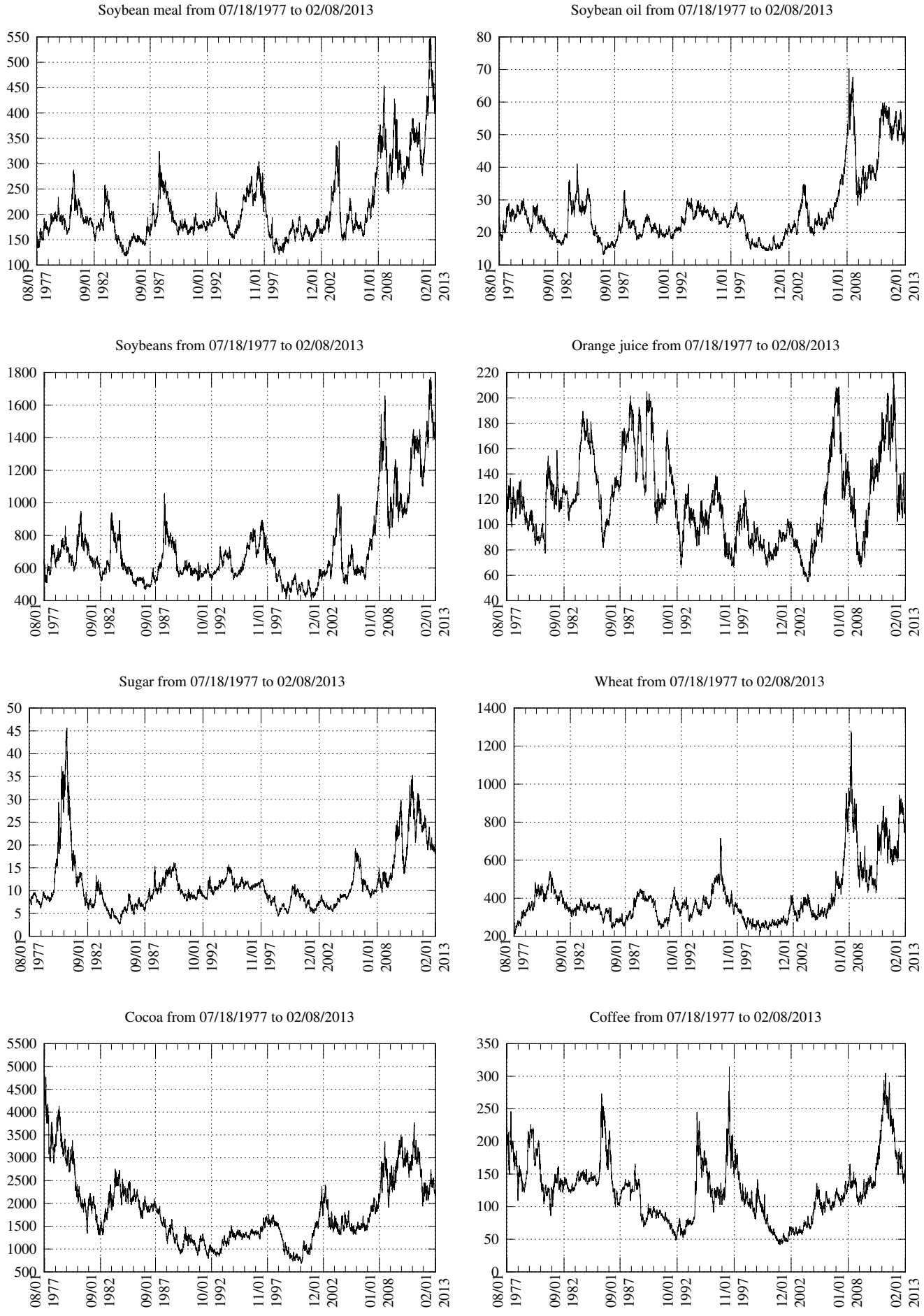


Figure 10.ii: Plots of daily continuous contract futures price level series

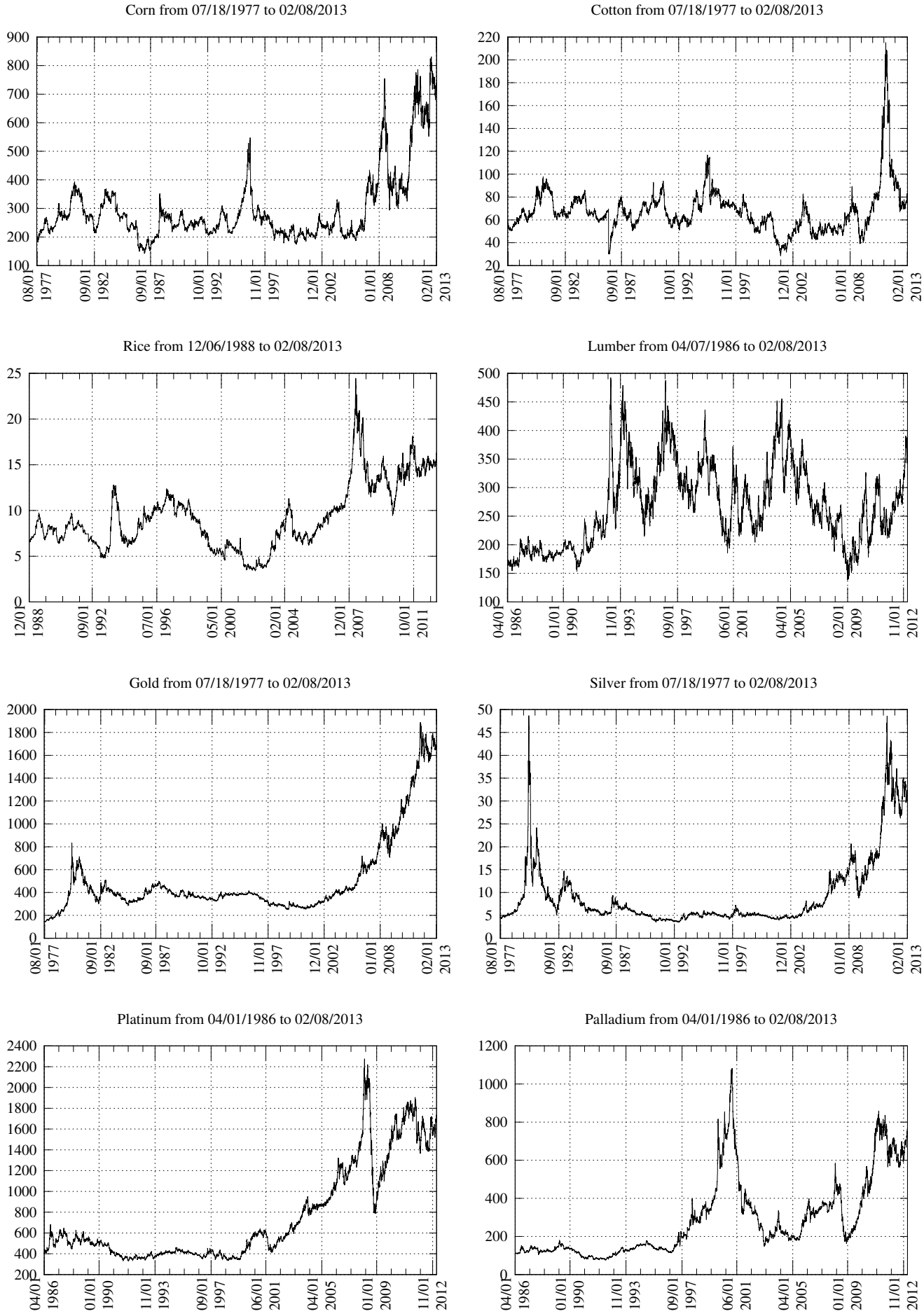
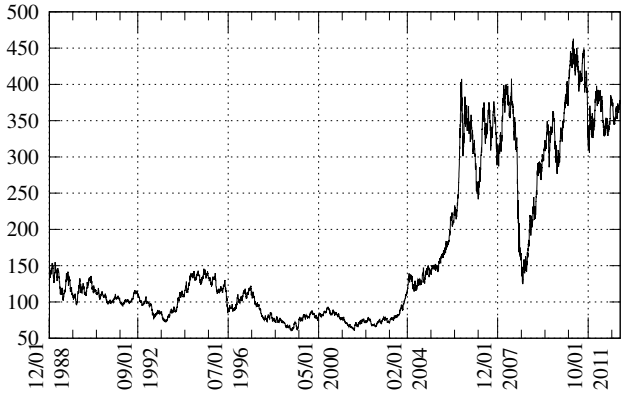


Figure 10.iii: Plots of daily continuous contract futures price level series

Copper from 12/06/1988 to 02/08/2013

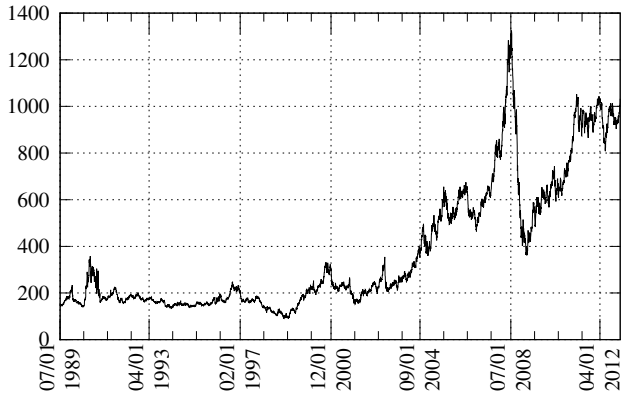


Light crude oil from 03/30/1983 to 02/08/2013

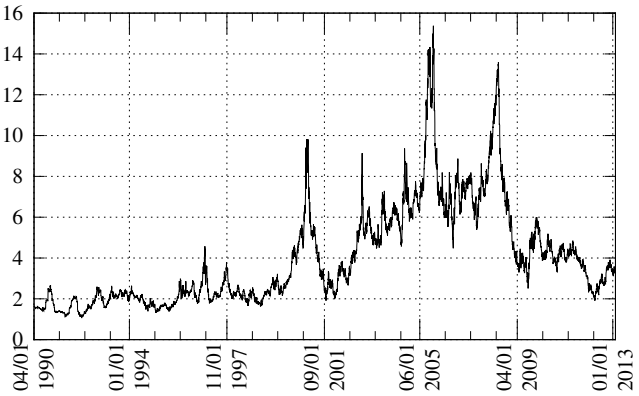
Heating oil from 07/01/1986 to 02/08/2013

Brent crude oil from 06/23/1988 to 02/08/2013

Gas oil from 07/03/1989 to 02/08/2013



Natural gas from 04/03/1990 to 02/08/2013



Gasoline RBOB from 10/04/2005 to 02/08/2013

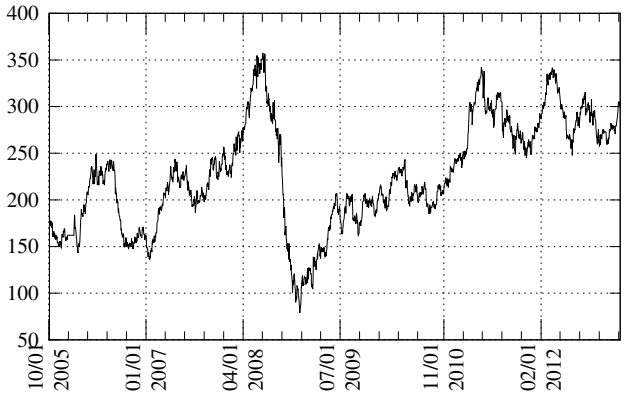
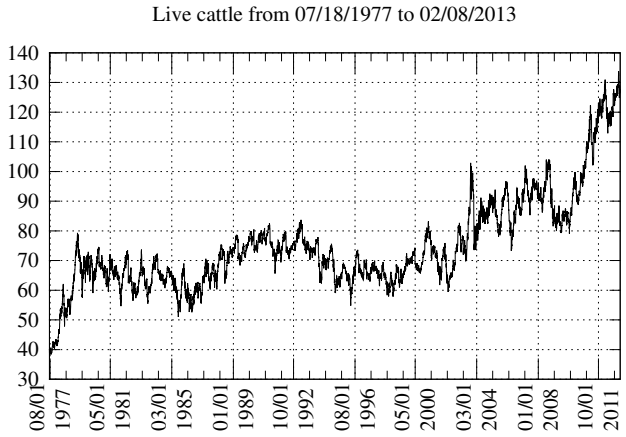


Figure 10.iv: Plots of daily continuous contract futures price level series



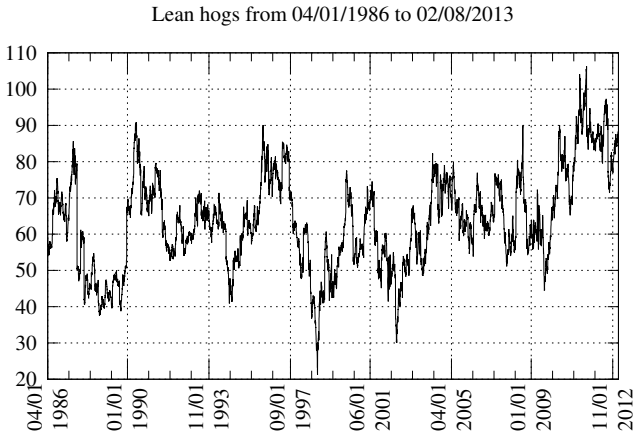


Figure 11.i: Histograms of daily continuous contract futures price level series

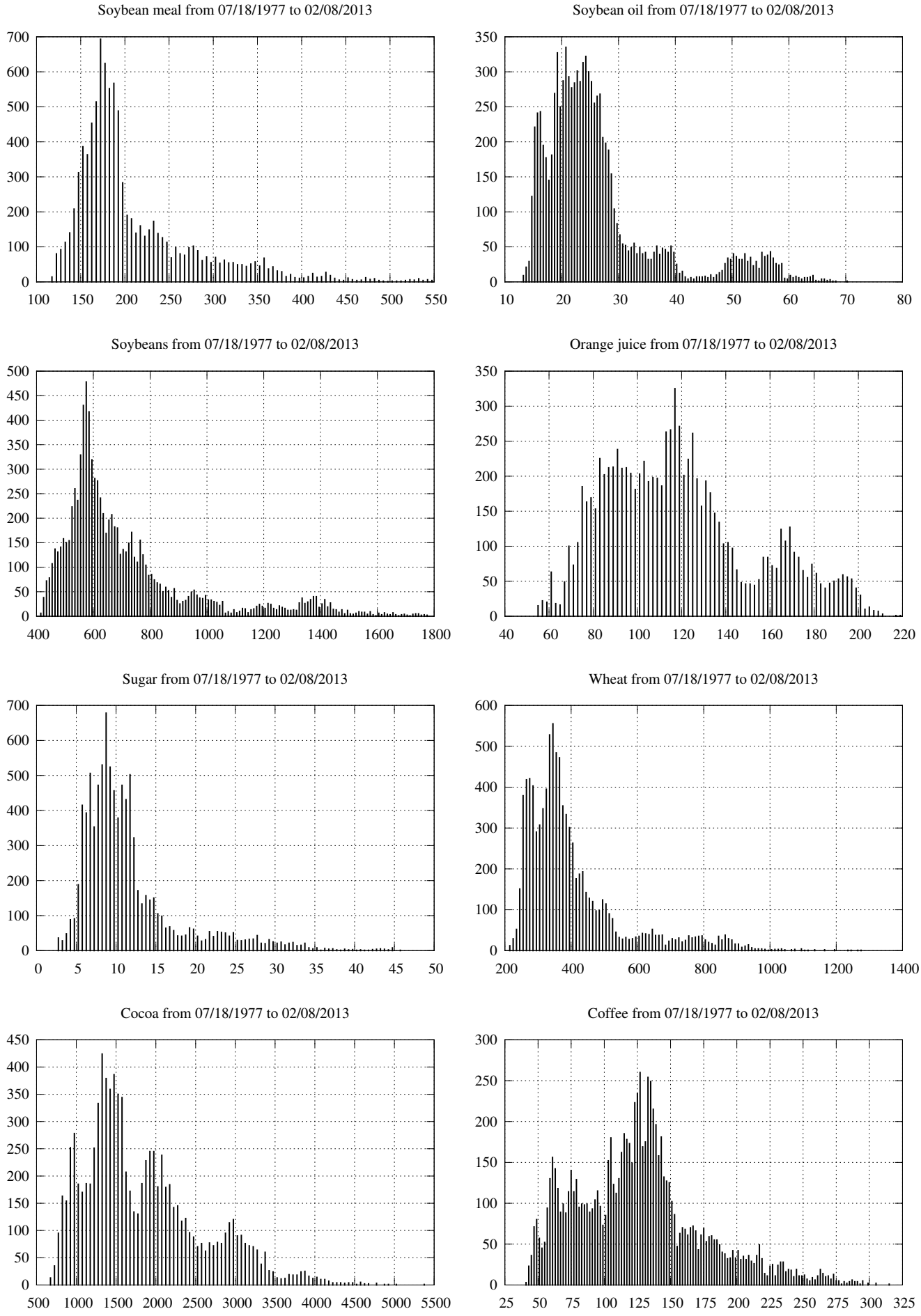
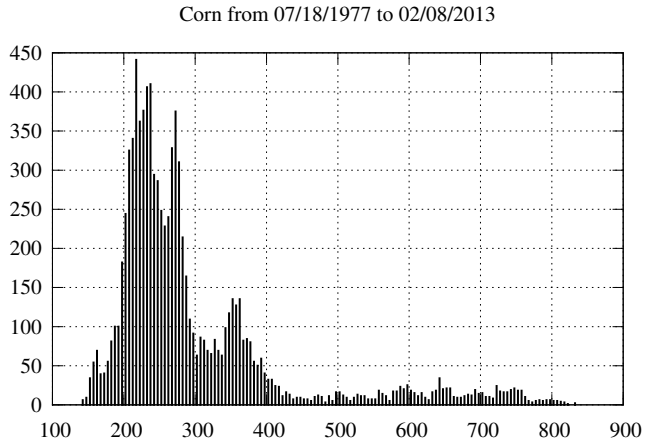
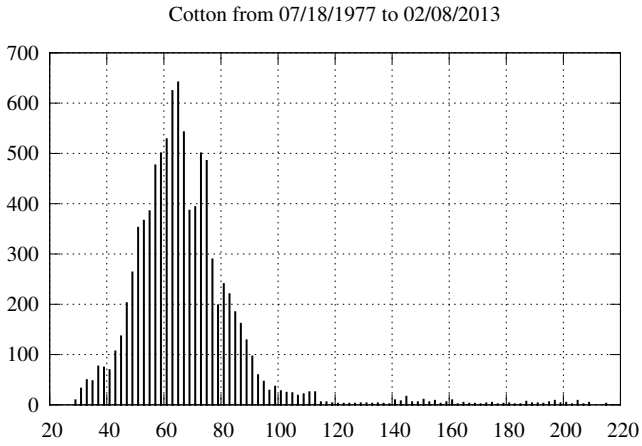
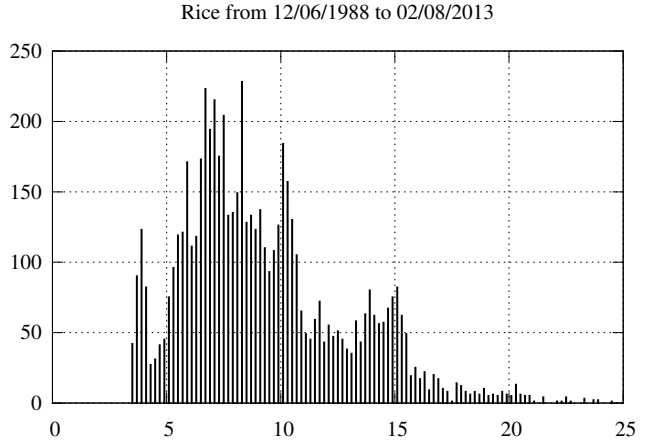
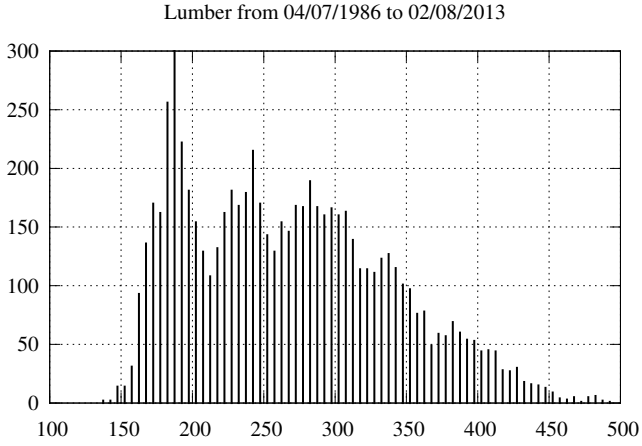


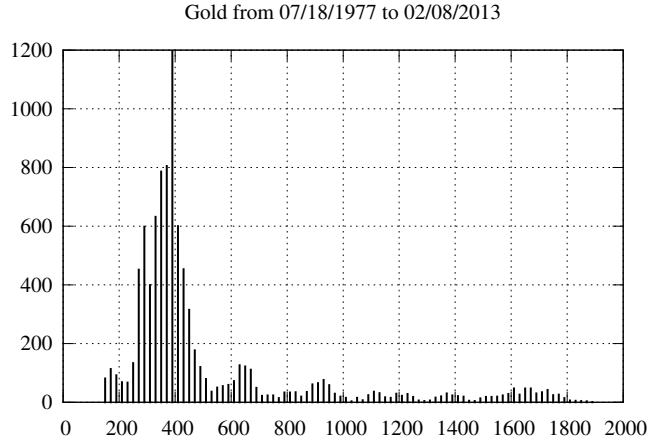
Figure 11.ii: Histograms of daily continuous contract futures price level series

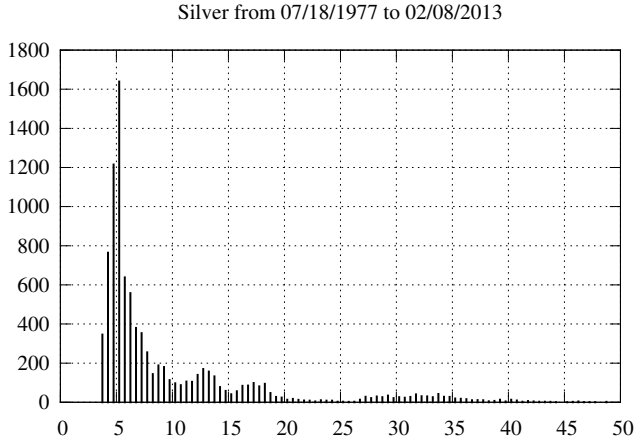


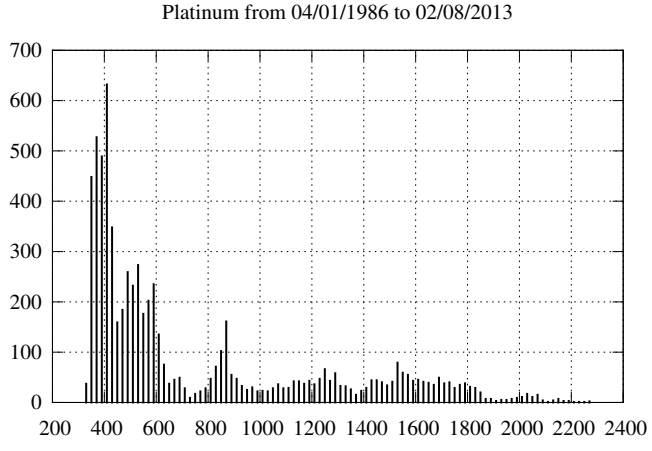












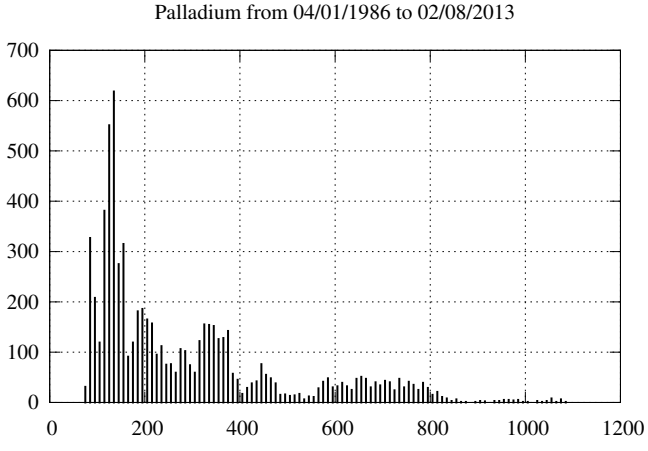
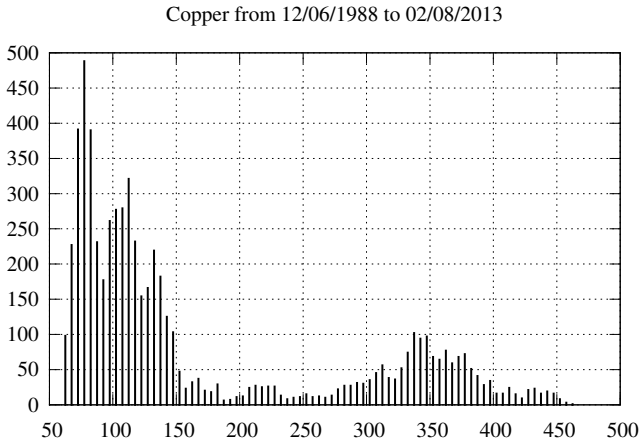
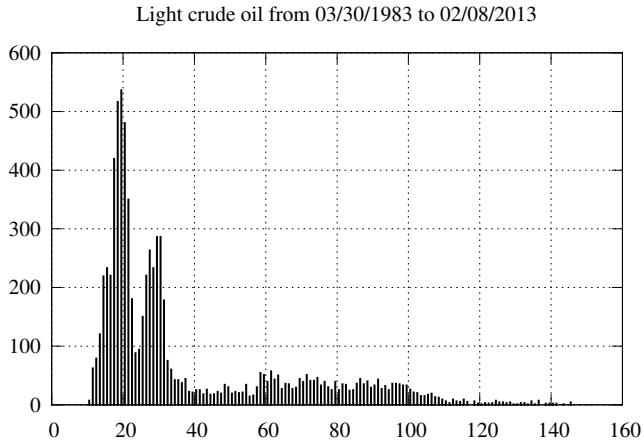
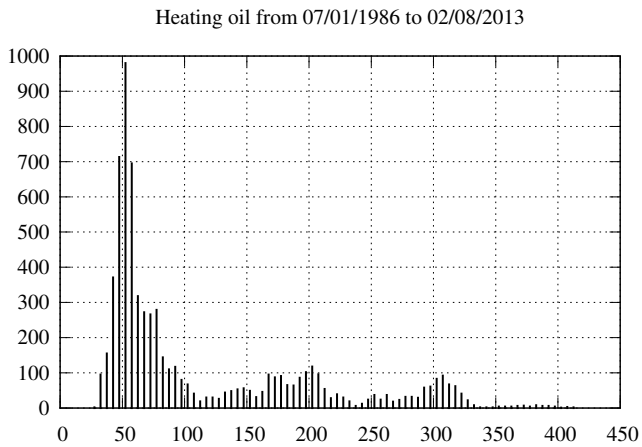
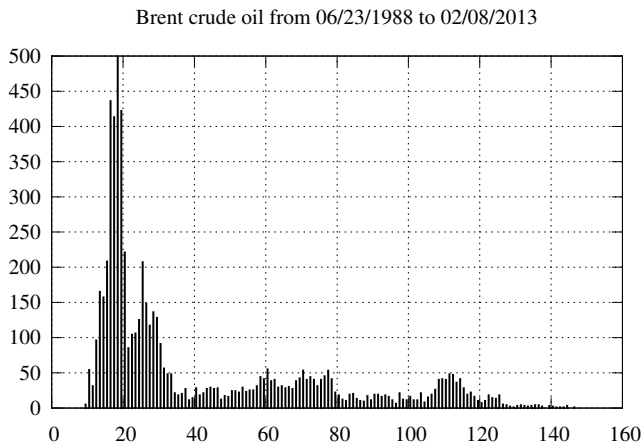


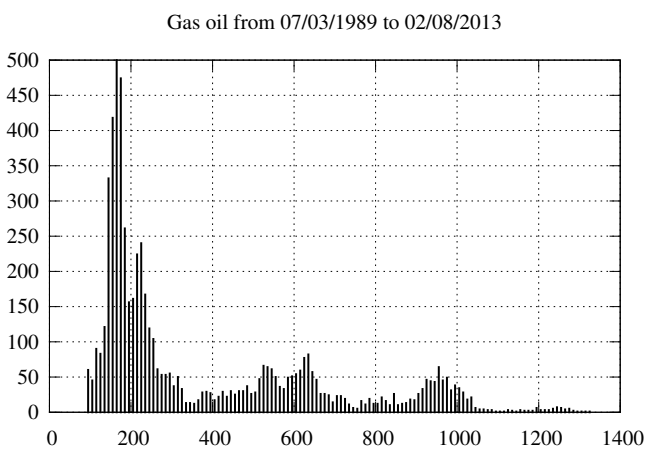
Figure 11.iii: Histograms of daily continuous contract futures price level series

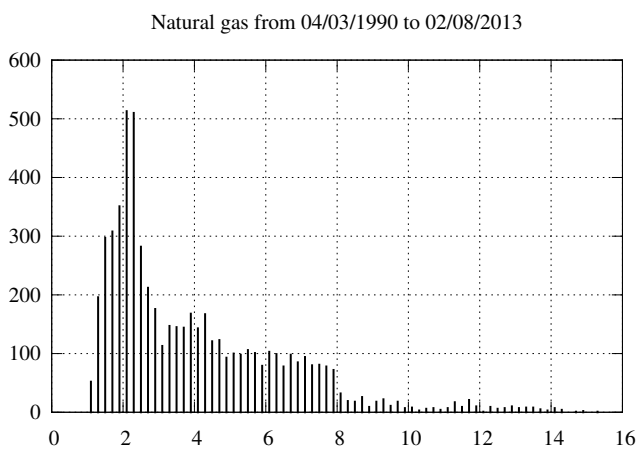












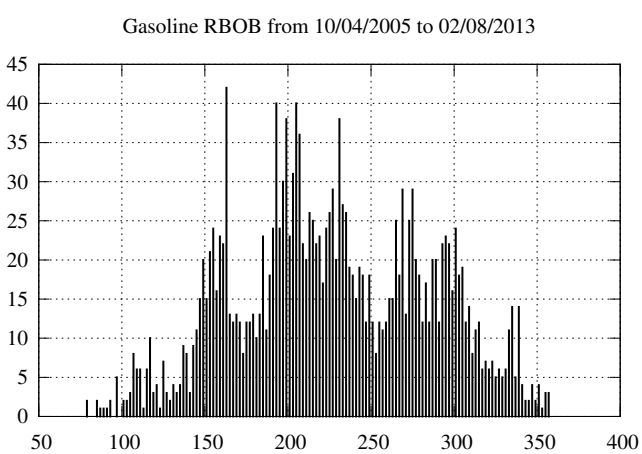


Figure 11.iv: Histograms of daily continuous contract futures price level series