Empirical evidence for nonlinearity and irreversibility of commodity futures prices

Karapanagiotidis, Paul

University of Toronto, Department of Economics

21 August 2013

Empirical evidence for nonlinearity and irreversibility of commodity futures prices*

Paul Karapanagiotidis†

Draft #8
August 21st, 2013

Abstract

Theory suggests that commodity futures price levels and returns data may exhibit both nonlinear and nonreversible features. This paper attempts to provide a thorough empirically investigation of these claims. The data set is composed of 25 individual continuous contract commodity futures series which fall within a number of industry sectors including softs, precious metals, energy, and livestock. Employing both time-domain and frequency-domain tests examining the higher order cumulant properties of these series, it is shown that they exhibit both nonlinearities and irreversibility differing across industry sector. Furthermore, in modeling these series I estimate a number of parametric models able to capture irreversibility such as the linear mixed causal/noncausal autoregressive model and various purely causal nonlinear models, since there is a close connection between these two classes of models. It is shown that the linear causal ARMA model is unable to adequately account for the features of the data and while the mixed causal/noncausal model improves model fit significantly by capturing latent irreversibility, the vast majority of the nonlinearity these series exhibit is of the "nonlinear in variance" type. Finally, out of sample forecasts and an evaluation of the estimated unconditional distribution of the mixed causal/noncausal models suggest that there may still exist model misspecification.

Keywords: mixed causal/noncausal autoregressions, nonlinear models, commodity futures, speculative price bubbles.
JEL: C22, C50, C51, C52, C58

1 Introduction

Recent theoretical arguments suggest that both nonlinearities and time irreversibility may prove important in describing the probabilistic nature of finance and macro time-series data. This paper seeks to investigate these claims empirically by applying some of the more recent approaches designed to accommodate these features to a data set comprised of 25 commodity

*I'd like to thank Christian Gourieroux for his helpful comments and suggestions.
†University of Toronto, Department of Economics, p.karapanagiotidis@utoronto.ca

futures contract prices ranging across various industry sectors including softs, precious metals, energy, and livestock. The main tool at my disposal is the model of a linear autoregressive stochastic process composed of a mix of both causal and noncausal components. That is, rather than assume that the process at any time period depends strictly on only the past shocks, the model allows it to depend on both the past and future shocks – that is, the process can be causal, noncausal, or both.

The paper is as follows. Section 2 will discuss some of the background literature related to the study of commodity futures prices. This will include a discussion of both the theory and empirical evidence, with an eye towards the concepts of nonlinear processes, time irreversibility, and speculative price "bubbles." Section 3 will discuss the details of the futures contracts including the underlying commodities, the markets they are traded in, and how the continuous contracts are reconstituted from individual contracts of varying maturities. Section 4 will discuss features of the data series themselves including tests suggesting nonlinear features. Section 5 will show that the linear causal ARMA models with Gaussian innovations fail to adequately capture the structure of the data. Section 6 will introduce the theory of mixed causal/noncausal autoregressions. Section 7 will discuss estimation of the mixed AR(r, s) models and discuss the empirical estimation results. Section 8 will also consider some purely causal models which are both nonlinear in mean or variance. Finally, section 9 will consider the results of an out of sample forecasting exercise amongst competing models.

2 Literature review

The study of commodity futures is not new. For example the seminal paper by 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.

More generally, existing research has examined the futures contract price dynamics as a function of either the price of commodity storage or as reflection of an underlying risk premia paid to speculators for taking on risk. One earlier example of the latter approach can be seen in the Keynesian notion of "backwardation"–that the expected future spot price should

be higher than the current futures contract price–reflecting the idea 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. In this way many studies have attempted to measure the risk premium in futures contract returns, if it exists. For example, Dusak (1973) 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. More recently, Fama and French (1987) considered both the cost of storage model *and* the risk premium model in studying the behaviour of futures prices for 21 commodities.

Other authors have focused exclusively on the cost of storage as an important contributor to the dynamics of future contracts prices. For example, Working (1949) pioneered the study of the cost of storage as an important component in the relationship between the current spot price of a commodity and the corresponding futures contract price for future delivery. More recently, in a series of papers throughout the 1990's, Deaton and Laroque modeled the cost of storage within a rational expectations framework. Perhaps the most interesting implication of this modeling exercise was the fact that the price process under rational expectations and storage tends to follow a piecewise linear stochastic process. That is, since storage cannot be negative (i.e. we cannot borrow that which hasn't yet been produced–and is not certain to be produced at all) we are presented with a natural asymmetry between speculators and hedgers (one that differs from the notion of backwardation proposed by Keynes); see Deaton and Laroque (1996). 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. This notion of price process asymmetry is important in particular to this paper since it provides a theoretical rationale for investigating the possibility that commodity price processes may possess nonlinear or time irreversible properties.

Empirical evidence also suggests commodity price processes may be better modeled as non-linear. For example, DeCoster, Labys, and Mitchell (1992) find strong evidence of nonlinear features in four commodity futures returns: coffee, sugar, silver, and gold. Specifically, they employ 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 *m*-space for various embedding dimensions, *m*. The results strongly suggest nonlinear structure and, importantly, the inspection of ARCH filtered residuals suggest the structure does not simply reflect heteroskedasticity. These results reinforce earlier work done by Blank (1991) who found nonlinear dynamics in soybean futures prices and Frank and Stengos (1989) who find similar results for gold and silver rates of return, again rejecting pure ARCH dynamics. Finally, Yang and Brorsen (1993) also find that GARCH(1,1) dynamics are not able to provide a complete picture of the nonlinear structure in a number of commodity futures returns series (although the inclusion of time varying conditional variance does improve model fit substantially). Of note, however, is the fact that these studies all employ the assumption of conditionally Gaussian GARCH innovations and it would be worthwhile to consider other, more leptokurtic, conditional distributions such as a t-distribution or even skewed t-distribution.

More recent evidence for nonlinear structure comes from Sigl-Grub and Schiereck (2010), who 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. In order to study this phenomenon they find a STAR model specification (i.e. smooth autoregressive transition), as discussed in Terasvirta (1994), works well in describing the data.

Nonlinear structure is closely related to the concept of time irreversibility of a stochastic process. We say that a process is reversible if the joint distribution of the process (X_τ), for $\tau = t_1, t_2, t_3, \ldots$, is the same as the joint distribution of the reversed process (X_{τ^*}) for $\tau^* = -t_1 + m, -t_2 + m, -t_3 + m$ for any integer m and set of time indices τ. Furthermore, any stochastic process can admit either a purely causal, purely noncausal, or mixed causality representation (that is, where the process is a function of past innovations, future innovations, or both, respectively). Consider then the strictly linear and causal class of ARMA processes $X_t = \sum_{j=0}^{\infty} b_j \epsilon_{t-j}$.[1] Weiss (1975) showed that when ϵ_t is Gaussian and uncorrelated the process X_t is time reversible (so that the joint density of X_t and $X_t^* = \sum_{j=0}^{\infty} b_j \epsilon_{t+j}$ are the same) and that the only time reversible non-Gaussian, i.i.d. ϵ_t, ARMA processes are

[1] Where b_j represents the coefficients of the rational lag polynomial defined by the ARMA, i.e. the coefficient from $\frac{(1+\beta_1 L+\ldots+\beta_q L^q)}{(1-\alpha_1 L-\ldots-\alpha_p L^p)}$.

the pure moving average processes $Z_t = \sum_{j=0}^{q} \beta_j \epsilon_{t-j}$ where the β_j's represent a symmetric or skew-symmetric set. Later, Findley (1986) extended this result to the more general mixed representation $Y_t = \sum_{j=-\infty}^{\infty} \gamma_j \epsilon_{t-j}$. Therefore, the Gaussian linear processes represent a very special case and in general these linear ARMA processes with non-Gaussian innovations are not time reversible.

Later, Ramsey and Rothman (1996) suggested that the notion of irreversibility of a stochastic process might be useful in modeling time asymmetries exhibited by business cycle movements. Seminal papers such as Burns and Mitchell (1946) and more recently, Neftci (1984), had done work in the area of business cycle asymmetries, noting that business cycle upswings were longer and slower than downswings and defining the notion of time asymmetry in terms of the transition probabilities on the signs of the series' first differences. In generalizing this work, Ramsey and Rothman (1996) defined stochastic process time asymmetry in terms of both "longitudinal" and "transversal" asymmetry. Longitudinal asymmetry refers to asymmetry where the 2nd derivative of the process is asymmetric about the zero of the 1st derivative; that is the process behaves differently when rising than when it is falling. Transversal asymmetry is characterized by different process dynamics above and below some horizontal plane in the time direction; that is, asymmetries in the vertical displacement of the series from its mean value. Of course, a series can be both longitudinally and transversely asymmetric. Therefore, the concept of time irreversibility captures the notion of a longitudinally asymmetric series and purely transversally asymmetric series are inherently time reversible – that is the asymmetry does not depend on the time direction as it does with longitudinal asymmetry. Interestingly, however, it can be shown that tranversally asymmetric, but longitudinally symmetric, series are time reversible, but their derivatives are time irreversible. This point is exploited by Ramsey and Rothman in attempting to detect and distinguish between the two types of asymmetry in an empirical study of an extended version of the renowned Nelson and Plosser (1982) dataset.

These two notions of time asymmetry can be shown to have implications for the functional form of the process. For example, a natural candidate for a transversally asymmetric process is the threshold autoregressive model of Tong and Lim (1980) since this process exhibits limit cycle behaviour consistent with this type of asymmetry. Piecewise linear processes of this type

were also found to be useful in modeling nonlinear series such as the famous Canadian Lynx data (see Priestley (1989) for further discussion). Ramsey and Rothman (1996) define time irreversibility as stemming from two sources: 1) the underlying model may be nonlinear even though the innovations are symmetrically (or perhaps Gaussian) distributed; or 2) the underlying innovations may be drawn from a non-Gaussian probability distribution while the model is linear. Importantly, it can also be shown that under the latter type of irreversibility, even though the causal (noncausal) represention is linear with i.i.d. non-Gaussian innovations, the equivalent noncausal (causal) representation, with i.i.d. shocks, must necessarily be nonlinear if the process is irreversible (Rosenblatt, 2000).

Consequently, Breidt et. al. (1991) derived the maximum likelihood estimator for the autoregressive mixed causal/noncausal model $X_t(1 - \alpha_1 L - \ldots - \alpha_s L^s) = \epsilon_t$, where ϵ_t is i.i.d. non-Gaussian and where some of the roots of the lag polynomial lie inside the unit circle. This is important since these mixed causality models provide a natural way to model longitudinal asymmetry in a linear fashion but where the equivalent, purely causal, i.i.d. shock representation is necessarily nonlinear.

The models of mixed causality from Breidt et. al. (1991) have been extended by Lanne and Saikkonen (2008) and applied to the U.S. inflation rate where the authors find evidence of noncausality. Moreover, Lof (2011) applies the mixed causal/noncausal model to the historical real S&P 500 dataset available from Robert Shiller and again finds evidence of noncausality.

In Breidt et. al. (2001) the authors study the so-called "All-Pass" model which is a special case of the mixed causal/noncausal model in that $(1 - \alpha L)X_t = (1 - \frac{1}{\alpha}L)\epsilon_t = -\alpha^{-1}L(1 - \alpha L^{-1})\epsilon_t$. This model is interesting since it is linear with i.i.d. non-Gaussian shocks, ϵ_t, but exhibits features similar to a GARCH model, since X_t can be shown to be zero mean, serially uncorrelated, heavy tailed (if ϵ_t is heavy tailed), but yet dependent through the higher moments. Therefore, the all-pass model represents a linear model with "nonlinear features." However, the authors suggest the all-pass model is much more constrained than the GARCH and should not be considered a serious alternative.

Furthermore, in developing a process useful for modeling price bubble behaviour, Gourieroux and Zakoian (2012) develop a special case of the mixed causal/noncausal autoregressive

model from Breidt et. al. (1991) which exploits the fact that the causal representation of the noncausal AR(1) model with i.i.d. Cauchy innovations, $X_t = \rho X_{t+1} + \epsilon_t$, is necessarily non-linear. Deriving the conditional Markov transition probabilities for the causal representation they show that both processes (X_t) and (X_t^2) admit semi-strong causal linear representations with causal innovations that are martingale difference sequences but no longer i.i.d. Interestingly, the weak innovations also display GARCH type effects. Moroever, the process remains stationarity despite the presence of a unit root; this unit root is expected since the unconditional moments of X_t do not exist. However, as discussed in their paper, typical linear unit root tests will fail in detecting the explosive bubbles of the noncausal AR(1) Cauchy process, even if the martingale property is satisfied. The noncausal AR(1) Cauchy model is also easily introduced in a mixed causal/noncausal framework, unlike previous bubble modeling strategies such as those discussed in Blanchard and Watson (1982) or Evans (1991).

Given the nature of the futures contract, expectations of the future clearly play a role in the current price. Of course, modeling a process as partly noncausal does not imply that agents somehow "know the future." Rather, modeling them this way allows for the possibility that the information set available to decision makers in the economy may be strictly larger than that available to the econometrician. For example, it can be shown that if agents in the commodity futures market have rational expectations, the innovations in the mixed causal/noncausal model may not represent the "true" *fundamental* shocks observed by decision makers, since the information set available to the econometrician is limited (Hansen and Sargent, 1991). Rather, they simply represent another equivalent linear representation with i.i.d. innovations.

In conclusion, this paper will attempt to bring these concepts of nonlinearity, time irreversibility, and price bubbles together in evaluating the statistical properties of commodity futures price series for 25 different physical commodities, across 5 different industry sectors. I hope to provide a robust empirical investigation of whether or not these series exhibit the aforementioned properties and how we might therefore best model and forecast these series. Before considering the models, however, let us first discuss the details of the data employed in the next section.

3 The futures contract

A futures contract is a contract between two parties to either buy or sell some good (in this case commodities) at a future date, given a predetermined "futures price" $F_{t,t+k}$ set today. This price is called the *delivery price*.[2] The commodity futures contract will specify terms of the agreement not limited to including:

- The quality of the good, often categorized by pre-specified "grades."
- The amount and units of the underlying good.
- The monetary denomination of the contract.
- Whether the good is to be delivered to the buyer upon the exercise date (otherwise the buyer will have to pick up the good themselves). It will also specify the location of delivery if applicable.
    - Contracts can alternatively be settled in cash if stipulated in advance.
- The date of delivery at some period in the future. Typically delivery is acceptable at any point during the specified month after the *last trading date*.
- The last trading date is the date upon which trading of the contract ceases.
- The delivery price $F_{t,t+k}$.
- A margin call rule.

Supposing that today is denoted as time t, the date of delivery is $t + h, h > 0$, and the price agreed upon at time t is to be paid at time $t + k, k > 0$ (i.e. $F_{t,t+k}$), we have that there are as many futures contracts at time t as there are (h, k, l, m), where l denotes delivery locations and m denotes a margin call rule.

[2] See J.C. Hull (2009) *Options, Futures, and Other Derivatives*, Prentice Hall, New Jersey, for an overview of futures contracts.

3.0.1 The margin call and "marking to market" process

Unlike forward contracts, which are typically highly customized instruments sold on over-the-counter (OTC) markets, futures are typically standardized instruments traded on organized futures exchanges. To avoid the counterparty risk associated with future settlement of forward contracts (considering the possibly large loss positions that may accrue to either side of such contracts), futures contracts are subject to a margin rule. An investor who wishes to trade futures contracts must first deposit collateral with the exchange called the *initial margin*. The clearing house (often a 3rd party institution, although the Chicago Mercantile Exchange (CME) and Intercontinental Exchange (ICE) both handle clearing house operations themselves) will then operate as an intermediate between buyers who wish to take on long positions and sellers taking on short positions. That is, the clearing house becomes the buyer to each seller and the seller to each buyer so that in the event of counterparty default the clearing house assumes the risk of loss. In order to minimize these potential losses, each day investors accounts are settled, a process known as "marking to market."

As an example of marking to market, consider a trader who goes long on a futures contract with a delivery price of $\$100$ specified immediately at time period t. Suppose later, at time period $t+j$ that a similar contract is now trading at $\$99$. Since the investor has taken the long position, they could "zero" out their position by simultaneously purchasing the same contract's short position. That is, since the original arrangement contracted them to purchase at some future date, they could simultaneously contract to sell under the same contractual arrangements. However, since the price has now changed, the investor has realized a loss of $(-\$1)e^{-r_f(k-j)} = (F_{t+j,t+k} - F_{t,t+k})e^{-r_f(k-j)}$ since the future price they would pay at time $t+k$ exceeds the price they would receive (where r_f is the unit time period risk free rate of interest).[3]

Each day the clearing house calculates investor's net financial position in this manner by comparing the prices of comparable contracts to each investor's net long or short position. Essentially, what this means is that each day a new futures contract is rewritten with a new

[3]Since the relationship above defines in essence the present value of the futures contract, it is clear that upon *open interest*, that is upon initial negotiation of the contract, its value must be zero. Moreover, this relationship also implies that upon the delivery date, the present value is equal to $F_{t+j,t+k} - F_{t,t+k} = S_{t+k} - F_{t,t+k}$ since $k = j$, where S_τ is the spot price of the underlying good at time τ. (Black, 1976)

contracted price to deliver equal to the corresponding current future's contract price. In this way a futures contract is like a series of forward contracts (since the delivery price doesn't change on a forward as time progresses). (Black, 1976)

If an investor is allowed to build up a substantial net loss over a time period, eventually their collateral (or initial margin) depletes beyond a pre-specified level. At this point the clearing house issues a "margin call," and in order to continue trading the investor is required to deposit more collateral. If they do not, the clearing house is obligated to liquidate the investor's assets.

Therefore, outstanding balances are tallied each day, a new contract is issued by the clearinghouse to reflect the changing value of the outstanding contract, and a margin call is issued if the level of an investor's collateral diminishes beyond a certain threshold.

Marking to market in practice requires that account values are reevaluated according to the current market price of a comparable futures contract (not the current spot price of the underlying good). Therefore, standardization is important in promoting a sufficient volume of trades as to allow for a competitive price.

3.1 The commodity futures

This paper will consider a number of physical commodity futures contracts from a broad range of categories. The categories are as follows:

Table 1: Commodities 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		

3.1.1 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.

3.1.2 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.

3.1.3 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.

3.1.4 Data specification

The following table outlines dates each commodity futures price series is available for, the time to maturity, currency denomination, commodity exchange and code, and basic unit/characteristics of the product traded.

Table 2: Commodities specification

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 of the price series are generated by Bloomberg as a "continuous rolling contract," whereby the prices for the individual outstanding contracts in existence at time t are combined into one using a type of formula described in the next section, 3.2.

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 (IntercontinentalExchange) 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.

3.2 Rolling over the futures contract

As futures contracts specify a set date upon which the underlying goods are to be exchanged, their price series can not continue indefinitely (as say could the underlying spot price of the commodity). For this reason, we need a method whereby we can "extend" the futures price series indefinitely by appending contracts for the same underlying good together across time. In the terminology of traders, this is called "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, no arbitrage should ensure that the following relationship between the futures and spot price of the underlying commodity holds at time t:

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

where r_f is the risk free annual rate of interest and $c \in [0, 1]$ is some adjustment parameter for cost of carry. 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 $F_{t,t} = S_t$.

Note that of course this relationship is an approximate one and it is clear that it will not hold exactly in reality (for example, when goods are perishable and cannot be stored indefinitely). None the less the relationship is useful for considering the rolling over of futures contracts since it becomes clear that an adjustment must be made to the price series if we want it to maintain the same units.

The reason is as follows: suppose for the sake of argument that the futures price does adhere to the spot-future parity theorem. Upon the approach of the futures maturity we wish to extend the price series. In order to do so as a trader we would have to close out our current position and 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:

$$(1 + \tfrac{k}{365}(r_f + c)) < (1 + \tfrac{k+j}{365}(r_f + c))$$

in the spot-futures parity theorem. This is known as *rollover risk* and the difference in the two prices is called *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 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: [4]

1. Just append together prices without any adjustment. This will clearly distort the series since it will now include 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. A similar method is employed by the CME and ICE in the construction of the data series employed in this paper. This method melds together the old and new prices in a continuous manner with more weight placed on the series with the longer time to maturity.

   As an example, consider the following two futures contracts on the same underlying

[4] See Bob Fulks (2000) "Back-Adjusting Futures Contracts," *Trading Recipes DB* a widely disseminated PDF document available on the world wide web. Alternatively Masteika, Rutkauskas, and Alexander (2012) "Continuous futures data series for back testing and technical analysis," *IPEDR,* 29, provides a more recent treatment of the relevant issues.

good, one with time to maturity $t+k$, the other with $t+k+j$:

$$\begin{aligned} F_{t,t+k} &= S_t(1+\frac{k}{365}(r_f+c)) + \epsilon_{1,t} \\ F_{t,t+k+j} &= S_t(1+\frac{k+j}{365}(r_f+c)) + \epsilon_{2,t} \\ P_t &= \alpha F_{t,t+k} + (1-\alpha)F_{t,t+k+j} \end{aligned} \tag{1a}$$

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. Therefore, we have that P_t represents the "merged" futures price that takes into account both contracts.

4 Features of the data

All continuous contract futures series represent daily closing prices for those days the commodities are traded on the exchange, ending on February 8th, 2013. The start dates for each of the series are given above in **Table 2**. Summary statistics for price level series are given in the following **Tables, 3 and 4**, and plots and histograms of each series are available in **Appendix A** (**Figures 22 to 29**). Furthermore, the equivalent tables to 3 and 4 for the continuously compounded returns series (i.e. $r_t = ln(P_t/P_{t-1})$) are also available in the Appendix A (**Tables 20 and 21**).

Table 3: Summary statistics - 1 of 2 - price level series

Series	Mean	Median	Stnd. Dev.	Skewness	Kurtosis
Soybean meal	210.347	185.800	70.151	1.729	6.190
Soybean oil	26.399	23.750	10.449	1.709	5.516
Soybeans	716.563	629.000	249.577	1.755	5.735
Orange juice	118.926	115.125	33.531	0.592	2.663
Sugar	11.586	9.830	6.343	1.946	7.283
Wheat	401.672	357.500	151.036	1.878	6.656
Cocoa	1835.268	1621.000	744.051	0.926	3.466
Coffee	126.325	124.450	48.051	0.699	3.495
Corn	298.578	258.250	126.933	2.097	7.126
Cotton	67.665	65.150	19.798	2.688	16.481
Rice	9.243	8.440	3.557	0.844	3.503
Lumber	267.773	261.700	70.562	0.463	2.458
Gold	510.664	385.400	351.245	2.202	7.139
Silver	9.406	6.037	7.680	2.272	7.910
Platinum	755.715	534.000	463.352	1.169	3.096
Palladium	286.657	206.150	203.778	1.303	3.935
Copper	168.275	115.400	111.428	1.060	2.562
Light crude oil	38.103	26.740	27.475	1.371	3.827
Heating oil	112.316	67.655	86.145	1.292	3.484
Brent crude oil	41.547	25.410	32.501	1.205	3.199
Gas oil	375.818	226.500	281.273	1.161	3.180
Natural gas	3.987	3.142	2.478	1.370	4.950
Gasoline RBOB	227.116	223.895	57.877	0.023	2.309
Live cattle	75.023	71.488	15.871	1.219	4.915
Lean hogs	63.726	63.345	13.133	0.165	2.830

* Note that the Kurtosis measure employed here is *not* an excess Kurtosis measure.

Table 4: Summary statistics - 2 of 2 - price level series

Series	T	Quantiles 0%	25%	50%	75%	100%	C.V.	$f(\omega)$
Soybean meal	9280	117.700	166.700	185.800	233.700	548.100	0.334	193.877
Soybean oil	9280	13.070	19.790	23.750	28.140	70.400	0.396	204.664
Soybeans	9280	410.000	562.750	629.000	774.313	1771.000	0.348	199.646
Orange juice	9280	54.650	92.800	115.125	136.600	219.950	0.282	191.674
Sugar	9280	2.690	7.690	9.830	12.623	45.640	0.547	198.670
Wheat	9280	214.750	305.500	357.500	436.000	1280.000	0.376	198.684
Cocoa	9280	682.000	1303.000	1621.000	2230.000	5368.000	0.405	204.087
Coffee	9280	41.500	91.500	124.450	148.293	314.800	0.380	194.797
Corn	9280	142.750	223.000	258.250	327.313	831.250	0.425	199.060
Cotton	9280	28.520	56.700	65.150	74.850	215.150	0.293	184.060
Rice	6309	3.430	6.730	8.440	11.030	24.460	0.385	201.108
Lumber	7005	138.100	205.500	261.700	315.800	492.400	0.264	180.010
Gold	9280	142.800	329.600	385.400	478.800	1888.700	0.688	212.997
Silver	9280	3.510	5.000	6.037	11.013	48.700	0.816	202.275
Platinum	7009	333.100	408.700	534.000	1012.700	2276.100	0.613	212.473
Palladium	7009	76.200	133.250	206.150	366.400	1082.800	0.711	207.577
Copper	6309	60.600	84.850	115.400	249.300	462.850	0.662	211.914
Light crude oil	7793	10.420	19.190	26.740	52.240	145.290	0.721	209.987
Heating oil	6944	29.520	51.948	67.655	168.920	410.600	0.767	209.983
Brent crude oil	6427	9.640	18.125	25.410	61.950	146.080	0.782	209.602
Gas oil	6160	91.250	167.688	226.500	560.125	1325.250	0.748	209.189
Natural gas	5964	1.079	2.129	3.142	5.430	15.378	0.621	193.989
Gasoline RBOB	1920	79.270	188.145	223.895	273.853	357.100	0.255	141.956
Live cattle	9280	38.400	65.050	71.488	81.650	133.850	0.212	199.270
Lean hogs	7009	21.200	54.800	63.345	72.466	106.275	0.206	166.416

* C.V. stands for coefficient of variation, which is equal to σ/μ. Also, $f(\omega)$ is the maximal element of the normalized spectral density. In all cases above the argmax frequency is $\omega = 0$ and this low frequency peak dominates the entire spectrum. Finally, T is the sample size.

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 the fact that most of the series are also positively skewed, again with a few exceptions in gasoline RBOB and lean hogs (and possibly orange juice). Examination of the histograms in **appendix A** yield similar interpretations. Moreover, some of the histograms indicate a bimodal structure, especially amongst those 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.

The series themselves all exhibit a very high level of persistence in the sense that (equivalently) their autocorrelation functions die off very slowly and their normalized spectral densities exhibit extremely sharp peaks at the zero frequency and are near zero elsewhere in the spectrum (see **table 4**). 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). This should not come as a surprise given what we know about the properties of some exotic parametric processes which are able to ellude detection but 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 6.4 (see Gourieroux and Zakoian, 2011). Clearly 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. Futhermore, a careful visual inspection of the time series themselves do not suggest the typical pattern associated with a linear unit root (even when the innovations do not represent weak white noise). Finally, linear unit root tests have been shown to have low power in the presence of nonlinearity (such as multiple regimes, for example).

Interestingly, a quick examination of the price level plots for many of the series suggests the presence of "bubble" like movements. That is, quick periods of dramatic price increases followed by an equally quick and precipitous decline. The question of what process might describe these bubble dynamics will be left for later examination when we describe the models,

but for now let us just say that a case can be made that at least for some of these bubble episodes speculative behaviour or the act of investors attempting to "corner the market" have driven their occurrences. For example, consider the large bubble in the silver price level series from **figure 22**, that takes place within the first 2000 days of 7/18/1977. This bubble reflects a famous episode of speculative excess after a failed effort by the brothers Nelson B. Hunt and William H. Hunt to corner the silver market in the late 1970's and early 1980's (Eichenwald, 1989), where at one point the combined holdings of the brothers included more than a half of the world's deliverable silver futures. Another example comes from the copper futures series, where rogue trader Yasuo Hamanaka of Sumitomo Corporation, attempted to corner the international copper market over a ten year period leading up to 1996 (Gettler, 2008). In this latter case, the bubble doesn't even stand out as a period of abnormal price dynamics. Given the number of these episodes that can be recounted from history, and the way in which they are difficult to identify by a simple visual inspection of the price series themselves, one might wonder whether they prove to be the exception or the rule in driving futures price level dynamics.

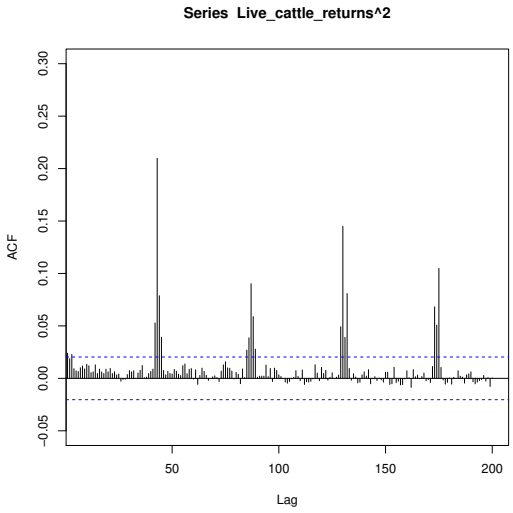
Inspection of the partial autocorrelation functions, $\rho(s)$, of the levels series reveals a general stylized fact of $\rho(1) \approx 1$ and small but significant $\rho(s)$ for some $s > 1$. This would suggest that if the model is linear then a causal AR(1) on the levels might suffice. In terms of the returns series, they exhibit small but significant (partial) autocorrelations; however, autocorrelation functions of the squares and absolute values are moderate and slowly declining to about the 50th lag. Moreover, a visual inspection of the returns series themselves reveals volatility clustering. These facts are both in line with the general stylized features of financial data and the possibility of modeling the returns as conditionally heteroskedastic (note that from **table 20** we see that the returns unconditional distributions are also leptokurtic and *negatively skewed*). Interestingly, both of the squared and absolute value of returns for lean hogs, live cattle, and the lumber exhibit no signicant autocorrelation except in spikes at periodic intervals in their autocorrelation functions, perhaps reflecting seasonal volatility spikes in their market – why this effect does not show up for other soft goods is curious, as is why the "in between spike" periods exhibit little autocorrelation. For reference I include the respective plots for the series which exhibits both the stylized features of the returns and the seasonally significant lags (I omit the other plots

given space considerations and their similarity; they are available upon request). See **figures 1 and 2**. Finally, of interest is the aberration in the cotton series returns which exhibit no autocorrelation at all in its squared values, but a significant amount in its absolute values.

4.1 Testing for nonlinearity

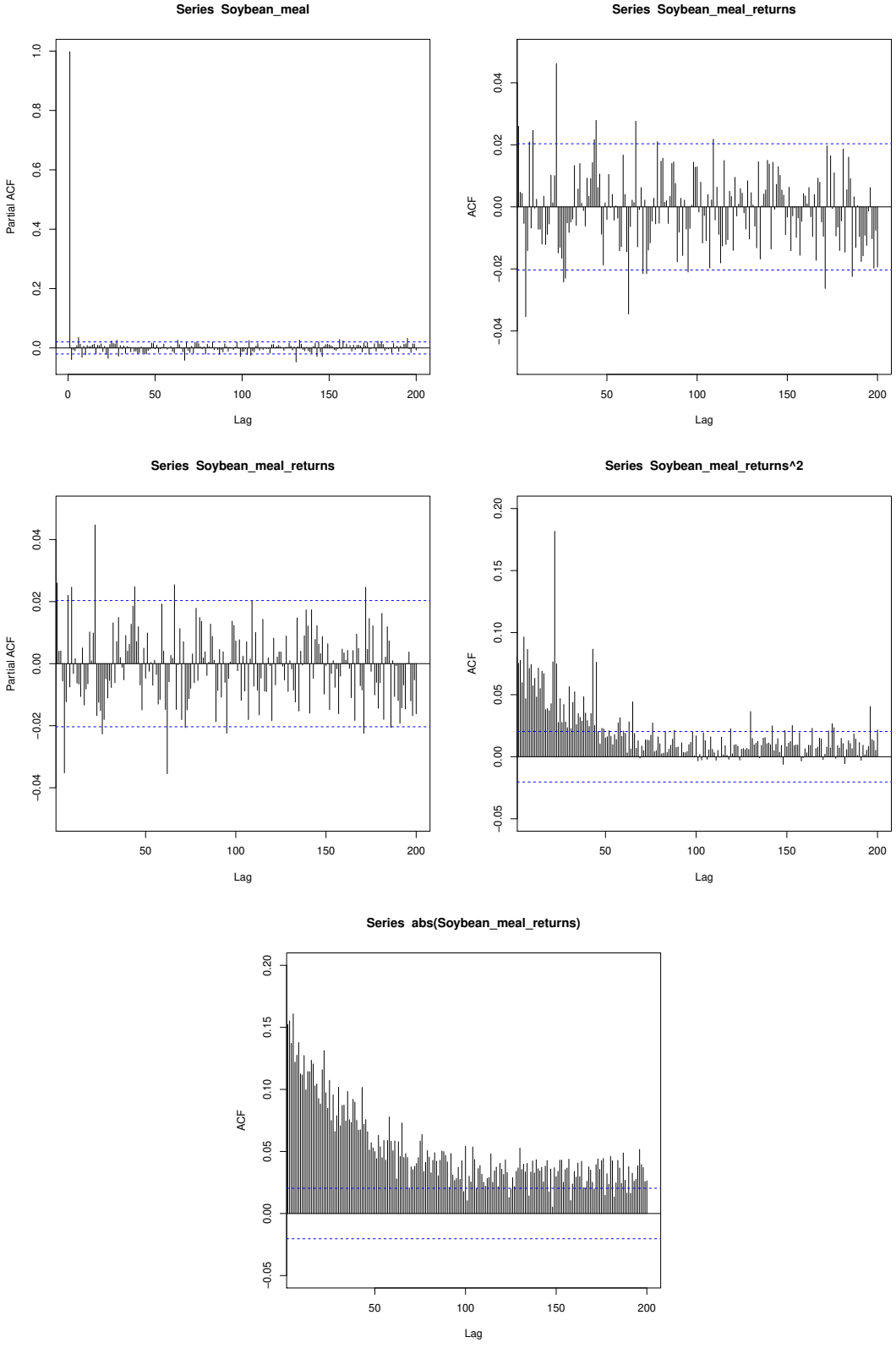
Of course, if this was all there was to say about the structure of the series, it would seem reasonable to simply fit a conditionally heteroskedastic model, such as an ARCH, to the returns and be done. However, as we shall see the autocovariance function is only useful for characterizing the process up to its 2nd order properties. That is, the autocovariance has nothing to say when it comes to higher order structure. It turns out that both the linear mixed causal/noncausal models with non-Gaussian innovations, and the causal nonlinear models more generally, are only indentified by higher order cumulant moments of order $k \geq 3$.

Figure 2: Autocorrelation function for Live cattle, squared returns



A number of tests for nonlinear structure have been suggested in the literature. For example, the BDS (Brock, Dechert and Scheinkman, 1987) test can be employed on the residuals of a best fitting ARMA (chosen according to say an information criterion like the AIC and a suitable test of weak white noise residuals) in order to look for deterministic chaos. 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

Figure 1: Autocorrelation functions for Soybean meal



in m-space. Of course, one alternative to deterministic chaos is stochastic nonlinearity and so the test can be considered more broadly as a test for nonlinearity.

From a frequency domain approach, the Hinich bispectrum test (1982) can test both Gaussianity and linearity and is related to the test developed by Subba Rao and Gabr (1984) and discussed in Priestley (1989) (see the rest of this section for more details). It has a less broad alternative than the BDS test since the alternative is linearity. Of course the bispectrum test has zero power against some types of nonlinearity since some processes may have zero bispectrum but non zero trispectrum and higher. Hinich (1996) also discusses a newer test which is the time domain analogue of the frequency domain bispectrum type tests.

For more details on various tests for nonlinearity the interested reader can consult Barnett et. al. (1996).

4.2 The BDS test

The BDS test is a test for independence. Therefore, we are required to first remove any linear dependence from the series before testing the residuals for possible nonlinear dependence. If the residuals from a best fitting linear causal model are not i.i.d., the linear causal model cannot be adequate in describing the dynamics of the data.

Testing of the innovations from the best fitting causal linear Gaussian ARMA model suggests nonlinear dynamics left uncaptured in both the levels and returns series. The 25 series are fit according to the AIC criterion to select a "best" fitting model and the residuals are then tested via the BDS test statistic up to an embedding dimension of $m = 15$.[5] Generally, for all the series (both returns and levels), the best fitting linear Gaussian ARMA model residuals are not i.i.d. at the 5% test significance level. However, there is one exception in the *lean hogs* price levels series, where for the largest value of ϵ (the parameter that defines "near points" in the m-dimensional space, i.e. $\|u - v\| < \epsilon$), we are not able to reject the null hypothesis of i.i.d. residuals. Interestingly, this result is despite remaining autocorrelation that exists in the

[5]In order to facilitate this search for a best fitting model I employ the *auto.arima()* function in R due to Hyndman, R.J. and Khandakar, Y. (2008) "Automatic time series forecasting: The forecast package for R", *Journal of Statistical Software*, 26(3). I employ the constraint of no differencing, and a maximum order of $p + q = 10$, $p \leq 11$ and $q \leq 3$. AIC's are specified not to be approximated and are generated directly from the MLE while the "stepwise" selection procedure is avoided to make sure all possible model combinations are tested.

residuals due to the seasonal patterns discussed earlier (the best model chosen does not include seasonal lags, given that the seasonal pattern tends to occur even 43rd day or so) – therefore, the result should be even more suggestive of i.i.d. innovations if the seasonality is accounted for. Note that this anomolous result for *lean hogs* will also be borne out when we inspect its bicoherence function. It should also be noted that the estimation selection process reveals a great difficulty in fitting a linear Gaussian ARMA model to the levels data and often models are rejected in that their parameters lie too close to the border of both stationarity and invertibility regions, resulting in estimator convergence problems. Finally, in section 5, I will also show that the causal linear ARMA Gaussian models can be improved upon by employing conditionally t-distributed innovations instead; although this still does not resolve the problem of uncaptured nonlinear structure.

4.3 The frequency domain Bispectrum test

The bispectrum type tests rest on the following logic from Subba Rao and Gabr (1984). Suppose (X_t) has the linear (possibly mixed causal/noncausal) representation,

$$X_t = \sum_{r=-\infty}^{\infty} a_r \epsilon_{t-r} \tag{2}$$

where (ϵ_t) is a sequence of i.i.d. variables, with $E[\epsilon_t] = 0, E[\epsilon_t^2] = \sigma_\epsilon^2, E[\epsilon_t^3] = \mu_3$. It can be shown that the autocovariance function of X_t is:

$$R(s) = \sigma_\epsilon^2 \sum_{r=-\infty}^{\infty} a_r a_{r+s} \tag{3}$$

and that the third order cumulant $k_3 = C(t_1, t_2)$ is:

$$C(t_1, t_2) = E[X_t X_{t+t_1} X_{t+t_2}] = \mu_3 \sum_{r=-\infty}^{\infty} a_r a_{r+t_1} a_{r+t_2} \tag{4}$$

Therefore, letting $A(\omega) = \sum_{j=-\infty}^{\infty} a_j e^{-ij\omega}$ be the transfer function we have the spectral

density of X_t as:

$$h(\omega) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} R(s) e^{-is\omega} \tag{5a}$$

$$= \frac{\sigma_\epsilon^2}{2\pi} \sum_{s=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} a_r a_{r+s} e^{-is\omega} \tag{5b}$$

$$= \frac{\sigma_\epsilon^2}{2\pi} |A(\omega)|^2 \tag{5c}$$

and likewise we have that the bispectrum is given as:

$$h_3(\omega_1, \omega_2) = \frac{1}{(2\pi)^2} \sum_{t_1=-\infty}^{\infty} \sum_{t_2=-\infty}^{\infty} C(t_1, t_2) e^{-i(t_1\omega_1 + t_2\omega_2)} \tag{6a}$$

$$= \frac{\mu_3}{(2\pi)^2} \sum_{t_1=-\infty}^{\infty} \sum_{t_2=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} a_r a_{r+t_1} a_{r+t_2} e^{-i(t_1\omega_1 + t_2\omega_2)} \tag{6b}$$

$$= \frac{\mu_3}{(2\pi)^2} A(\omega_1) A(\omega_2) A^*(\omega_1 + \omega_2) \tag{6c}$$

where the asterisk denotes complex conjugate. Therefore, for the linear model we have that:

$$X_{ij} = \frac{|h_3(\omega_i, \omega_j)|^2}{h(\omega_i) h(\omega_j) h(\omega_i + \omega_j)} = \frac{\mu_3^2}{2\pi\sigma_\epsilon^2}, \quad \text{for all} \quad (i, j) \tag{7}$$

which is called the *bicoherence* between frequencies ω_i and ω_j. That is, for the linear model, the bicoherence should be constant across bifrequencies. Moreover, for models where ϵ_t is Gaussian, we have that μ_3 is necessarily zero and so the bispectrum of X_t is zero as well (under suitable conditions imposed on the sequence of a_r's). The essence of the Hinich and Subba Rao & Gabr bispectral tests is therefore to test where or not 1) the bispectrum of X_t is significantly different from zero, as measured by a test statistic related to Hotelling's T^2 distribution and 2) if the process X_t is non-Gaussian test whether X_{ij} is non-constant over a selected grid of bifrequencies. Again, note that the derivation above is quite general and applies to the processes of mixed causality.

4.4 The estimated bispectral modulus and bicoherence functions

4.4.1 The Bilinear model

While these statistical tests are useful, it is often the case that a simple inspection of the estimated bispectrum and bicoherence is enough to rule out either Gaussianity and linearity given the dramatic shapes of the bispectrum and bicoherence function plots. For example, consider first the particular bilinear process of Granger and Andersen (1978) defined as:

$$X_t = 0.75X_{t-1}\epsilon_{t-1} + \epsilon_t, \quad \text{where} \quad \epsilon_t \sim N(0,1) \tag{8}$$

It is clear that X_t represents a nonlinear process (we say the process is "nonlinear in mean"). A plot of the sample path of the process is given in **Figure 13** within **Appendix A**, for the first 1000 values from a total sample path of $T = 30000$. Also included is the bicovariance function $E[(X_t - \mu_x)(X_{t-t_1} - \mu_x)(X_{t-t_2} - \mu_x)]$ for $(t_1, t_2) \geq 0$ and both the bispectrum (6c) and bicoherence (7) for bifrequencies $(\omega_1, \omega_2) \geq 0$ (with a spectral period of 1).

Notice that the autocovariance and spectral density functions are not helpful here since they are unable to identify departures from Gaussianity and their shapes do not suggest anything out of the ordinary. However, the shapes of the bicovariance and bispectral modulus suggest that the process is definitely non-Gaussian and the curving, wing-like, segments of the bicoherence reveal that the process is also nonlinear.

4.4.2 The GARCH model

As another example, consider the pure GARCH(1,1) process of Bollerslev (1986); that is $(\epsilon_t = \sigma_t z_t)|F_{t-1} \sim N(0, \sigma_t^2)$ with $\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2$ and $F_{t-1} = \{z_{t-1}, z_{t-2}\ldots\}$, where $z_t \sim N(0,1)$. The ARCH and GARCH models form a special type of nonlinear process known as *multiplicatively nonlinear*, or "nonlinear in variance" (as opposed to "in mean" as was the case with the bilinear process above). From Hinich (2009) we know that since z_t is i.i.d. and symmetric we have that $E[z_t^3] = \mu_3 = 0$ and so the bispectrum of the GARCH(1,1) is zero for all bifrequencies (see (6c)). Moreover if z_t is not symmetric then $\mu_3 \neq 0$ and so the bispectrum is a real constant for all bifrequencies. Of course, the interesting feature of the ARCH/GARCH

models is the fact that up to 2nd order properties, they appear to be white noise (and so they represent processes which are weak white noise, but not i.i.d.). **Figure 14** within **Appendix A** demonstrates the results discussed with parameters $\alpha_0 = \alpha_1 = 0.1$, $\beta_1 = 0.89$ and sample size $T = 30000$. Note that given the noise in the bicovariance estimate given a finite sample the esimated bispectrum modulus trends to have a high variance as well. Moreover, it exhibits a positive level since despite the symmetric distribution of z_t, the finite sample r_t skewness is not quite zero (it is 0.1062087).

The fact that the shape of the bispectrum is unaffected by nonlinearity in variance shouldn't come as a surprise since the variance is an *even* moment. Therefore the GARCH model only shows up in the even moments, such as the variance and kurtosis (unless the distribution of z_t is asymmetric, in which case the levels process unconditional mean and skewness is affected directly). Likewise, models that are nonlinear in mean are likely to have an effect on the *odd* moments of the unconditional distribution, such as the mean and the skewness.

Moreover, from Campbell, Lo, and MacKinlay (1996, Ch12.1) we know that models which are nonlinear in mean have nonzero bicovariance (and higher) for all $(t_1, t_2, \ldots)$ Models that are nonlinear in variance, but obey the martingale property (e.g. GARCH) have zero bicovariance (and higher) unless one of the elements of $(t_1, t_2, \ldots)$ is equal to another. For example, they show that the ARCH(1) has zero bicovariance but for $t_1 = t_2$ and $t_3 = 0$ the tricovariance is nonzero. This fact explains why testing the autocovariance function of the squared returns, i.e. $E[r_t^2 r_{t-t_1}^2]$, for ARCH effects has become so commonplace (the McLeod and Li test (1983)). Note that these features of the bicovariance function will be exploited in section 8 when I employ a statistical test from Hsieh (1989), section 5, in determining more specifically the functional form of any nonlinear structure.

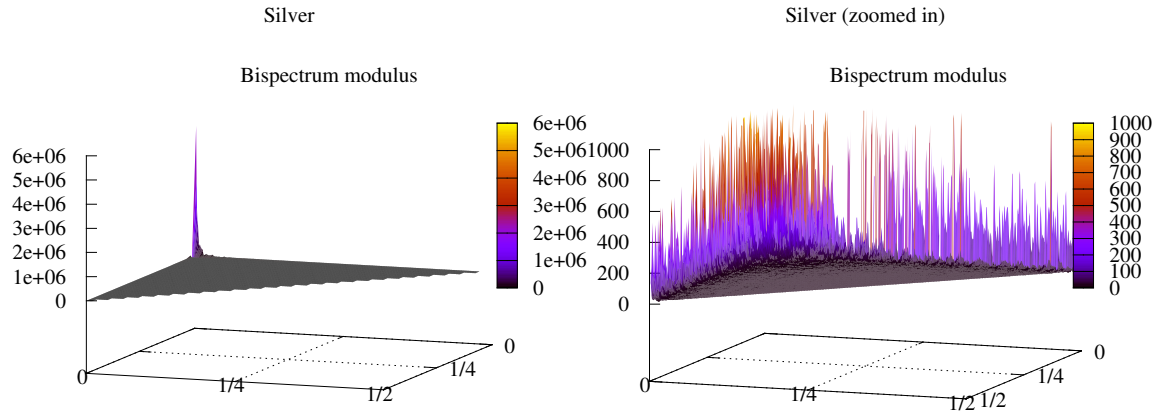
4.4.3 The commodity futures data

For now, consider the shapes of the bispectrum modulus and bicoherence for both the levels and returns of the 25 futures price series. An initial inspection of the estimated bispectrum modulus and bicoherence function plots for the returns series reveals mostly noise centered around zero (with the exception of a spike at the origin reflecting the non-zero skewness of the returns

series)[6] which suggests these series are not nonlinear in mean, although the ARCH effects suggests they are strongly nonlinear in variance. Of course, this nonlinearity would show up in the trispectrum but not the bispectrum. While some of the bicoherence plots suggest some subtle structure (for example cotton, live cattle, and lean hogs), most of this could be due to seasonality creating periodic skewness.

As for the price levels themselves, the bicovariance, bispectrum modulus, and bicoherence all suggest definite nonlinearity in mean, with possibly the exception of the *lean hogs* series (which corroborates with the BDS test results earlier). First, all of the bicovariance functions are slowly declining and share the same shape with the exception of natural gas, gasoline RBOB, lean hogs, and lumber; see **Figure 15** within **Appendix A**. I've included the silver price level series bicovariance plot as well, since it exhibits the representative shape of all the others (although their heights differ according to the level of κ_3 of the series, where κ_i is the ith estimated cumulant). Finally, notice that the shape of all the bicovariance estimates implies the obvious fact that the unconditional distributions of the price level series are not Gaussian.

Figure 3: Bispectrum, price levels



Of course, this suggests that the bispectrum modulus will be similar for all the series except the above mentioned four: natural gas, gasoline RBOB, lean hogs, and lumber. Since their spectral densities all differ amongst the higher frequencies (recall that they exhibited large

[6]These series are skewed but not "auto-coskewed" as a process nonlinear in mean would exhibit.

peaks at the zero frequency), it is not clear that their bicoherence functions will all be the same, however. Inspection of the bispectrum modulus plots suggests they all exhibit the same overall peaked shape exhibited in **Figure 3**. While the four anomalous series mentioned above do exhibit slight deviations especially with regards to their "edges" (where either frequency $\omega_1 \approx 0$ or $\omega_2 \approx 0$) they do not suggest any dramatically different features. It is interesting to note by analogy that just as the area under the spectral density within any frequency interval represents the proportion of the overall unconditional variance of the process explained by the frequencies in the interval, the same applies to the area under the bispectral modulus – that is, the area within any plane in the bifrequency domain represents the proportion of the overall unnormalized skewness of the process explained by the bifrequencies in the plane. Therefore, not only are the processes skewed (and therefore non-Gaussian), but the vast majority of this skewness is generated by the extreme low frequency components of the process (i.e. the low frequency trend movements). Of course, in this context most of the low frequency movements that cause skewness in the unconditional process are the *bubble movements*.

The bicoherence plots are also quite revealing. With the exception of orange juice, cocoa, coffee, cotton, gasoliner RBOB, lean hogs, and lumber, all of the bicoherence functions share the same exponentially increasing shape (see **Figures 16 and 17** within **Appendix A** for which i've provided the biocoherence of the silver price level series as an example). As for the other series, the bicoherence grows but at a much slower rate, with the exception of lean hogs, gasoline RBOB, and possibly cotton which suggests these series may be linear in mean. Another notable feature is the fact that cocoa's bicoherence seems to grow more linearly than exponentially. Either way, probably the most interesting fact of these bicoherence plots is that they are growing in bifrequencies close to each other; that is, they are growing in $\omega_1 \approx \omega_2 \neq 0$.

4.4.4 Conclusions

All the results above tend to suggest the presence of nonlinear structure in the price levels (through perhaps both the mean and variance if we are to decompose the process as its first order approximation $X_t = f(\epsilon_t, \epsilon_{t-1}, \ldots) \approx f(0, \epsilon_{t-1}, \ldots) + \epsilon_t f_1(0, \epsilon_{t-1}, \ldots) \equiv g(\epsilon_{t-1}, \ldots) + \epsilon_t h(\epsilon_{t-1}, \ldots)$), and nonlinear structure in the returns through their variance. However, we will

have to wait until **section 8** in order to further determine the functional form of any nonlinearity since we must first discuss the mixed causal/noncausal autoregressive model, and the concept of reversibility of a time series, in **section 6**.

5 The linear causal ARMA model

In this section I will now show that the causal linear ARMA model with both Gaussian and non-Gaussian innovations is unable to adequately capture the features of the futures price level data. Recall that the returns data exhibited very little linear autocorrelation; however, upon fitting the best Gaussian ARMA model, the innovations exhibited features suggestive of nonlinearity such as a rejection of the i.i.d. hypothesis of the BDS test statistic and the presence of ARCH effects. Moreover, an inspection of the bicovariance of returns suggested little in terms of nonlinearity or irreversibility (although that is not to say that these returns series may not be nonlinear or irreversible through their higher order cumulants beyond κ_3). For this reason I will leave modeling the returns for later, when I consider models that are nonlinear in variance in section 8.

In order to assess the ARMA model's ability to fit the price level data, I employ the AIC criterion, along with Ljung-Box statistics testing the hypothesis that both the innovations and their squares exhibit no linear autocorrelation.

The standard Gaussian ARMA model is considered, as is the case where the innovations are both t-distributed and skew t-distributed i.i.d. random variables. The skew t-distribution is entertained as an option considering the skewed nature of the data series histograms. The skewed t-distribution employed is of the two parameter type as described in Jones (2010). As will be shown, while the t-distributed innovations improve model fit dramatically, they are ultimately unable to generate i.i.d. residuals given a causal, linear ARMA model.

Table 5 presents the results of numerical maximum likelihood estimation. The model parameters, (p, q), were selected via normalized AIC (the AIC divided by the number of elements in the log-likelihood summation) amongst a set of of possible set of (p, q) values such that $p \leq 10$ and $q \leq 2$. The first row of the results for each series represents the Gaussian error term

case, while the second and third rows represent the t-distributed and skew t-distributed cases respectively. The lags column represents the number of lags (the natural log of the sample size $\ln(T)$) included in both the Ljung-Box and McLeod-Li test statistics, where p-values are provided in their respective columns. Finally, an 'x' marks the model with the lowest normalized AIC.

A few things are of note here. First, in every case, the addition of t-distributed errors improves model fit dramatically, over the Gaussian case. Since the ARMA model with Gaussian errors is nested (the t-distribution with infinite degrees of freedom is Gaussian) I employ likelihood ratio (LR) statistics to determine if the improvement in fit is statistically significant. In every case the addition of t-distributed innovations passes a LR test at the 1% significance level.

However, in comparing the skewed t-distributed model to the standard t-distribution, the results are mixed. In most cases the skewed t-distribution improves model fit and passes a LR test at the 1% level. Likewise, corn and rice also pass at the 5% level and sugar and heating oil at the 10% level. Exceptions to these results are found with natural gas, Brent crude oil, light crude oil, gold, and coffee which all fail LR tests at the 10% level. This suggests that there is little gain in employing a skewed t-distribution on the innovations of their ARMA representations.

Moreover, even in the cases where the skewed t-distributed is found to be significant, the level of skew is very small. For example, in the skewed-t case, the distribution is defined by two skew parameters, α and β where their sum equals the degrees of freedom of the distribution. If $\alpha > \beta$ then the distribution is positively skewed and vice versa if it is negatively skewed. In most cases the skew is positive, however, for platinum, silver, orange juice, soybeans, lean hogs, live cattle, gasonline RBOB, and gas oil the skew is negative. The fact that the skew is negative for soybeans, but not for soybean meal or soybean oil seems strange. Since the conditional distribution of the series price levels (for the causal model based on information at $t-1$) is defined by that of the innovations, it is interesting to note how the unconditional skewness differs from the conditional: recall that all the unconditional distributions were positively skewed. This inconsistency is most likely due to the uncaptured nonlinear structure. The ratio of the estimated skew parameters α/β (and its inverse for negatively skewed series) ranges

from 1.009 for copper to 1.151 for lumber, however most range around 1.035 to 1.065. Finally, it is interesting to note how small the estimated t-distribution degrees of freedom parameter is. For most series it ranges from 1.056 (nearly Cauchy distributed) to 2.73 for most series, with exceptions: live cattle at 3.4, lumber at 4.23, and gasoline RBOB at 4.92.

Ultimately, however, the causal linear ARMA model is not able to fully capture the structure of the data. The McLeod-Li test of autocorrelation in the squared residuals suggests uncaptured nonlinearity in all cases. Plots of the residuals series also suggest ARCH effects. Interestingly, in some cases (especially the Gaussian) the Ljung-Box test is unable to reject the null of no autocorrelation in the residuals' levels, despite autocorrelation in their squares. Either way, the evidence presented in this section suggests that we need a better model if we are to adequately capture the features of the futures price level data.

Table 5: Estimation results of ARMA models - 1 of 2

		p	q	Log-likelihood	AICn	Lag	Ljung-Box	McLeod-Li
Soybean meal		10	0	-26155.680	5.64567	9	1.000	0.000
		2	1	-24203.947	5.21879	9	0.000	0.000
	x	10	0	-24171.704	5.21806	9	0.000	0.000
Soybean oil		10	0	-5911.793	1.27806	9	1.000	0.000
		5	0	-4653.824	1.00525	9	0.001	0.000
	x	10	0	-4639.175	1.00392	9	0.007	0.000
Soybeans		10	0	-36719.741	7.92486	9	1.000	0.000
		5	0	-34817.241	7.50949	9	0.000	0.000
	x	10	0	-34779.631	7.50672	9	0.000	0.000
Orange juice		1	2	-21058.066	4.53994	9	0.429	0.000
		3	0	-19408.642	4.18554	9	0.033	0.000
	x	10	0	-19382.514	4.18479	9	0.070	0.000
Sugar		10	0	-3934.267	0.85141	9	1.000	0.000
		3	0	-796.395	0.17299	9	0.000	0.000
	x	1	2	-794.876	0.17284	9	0.000	0.000
Wheat		10	0	-33493.832	7.22887	9	1.000	0.000
		5	0	-30990.145	6.68424	9	0.000	0.000
	x	5	0	-30983.923	6.68311	9	0.000	0.000
Cocoa		10	0	-47045.067	10.15255	9	1.000	0.000
		1	0	-45932.361	9.90114	9	0.002	0.000
	x	10	0	-45847.856	9.89468	9	0.000	0.000
Coffee		10	0	-24225.191	5.22917	9	0.999	0.000
	x	10	0	-21901.578	4.72806	9	0.000	0.000
		10	0	-21901.239	4.72821	9	0.000	0.000
Corn		1	1	-29700.534	6.40253	9	0.706	0.000
		2	1	-26849.379	5.78904	9	0.438	0.000
	x	2	0	-26847.118	5.78856	9	0.419	0.000
Cotton		10	0	-16352.151	3.53056	9	1.000	0.000
		3	0	-13515.309	2.91502	9	0.000	0.000
	x	10	0	-13494.465	2.91445	9	0.000	0.000
Rice		2	1	2398.111	-0.75887	8	0.295	0.000
		1	1	3554.157	-1.12529	8	0.000	0.000
	x	2	2	3559.414	-1.12618	8	0.000	0.000
Lumber		1	1	-22031.108	6.29215	8	0.478	0.000
		1	1	-21541.809	6.15272	8	0.484	0.000
	x	1	1	-21531.166	6.14996	8	0.384	0.000
Gold		10	0	-32334.783	6.97881	9	1.000	0.000
		3	0	-28530.443	6.15208	9	0.000	0.000
	x	2	2	-28528.430	6.15142	9	0.000	0.000
Silver		10	0	-3714.602	0.80401	9	0.999	0.000
		5	0	3469.612	-0.74644	9	0.000	0.000
	x	10	0	3479.433	-0.74767	9	0.000	0.000

Table 6: Estimation results of ARMA models - 2 of 2

		p	q	Log-likelihood	AICn	Lag	Ljung-Box	McLeod-Li
Platinum		1	1	-27963.539	7.98160	8	0.237	0.000
		5	0	-25925.178	7.40525	8	0.000	0.000
	x	10	0	-25897.058	7.40422	8	0.000	0.000
Palladium		5	0	-24089.578	6.88080	8	0.257	0.000
		10	0	-21354.485	6.10587	8	0.000	0.000
	x	5	0	-21367.921	6.10420	8	0.000	0.000
Copper		10	0	-17319.500	5.50294	8	1.000	0.000
		5	0	-15333.940	4.86737	8	0.000	0.000
	x	10	0	-15311.933	4.86615	8	0.000	0.000
Light crude oil		5	0	-11110.308	2.85498	8	0.664	0.000
		2	0	-8663.663	2.22530	8	0.001	0.000
	x	1	1	-8663.256	2.22517	8	0.002	0.000
Heating oil		1	2	-17228.339	4.96423	8	0.255	0.000
	x	2	0	-15489.176	4.46389	8	0.064	0.000
		1	1	-15490.537	4.46393	8	0.070	0.000
Brent crude oil		2	2	-9405.228	2.92957	8	0.007	0.000
	x	2	2	-7617.552	2.37340	8	0.001	0.000
		2	2	-7616.836	2.37349	8	0.002	0.000
Gas oil		2	1	-22060.940	7.16659	8	0.487	0.000
		1	0	-20644.213	6.70505	8	0.640	0.000
	x	1	0	-20640.309	6.70411	8	0.532	0.000
Natural gas		2	2	2089.730	-0.69900	8	0.160	0.000
	x	1	2	3892.707	-1.30361	8	0.014	0.000
		1	1	3892.495	-1.30354	8	0.010	0.000
Gasoline RBOB		5	0	-5839.139	6.10563	7	0.814	0.000
		3	0	-5765.853	6.02176	7	0.060	0.000
	x	3	0	-5760.988	6.01772	7	0.070	0.000
Live cattle		10	0	-11363.842	2.45433	9	1.000	0.000
		10	0	-10215.703	2.20684	9	0.721	0.000
	x	10	0	-10210.809	2.20600	9	0.750	0.001
Lean hogs		1	1	-11779.078	3.36275	8	0.601	0.996
		1	0	-9478.016	2.70605	8	0.221	0.994
	x	5	0	-9462.261	2.70453	8	0.676	0.994

* Lags refers to the number of lags incorporated into both the Ljung-Box and McLeod-Li test statistics. This value is $\ln(T)$. Moreover, the values listed of the aforementioned statistics are the p-values. The AICn value is the AIC normalized by the number of elements in the log-likelihood summation.

6 The mixed causal/noncausal autoregressive model

6.1 The general form

Definition 6.1. *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:

$$\phi(L)\varphi(L^{-1})x_t = \epsilon_t, \tag{9a}$$

$$\text{where} \quad \varphi(L^{-1}) = 1 - \varphi_1 L^{-1} - \varphi_2 L^{-2} - \ldots - \varphi_s L^{-s}, \tag{9b}$$

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_r L^r, \tag{9c}$$

and (ϵ_t) is a sequence of i.i.d. random variables.

L is the lag operator where $Lx_t = x_{t-1}$ and $L^{-1}x_t = x_{t+1}$. Moreover, we assume that the lag polynomial operators $\varphi(z^{-1})$ and $\phi(z)$ have their roots strictly outside the complex unit circle. In other words,

$$\varphi(z) \neq 0 \quad \text{and} \quad \phi(z) \neq 0 \quad \text{for complex} \quad z \quad \text{where} \quad |z| \leq 1. \tag{10}$$

Another way to write the model in (9), employed in Breidt et. al. (1991), is to consider an autoregression of order $k = r + s$:

$$\alpha(L)x_t = u_t, \tag{11a}$$

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

where the operator $\alpha(z)$ can be factorized as $\alpha(z) = \alpha_1(z)\alpha_2(z)$, where $\alpha_1(z)$ (of order r) contains all its roots strictly outside the complex unit circle and $\alpha_2(z)$ (of order s) contains all its roots strictly inside the unit circle. Of course, we can always map the parameters from model (9) to (11) since we have that $-(1/\alpha_s L^s)\alpha_2(L) = \varphi(L^{-1})$, where the coefficients of $\varphi(L^{-1})$ are given as $\varphi_i = -\alpha_{s-i}/\alpha_s$ for $i = 1, \ldots, s-1$ and $\varphi_s = 1/\alpha_s$ for $i = s$, and the roots of $\alpha_2(L)$ and $\varphi(L^{-1})$ are inverses (in the sense that $\alpha_2(z) = \varphi(1/z) = 0$ for some complex z where $|z| < 1$). The mapping places the constraint on the innovations of both models such that $\epsilon_t = -(1/\alpha_s L^s)u_t = -(1/\alpha_s)u_{t+s}$.

6.2 Special cases and the linear moving average representation

The purely causal and purely noncausal autoregressive models are special cases in which the process error terms (ϵ_t) can be interpreted in terms of innovations.

6.2.1 The purely causal autoregressive process

When the noncausal operator in model (9), $\varphi(L^{-1})$ is equal to identity, that is, if $\varphi_i = 0$ for all $i = 1, \ldots, s$, so that $\phi(L)x_t = \epsilon_t$ where $\phi_j \neq 0$ for some $j = 1, \ldots, r$, we say that the process x_t is *purely causal* and we have that ϵ_t is now equal to a causal innovation. In other words, x_t is strictly a function of the past of the process, $\mathcal{F}_{t-1} = \{x_{t-1}, x_{t-2}, \ldots, x_{t-r}\}$, and the shocks ϵ_t are independent of the past. When ϵ_t admits second order moments and is zero mean then $\epsilon_t = x_t - E[x_t|\mathcal{F}_{t-1}]$. If the lag polynomial operator $\phi(L)$ is invertible then (x_t) has the equivalent epresentation given as

$$x_t = \phi(L)^{-1}\epsilon_t = \sum_{j=0}^{\infty} \phi_j^* \epsilon_{t-j} \tag{12}$$

where $\phi_0^* = 1$ and the right hand side converges in mean squared (i.e. is covariance stationary).

6.2.2 The purely noncausal autoregressive process

Analogously, when the causal operator in model (9), $\phi(L)$ is equal to identity, that is, if $\phi_i = 0$ for all $i = 1, \ldots, r$, so that $\varphi(L^{-1})x_t = \epsilon_t$ where $\varphi_j \neq 0$ for some $j = 1, \ldots, s$, we say that the process x_t is *purely noncausal* and we have that ϵ_t is now equal to a noncausal innovation. In other words, x_t is strictly a function of the future of the process, $\mathcal{F}_{t+1} = \{x_{t+1}, x_{t+2}, \ldots, x_{t+s}\}$, and the shocks ϵ_t are independent of the future. When ϵ_t admits second order moments and is zero mean then $\epsilon_t = x_t - E[x_t|\mathcal{F}_{t+1}]$. If the lag polynomial operator $\varphi(L^{-1})$ is invertible then (x_t) has the equivalent representation given as

$$x_t = \varphi(L^{-1})^{-1}\epsilon_t = \sum_{j=0}^{\infty} \varphi_j^* \epsilon_{t+j} \tag{13}$$

where $\varphi_0^* = 1$ and the right hand side converges in mean squared (i.e. is covariance stationary).

6.2.3 The mixed causal/noncausal autoregressive process

Finally, again given the model in (9), if we have that $\varphi_i \neq 0$ for some $i = 1, \ldots, s$, and $\phi_j \neq 0$ for some $j = 1, \ldots, r$, so that neither of the lag polynomial operators are equal to identity, then (9a) defines both a causal and noncausal autoregression and we say that the process is a mixed causal/noncausal autoregression of order (r, s).

From (12) and (13) we saw that if the process (x_t) in (9a) is either purely causal or purely noncausal then it admits a purely backward or purely forward representation. If the model is of mixed causality then we say that the process has the *two-sided* moving average representation given by

$$x_t = \varphi(L^{-1})^{-1}\phi(L)^{-1}\epsilon_t = \sum_{j=-\infty}^{\infty} \theta_j \epsilon_{t-j}. \tag{14}$$

where in this general case the error term ϵ_t is neither a causal or a noncausal innovation.

It will also be useful to define the following processes u_t and v_t. From (9a) and (10) let u_t be defined as

$$u_t = \phi(L)x_t = \varphi(L^{-1})^{-1}\epsilon_t = \sum_{j=0}^{\infty} \varphi_j^* \epsilon_{t+j}, \tag{15}$$

where $\varphi_0^* = 1$ and the right hand side converges in mean squared (i.e. is covariance stationary). We call (15) the *forward looking* representation.

Moreover, also from (9a) and (10) let v_t be defined as

$$v_t = \varphi(L^{-1})x_t = \phi(L)^{-1}\epsilon_t = \sum_{j=0}^{\infty} \phi_j^* \epsilon_{t-j}, \tag{16}$$

where $\phi_0^* = 1$ and the right hand side converges in mean squared (i.e. is covariance stationary). We call (16) the *backward looking* representation.

6.3 Reversibility and the strong linear processes

The nonlinear features of the majority of strong moving average processes, and in particular the mixed causal/noncausal processes, is due to their irreversibility properties. Let us first recall the definition of a *reversible* process is given as follows (e.g. Ramsey & Rothman (1996))

Definition 6.2. *Time reversible process*

A time series (X_t) is time reversible if for every positive integer n, every $t_1, t_2, \ldots, t_n \in \mathbf{R}$, and all $m \in \mathbf{N}$, the vectors $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ and $(X_{-t_1+m}, X_{-t_2+m}, \ldots, X_{-t_n+m})$ have the same joint probability distributions.

Of course, this is the same as the condition from Brillinger and Rosenblatt (1967, p.210) that a time reversible process must admit the same ith cumulant κ_i both backwards and forwards–that is, we must have $C(t_1, t_2, \ldots, t_{i-1}) = C(-t_1, -t_2, \ldots, -t_{i-1})$ for all i. [7]

Note that if the process (X_t) is both Markov and strong stationary, then the condition in Definition 1 reduces to $f_{t|t-1}(X_t|X_{t-1}) = f_{t-1|t}(X_t|X_{t-1})$ where $f_{t|t-1}(\cdot)$ is the conditional density of X_t, conditional on time $t-1$ information. In other words, the transition probability from state X_{t-1} into state X_t does not depend on whether time is flowing forward or backward.

Intuitively, the notion of reversibility implies that if we take some process (X_t), and reverse its time direction, that the probability law defining the joint process (or transition probability for a stationary Markov process) remains unchanged. This of course implies a sort of "time directional symmetry" both forwards and backwards in time. Therefore, the notion of reversible processes is important in considering the process of mixed causality in (9a) since this process implies both a forward and backward representation through the terms v_t and u_t in (15) and (16).

Intuitively, the process (x_t) is reversible if its distributional properties are the same in both direct and reverse time. In general a strong linear process is not reversible. More precisely we have the following proposition,

Proposition 6.3. *Properties of the strong Gaussian process*

i) A process admits both strong linear causal and nonlinear causal representations if and only if the process is Gaussian.

ii) If so, this process is also reversible.

The first results in (i) is due to the fact that the distribution of a Gaussian process is characterized entirely by its mean and its covariance function, $R_x(s) = Cov(x_t, x_{t+s})$ and since $R_x(s) = R_x(-s)$, the result in (ii) follows. An important consequence of proposition 6.3 is

[7] see equation (4).

that a general mixed causal/noncausal autoregressive process with non-Gaussian error terms will feature nonlinear dynamics both in the direct and reverse time directions.

6.3.1 Example: Gaussian innovations

To illustrate proposition 6.3 let us consider the following example of a simulated AR(1) process with standard Gaussian innovations given in **Figure 4**; that is a process defined by $x_t = \rho x_{t-1} + \epsilon_t$ where $\epsilon_t \sim N(0,1)$ and $\rho = 0.8$.

Figure 4: The Gaussian AR(1) process

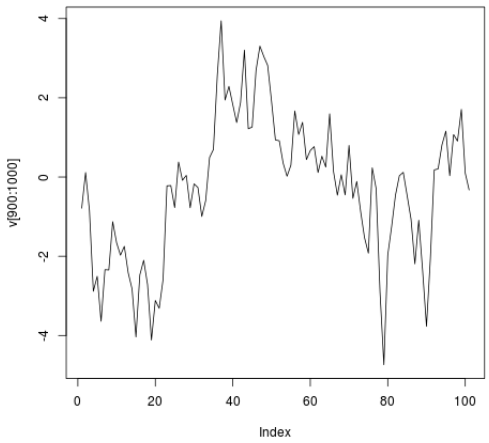
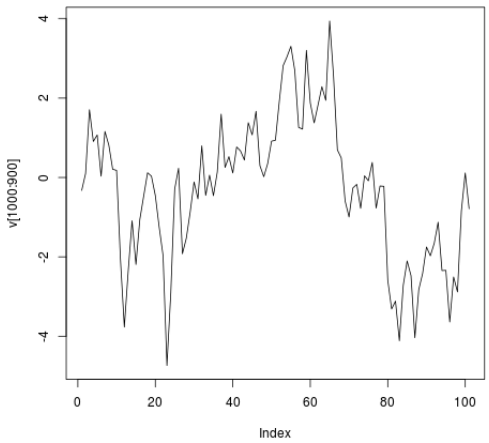


Figure 5: The Gaussian AR(1) process in reversed time



Now consider the same sample path but in reversed time shown in **figure 5**. By proposition 6.3 we know that this simulated path must come from the same Gaussian AR(1) process as

the original. This example illustrates a crucial point which is the linear ARMA(p,q) models with Gaussian noise are reversible and their forward and backward probability laws both are the same. Therefore, it is impossible to statistically identify between the causal and noncausal representations of such a Gaussian process. However, what is most interesting about this reversibility is how special it is. In fact, reversibility of the ARMA processes represents the exception rather than the rule.

To reinforce this connection, consider the following example. Given the mixed linear AR(1,1), that is $(1-\varphi_1 L^{-1})(1-\phi_1 L)x_t = \epsilon_t$, with non-Gaussian innovations ϵ_t, it is impossible to distinguish between this model and the strictly causal linear AR(2) model, $(1-\varphi_1 L)(1-\phi_1 L)x_t = \epsilon_t$, by appealing only to the 2nd order properties of the process (i.e. autocovariance function) since they imply the same spectral density. Denoting $h_{x,m}(\omega)$ as the unnormalized spectral density of the mixed process and $h_{x,c}(\omega)$ as that of the strictly causal process, we have

$$h_{x,m}(\omega) = |\varphi(\left[e^{-i\omega}\right]^{-1})|^{-2}|\phi(e^{-i\omega})|^{-2}\sigma_\epsilon^2/2\pi \tag{17a}$$

$$= \varphi(e^{+i\omega})^{-1}\varphi(e^{-i\omega})^{-1}\phi(e^{-i\omega})^{-1}\phi(e^{+i\omega})^{-1}\sigma_\epsilon^2/2\pi \tag{17b}$$

$$= |\varphi(e^{-i\omega})|^{-2}|\phi(e^{-i\omega})|^{-2}\sigma_\epsilon^2/2\pi = h_{x,c}(\omega) \tag{17c}$$

Of course, this also suggests we cannot identify between the linear mixed causal/noncausal models and the strictly causal or noncausal models with Gaussian innovations, since the Normal distribution is entirely characterized by its mean and variance and thus the linear (mixed or otherwise) Gaussian process has higher order polyspectra equal to zero everywhere (i.e. its bicovariance function $E[(x_t-\mu)(x_{t-s_1}-\mu)(x_{t-s_2}-\mu)] = R_{x,2}(s_1,s_2) = 0$ for all s_1, s_2).

It is also interesting to note that the mixed causal/noncausal autoregressive model from (9a) is effectively doubly filtered strong white noise, where the noise is first filtered in direct time and the resulting process is then filtered again in reverse time. As an example of how this fact can affect the the final process, x_t, let us consider the example of the mixed causality AR(1)

model with t-distributed innovations, $\epsilon_t \sim t(0, \sigma, \gamma)$. First, we have that from (16),

$$v_t = \phi(L)^{-1}\epsilon_t = \frac{\epsilon_t}{(1-\phi_1 L)} \tag{18}$$

and so the variance and spectral density of the direct time (i.e. backward) filtered noise, v_t, is given respectively as

$$\sigma_v^2 = \frac{\sigma_\epsilon^2}{1-\phi_1^2} = \frac{\sigma^2}{(1-\phi_1^2)}\frac{\gamma}{(\gamma-2)} \quad \text{for} \quad \gamma > 2 \quad \text{and} \tag{19a}$$

$$h_v(\omega) = \frac{\sigma_\epsilon^2}{2\pi}\frac{1}{1-2\phi_1\cos\omega+\phi_1^2}. \tag{19b}$$

Therefore, the lag polynomial $\phi(L)$ represents a low-pass filter, with the emphasis on passing along the low frequencies of ϵ_t increasing in ϕ_1.

Next, x_t represents the *forward* filtered output of $\varphi(L^{-1})^{-1}v_t$ and so we have that

$$x_t = \varphi(L^{-1})^{-1}v_t = \sum_{j=0}^{\infty}\varphi_1^j v_{t+j}. \tag{20}$$

Of course, the spectral density of x_t is simply that of the AR(2) model:

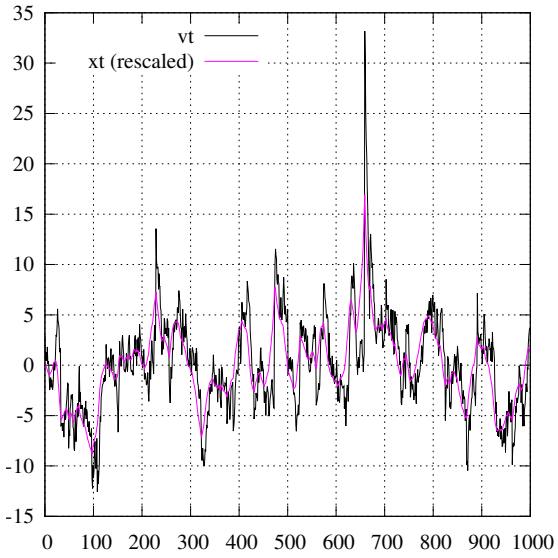
$$h_x(\omega) = \frac{\sigma_\epsilon^2}{2\pi}\frac{1}{|\phi(e^{-i\omega})|^2}\frac{1}{|\varphi(e^{+i\omega})|^2}, \tag{21}$$

and so it is unidentified up to 2nd order properties. Finally, the variance of x_t is given as

$$\sigma_x^2 = \frac{\sigma_v^2}{1-\varphi_1^2} = \frac{\sigma_\epsilon^2}{(1-\varphi_1^2)(1-\phi_1^2)} = \frac{\sigma^2}{(1-\varphi_1^2)(1-\phi_1^2)}\frac{\gamma}{(\gamma-2)} \tag{22}$$

Therefore, given the mixed AR(1) model, with $\varphi_1 \approx 1$ and $\phi_1 \approx 1$ the process x_t will represent a highly smoothed version of the original direct time process, v_t, scaled by $1/\sqrt{1-\varphi_1^2}$. See **figure 6** for a plot of the simulated processes, v_t and x_t, given $\varphi_1 = \phi_1 = 0.9$, $\gamma = 3$, and $\sigma = 1$.

Figure 6: Simulated processes v_t and x_t



Generally, ARMA processes driven by non-Gaussian innovations are not reversible (except for the unique case of symmetric or skew-symmetric coefficients in the Wold representation of the process – see Hallin et. al. (1998) for a discussion). Therefore, in order to model possible time asymmetries in commodity futures prices, we need to impose a non-Gaussian error term, ϵ_t, in (9a).

6.3.2 Example: Cauchy innovations

Let us now consider a simulated path from the AR(1) process $x_t = \rho x_{t-1} + \epsilon_t$ with Cauchy innovations, ϵ_t, where $\rho = 0.8$ as given in **Figure 7**,

Figure 7: The Cauchy AR(1) process

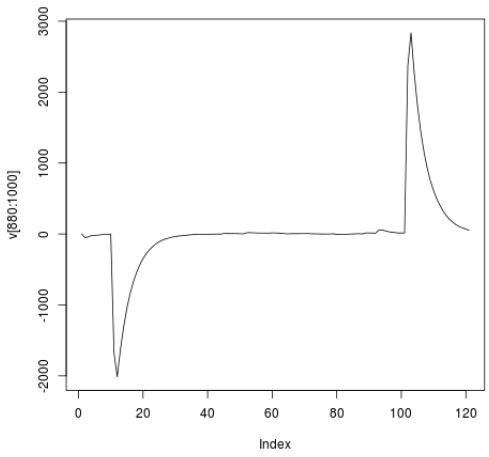
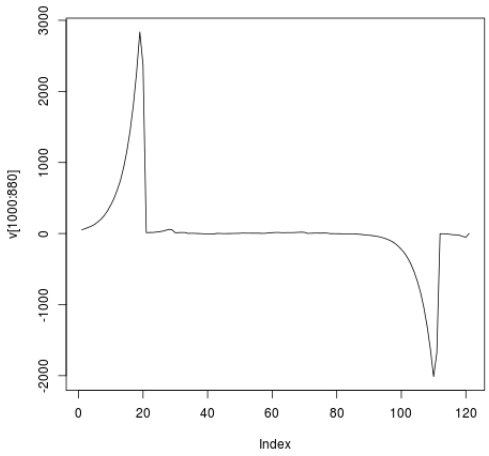


Figure 8: The Cauchy AR(1) process in reversed time



and again, the path in reversed time is given in **Figure 8**. It is now clear that the process is irreversible due to the time asymmetry of the process. **Figure 7** provides a typical path from the causal AR(1) process with Cauchy causal innovations; that is, it always includes some jumps followed by an exponential decrease/increase after the jumps. However, in reverse time the peaks follow the reverse pattern exhibiting exponential increases followed by a crash. In this case, the time irreversibility is easily detected by examining the opposite natures of these extreme behaviours.

Finally, consider the simulated mixed causal/noncausal process $(1-\rho L)(1-\rho L^{-1})x_t = \epsilon_t$ with Cauchy innovations as given in **Figure 9**.

Figure 9: The mixed Cauchy AR(1) process

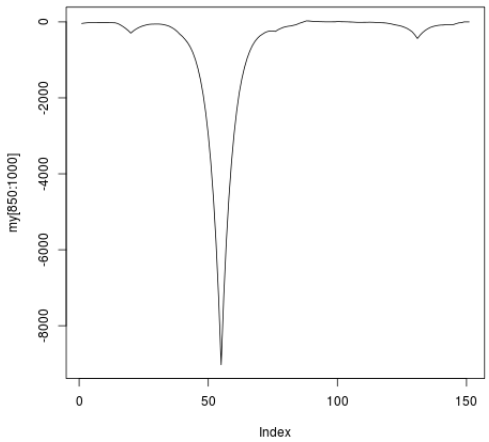
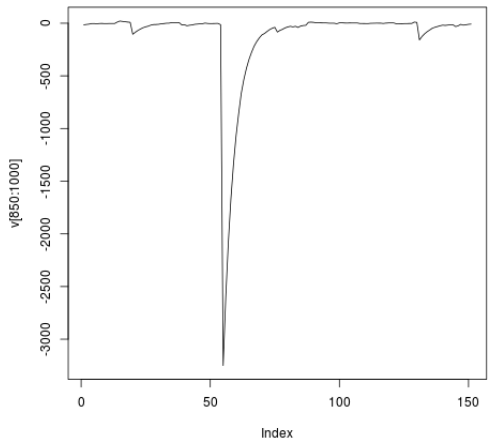


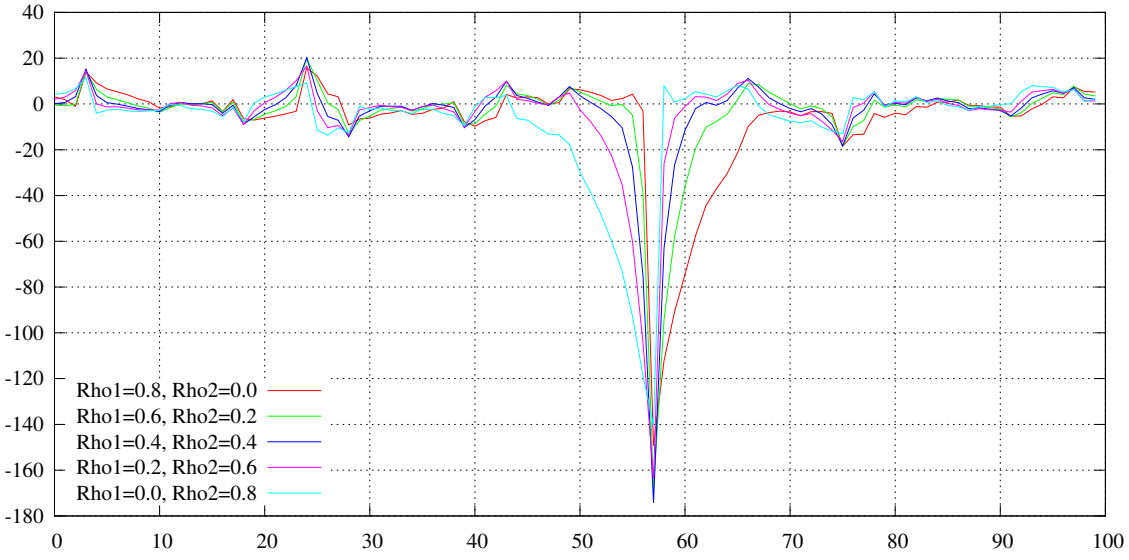
Figure 10: The causal Cauchy AR(1) process



And compare the process above to the strictly causal version of the same process, that is $(1 - \rho L)x_t = \epsilon_t$, as given in **Figure 10**.

Moreover, consider **Figure 11** which generates sample paths according to various values of ρ_1 and ρ_2 in $(1 - \rho_1 L)(1 - \rho_2 L^{-1})x_t = \epsilon_t$ given the same set of simulated shocks (ϵ_t). Notice the degree and time direction of the booms and crashes depends on the parameters ρ_1 and ρ_2.

Figure 11: The mixed causal/noncausal Cauchy model



Now we can see how adding a symmetric noncausal lag polynomial to the autoregression restores the time symmetry. Moreover, given these symmetric coefficients we chose, the model $x_t = (1 - \rho L)^{-1}(1 - \rho L^{-1})^{-1}\epsilon_t$ now satisfies reversibility despite the non-Gaussian nature of the innovations (since the coefficients of the mixed causal/noncausal Wold representation form

a symmetric set). However, for any other asymmetric choices of the mixed AR coefficients the process would be irreversible. Such processes are useful in that they can capture all sorts of time asymmetries we see in the data.

The linear forward looking autoregressive Cauchy process of order 1 arises then as special case of the mixed model in (9a). That is, we have $\phi_j = 0$ for all $j = 1, \ldots, r$, $s = 1$, and (ϵ_t) is an i.i.d. Cauchy sequence. Again, since it is known from proposition 6.3 that only the Gaussian processes possess both causal and noncausal strong form linear autoregressive representations, the causal form (i.e. backward looking or reversed process representation) of the AR(1) Cauchy process is necessarily nonlinear. Moreover, the AR(1) Cauchy process is also time irreversible. The theoretical properties of the AR(1) Cauchy process are reviewed next in section 6.4 (see also Gourieroux and Zakoian (2012)).

6.4 The Cauchy autoregressive process of order 1

6.4.1 The noncausal strong form

As previously discussed, a special case of the mixed process of (9a) is the strictly noncausal linear autoregressive process of order 1 with i.i.d. Cauchy innovations (recall the t-distribution with 1 degree of freedom is equivalent to the Cauchy distribution). While this represents an extreme case, it is a useful model to consider since it can serve as an example for understanding the nature of the one-step ahead predictive density based on past information. Of course, from Rosenblatt (2000) we know that this predictive density for a linear mixed (or strictly noncausal) process is necessarily nonlinear when the innovations are not Gaussian.

To begin, consider the linear strong form of the autoregressive Cauchy process of order 1, with backwards innovations (i.e. the forward looking process) defined as:

$$x_t = \rho x_{t+1} + \epsilon_t, \tag{23}$$

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

It can be shown that the unconditional distribution of the process in equation (23) is given as

$$f_t(x_t) = \frac{1-|\rho|}{\sigma_\epsilon \pi} \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + (1-|\rho|)^2 x_t^2}. \tag{24}$$

See Proposition 1 in Gourieroux and Zakoian (2012) for more details. Note that since, for $|\rho| < 1$, the process in equation (23) is strong form stationary, we can be assured that $f_t(x_\tau) = f_{t+h}(x_\tau)$ for all τ and h both integers.

Moreover, the Markov transition distribution (conditional density) of the forward-looking process is given as

$$f_{t|t+1}(x_t|x_{t+1}) = \frac{1}{\sigma_\epsilon \pi} \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + z_t^2}, \quad \text{where} \quad z_t = \frac{x_t - \rho x_{t+1}}{\sigma_\epsilon}, \tag{25}$$

which follows from the definition of the standard Cauchy distribution.

Therefore, from Bayes theorem along with equations (24) and (25), we have that

$$\begin{aligned} f_{t+1|t}(x_{t+1}|x_t) &= f_{t|t+1}(x_t|x_{t+1}) f_{t+1}(x_{t+1}) / f_t(x_t) && \text{(26a)} \\ &= \frac{1}{\sigma_\epsilon \pi} \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + z_t^2} \frac{\frac{1-|\rho|}{\sigma_\epsilon \pi} \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + (1-|\rho|)^2 x_{t+1}^2}}{\frac{1-|\rho|}{\sigma_\epsilon \pi} \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + (1-|\rho|)^2 x_t^2}} && \text{(26b)} \\ &= \frac{1}{\sigma_\epsilon \pi} \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + z_t^2} \frac{\sigma_\epsilon^2 + (1-|\rho|)^2 x_t^2}{\sigma_\epsilon^2 + (1-|\rho|)^2 x_{t+1}^2}, && \text{(26c)} \end{aligned}$$

which provides the causal transition density of the reversed process from equation (23) (i.e. the one-step ahead predictive density).

Note that since in general we have

$$f_{t|t+1}(x_t|x_{t+1}) = \frac{1}{\sigma_\epsilon \pi} \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + z_t^2} \neq \frac{1}{\sigma_\epsilon \pi} \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + z_t^2} \frac{\sigma_\epsilon^2 + (1-|\rho|)^2 x_{t+1}^2}{\sigma_\epsilon^2 + (1-|\rho|)^2 x_t^2} = f_{t+1|t}(x_t|x_{t+1}), \tag{27}$$

the process defined in equation (23) is irreversible. That is, the transition probabilities of the Markov process backward through time are not the same as those forward in time.

6.4.2 The causal weak form represention

Again, only the Gaussian autoregressive processes possess both causal and noncausal strong form linear autoregressive representations. However, the noncausal AR(1) Cauchy model admits a weak form linear representation given as

$$x_t = E_{t|t-1}[x_t|x_{t-1}] + \eta_t\sqrt{Var_{t|t-1}[x_t|x_{t-1}]}, \quad \text{where} \tag{28a}$$

$$E_{t|t-1}[x_t|x_{t-1}] = sign(\rho)x_{t-1} \quad \text{and} \tag{28b}$$

$$E_{t|t-1}[x_t^2|x_{t-1}] = \frac{1}{|\rho|}x_{t-1}^2 + \frac{\sigma_\epsilon^2}{|\rho|(1-|\rho|)}, \tag{28c}$$

which would otherwise be a weak double AR(1) process of the Ling & Li (2008) type, except for the fact that η_t is not i.i.d. We say the representation is a weak one since both 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.

Interestingly, the process remains stationarity despite the presence of a unit root; this unit root is expected since the unconditional moments of x_t do not exist. However, as discussed in Gourieroux and Zakoian (2012), typical linear unit root tests will fail in detecting the explosive bubbles of the noncausal AR(1) Cauchy process, even if the martingale property is satisfied.

6.4.3 The causal strong form represention

In order to represent the causal, reversed process, from equation (23) 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 (2003), section 2.1).

Intuitively, a nonlinear error term, (γ_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 γ_t, say $x_t = G(x_{t-1}, \gamma_t)$, and x_t and γ_t satisfy a continuous one-to-one relationship given any x_{t-1}.

Moreover, since the conditional cumulative distribution function (CCDF) of $x_t|x_{t-1}$ is

strictly monotone increasing and continuous, it has an inverse so we can write

$$x_t = G(x_{t-1}, \gamma_t) \tag{29a}$$

$$= F^{-1}(\varphi(\gamma_t)|x_{t-1}) \tag{29b}$$

$$\Leftrightarrow \gamma_t = \varphi^{-1}(F(x_t|x_{t-1})), \quad \text{where} \quad \gamma_t \sim IIN(0,1) \tag{29c}$$

and $F(\cdot|x_{t-1})$ is the CCDF of x_t while $\varphi(\cdot)$ is the CDF of the standard Normal distribution. Note that the choice of the Normal distribution as our strong form error term is purely conventional; however, it has the advantage of being able to derive directly the Volterra representation of the nonlinear time series, which defines x_t as the limit of polynomials in the current and lagged values of Gaussian white noise (see Gourieroux and Jasiak (2003), section 2.2, Property 5 and also section 8).

Finally, it should be noted that the AR(1) Cauchy model is easily introduced in a mixed causal/noncausal framework unlike previous bubble modeling strategies described in the next section.

6.5 Other bubble like processes

As described in Gourieroux and Zakoian (2011), 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 mixed causal/noncausal model as the Cauchy AR(1) model. Interestingly, these previous bubble processes also represent processes nonlinear in mean (see section 8), but which still maintain the martingale property. For example, the bubble process of Blanchard and Watson (1982) given as

$$x_{t+1} = \begin{cases} \frac{1}{\pi}x_t + \epsilon_{t+1} & \text{with probability} \quad \pi \\ \epsilon_{t+1} & \text{with probability} \quad (1-\pi) \end{cases} \tag{30a}$$

represents a martingale process but is piecewise nonlinear in that given the latent state, the parameter on the autoregression switches between zero and $1/\pi$. Note that (ϵ_t) is a nonlinear

Gaussian error term.

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

$$x_{t+1} = \begin{cases} \left(\delta + \frac{1}{\pi}(1+r)\theta_{t+1}\left(x_t - \frac{\delta}{(1+r)}\right)\right)u_{t+1} & \text{if} \quad x_t > \alpha \\ (1+r)x_t u_{t+1} & \text{if} \quad x_t \leq \alpha. \end{cases} \tag{31a}$$

Notice that in this case the state is not latent, but observable in x_t. In this way, the process is similar to the self-exciting threshold autoregression of Tong and Lim (1980).

Ultimately, these two bubble processes 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 expectations hypothesis and the Evans (1991) process is given as an example of a stationary bubble process that defies 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.

6.6 Testing for reversibility in the data

There is clearly an intimate connection between the autoregressive mixed causal/noncausal model from (9) and the class of time irreversible processes. It stands to reason that if a process is time irreversible then it may be better modeled by a mixed causal/noncausal model than one which is strictly causal.

In Ramsey and Rothman (1996) the authors suggest time irreversibility may be useful in modeling the asymmetric nature of business cycles. They suggest a test for the time irreversibility of a process, based on the following theorem and the notion of a symmetric bicovariance function $E[x_t x_{t-s} x_{t-k}]$,

Theorem 6.4. *Reversibility of a time-series*

Let (x_t) be a stationary time series with mean zero and assume the multivariate characteristic generating functions of $\{x_t, x_{t-k}\}$ and $\{x_{t-k}, x_t\}$ can be expanded as a convergent series in the moments and cross moments of the respective joint probability distributions; that is, it is assumed that the joint probability distributions are uniquely characterized by their respective sequence of moments and cross moments. Then (x_t) is time reversible only if:

$$E[x_t^i x_{t-k}^j] = E[x_t^j x_{t-k}^i] \quad \textit{for all} \quad j, k, i \in \mathbf{N}$$

where the expectation is taken with respect to each respective joint distribution.

That is, (x_t) is time reversible only if the "higher order" autocovariance functions of $\{x_t, x_{t-k}\}$ and $\{x_{t-k}, x_t\}$ are the same. Therefore, just as it is impossible to identify the mixed model from the strictly causal or noncausal models based on 2nd order properties of the process, it is also impossible to determine whether the process is reversible or not, since we clearly have that if $i = j = 1$ the condition in theorem 6.4 holds trivially for real processes. Moreover, since the linear processes with Gaussian innovations have their "higher order" covariances equal to zero everywhere (again since the Normal distribution is characterized by only its mean and variance) it is clear that they are automatically time reversible and so again the theorem above holds trivially. Of course, if the process (x_t) is nonlinear with Gaussian innovations it is quite possible that x_t may then be time irreversible since the process x_t itself is generally not Gaussian (Ramsey and Rothman (1996) provide the example of the Bilinear process $x_t = \alpha_1 x_{t-1} + \beta_1 \epsilon_{t-1} + \gamma_1 x_{t-1}\epsilon_{t-1} + \epsilon_t$ where $\epsilon_t \sim N(0,1)$ as an example). Ramsey and Rothman refer to the latter type of time irreversibility as "Type 1" irreversibility and the former type (where the model is linear, with non-Gaussian innovations) as "Type 2" irreversibility.

6.6.1 The TR test applied to the data

From Brillinger and Rosenblatt (1967, p.210) we know that a stationary time reversible process must admit the same ith cumulant κ_i both backwards and forwards–that is, we must have $C(t_1, t_2, \ldots, t_{i-1}) = C(-t_1, -t_2, \ldots, -t_{i-1})$ for all i [8]. Of course, this implies that the imaginary part of the bispectrum must be zero, a fact exploited by Hinich and Rothman (1998) in

[8] see equation (4).

developing a frequency domain statistical test of time reversibility.

Ramsey and Rothman (1996) also define a statistical time-domain test, called the "TR test," based on the difference between the 3rd order cumulants of the direct time and reversed stationary processes,

$$\gamma_{2,1}(k) = E[(x_t - \mu_X)^2(x_{t-k} - \mu_X)] - E[(x_t - \mu_X)(x_{t-k} - \mu_X)^2] \tag{32a}$$

$$= C(0,k) - C(k,k) \tag{32b}$$

$$= C(0,k) - C(0,-k) \tag{32c}$$

$$= C(-k,-k) - C(k,k). \tag{32d}$$

Of course, this is equivalent to comparing the values of the bicovariances along the "main-diagonal" or where $t_1 = t_2 = k$. Therefore, in light of the definition of reversibility above, under the null hypothesis we have that $\gamma_{2,1}(k) = 0$ for all $k > 0$.

The population characteristic in expression (32a) can be estimated by sample moment counterpart. Moreover, under some mixing conditions (see Ramsey and Rothman (1996)) the statistic is distributed asymptotically Normal. When (x_t) is a stationary sequence of i.i.d. variables then Theorem 2 from Ramsey and Rothman (1996) provides an expression for the exact finite sample variance of this statistic under the null of reversibility. However, when (x_t) is not i.i.d. we must simulate the statistic's variance under the null. Recall from above there can exist two types of irreversible processes: type 1 processes where the model is nonlinear but with shocks that may or may not be Gaussian, and type 2 where the model is linear but with non-Gaussian shocks. Therefore there exist two cases under the null of reversibility: the model is linear with Gaussian shocks or the model is nonlinear but reversible.

In order to simulate the finite sample variance of the TR statistic under the null Ramsey and Rothman (1996) suggest we estimate the parameters of the best fitting (according to some information criterion) linear causal ARMA model with Gaussian innovations. We then simulate sample paths from this estimated process a number of times, each time computing the TR statistic. After a sufficient number of iterations, the variance of these statistics is estimated by sample moment. Of course, if the model is truly linear with Gaussian innovations then this is

an exact simulation of the null; however, if the model is nonlinear but reversible then this is only an approximation.

Monte carlo simulations are performed for each estimated linear causal ARMA model with Gaussian innovations from section 5. $N = 1500$ individual sample paths of the same length as the respective data series are generated from these models and the TR test statistic $\gamma_{2,1}(k)$ is then computed for the more general case where $\gamma_{2,1}(k,s) = C(-k,-s) - C(k,s)$. In this way we generate the finite sample distributions for each coordinate (k,s) of the TR statistic, up to N observations. Inspection of these finite sample distributions suggest they differ substantially from Normality in that they are both leptokurtic and skewed and exhibit a "convex" shape similar in appearance to a α-stable distribution with $\alpha = 1/2$. Therefore, rather than assume standard Normal statistical significance cutoff values, sample quantiles are estimated accordingly for a two-sided test given 20%, 10%, 5%, and 1% significance levels of commiting a type 1 error.

Table 7 presents the results of the estimated TR test statistics for the various commodity futures price level series. The various percentage values indicate the significance level at which we are able to reject the null hypothesis of a two-sided test – the absence of a value suggests we are not able to reject the null hypothesis of reversibility for at least the 20% significance level. Interestingly, we are unable to reject the null for most of the metals and energy commodity futures, as well as for orange juice and sugar,[9] and the strongest evidence for irreversibility seems present in the soft goods and livestock, specifically the cocoa, coffee, and cotton futures series.

Both Hinich and Rothman (1998) and Ramsey and Rothman (1996) apply their respective statistical tests to a number of economic time series (for example the Nelson and Plosser (1982) dataset employed in their study of stochastic trends) and find evidence of irreversibility. Ramsey and Rothman (1996) also suggests that the "sign" of $\gamma_{2,1}(k)$ across k tells us something about the nature of the longitudinal asymmetry (for example, considering whether a business cycle series rises slowly and falls quickly or vice versa) and so it isn't a stretch to imagine that the "shape" of $\gamma_{2,1}(k)$ (e.g. its rate of change or prominent features) might also

[9] At least at the 10% or higher significance level.

Table 7: TR test results

Series	**Lag** (k, s)									
	(1,1)	(1,2)	(1,3)	(1,4)	(2,2)	(2,3)	(2,4)	(3,3)	(3,4)	(4,4)
Soybean meal	5%		5%	5%	5%	5%	10%	5%	5%	5%
Soybean oil		20%								
Soybeans	20%		20%	20%	20%	20%		20%	20%	20%
Orange juice										
Sugar		20%					20%			
Wheat	20%		20%	20%	20%	20%		20%	20%	20%
Cocoa	1%	5%	1%	1%	1%	1%	10%	1%	1%	1%
Coffee	1%	10%	5%	5%	5%	5%	10%	5%	5%	5%
Corn	20%	10%	20%	20%	20%	20%	10%	20%	20%	20%
Cotton	5%	5%	5%	5%	5%	5%	1%	5%	5%	5%
Rice		10%					20%			
Lumber	10%	20%	20%	20%	20%	20%		20%	20%	20%
Gold										
Silver	10%	5%	5%	5%	5%	5%	1%	5%	5%	5%
Platinum										
Palladium		20%					5%			
Copper							20%			
Light crude oil		20%					20%			
Heating oil										
Brent crude oil										
Gas oil										
Naturas gas		1%								
Gasoline RBOB										
Live cattle	20%		10%	10%	10%	10%		10%	10%	10%
Lean hogs	5%	10%	5%	5%	5%	5%	20%	5%	5%	5%

tell us something.

Note that just because a series doesn't suggest irreversibility through its third cumulant does not mean it may not be irreversible through higher order cumulants. Considering the strong nature of the constraint imposed on the probability structure, reversibility is probably the exception rather than the rule for most time series we encounter in economics.

7 Estimating the mixed causal/noncausal model

Estimation of the model parameters in (9) can be accomplished by maximizing the approximate likelihood function with respect to the parameters (MLE). Breidt et. al. (1991) studies the properties of the maximum likelihood estimator for model (11). Moreover, Lanne et. al. (2008,2011) extend these results to the model in (9), where the distribution of the error term is allowed to depend on a degrees of freedom parameter.

As in Lanne et. al. we assume the regularity conditions of Andrews et. al. (2006) which require the likelihood to be twice differentiable with respect to (y, λ) (where y is the complete data vector and λ is a vector of parameters).

The rationale for approximating the likelihood centers on the fact that because of the (r, s) lags in the mixed model, we effectively have no way of describing the density of the "tail ends" of the data, both at the beginning and end. Moreover, we need to use a change of variables to go between the density of ϵ_t and x_t since the Jacobian determinant is not unitary.

The change of variables is as follows. Consider the time series x_t for $t = 1, \ldots, T$. Using the expressions from (15) and (16) we have $u_t = \phi(L)x_t$ and $v_t = \varphi(L^{-1})x_t$ and so we can we write the matrix relation:

$$\begin{bmatrix} v_1 \\ \vdots \\ v_{T-s} \\ u_{T-s+1} \\ \vdots \\ u_T \end{bmatrix} = \begin{bmatrix} x_1 - \varphi_1 x_2 - \cdots - \varphi_s x_{1+s} \\ \vdots \\ x_{T-s} - \varphi_1 x_{T-s+1} - \cdots - \varphi_s x_T \\ x_{T-s+1} - \phi_1 x_{T-s} - \cdots - \phi_r x_{T-s+1-r} \\ \vdots \\ x_T - \phi_1 x_{T-1} - \cdots - \phi_r x_{T-r} \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} x_1 \\ \vdots \\ x_{T-s} \\ x_{T-s+1} \\ \vdots \\ x_T \end{bmatrix} \tag{33}$$

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

$$\begin{bmatrix} v_1 \\ \vdots \\ v_r \\ \epsilon_{r+1} \\ \vdots \\ \epsilon_{T-s} \\ u_{T-s+1} \\ \vdots \\ u_T \end{bmatrix} = \begin{bmatrix} v_1 \\ \vdots \\ v_r \\ v_{r+1} - \phi_1 v_r - \cdots - \phi_r v_1 \\ \vdots \\ v_{T-s} - \phi_1 v_{T-s-1} - \cdots - \phi_r v_{T-s-r} \\ u_{T-s+1} \\ \vdots \\ u_T \end{bmatrix} = \boldsymbol{B} \begin{bmatrix} v_1 \\ \vdots \\ v_{T-s} \\ u_{T-s+1} \\ \vdots \\ u_T \end{bmatrix} \tag{34}$$

And so we have the transformation $\boldsymbol{\epsilon} = \boldsymbol{BAx}$. Note that by symmetry we could have started the other way around, deriving a relationship between $\boldsymbol{x}$ and $\boldsymbol{Ax}$ by first taking $u_t = \phi(L)x_t$ and then secondly generating $\boldsymbol{BAx}$ by $\epsilon_t = \varphi(L^{-1})u_t$.

Either way, we have that the elements of $\boldsymbol{\epsilon}$ are independent of each other and the joint density of $\boldsymbol{\epsilon}$ under the true parameter vector is given as:

$$\mathcal{L}_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}|\Omega) = f_v(v_1, \ldots, v_r) \left(\prod_{t=r+1}^{T-s} f_\epsilon(\epsilon_t; \lambda) \right) f_u(u_{T-s+1}, \ldots, u_T) \tag{35}$$

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

It can also be shown that the determinant of the $\boldsymbol{B}$ matrix is unity so that we can express the joint density in terms of $\boldsymbol{x}$ as:

$$\mathcal{L}_{\boldsymbol{x}}(\boldsymbol{x}|\Omega) = f_v(\varphi(L^{-1})x_1, \ldots, \varphi(L^{-1})x_r) \left(\prod_{t=r+1}^{T-s} f_x(\varphi(L^{-1})\phi(L)x_t; \lambda) \right) f_u(\phi(L)x_{T-s+1}, \ldots, \phi(L)x_T)|det(\boldsymbol{A})| \tag{36}$$

Since the determinant of $\boldsymbol{A}$ is independent of sample size,[10] we can approximate the likelihood

[10]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.

by using the second factor in the above expression. Bear in mind that for large samples, T will dwarf $r + s$ and so the approximation will be a good one.

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

7.1 Estimation results

In this section I will now compare estimation results from the mixed autoregressive model of order (r, s) in (9a) as applied to the 25 futures price level data series. Estimation of the model parameters numerically optimizes the likelihood function discussed in the last section.

As in section 5 where the linear causal ARMA model was shown to inadequately capture the features of the price level data, I will again employ the AIC criterion, along with Ljung-Box statistics testing the hypothesis that both the innovations and their squares exhibit no linear autocorrelation. Similarly, I will consider the best fitting linear causal ARMA model from section 5 as a benchmark model, where the innovations will be either t-distributed or skewed-t, depending on which version of the model did better in that section. Please see section 5 for details of the methodology employed again here.

Table 8 presents the results of numerical maximum likelihood estimation. The mixed AR model orders, (r, s), were selected via normalized AIC (the AIC divided by the number of elements in the log-likelihood summation) amongst a set of of 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 from section 5, while the second and third rows represent the mixed AR model with t-distributed and skew t-distributed innovations respectively. The lags column represents the number of lags (the natural log of the sample size $\ln(T)$) included in both the Ljung-Box and McLeod-Li test statistics, 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 AR(r, s) model improves model fit over the baseline causal ARMA model with t-distributed innovations selected from section 5. In the cases where the models are nested, I again employ LR tests. In most cases

the mixed causal model improves model fit significantly at the 1% significance level. However, both cotton and natural gas seem to suggest that a strictly causal model may be sufficient. While LR tests aren't useful here as the relevant models are non-nested, a comparison of the log-likelihoods between the causal and mixed model options suggests little difference.

Again in comparing the skewed t-distributed model to the standard t-distribution, 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 parameters from the mixed AR model innovations are quite close to those of the benchmark causal ARMA model and range from near 1 (i.e. Cauchy distributed) to around 3 or so in most cases, which suggests bubble like behaviour as defined by Gourieroux and Zakoian (2011).

Despite the improved model fit, however, it is clear that the mixed AR model is still unable to adequately capture the dynamics of the series. McLeod-Li tests of the null that the autocovariance of the squared residuals is zero is rejected in all series, suggesting they are not i.i.d. Of course, this is evidence that there still exists uncaptured nonlinear structure in the residuals. Interestingly, this is despite the fact that the mixed causal/noncausal AR(r, s) model can be shown to have a nonlinear strictly causal equivalent representation, which suggests the causal representation implied by the mixed AR(r, s) model must be misspecified.

Interestingly, an examination of the roots of the lag polynomials implied by the estimated parameters also confirms the mixed causality nature of the series, at least as it applies to the strong form linear representation. If we accept only the statistically significant estimated parameters [11] and solve for the roots of the implied causal and noncausal lag polynomials, $\phi(L)$ and $\varphi(L^{-1})$ (from (9a)), we find that that the roots of both appropriately lie outside the unit circle (which implies that from (11a), $\alpha(L) = \alpha_1(L)\alpha_2(L)$, is such that the roots of $\alpha_1(L)$ lie

[11] 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.

strictly outside the complex unit circle while those of $\alpha_2(L)$ lie strictly inside). Of course, if a purely causal model fit the data better, none of the lags of the noncausal polynomial, $\varphi(L^{-1})$, should be statistically significant yet they are.

For reference I have constructed tables with all of the roots of the lag polynomials of both the benchmark causal ARMA models of order (p, q) and the mixed causal/noncausal AR models of order (r, s); see **Tables 22 to 24** within **Appendix A**. For the case of ARMA models with $q \neq 0$, the roots reported are those of $\frac{\alpha(L)}{\beta(L)} = \delta(L)$ in $\delta(L)x_t = \epsilon_t$, where $\alpha(L)$ and $\beta(L)$ are the AR and MA lag polynomials respectively.

Notice first that in some mixed model cases that one of the roots of the noncausal polynomial $\varphi(L^{-1})$ lies *inside* the unit circle (this applies to all the metals and gas oil). This poses no problem of course to the validity of the mixed model representation, since we can always factor out that particular root and absorb it into the causal polynomial $\phi(L)$. That is, consider the estimated model from (9a) where we have factorized both the causal and noncausal lag polynomials as,

$$(1 - \lambda_{1,c}L)\dots(1 - \lambda_{r,c}L)(1 - \lambda_{1,nc}L^{-1})\dots(1 - \lambda_{s,nc}L^{-1})x_t = \epsilon_t \quad (37a)$$

$$\Leftrightarrow (1 - \lambda_{1,c}L)\dots(1 - \lambda_{r,c}L)(1 - \lambda_{r+1,c}L)(1 - \lambda_{2,nc}L^{-1})\dots(1 - \lambda_{s,nc}L^{-1})x_t = u_t \quad (37b)$$

where $\lambda_{i,c}$ represents the inverse of the ith causal root (nc for noncausal) and without loss of generality we assume that $|1/\lambda_{1,nc}| < 1$. Therefore, $u_t = -\frac{1}{\lambda_{1,nc}}\epsilon_{t-1}$ and $\lambda_{1,nc} = 1/\lambda_{r+1,c}$.

Perhaps most interesting is the fact that for many of the causal ARMA models, the roots actually imply a mixed representation! For example the ARMA representations of 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 $\delta(L)$ polynomial, suggesting that this polynomial could be factored and then estimated as a mixed causal/noncausal model (instead of the traditional differencing technique employed). In fact, for light crude oil, gas oil, and heating oil, the very large value (or lack of) of the other roots suggests that these series may be better represented as nearly purely noncausal (this result is confirmed by the mixed roots of light crude oil, but not for gas or heating oil which have causal roots relatively close to 1). Finally,

note that the mixed representation for soybeans suggests 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, the results so far suggest that most of the futures price levels series are better fitted by mixed causal/noncausal AR(r, s) models that take into account their time irreversible nature. Exceptions to this are found in cotton and natural gas, which fail likelihood ratio tests of model fit improvement over the causal ARMA benchmark. However, despite the mixed AR model's improved model fit, there still exists uncaptured nonlinearity in the model residuals (with perhaps the exception of the lean hogs series). In the next section I will discuss possible nonlinear models to account for this.

Table 8: Estimation results of mixed causal/noncausal AR(r, s) models - 1 of 2

		p/r	q/s	Log-likelihood	AICn	Lag	Ljung-Box	McLeod-Li
Soybean meal		10	0	-24171.704	5.21806	9	0.000	0.000
	x	10	10	-24081.130	5.20608	9	0.007	0.000
		10	10	-24081.059	5.20628	9	0.007	0.000
Soybean oil		10	0	-4639.175	1.00392	9	0.007	0.000
	x	10	10	-4582.938	0.99480	9	0.135	0.000
		10	10	-4582.844	0.99500	9	0.138	0.000
Soybeans		10	0	-34779.631	7.50672	9	0.000	0.000
		10	10	-34699.422	7.49944	9	0.000	0.000
	x	10	10	-34695.177	7.49874	9	0.000	0.000
Orange juice		10	0	-19382.514	4.18479	9	0.070	0.000
		10	10	-19320.480	4.17786	9	0.378	0.000
	x	10	10	-19317.960	4.17753	9	0.389	0.000
Sugar		1	2	-794.876	0.17284	9	0.000	0.000
	x	2	2	-767.750	0.16704	9	0.000	0.000
		2	2	-767.645	0.16724	9	0.000	0.000
Wheat		5	0	-30983.923	6.68311	9	0.000	0.000
		5	5	-30935.424	6.67711	9	0.000	0.000
	x	5	5	-30926.145	6.67533	9	0.000	0.000
Cocoa		10	0	-45847.856	9.89468	9	0.000	0.000
		2	1	-45896.441	9.89597	9	0.000	0.000
	x	10	10	-45769.055	9.89051	9	0.003	0.000
Coffee		10	0	-21901.578	4.72806	9	0.000	0.000
		10	10	-21842.943	4.72267	9	0.014	0.000
	x	10	10	-21841.150	4.72249	9	0.012	0.000
Corn		2	0	-26847.118	5.78856	9	0.419	0.000
	x	2	3	-26815.913	5.78413	9	0.776	0.000
		2	3	-26815.622	5.78428	9	0.783	0.000
Cotton		10	0	-13494.465	2.91445	9	0.000	0.000
	x	1	3	-13495.916	2.91137	9	0.000	0.000
		1	3	-13495.906	2.91158	9	0.000	0.000
Rice		2	2	3559.414	-1.12618	8	0.000	0.000
	x	1	3	3593.842	-1.13778	8	0.013	0.000
		1	3	3594.173	-1.13756	8	0.013	0.000
Lumber		1	1	-21531.166	6.14996	8	0.384	0.000
		10	10	-21446.974	6.14745	8	0.562	0.000
	x	10	10	-21444.622	6.14706	8	0.546	0.000
Gold		3	0	-28530.443	6.15208	9	0.000	0.000
	x	10	10	-28435.869	6.14662	9	0.000	0.000
		10	10	-28435.811	6.14683	9	0.000	0.000
Silver		10	0	3479.433	-0.74767	9	0.000	0.000
		10	10	3549.149	-0.76159	9	0.000	0.000
	x	10	10	3552.141	-0.76202	9	0.000	0.000

Table 9: Estimation results of mixed causal/noncausal AR(r, s) models - 2 of 2

		p/r	q/s	Log-likelihood	AICn	Lag	Ljung-Box	McLeod-Li
Platinum		10	0	-25897.058	7.40422	8	0.000	0.000
		10	10	-25810.911	7.39273	8	0.000	0.000
	x	10	10	-25798.400	7.38944	8	0.000	0.000
Palladium		5	0	-21367.921	6.10420	8	0.000	0.000
	x	8	8	-21265.772	6.08745	8	0.000	0.000
		8	8	-21265.746	6.08773	8	0.000	0.000
Copper		10	0	-15311.933	4.86615	8	0.000	0.000
		10	10	-15243.741	4.85506	8	0.000	0.000
	x	10	10	-15240.889	4.85447	8	0.000	0.000
Light crude oil		2	0	-8663.663	2.22530	8	0.001	0.000
		1	3	-8641.851	2.22079	8	0.015	0.000
	x	1	3	-8639.603	2.22047	8	0.014	0.000
Heating oil		2	0	-15489.176	4.46389	8	0.064	0.000
	x	10	10	-15381.000	4.44945	8	0.042	0.000
		8	8	-15400.897	4.45176	8	0.000	0.000
Brent crude oil		2	2	-7617.552	2.37340	8	0.001	0.000
		10	10	-7517.822	2.35393	8	0.000	0.000
	x	10	10	-7512.764	2.35267	8	0.000	0.000
Gas oil		1	0	-20640.309	6.70411	8	0.532	0.000
		10	10	-20535.023	6.69642	8	0.259	0.000
	x	10	10	-20532.228	6.69584	8	0.261	0.000
Natural gas		1	2	3892.707	-1.30361	8	0.014	0.000
	x	1	1	3891.158	-1.30364	8	0.017	0.000
		1	1	3891.727	-1.30350	8	0.018	0.000
Gasoline RBOB		3	0	-5760.988	6.01772	7	0.070	0.000
		2	1	-5761.929	6.01766	7	0.050	0.000
	x	2	1	-5756.133	6.01266	7	0.056	0.000
Live cattle		10	0	-10210.809	2.20600	9	0.750	0.001
		10	10	-10190.943	2.20604	9	0.915	0.000
	x	8	8	-10193.265	2.20494	9	0.873	0.000
Lean hogs		5	0	-9462.261	2.70453	8	0.676	0.994
		0	2	-9459.574	2.70146	8	0.572	0.993
	x	0	2	-9455.188	2.70050	8	0.570	0.993

* Lags refers to the number of lags incorporated into both the Ljung-Box and McLeod-Li test statistics. This value is $\ln(T)$. Moreover, the values listed of the aforementioned statistics are the p-values. The AICn value is the AIC normalized by the number of elements in the log-likelihood summation.

8 Nonlinear models

The results from the last section suggest that most of the futures price levels series are better fitted by mixed causal/noncausal AR(r, s) models that take into account their time irreversible nature. However, despite the mixed AR model's improved model fit, there still exists uncaptured nonlinearity in the model residuals (with perhaps the exception of the lean hogs series). In this section I will try to establish more specifically the functional form of the nonlinear structure in both the price levels and returns series.

8.1 The general form

Consider the most general form that a discrete time model can take

$$\epsilon_t = f(\dots, X_{t-2}, X_{t-1}, X_t, X_{t+1}, X_{t+2}, \dots), \quad \text{where} \quad \epsilon_t \sim i.i.d. \tag{38}$$

Therefore this model takes some input, a stochastic process, (X_t), and reduces it to a mean zero strong white noise process, ϵ_t, where $f(\cdot)$ is some as of yet unspecified function. Of course, if $f(\cdot)$ is a purely linear function and we care only about the second order properties of ϵ_t, then it is sufficient that ϵ_t is only a weak white noise (and if ϵ_t is Gaussian then this implies it is also i.i.d.). If this is not the case then clearly the function $f(\cdot)$ is misspecified since there still remains time dependent structure in ϵ_t left unaccounted for.

If $f(\cdot)$ is an invertible function then we can write the equivalent representation in terms of the strong white noise as

$$X_t = h(\dots, \epsilon_{t-2}, \epsilon_{t-1}, \epsilon_t, \epsilon_{t+1}, \epsilon_{t+2}, \dots), \quad \text{where} \quad \epsilon_t \sim i.i.d., \tag{39}$$

and $h(\cdot) = f^{-1}(\cdot)$. Of course, (39) represents the most general mixed causal/noncausal model and if we impose the restriction that $h(\cdot)$ is linear then the mixed causality AR(r, s) model is a special case.

Now, suppose that (39) represents a purely causal process (X_t) and that $h(\cdot)$ is sufficiently well behaved as to permit a Taylor series expansion about the point $\mathbf{0} = \{0, 0, 0, \dots\}$. That is

we have,

$$X_t = \mu + \sum_{u=0}^{\infty} g_u \epsilon_{t-u} + \sum_{u=0}^{\infty}\sum_{v=0}^{\infty} g_{uv}\epsilon_{t-u}\epsilon_{t-v} + \sum_{u=0}^{\infty}\sum_{v=0}^{\infty}\sum_{w=0}^{\infty} g_{uvw}\epsilon_{t-u}\epsilon_{t-v}\epsilon_{t-w} + \dots \tag{40a}$$

$$\text{where} \quad \mu = h(\mathbf{0}), \quad g_u = \frac{\partial h(\mathbf{0})}{\partial \epsilon_{t-u}}, \quad g_{uv} = \frac{\partial^2 h(\mathbf{0})}{\partial \epsilon_{t-u}\partial \epsilon_{t-v}}, \quad g_{uvw} = \frac{\partial^3 h(\mathbf{0})}{\partial \epsilon_{t-u}\partial \epsilon_{t-v}\partial \epsilon_{t-w}}, \quad \text{etc.} \tag{40b}$$

The expansion above in (40a) is well known as the discrete time Volterra series. In its current form it would prove very difficult to practically implment this form as a statistical model given the large number of possible parameter configurations.

8.2 The causal truncated model

Now suppose that instead of an expansion at $\mathbf{0} = \{0, 0, 0, \dots\}$ we expand (39) at the point $\{0, \epsilon_{t-1}, \epsilon_{t-2}, \dots\}$ and with respect to ϵ_t only so we have,

$$X_t = h(0, \epsilon_{t-1}, \epsilon_{t-2}, \dots) + \epsilon_t h_1(0, \epsilon_{t-1}, \epsilon_{t-2}, \dots) + \epsilon_t^2 h_{11}(0, \epsilon_{t-1}, \epsilon_{t-2}, \dots)/2 + \dots \tag{41}$$

In nearly every nonlinear specification employed in economics, we drop all the terms higher than the first order from (41) and approximate the true (purely causal) nonlinear model in (39) by

$$X_t = m(\epsilon_{t-1}, \epsilon_{t-2}, \dots) + s(\epsilon_{t-1}, \epsilon_{t-2}, \dots)\epsilon_t \tag{42}$$

where $m(\epsilon_{t-1}, \epsilon_{t-2}, \dots) \equiv h(0, \epsilon_{t-1}, \epsilon_{t-2}, \dots)$ is the conditional mean equation and $s(\epsilon_{t-1}, \epsilon_{t-2}, \dots)$ $\equiv$ $h_1(0, \epsilon_{t-1}, \epsilon_{t-2}, \dots)$ is the conditional scale equation, and since $E_{t-1}[(X_t - m(\cdot))^2] = Var_{t-1}(X_t)$ we have that $s(\epsilon_{t-1}, \epsilon_{t-2}, \dots)^2$ is the conditional variance equation (where (ϵ_t) is i.i.d. with $E[\epsilon_t^2] = 1$). Note that by dropping the higher order terms from (41) we link the higher order conditional moments of the process to the second conditional moment since we have that for $k \geq 2$, $E_{t-1}[(X_t - m(\cdot))^k] = s(\cdot)^k E[\epsilon_t]^k$.

All of the nonlinear models we are familar with fall into either of two categories: those that are nonlinear in mean and those that are nonlinear in variance. For example, the TAR (and STAR models) of Tong and Lim (1980) , the Bilinear model of Granger and Andersen (1978), and the nonlinear AR and MA models $X_t = \epsilon_t + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{i=1}^{p}\sum_{j=1}^{q} \alpha_{i,j} X_{t-i}X_{t-j} + \dots$

and $X_t = \epsilon_t + \sum_{i=1}^{p} \beta_i \epsilon_{t-i} + \sum_{i=1}^{p} \sum_{j=1}^{q} \beta_{i,j} \epsilon_{t-i} \epsilon_{t-j} + \dots$ respectively, are all examples of models that are nonlinear in mean. As an example of a model that is nonlinear in variance one need not look further than the celebrated ARCH model of Engle (1982). Therefore, the distinguishing feature of each type of model is whether $m(\cdot)$ or $s(\cdot)\epsilon_t$ are linear or nonlinear functions of the current and past innovations. Notice that even if $s(\cdot)$ is a linear function of past innovations, this still implies that $s(\cdot)\epsilon_t$ is a nonlinear function of the current and past innovations.

8.3 The Hsieh test for causal nonlinear functional form

Hsieh (1989), section 5, proposes a useful statistical test that attempts to distinguish between the two types of functional nonlinearity. Let (e_t) be the sequence of residuals from the best fitting estimated linear causal model of some stochastic process (X_t) (and assume the special causal form of (39) given as (42)). Now suppose that (e_t) is not suggestive of an i.i.d. residual (e.g. McLeod-Li tests or BDS statistics from previous sections); this implies that there remains nonlinear structure left unaccounted for. Since all the linear structure is accounted for, the residuals (e_t) should be weak white noise. Moreover, if $m(\cdot)$ was purely linear to begin with then $e_t = s(\epsilon_{t-1}, \epsilon_{t-2}, \dots)\epsilon_t$ and so there is no higher order dependence through the conditional mean and any remaining dependence must then be through the higher order moments such as the conditional variance. However, if $m(\cdot)$ was not linear, this implies that there exists a purely nonlinear function of the *past* residuals which can explain e_t. This suggests a statistical test of the following null hypothesis of nonlinearity through the conditional mean:

$$H_0 : \quad E[e_t | e_{t-1}, e_{t-2}, \dots, e_{t-k}] = 0 \tag{43a}$$

$$H_1 : \quad E[e_t | e_{t-1}, e_{t-2}, \dots, e_{t-k}] \neq 0 \tag{43b}$$

Of course, if there exists a purely nonlinear function of the past residuals which explains the current residual, then this function itself can be expanded in another Taylor series expansion about the point $\mathbf{0} = \{0, 0, \dots\}$, which will include the cross products $e_{t-i}e_{t-j}$ for all $i, j > 0$ up to some finite k. Therefore, we should have that if the null hypothesis is false, then the

bicovariance of the mean zero process e_t, $E[e_t e_{t-i} e_{t-j}]$, should be non-zero for at least some i, j (although there do exist processes nonlinear in mean whose odd product moments are zero– see Pemberton and Tong (1981)).

Hsieh proposes to estimate the normalized bicovariance $E[e_t e_{t-i} e_{t-j}]/E[e_t^2]^{1.5}$ by sample moment equivalents and then under suitable assumptions about the behaviour of (e_t), we have that the estimator is asymptotically Gaussian with mean zero and a variance which can be consistently estimated by $\frac{1}{T}\sum_{t=\max\{i,j\}+1}^{T} e_t^2 e_{t-i}^2 e_{t-j}^2] / \left(\frac{1}{T}\sum_{t=1}^{T} e_t^2\right)^3$.

8.3.1 Monte carlo exercise

Hsieh (1989) conducts a monte carlo simulation exercise which demonstrates the test has good statistical power in rejecting the null hypothesis for a variety of models which are nonlinear in mean such as the nonlinear MA, TAR, Tent map (i.e. an example of a deterministic chaos process), and ARCH-M (i.e. ARCH in mean model). However, in these simulations Hsieh generates sample paths according to Gaussian innovations. Of course, if the innovations are t-distributed instead then it is not clear how many samples, T, are required before the test statistic converges to an asymptotically Gaussian distribution (e.g. if the innovations are Cauchy then standard central limit theorems no longer apply since they require a finite variance assumption). If we use standard Gaussian p-values in determining the rejection region of the test statistic we are likely to under reject the null hypothesis, leading to lower test power. My **Tables 10 and 11** duplicate the monte carlo experiments of Table 11,[12] from Hsieh (1989), section 5, but where I've also included sample paths generated by processes with t-distributed innovations of various degrees of freedom. My **Table 10** directly duplicates Hsieh's Table 11, while my **Table 11** generates simulated sample paths of $T = 10,000$ values instead of only $T = 1,000$.

It is clear from these results that the test suffers from poor power when the innovations are t-distributed with low degrees of freedom and the sample paths are of length $T = 10,000$. For example, at degree of freedom $v = 3$, the test rejects the nonlinear MA model at the (2,1) lag only %35.75 of the time. Considering that the estimation results suggest from section 7.1 suggest a low degree of freedom parameter on the t-distributed innovations of the mixed

[12]That is, they give the percentage of null rejections at the 1% level of a two-tailed test assuming a standard Normal distribution.

Table 10: Power of bicovariance test - % of rejections - 2,000 iterations, T=1,000

Lag		AR(1) $\alpha = .5$			MA(1) $\beta = .5$			Nonlinear MA $\gamma = .8$			Threshold AR(1) $\alpha_i \in \{-.5, .4\}$			ARCH(1) $\alpha = 1, \beta = .5$		
1	1	0.05	0.05	0.15	0.15	0.10	0.30	0.15	0.05	0.00	100.00	97.35	65.95	0.15	0.15	0.00
2	1	0.80	0.70	0.60	0.70	0.75	0.60	99.75	81.40	23.20	1.55	0.65	0.15	0.50	0.30	0.05
2	2	0.15	0.30	0.00	0.05	0.10	0.00	0.20	0.05	0.05	0.15	1.70	7.85	0.15	0.25	0.10
3	1	1.65	0.70	0.75	1.35	0.80	0.60	0.30	0.20	0.25	1.15	0.85	0.65	0.50	0.30	0.00
3	2	1.55	0.75	0.10	1.00	0.70	0.85	0.40	0.35	0.15	0.35	0.35	0.25	0.40	0.30	0.05
3	3	0.35	0.20	0.15	0.00	0.20	0.20	0.20	0.20	0.20	0.05	0.05	0.20	0.20	0.10	0.00
4	1	1.15	1.05	0.65	1.20	0.70	0.80	0.85	0.85	0.35	1.20	1.40	0.60	0.80	0.55	0.30
4	2	1.80	1.20	0.07	0.70	1.00	0.45	0.55	0.70	0.45	0.75	0.80	0.50	0.60	0.35	0.10
4	3	1.80	0.90	0.40	0.80	0.70	0.85	0.70	0.60	0.55	0.60	0.55	0.10	0.90	0.10	0.25
4	4	0.30	0.15	0.15	0.20	0.25	0.05	0.10	0.15	0.25	0.15	0.20	0.15	0.15	0.10	0.05

* The first column for each model represents the Gaussian error term case, the second represents the t-distribution case with degree of freedom $v = 6$ and the third represents degree of freedom $v = 3$.

Table 11: Power of bicovariance test - % of rejections - 2,000 iterations, T=10,000

Lag		AR(1) $\alpha = .5$			MA(1) $\beta = .5$			Nonlinear MA $\gamma = .8$			Threshold AR(1) $\alpha_i \in \{-.5, .4\}$			ARCH(1) $\alpha = 1, \beta = .5$		
1	1	0.10	0.20	0.30	0.25	0.25	0.30	0.05	0.25	0.05	100.00	99.65	78.90	0.15	0.15	0.00
2	1	1.05	1.30	0.90	1.15	0.90	0.70	100.00	98.55	35.75	20.00	3.10	0.10	0.75	0.05	0.20
2	2	0.05	0.30	0.10	0.25	0.25	0.25	0.30	0.40	0.15	9.00	76.80	68.6	0.55	0.20	0.05
3	1	0.75	0.90	0.65	0.75	0.65	0.75	0.65	0.35	0.50	1.00	0.95	0.50	0.60	0.30	0.10
3	2	1.15	0.80	0.50	1.15	0.85	1.05	0.30	0.60	0.45	0.55	0.20	0.05	0.50	0.35	0.15
3	3	0.05	0.15	0.40	0.25	0.20	0.25	0.30	0.40	0.10	0.05	0.60	5.50	0.25	0.05	0.10
4	1	0.80	1.00	0.60	1.20	0.70	1.10	1.10	1.15	0.35	0.65	1.25	0.50	0.55	0.10	0.10
4	2	1.05	1.00	1.10	1.25	1.10	0.45	0.75	0.85	0.45	1.15	0.80	0.55	0.75	0.15	0.15
4	3	0.75	1.10	0.90	1.25	0.80	0.85	0.30	0.75	0.70	0.95	0.35	0.10	0.60	0.20	0.00
4	4	0.10	0.25	0.60	0.15	0.30	0.40	0.25	0.60	0.15	0.01	0.20	0.35	0.25	0.20	0.25

* The first column for each model represents the Gaussian error term case, the second represents the t-distribution case with degree of freedom $v = 6$ and the third represents degree of freedom $v = 3$.

causal/noncausal AR(r, s) processes, if we are to test the futures series for the functional form of remaining nonlinearity, we cannot assume that the Hsiesh (1989) test statistic converges quickly to asymptotic Normality.

Table 12 checks the power of the Hsieh statistic for various other models which are nonlinear in mean. The first two models are the stochastic parameter AR(1) and the exponential AR(1) model which are chosen given their ability to represent bubble phenomenon. The specific form of the true data generated processes (DGP) are given as:

Stochastic parameter AR(1)

$$X_t = 0.1 + \alpha_t X_{t-1} + \epsilon_t, \quad \text{where} \quad \epsilon_t \sim \text{skew-t}(0, 0.1, v_1 = 6, v_2 = 5) \quad \text{and} \tag{44a}$$

$$\alpha_t = 0.089 + 0.007\epsilon_{t-1} + 0.99\alpha_{t-1} \tag{44b}$$

Exponential AR(1)

$$X_t = \gamma_t X_{t-1} + \epsilon_t, \quad \text{where} \quad \epsilon_t \sim \text{t}(0, 0.125, v = 10) \quad \text{and} \tag{44c}$$

$$\gamma_t = 0.99 + 0.2 \exp\left\{-1.0\left(0.99X_{t-30}^2 + 0.99^2 X_{t-40}^2 + 0.99^3 X_{t-50}^2\right)\right\} \tag{44d}$$

Sample path simulations from the models in (44) appear similar to the futures price levels series (see **Figure 12**).

Also included in **Table 12** are the nonlinear AR(2) and Bilinear(1,0,1,1) models along with the All-pass model which is not nonlinear yet exhibits features similar to the GARCH model. The All-pass is included as a benchmark model. The true DGP are given as:

Nonlinear AR(2)

$$X_t = 0.1 + 0.8X_{t-1} - 0.3X_{t-2} + 0.1X_{t-1}X_{t-2} + 0.2X_{t-1}^2 - 0.2X_{t-2}^2 + \epsilon_t \tag{45a}$$

$$\text{where} \quad \epsilon_t \sim \text{t}(0, 0.1, v = 10) \tag{45b}$$

Bilinear(1,0,1,1)

$$X_t = 0.8X_{t-1} + 0.45X_{t-1}\epsilon_{t-1} + \epsilon_t, \quad \text{where} \quad \epsilon_t \sim \text{t}(0, 1, v = 10) \tag{45c}$$

All-pass

$$X_t = 0.5X_{t-1} - (1/0.5)\epsilon_{t-1} + \epsilon_t, \quad \text{where} \quad \epsilon_t \sim \text{t}(0, 1, v = 3) \tag{45d}$$

Figure 12: Simulated SPAR(1) and EAR(1) processes

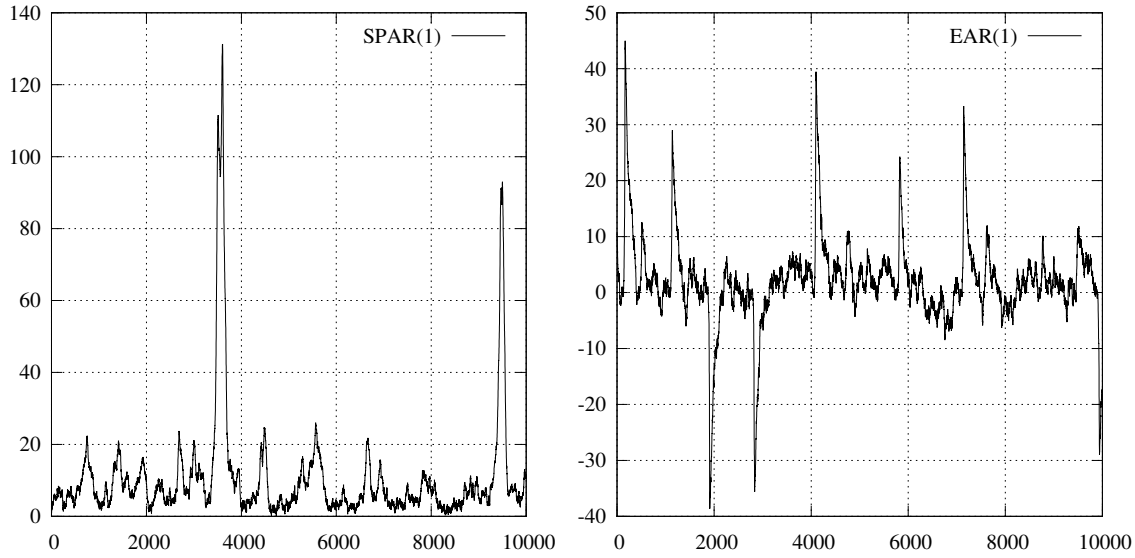


Table 12: Power of bicovariance test - % of rejections - 2,000 iterations, T=10,000

Lag		SPAR(1)	EAR(1)	NLAR(2)	Bilinear	All-pass
1	1	0.75	9.50	99.50	0.35	4.30
2	1	4.55	15.50	99.20	0.30	0.00
2	2	0.45	4.85	99.70	0.10	3.60
3	1	5.70	14.70	34.45	0.70	0.15
3	2	5.30	10.60	99.50	0.05	0.05
3	3	0.45	2.20	98.80	0.05	2.15
4	1	6.05	15.70	1.35	0.75	0.30
4	2	5.35	8.95	98.60	0.00	0.25
4	3	4.70	5.65	69.30	0.00	0.10
4	4	0.25	0.90	5.15	0.10	1.05

Interestingly, the test exhibits low power even when testing models where under the null hypothesis the error term ϵ_t is relatively close to Gaussian. For example, both of the models which can drive "bubble" type phenomenon (the SPAR(1) and EAR(1)) exhibit low rejection rates where the degree of freedom of the t-distributed error term, v, is around 10. Moreover, the Bilinear model is not rejected for lags $(i, j) \geq 1$ but it *is* rejected when at least one lag is equal to zero at rates ranging from 56% to 61% (not shown in table). Moreover, while the power of the test is low among many alternative models nonlinear in mean, the rate of rejecting when the null hypothesis is true is also very low. For example, the AR(1), MA(1), ARCH(1), and All-pass all exhibit very low rejection levels and so the probability of commiting a type I

error is low. Given these facts it is clear that use of the Hsieh test statistic must be interpreted cautiously and that we should most likely reject the null hypothesis at a much lower significance level than 1% in most cases. These results also suggest a reinterpretation of Hsieh (1989) that found nonlinearity in forex returns to be generated strictly through the conditional variance and not the conditional mean, since it is likely that these forex returns are better modeled by a conditionally fat tailed distribution such as the t-distribution with low degrees of freedom; in fact Hsieh himself admits as much on page 364 of Hsieh (1989), but fails to acknowledge the poor power of the test under such a change.

8.3.2 Testing for the functional form of nonlinearity in the commodity futures data

I now employ the Hsieh statistic in an attempt to interpret the functional form of the nonlinearity that exists in the futures price levels series. Recall that the Hsiesh statistic is equivalent to the rescaled bicovariance measure applied to the residuals of the best linear approximate causal model, where the scaling applied is the asymptotic standard deviation. That is, we employ the estimator

$$HS(i,j) = \left(\frac{1}{T} \sum_{t=\max\{i,j\}+1}^{T} e_t e_{t-i} e_{t-j} \right) / \left(\frac{1}{T} \sum_{t=1}^{T} e_t^2 \right)^{1.5} \tag{46}$$

which is asymptotically Gaussian distributed with asymptotic variance

$$var(HS(i,j)) = \left(\frac{1}{T} \sum_{t=\max\{i,j\}+1}^{T} e_t^2 e_{t-i}^2 e_{t-j}^2 \right) / \left(\frac{1}{T} \sum_{t=1}^{T} e_t^2 \right)^{3} \tag{47}$$

where without loss of generality we assume that e_t has unconditional mean of zero. Therefore, we reject the null hypothesis in (43a) of no nonlinearity through the conditional mean of e_t at lag $(i,j) \geq 0$, if $\sqrt{T}\,(HS(i,j))\,/\sqrt{Var(HS(i,j))}$ is greater in absolute value than some critical quantile value of the standard Gaussian distribution.

Figures 18 - 21 in **Appendix A** provide a two-dimensional map of the normalized statistic results for all lags from $(30,30) \geq (i,j) \geq (0,0)$. The different colors indicate the critical value rejection levels respectively: red indicates a rejection at the 1% level, orange at the 5% level, and yellow at the 10% level. Inspection of the figures suggests that all of the series may be better modeled by including a nonlinear component in the conditional mean equation,

however, it is difficult to know in what form this nonlinear conditional mean should manifest.

Interestingly, the sporadic spattering of rejections in the Hsieh figures is consistent with a process where the nonlinearity in the mean equation is not persistent (i.e. the effect of nonlinear shocks die out immediately and not slowly over time). Therefore, the patterns are more consistent with a nonlinear moving average type representation than a nonlinear autoregressive model. For example, tests show that estimating the additional cross product term $\epsilon_{t-i}\epsilon_{t-j}$ in the conditional mean, where (i, j) are set to lags where a sigificant rejection occurs, improves model fit beyond the best causal, linear, approximate model. Of course, the trouble in formulating such a model is that there does not appear to be any consistent pattern in the nature of the rejection lags. Also of interest is the fact that spikes in the bicovariance do not seem to match up with the odd spikes in the ACF of the squared innovations that were observed when examining certain returns series (see **Figure 2**), suggesting these large spikes are entering through the variance and not the mean (the best approximate causal linear model in most cases includes a root very close to unity, and so the innovations from these models exhibit properties quite similar to the returns series).

8.4 Estimation of nonlinear causal models

In order to be more definitive, I estimate all the nonlinear models listed above from equations (44a) to (45d) by maximum likelihood and compare the results to both the benchmark best mixed causal/noncausal AR(r, s) model from section 7.1 and the same benchmark mixed model but with GARCH(1,1) innovations. The GARCH innovations are meant only to capture any possible nonlinearity that is entering through the conditional variance and is not guaranteed be the best specification of this type. Finally, the SPAR and EAR specifications employed include two lags rather than one to allow for more interesting dynamics and I also include a TAR(3,3) model (which includes three lags in both the high and low states, with a threshold variable of X_{t-30}) and an All-pass model which is a linear model that exhibits GARCH like properties.

The results suggest that the mixed model with GARCH(1,1) innovations provides far superior model fit to any of the other strictly causal nonlinear models which formulate the nonlinearity through the conditional mean. The GARCH(1,1) model beats the other models according to

the normalized AIC criterion for every futures price level series, with the exception of the lean hogs series where adding GARCH(1,1) innovations adds little. Furthermore, both the Ljung-Box and McLeod-Li type statistics suggest the innovations of this nonlinear in variance model are much closer to i.i.d. (only in the case of the metals and some of the energy series do we find the results lacking in this regard but this is due only to lag 1 serial correlation in the rescaled residuals which most likely could be corrected by a more appropriate nonlinear in variance type specification).

Models such as the SPAR(2) or EAR(2) which are able to generate bubble like behaviour through nonlinear conditional mean equations appear inadequate. Of particular interest is the fact that the coefficients on the innovations which enter into the stochastic autoregressive parameters of the former model, i.e. the γ_1 in $\alpha_t = \gamma_0 + \gamma_1\epsilon_{t-1} + \alpha_{t-1}$ are not found to be significantly different from zero, suggesting the past lagged innovations $\epsilon_{t-1}, \epsilon_{t-2}, \ldots$ do not affect the current autoregressive persistence of the price levels series (at least within the scope of this specification).

Morever, the mixed causal/noncausal AR(r, s) estimation results from section 7.1, which implied bubble like behaviour through the conditional mean via t-distributed error terms with low degrees of freedom parameters seem inferior to a specification with t-distributed GARCH type innovations according to measures of model fit and residual whiteness. Interestingly, the GARCH specification is estimated with much larger degree of freedom parameters on the t-distributed error terms. Estimation results from section 7.1 suggested t-distributed degree of freedom parameters somewhere between just over 1 and 3, while under the GARCH(1,1) error term specification we see these parameters roughly double. Therefore, despite the nonlinear purely causal equivalent representation of the linear mixed causal/noncausal AR(r, s) models employing t-distributed error terms with low degrees of freedom parameters, it appears they are unable to account as effectively for the type of nonlinear structure that causes bubbles in the commodity futures price levels as is a specification which is nonlinear in variance.

8.4.1 Stylized facts

What is perhaps most surprisingly is the consistency of these results across futures price level series. Therefore, the results from this section, as well as the previous ones, suggests a set of "stylized facts" that describe this particular futures price level data set, if we are to assume a model of the type from equation (42):

1. The vast majority of the nonlinear structure in the futures price levels series is generated through the changing conditional variance (the structure is nonlinear through the variance).

2. An approximately linear model captures the vast majority of any persistence suggested by the autocovariance and bicovariance. What remains shows up mainly in the tricovariance and higher.

3. There exists remaining nonlinear structure through the conditional mean of the residuals of the approximate linear model; however, this structure is not "persistent" in the sense that innovations die out quickly and do not persist in effect for more than one period.

4. There appears to exist two types of nonlinear structure:

    - That which is "persistent" as exhibited by the tricovariance (but where this persistence does not show up in the autocovariance and bicovariance of the linear model residuals, i.e. structure that is captured reasonably well by a GARCH type model).
    - That which is periodic in the tricovariance and appears as periodic "spike-like" patterns in the ACF of the squared innovations of the linear model (or the returns– see below). I call this "seasonal" nonlinearity through variance. Series of this type are orange juice, live cattle, lean hogs, and lumber.

5. The price levels series exhibit irreversibility and an AR model of mixed causality improves upon a strictly causal model by standard measures of model fit. However, the amount of irreversibility detectable by the bicovariance estimator varies across industry sector, with the most significant appearing in some of the soft goods such as cocoa and coffee and the livestock goods.

6. The conditional distribution of both the price levels series and the returns series is better modeled as student t-distributed rather than Gaussian. Moreover, the use of a parametric skewed student t-distributed error term adds little in most cases.

7. Since the linear approximate model with t-distributed error terms nearly always exhibits a root close to unity, the residuals from this model exhibit very similar properties to the returns (the returns have a normalized martingale structure imposed upon them). Therefore the facts above also apply to the returns series to more or less the same degree.

Table 13: Estimation results of nonlinear models - 1 of 5

	Model	Log-likelihood	AICn	Lag	Ljung-Box	McLeod-Li
Soybean meal	SPAR(2)	-24504.380	5.28441	9	0.000	0.000
	EAR(2)	-24072.060	5.21821	9	0.000	0.000
	All-pass	-46979.330	10.12702	9	0.000	0.000
	NLAR(2)	-24945.070	5.37898	9	0.000	0.000
	TAR(3,3)	-24116.740	5.21681	9	0.000	0.000
	Bilinear(2,2)	-24204.230	5.21906	9	0.000	0.000
	GARCH(1,1)	-22937.117	4.95942	9	0.751	0.708
	Benchmark	-24081.130	5.20608	9	0.007	0.000
Soybean oil	SPAR(2)	-4656.400	1.00591	9	0.002	0.000
	EAR(2)	-4618.474	1.00292	9	0.020	0.000
	All-pass	-28967.230	6.24469	9	0.000	0.000
	NLAR(2)	-4871.949	1.05194	9	0.000	0.000
	TAR(3,3)	-4620.182	1.00134	9	0.002	0.000
	Bilinear(2,2)	-4658.458	1.00570	9	0.005	0.000
	GARCH(1,1)	-3512.315	0.76400	9	0.734	0.760
	Benchmark	-4582.938	0.99480	9	0.135	0.000
Soybeans	SPAR(2)	-45145.660	9.73392	9	0.000	0.000
	EAR(2)	-34620.300	7.50386	9	0.000	0.000
	All-pass	-60764.490	13.09828	9	0.000	0.000
	NLAR(2)	-46590.610	10.04497	9	0.000	0.000
	TAR(3,3)	-34680.750	7.50092	9	0.000	0.000
	Bilinear(2,2)	-35863.470	7.73237	9	0.000	0.000
	GARCH(1,1)	-33444.821	7.22912	9	0.892	0.434
	Benchmark	-34695.177	7.49874	9	0.000	0.000
Orange juice	SPAR(2)	-20639.290	4.45124	9	0.000	0.000
	EAR(2)	-19288.700	4.18173	9	0.023	0.000
	All-pass	-45570.890	9.82345	9	0.000	0.000
	NLAR(2)	-21590.930	4.65594	9	0.000	0.000
	TAR(3,3)	-19333.580	4.18261	9	0.081	0.000
	Bilinear(2,2)	-19411.960	4.18602	9	0.076	0.000
	GARCH(1,1)	-18861.845	4.06896	9	0.153	0.146
	Benchmark	-19317.960	4.17753	9	0.389	0.000
Sugar	SPAR(2)	-2139.321	0.46332	9	0.000	0.000
	EAR(2)	-827.275	0.18142	9	0.000	0.000
	All-pass	-27336.900	5.89329	9	0.000	0.000
	NLAR(2)	-1150.396	0.24971	9	0.000	0.000
	TAR(3,3)	-804.275	0.17628	9	0.000	0.000
	Bilinear(2,2)	-801.858	0.17436	9	0.000	0.000
	GARCH(1,1)	647.095	-0.13758	9	0.192	0.504
	Benchmark	-767.750	0.16704	9	0.000	0.000

Table 14: Estimation results of nonlinear models - 2 of 5

	Model	Log-likelihood	AICn	Lag	Ljung-Box	McLeod-Li
Wheat	SPAR(2)	-32535.590	7.01565	9	0.000	0.000
	EAR(2)	-30885.460	6.69457	9	0.000	0.000
	All-pass	-54095.140	11.66077	9	0.000	0.000
	NLAR(2)	-33380.030	7.19725	9	0.000	0.000
	TAR(3,3)	-30926.310	6.68915	9	0.000	0.000
	Bilinear(2,2)	-31221.280	6.73168	9	0.000	0.000
	GARCH(1,1)	-29671.139	6.40137	9	0.796	0.343
	Benchmark	-30926.145	6.67533	9	0.000	0.000
Cocoa	SPAR(2)	-49865.960	10.75145	9	0.000	0.000
	EAR(2)	-45570.540	9.87661	9	0.000	0.000
	All-pass	-73788.550	15.90550	9	0.000	0.000
	NLAR(2)	-94362.340	20.34282	9	0.000	0.000
	TAR(3,3)	-45712.130	9.88608	9	0.000	0.000
	Bilinear(2,2)	-50781.320	10.94811	9	0.000	0.000
	GARCH(1,1)	-44888.458	9.70075	9	0.590	0.304
	Benchmark	-45769.055	9.89051	9	0.003	0.000
Coffee	SPAR(2)	-22925.260	4.94401	9	0.000	0.000
	EAR(2)	-21772.590	4.71995	9	0.000	0.000
	All-pass	-48397.790	10.43276	9	0.000	0.000
	NLAR(2)	-22577.010	4.86851	9	0.000	0.000
	TAR(3,3)	-21828.080	4.72196	9	0.000	0.000
	Bilinear(2,2)	-21965.270	4.73642	9	0.000	0.000
	GARCH(1,1)	-20775.766	4.49282	9	0.072	0.296
	Benchmark	-21841.150	4.72249	9	0.012	0.000
Corn	SPAR(2)	-26969.930	5.81589	9	0.383	0.000
	EAR(2)	-26732.280	5.79464	9	0.358	0.000
	All-pass	-53238.200	11.47606	9	0.000	0.000
	NLAR(2)	-27762.680	5.98635	9	0.000	0.000
	TAR(3,3)	-26781.970	5.79308	9	0.458	0.000
	Bilinear(2,2)	-26899.520	5.80007	9	0.000	0.000
	GARCH(1,1)	-25319.245	5.46103	9	0.674	0.911
	Benchmark	-26815.913	5.78413	9	0.776	0.000
Cotton	SPAR(2)	-13733.350	2.96257	9	0.000	0.000
	EAR(2)	-13474.960	2.92198	9	0.000	0.000
	All-pass	-38875.670	8.38036	9	0.000	0.000
	NLAR(2)	-14774.740	3.18662	9	0.000	0.000
	TAR(3,3)	-13479.650	2.91690	9	0.000	0.000
	Bilinear(2,2)	-13511.260	2.91405	9	0.000	0.000
	GARCH(1,1)	-12757.530	2.75259	9	0.570	1.000
	Benchmark	-13495.916	2.91137	9	0.000	0.000

Table 15: Estimation results of nonlinear models - 3 of 5

	Model	Log-likelihood	AICn	Lag	Ljung-Box	McLeod-Li
Rice	SPAR(2)	3556.749	-1.12470	9	0.000	0.000
	EAR(2)	3495.877	-1.11388	9	0.000	0.000
	All-pass	-14455.260	4.58474	9	0.000	0.000
	NLAR(2)	3441.419	-1.08876	9	0.000	0.000
	TAR(3,3)	3520.635	-1.11790	9	0.000	0.000
	Bilinear(2,2)	3553.963	-1.12477	9	0.000	0.000
	GARCH(1,1)	4368.026	-1.38272	8	0.191	0.000
	Benchmark	3593.842	-1.13778	8	0.013	0.000
Lumber	SPAR(2)	-21625.670	6.17897	9	0.583	0.000
	EAR(2)	-21419.520	6.16234	9	0.417	0.000
	All-pass	-39866.820	11.38544	9	0.000	0.000
	NLAR(2)	-22534.870	6.43806	9	0.000	0.000
	TAR(3,3)	-21459.780	6.15649	9	0.434	0.000
	Bilinear(2,2)	-21536.720	6.15271	9	0.000	0.000
	GARCH(1,1)	-20674.554	5.92261	8	0.977	0.587
	Benchmark	-21444.622	6.14706	8	0.546	0.000
Gold	SPAR(2)	-29566.970	6.37572	9	0.000	0.000
	EAR(2)	-28457.470	6.16847	9	0.000	0.000
	All-pass	-55169.260	11.89228	9	0.000	0.000
	NLAR(2)	-29556.980	6.37314	9	0.000	0.000
	TAR(3,3)	-28452.750	6.15433	9	0.000	0.000
	Bilinear(2,2)	-28819.910	6.21404	9	0.000	0.000
	GARCH(1,1)	-26328.729	5.68365	9	0.000	0.053
	Benchmark	-28435.869	6.14662	9	0.000	0.000
Silver	SPAR(2)	2903.283	-0.62369	9	0.000	0.000
	EAR(2)	3365.470	-0.72708	9	0.000	0.000
	All-pass	-24813.100	5.34931	9	0.000	0.000
	NLAR(2)	3048.051	-0.65532	9	0.000	0.000
	TAR(3,3)	3428.187	-0.73885	9	0.000	0.000
	Bilinear(2,2)	3462.235	-0.74482	9	0.000	0.000
	GARCH(1,1)	6259.295	-1.34727	9	0.010	0.000
	Benchmark	3552.141	-0.76202	9	0.000	0.000
Platinum	SPAR(2)	-28328.710	8.08869	9	0.000	0.000
	EAR(2)	-25834.250	7.42758	9	0.000	0.000
	All-pass	-46273.850	13.20743	9	0.000	0.000
	NLAR(2)	-29268.830	8.35645	9	0.000	0.000
	TAR(3,3)	-25840.610	7.40840	9	0.000	0.000
	Bilinear(2,2)	-26376.620	7.53065	9	0.000	0.000
	GARCH(1,1)	-24335.574	6.95859	8	0.199	0.066
	Benchmark	-25798.400	7.38944	8	0.000	0.000

Table 16: Estimation results of nonlinear models - 4 of 5

	Model	Log-likelihood	AICn	Lag	Ljung-Box	McLeod-Li
Palladium	SPAR(2)	-21440.460	6.12258	9	0.000	0.000
	EAR(2)	-21297.540	6.12374	9	0.000	0.000
	All-pass	-43885.670	12.52588	9	0.000	0.000
	NLAR(2)	-23227.250	6.63201	9	0.000	0.000
	TAR(3,3)	-21320.970	6.11319	9	0.000	0.000
	Bilinear(2,2)	-21457.230	6.12651	9	0.000	0.000
	GARCH(1,1)	-19180.988	5.48535	8	0.003	0.176
	Benchmark	-21265.772	6.08745	8	0.000	0.000
Copper	SPAR(2)	-16441.300	5.21684	9	0.000	0.000
	EAR(2)	-15223.650	4.86776	9	0.000	0.000
	All-pass	-35168.300	11.15196	9	0.000	0.000
	NLAR(2)	-16934.920	5.37274	9	0.000	0.000
	TAR(3,3)	-15266.520	4.86623	9	0.000	0.000
	Bilinear(2,2)	-15344.670	4.86814	9	0.000	0.000
	GARCH(1,1)	-13610.866	4.33674	8	0.567	0.001
	Benchmark	-15240.889	4.85447	8	0.000	0.000
Light crude oiil	SPAR(2)	-9367.224	2.40719	9	0.000	0.000
	EAR(2)	-8661.489	2.23983	9	0.000	0.000
	All-pass	-29975.810	7.69528	9	0.000	0.000
	NLAR(2)	-10224.860	2.62684	9	0.000	0.000
	TAR(3,3)	-8651.234	2.23167	9	0.000	0.000
	Bilinear(2,2)	-8662.182	2.22543	9	0.000	0.000
	GARCH(1,1)	-6605.455	1.69867	8	0.702	0.103
	Benchmark	-8639.603	2.22047	8	0.014	0.000
Heating oil	SPAR(2)	-15488.810	4.46523	9	0.016	0.000
	EAR(2)	-15405.810	4.47224	9	0.022	0.000
	All-pass	-32778.580	9.44364	9	0.000	0.000
	NLAR(2)	-17069.280	4.91999	9	0.000	0.000
	TAR(3,3)	-15432.210	4.46723	9	0.095	0.000
	Bilinear(2,2)	-15488.700	4.46433	9	0.000	0.000
	GARCH(1,1)	-13772.541	3.97077	8	0.132	0.000
	Benchmark	-15381.000	4.44945	8	0.042	0.000
Brent crude oil	SPAR(2)	-7636.773	2.38032	9	0.000	0.000
	EAR(2)	-7615.577	2.39159	9	0.004	0.000
	All-pass	-23826.660	7.41726	9	0.000	0.000
	NLAR(2)	-8129.052	2.53293	9	0.000	0.000
	TAR(3,3)	-7597.920	2.37890	9	0.000	0.000
	Bilinear(2,2)	-7624.903	2.37569	9	0.000	0.000
	GARCH(1,1)	-5807.002	1.81130	8	0.444	0.019
	Benchmark	-7512.764	2.35267	8	0.000	0.000

Table 17: Estimation results of nonlinear models - 5 of 5

	Model	Log-likelihood	AICn	Lag	Ljung-Box	McLeod-Li
Gas oil	SPAR(2)	-21630.640	7.02846	9	0.210	0.000
	EAR(2)	-20533.240	6.72447	9	0.842	0.000
	All-pass	-36860.730	11.97134	9	0.000	0.000
	NLAR(2)	-20881.070	6.78437	9	0.000	0.000
	TAR(3,3)	-20565.270	6.71330	9	0.860	0.000
	Bilinear(2,2)	-20653.250	6.71005	9	0.000	0.000
	GARCH(1,1)	-18993.143	6.17121	8	0.997	0.001
	Benchmark	-20532.228	6.69584	8	0.261	0.000
Natural gas	SPAR(2)	3903.088	-1.30597	9	0.000	0.000
	EAR(2)	3819.260	-1.28822	9	0.000	0.000
	All-pass	-12712.390	4.26543	9	0.000	0.000
	NLAR(2)	3731.592	-1.24911	9	0.000	0.000
	TAR(3,3)	3848.145	-1.29327	9	0.000	0.000
	Bilinear(2,2)	3891.110	-1.30296	9	0.000	0.000
	GARCH(1,1)	5236.231	-1.75415	8	0.538	0.422
	Benchmark	3891.158	-1.30364	8	0.017	0.000
Gasoline RBOB	SPAR(2)	-5807.628	6.06635	8	0.000	0.000
	EAR(2)	-5639.084	6.04180	8	0.000	0.000
	All-pass	-11610.390	12.10567	8	0.000	0.000
	NLAR(2)	-5931.780	6.19372	8	0.000	0.000
	TAR(3,3)	-5678.679	6.02083	8	0.000	0.000
	Bilinear(2,2)	-5769.072	6.02301	8	0.000	0.000
	GARCH(1,1)	-5693.417	5.94931	7	0.883	0.976
	Benchmark	-5756.133	6.01266	7	0.056	0.000
Live cattle	SPAR(2)	-10370.960	2.23776	9	0.005	0.000
	EAR(2)	-10219.650	2.21661	9	0.000	0.000
	All-pass	-35258.520	7.60072	9	0.000	0.000
	NLAR(2)	-17638.990	3.80405	9	0.000	0.000
	TAR(3,3)	-10217.490	2.21157	9	0.000	0.000
	Bilinear(2,2)	-10242.390	2.20940	9	0.000	0.001
	GARCH(1,1)	-9947.719	2.15393	9	0.991	0.867
	Benchmark	-10190.943	2.20604	9	0.915	0.000
Lean hogs	SPAR(2)	-9598.605	2.74257	9	0.000	1.000
	EAR(2)	-9431.913	2.71358	9	0.279	1.000
	All-pass	-26839.060	7.66098	9	0.000	0.000
	NLAR(2)	-9974.069	2.84917	9	0.000	0.875
	TAR(3,3)	-9438.812	2.70807	9	0.349	1.000
	Bilinear(2,2)	-9474.789	2.70638	9	0.339	1.000
	GARCH(1,1)	-9452.342	2.70025	8	0.600	0.985
	Benchmark	-9455.188	2.70050	8	0.570	0.993

9 Forecasting

Up until this point the entire discussion has been focused on evaluating within-sample model fit. This section will therefore attempt to measure the ability of the various models to not only fit the training sample, but also to be able to forecast ahead out of sample within a reasonable horizon. In order to facilitate forecast comparisons I have set aside an additional 107 sample data points beyond the most recent date available within-sample, which is February 8th, 2013 (recall the the discussion from section 3.1.4). Therefore, this out of sample period extends between February 11th to July 15th, 2013. [13]

9.1 The optimal MMSE forecast

Given the general stochastic process from equation (39), let $\mathcal{F}_t$ represent the information set generated by the stochastic process at time t. It can be shown that in the general case, the optimal one step ahead predictor of X_t (denoted $\hat{X}_{t|t-1} = g(\mathcal{F}_{t-1})$) under the minimum mean squared error (MMSE) criterion is the *conditional expectation*, $E[X_t|\mathcal{F}_{t-1}]$. That is, the argmin of the quadratic loss function, the expected mean squared error $E\left[\left(X_t - \hat{X}_{t|t-1}\right)^2\right]$, can be shown to be $g(\mathcal{F}_{t-1}) = E[X_t|\mathcal{F}_{t-1}]$. See see Priestley (1981, pg. 76) for proof.

Of course, in the special case where the ϵ_t's in (39) are uncorrelated and Gaussian, and $h(\cdot)$ is a purely linear function then we have from Rosenblatt (2000) that the one step ahead predictor is linear and so coincides with the orthogonal linear projection of $\mathcal{F}_{t-1}$ onto the space spanned by X_t. Moreover, if we restrict equation (39) to be of the purely causal linear form then the same result applies, even if the innovations are non-Gaussian.

Given these results forecasts from the strictly causal linear ARMA models with t-distributed innovations, ϵ_t, from section 5 are easy to generate recursively. However, for the AR models of mixed causality from section 7.1 the one step ahead predictor is not so easily calculated since

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

we have from equation (9a) that: [14]

$$E_T[x_{T+1}] = \phi_1 x_T + \phi_2 x_{T-1} + \cdots + \phi_r x_{T-r+1} + E_T[u_{T+1}] \tag{48a}$$

$$\text{where} \quad E_T[u_{T+1}] \neq 0 \tag{48b}$$

and where u_{T+1} is given as in (15).

9.1.1 A numerical algorithm

Despite this difficulty Lanne, Luoto, and Saikkonen (2010) suggest a means whereby we can simulate the one step ahead predictor. What follows is an abridged version of their discussion. Please see their paper for more details.

Since $E_T[u_{T+1}] \neq 0$ we can therefore simulate this expectation as follows:

1. First we require the density of $\boldsymbol{\epsilon}^+ = \{\epsilon_{T+1}, \epsilon_{T+2}, \dots\}$, conditional on the data $\boldsymbol{x} = \{x_T, x_{T-1}, \dots\}$.

2. From equation (36) it can be shown that:

$$\frac{\mathcal{L}_{\boldsymbol{x},\boldsymbol{\epsilon}^+}(\boldsymbol{x}, \boldsymbol{\epsilon}^+|\Omega)}{\mathcal{L}_{\boldsymbol{x}}(\boldsymbol{x}|\Omega)} = p(\boldsymbol{\epsilon}^+|\boldsymbol{x}; \Omega) = \frac{f_{u,\epsilon^+}(\boldsymbol{u}(\boldsymbol{\phi}), \boldsymbol{\epsilon}^+)}{f_u(\boldsymbol{u}(\boldsymbol{\phi}))} \tag{49}$$

   where Ω represents the parameters of the mixed causal/noncausal AR(r, s) model and $\boldsymbol{u}(\boldsymbol{\phi}) = \{\phi(L)x_{T-s+1}, \dots, \phi(L)x_T\} = \{u_{T-s+1}, \dots, u_T\}$. Therefore, we can use monte carlo simulations to approximate both the numerator and denominator of (49) in order to approximate the desired conditional expectation:

$$E_T[u_{T+1}] \approx \frac{1}{f_u(\boldsymbol{u}(\boldsymbol{\phi}))} \int \left(\sum_{j=0}^{M-1} \varphi_j^* \epsilon_{T+1+j} \right) f_{u,\epsilon^+}(\boldsymbol{u}(\boldsymbol{\phi}), \boldsymbol{\epsilon}^+) d\boldsymbol{\epsilon}^+ \tag{50}$$

   under the assumption of some finite M, and that as $M \to \infty$, $(\varphi_j^*) \to 0$, and where $u_{T+1} \approx \sum_{j=0}^{M-1} \varphi_j^* \epsilon_{T+1+j}$ from (15).

[14] Note that since the strictly causal nonlinear in mean models from section 8 were not found to improve model fit–with the exception of adding nonlinear MA terms–I will not compare their point forecasts; although later I will consider the unconditional distribution of the simulated process that is nonlinear in variance.

3. In order to do this, however, we need to accomplish a change of variables between $(\boldsymbol{u}(\boldsymbol{\phi}), \boldsymbol{\epsilon}^+)$ and $(\{\epsilon_{T-s+1}, \ldots, \epsilon_T\}, \boldsymbol{\epsilon}^+)$. Given (15) we have that an approximate mapping between these two sets is given as:

$$\begin{bmatrix} 1 & \varphi_1^* & \ldots & \ldots & \ldots & \ldots & \varphi_{M+s-1}^* \\ 0 & \ddots & \ddots & & & & \vdots \\ \vdots & \ddots & 1 & \varphi_1^* & \ldots & \ldots & \varphi_M^* \\ \vdots & & \ddots & 1 & 0 & \ldots & 0 \\ \vdots & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & \ddots & \ddots & 0 \\ 0 & \ldots & \ldots & \ldots & \ldots & 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{T-s+1} \\ \vdots \\ \epsilon_T \\ \epsilon_{T+1} \\ \vdots \\ \epsilon_{T+M} \end{bmatrix} \approx \begin{bmatrix} u_{T-s+1} \\ \vdots \\ u_T \\ \epsilon_{T+1} \\ \vdots \\ \epsilon_{T+M} \end{bmatrix} \tag{51}$$

which can be written as $\boldsymbol{C}\boldsymbol{e} \approx \boldsymbol{w}$. Therefore, inverting $\boldsymbol{C}$ and noting that its determinant is unity we can write the numerator in (49) as:

$$f_{u,\epsilon^+}(\boldsymbol{u}(\boldsymbol{\phi}), \boldsymbol{\epsilon}^+) \approx \prod_{j=1}^{s} f_\epsilon(\epsilon_{T-s+j}(\boldsymbol{u}(\boldsymbol{\phi}), \boldsymbol{\epsilon}^+)) \prod_{t=T+1}^{T+M} f_\epsilon(\epsilon_t) \tag{52}$$

where i've written the elements $\epsilon_{T-s+j}(\boldsymbol{u}(\boldsymbol{\phi}), \boldsymbol{\epsilon}^+)$ as such to indicate that they are functions of both $\boldsymbol{u}(\boldsymbol{\phi})$ and $\boldsymbol{\epsilon}^+$.

4. Therefore, if we simulate N i.i.d. draws of the M length vector $\boldsymbol{\epsilon}_{\boldsymbol{i}}^+$ (i.e. for $i = 1, \ldots, N$) according to $f_\epsilon(\cdot)$, an approximation to the desired expectation in (50) is given as:

$$E_T[u_{T+1}] \approx \frac{N^{-1} \sum_{i=1}^{N} \left(\sum_{j=0}^{M-1} \varphi_j^* \epsilon_{i,T+1+j} \right) \prod_{j=1}^{s} f_\epsilon(\epsilon_{T-s+j}(\boldsymbol{u}(\boldsymbol{\phi}), \boldsymbol{\epsilon}_{\boldsymbol{i}}^+))}{N^{-1} \sum_{i=1}^{N} \prod_{j=1}^{s} f_\epsilon(\epsilon_{T-s+j}(\boldsymbol{u}(\boldsymbol{\phi}), \boldsymbol{\epsilon}_{\boldsymbol{i}}^+))} \tag{53}$$

- Notice that conditional on time T information, we can also easily compute the expectation $E_T[u_{T+h}]$ for any horizon h by simply replacing the term $\sum_{j=0}^{M-1} \varphi_j^* \epsilon_{i,T+1+j}$ in (53) with $\sum_{j=0}^{M-h} \varphi_j^* \epsilon_{i,T+h+j}$ for any integer $h \geq 1$. Therefore, once we have computed the denominator in (53) we need not calculate it again in computing expectations across horizons h.

- Moreover, if we so desire, we can also compute interval forecasts by replacing $\sum_{j=0}^{M-1} \varphi_j^* \epsilon_{i,T+1+j}$ with $\mathbf{1}\left(\sum_{j=0}^{M-1} \varphi_j^* \epsilon_{i,T+1+j} < \alpha\right)$; that is, we take the conditional expectation, given time T information, of the indicator function of u_{T+1} so that if $u_{T+1} < \alpha$ we return 1 and 0 otherwise. Therefore, with this change (53) then becomes an estimate of the conditional cummulative distribution function (CDF) at α, or the probability that $u_{T+1} < \alpha$. Given an appropriately chosen grid of α_i's we can then generate an approximation to the shape of the CDF across its support.

- While the methods described above are computationally intensive, it is clear that they are ripe for parallelization since we can potentially draw each of the N i.i.d. samples of $\boldsymbol{\epsilon}_{\boldsymbol{i}}^{+}$ at the same time. In order to save time in this way 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.

9.2 Monte carlo results

Simulated mixed AR(r, s) models of the types from section 7.1 were generated, with $T = 10000$, according to a number of different parameter choices well within the stationarity region. Both an equivalent causal ARMA model of the type from section 5 and a mixed AR(r, s) model were then estimated in-sample and then forecast out-of-sample, generating $T_0 = 500$ forecast errors. These results were consistent in demonstrating an improvement in forecasts under the mixed AR(r, s) specification as opposed to the best causal ARMA, especially when the simulated process was mostly noncausal. Surprisingly, even when the t-distributed innovations were not generated i.i.d. but rather according to a GARCH(1,1) specification, the mixed AR(r, s) forecasts still proved superior although their performance was significantly degraded.

9.3 Forecasting the commodity futures price level data

In comparing point forecasts, the out of sample data set is used to generate forecast errors according to

$$\zeta_{T+h} = E[X_{T+h}|\mathcal{F}_T] - x^*_{T+h}, \quad \text{for} \quad 10 \geq h \geq 1, \tag{54}$$

where x^*_{T+h} is the true out of sample outcome for a particular series at time $T+h$. Next, the root mean squared errors are computed as the root of the average squared errors for each series i at horizon h as

$$RMSE_{i,h} = \sqrt{\frac{1}{T_0}\sum_{j=1}^{T_0}\zeta^2_{T+h}}. \tag{55}$$

Table 18 presents the ratios of the RMSE's for each of the futures price level data series for horizons $h = \{1, 5, 10\}$ with $T_0 = 98$ forecast errors generated from the 107 out-of-sample data points. The numerator is the purely causal ARMA model forecasts with t-distributed errors from section 5 and the denominator represents the forecast errors from the mixed AR(r, s) models of section 7.1, so ratios that are greater than 1 suggest smaller forecast errors amongst the mixed AR(r, s) models.

The results of the forecasting exercise seem somewhat consistent with **Table 7** in that series which exhibited the most significant levels of time irreversibility in their estimated bicovariance measures tended to gain the most from partly noncausal point forecasts. For example, from the softs category, cocoa, coffee, cotton, and lumber all exhibit improved point forecasts and these series were found to be significantly time irreversible. Contrastingly, orange juice and sugar were not found to reject the null of reversibility according to the TR test and their forecasts proved more accurate when modeled as strictly causal. Also consistent with **Table 7** is the poor mixed causal/noncausal point forecasts for the metals and energy sectors which failed to reject the null of reversibility. However, despite these consistencies, the poor mixed causality forecasts for both the soybean products and livestock categories are anomalous. Particularly surprising in light of their TR test results are the poor mixed causal/noncausal forecasts of soybeanmeal and lean hogs and the very much improved mixed forecasts of natural gas and gasoline RBOB. In light of these results it would be desirable to have access to a larger out of sample pool with which to forecast, as 107 elements may represent too small a sample to draw

Table 18: Ratio of out of sample forecast RMSE's

Series	**Horizon**		
	1	**5**	**10**
Soybean meal	00.784	00.849	00.897
Soybean oil	00.801	00.843	00.806
Soybeans	00.271	00.403	00.447
Orange juice	00.998	00.988	00.994
Sugar	00.919	00.954	00.971
Wheat	00.945	00.951	00.945
Cocoa	**01.029**	**01.010**	**01.018**
Coffee	00.988	**01.005**	**01.016**
Corn	00.475	00.570	00.624
Cotton	00.998	**01.007**	**01.011**
Rice	00.884	00.960	**01.004**
Lumber	00.989	**01.003**	**01.003**
Gold	00.076	00.121	00.159
Silver	00.357	00.448	00.613
Platinum	00.382	00.496	00.655
Palladium	00.334	00.471	00.542
Copper	00.425	00.537	00.596
Light crude oil	00.879	00.944	00.955
Heating oil	00.912	00.855	00.807
Brent crude oil	00.529	00.468	00.554
Gas oil	00.331	00.433	00.521
Natural gas	**01.002**	00.992	**01.001**
Gasoline RBOB	**01.051**	**01.090**	**01.138**
Live Cattle	00.999	00.999	**01.000**
Lean hogs	00.011	00.016	00.020

any strong conclusions.

Interestingly, forecasts of $E_T[u_{T+h}]$ generated from the simulation algorithm described in Lanne, Luoto, and Saikkonen (2010) tend to be sensitive to model misspecification. In many of the troublesome cases, especially the soy products, metals, and to a lesser extent the energy sector, these forecasts fluctuated wildly from one sample of $\boldsymbol{\epsilon}_{\boldsymbol{i}}^{+}$, $i = 1, \ldots, N$ to the next. Moreover, increasing N to reasonable levels does not seem to improve the situation. Examination of the forecast errors suggests this is a problem specific to these series and their estimated models, as other series such as sugar or live cattle do not exhibit this problem at all.

There is evidence in the literature of similarly varied point forecasting results when attempting to model financial data by the mixed causal/noncausal model from section 6. Lanne, Nyberg, and Saarinen (2011) estimated mixed causal/noncausal models for the Marcellino, Stock, and Watson (2006) dataset which includes 170 monthly U.S. macroeconomic and financial time series. The dataset is split into 5 categories of variables: A) income, output, sales and capacity utilization B) employment and unemployment related C) construction, inventories and orders D) interest rates and asset prices and E) nominal prices, wages, and money. However, unlike the current study they employ a number of different transformations designed to ensure stationarity, such as log differencing. Within the (D) category, they forecast log differenced exchange rates, S&P 500 and NYSE indices, interest rates, and bond yields. Generally for all of the categories with the exception of (D), the authors find that, on average, modeling the series as partly noncausal improves point forecasts. However, the series in (D) prove more troubling for the mixed causal/noncausal representation to generate superior point forecasts to the strictly causal model. The authors attribute this failure to both the low estimated degrees of freedom of the t-distributed error terms and the presence of conditional heteroskedasticity, suggesting that perhaps another parametric distribution may prove a better choice. Interestingly, these results prove similar to **Table 18**, despite the lack of transformations and the difference in type of financial asset. Some other points of note are that the authors suggest that models selected under the BIC (Bayesian information criterion) as opposed to the AIC tended to forecast better and that the MAFE (Mean absolute forecast error) proves less sensitive to dramatic outliers than the MSFE employed here.

With regards to conditional heteroskedasticity, in section 8 I showed that a model which is nonlinear through its variance appears to improve model fit. However, simulated mixed AR(r, s) series under the assumption of GARCH(1,1) innovations failed to generate the same dramatic variability in the forecast errors of $E_T[u_{T+h}]$ as is exhibited by the data for the soy, metals, and energy sectors. Unfortunately, it is not possible to modify the algorithm of Lanne, Luoto, and Saikkonen (2010) to account for non-i.i.d. innovations of the GARCH type, since these innovations depend on past information through their scale parameter. Specifically, in simulating i.i.d. innovations in step (4) of the algorithm above we'd need to normalize the heteroskedastic errors by their their time dependent scale parameters which are functions of past shocks $\{\epsilon_{T-s+1}, \ldots, \epsilon_T\}$. However, this poses a problem since $\{\epsilon_{T-s+1}, \ldots, \epsilon_T\}$ is exactly that which we wish to compute based on draws of $\boldsymbol{\epsilon}_{\boldsymbol{i}}^{+}$.

9.4 Comparison of the estimated unconditional distributions

Another means by which we can evaluate the various models is by comparing their unconditional distributions estimated by sample histogram. Histograms are computed for each of the models from section 5, 7.1, and the mixed AR(r, s) model with GARCH(1,1) innovations from 8, by simulating them for a sufficiently large number of points. Then a histogram is computed from the true futures price level series.

The metric employed in comparing the estimated unconditional distributions is the Kullback-Leibler divergence measure which represents a non-symmetric measure of the difference between two probability distributions P and Q (i.e. the $KL(P, Q) \neq KL(Q, P)$ in general). 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. [15] 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 19** 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 estimated from the simulated linear causal ARMA, mixed AR(r, s), and mixed AR(r, s) (with

[15] 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}.

GARCH(1,1) innovations) models from sections 5, 7.1, and 8, respectively.

Table 19: Kullback-Leibler divergence measures

	KL(Q,P)			KL(P,Q)		
Series	**ARMA**	**MIXED**	**MIX/GARCH**	**ARMA**	**MIXED**	**MIX/GARCH**
Soybean meal	n.s.	00.329	01.370	n.s.	97.216	263.671
Soybean oil	01.965	00.316	00.284	495.751	55.752	71.478
Soybeans	n.s.	00.310	00.503	n.s.	49.584	124.888
Orange juice	00.976	00.216	00.487	351.966	60.033	229.859
Sugar	01.768	00.500	00.286	326.343	168.821	96.815
Wheat	00.535	00.427	00.211	44.699	32.956	06.182
Cocoa	00.625	01.247	00.297	230.260	37.961	107.290
Coffee	**04.519**	00.216	00.410	703.097	81.218	192.468
Corn	01.526	00.549	00.339	185.980	144.244	82.144
Cotton	00.808	**12.710**	02.349	114.104	25.918	16.083
Rice	00.429	00.311	uns.	59.220	123.030	uns.
Lumber	00.149	00.136	00.211	07.610	08.477	81.658
Gold	n.s.	uns.	uns.	n.s.	uns.	uns.
Silver	n.s.	uns.	uns.	n.s.	uns.	uns.
Platinum	n.s.	00.662	00.633	n.s.	96.789	140.821
Palladium	n.s.	01.368	uns.	n.s.	440.585	uns.
Copper	n.s.	00.832	00.943	n.s.	173.295	206.154
Light crude oil	n.s.	00.813	00.759	n.s.	202.916	301.470
Heating oil	n.s.	01.043	**29.883**	n.s.	326.858	159.701
Brent crude oil	n.s.	00.759	**138.219**	n.s.	118.503	01.424
Gas oil	n.s.	00.709	00.506	n.s.	132.528	86.560
Natural gas	00.906	00.753	00.404	303.694	325.575	143.436
Gasoline RBOB	01.429	00.261	00.280	483.674	08.649	19.078
Live Cattle	00.562	**18.227**	00.701	31.469	76.491	34.953
Lean hogs	02.649	00.032	00.044	640.295	03.308	03.313
average	**01.346**	01.858	08.529	284.154	121.335	**112.831**
selective average	01.206	**00.650**	01.029			

* P represents the sample data. "n.s." stands for non-stationary, i.e. the simulations from the causal linear model were explosive. "uns." within the context of the MIX/GARCH models implies that the simulated sample paths were for a lack of better words "unstable": highly eratic with extremely long tails and extremely irregular, almost "chaotic" type behaviour. In general, while stationary, models with "uns." listed represented poor possibilities as having come from the sample data DGP. Finally, the selective average omits the outliers in bold.

Table 19 is broken into two sections: the three left columns report the Kullback-Leibler measure where the three prospective models are used to approximate the sample data; in this case if the sample data density has zero support, it does not punish the prospective density for allocating too much probability to this region since this component of the Kullback-Leibler sum is zero. On the other hand, the three right columns report the Kullback-Leibler measure

when the sample density is used to approximate each of the three prospective models; in this case the opposite occurs if the prospective density has zero support. Finally, smaller values indicate less information lost by the approximation and are preferred.

Either way, the results of these comparisons suggest the following. First, of the estimated linear causal ARMA models where the unconditional distribution exists, the Kullback-Leibler measures suggest that it is a poor fit to the sample data. The linear causal 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.

In terms of a comparison between the mixed causal/noncausal AR(r, s) model with t-distributed innovations and the modification where I allow these innovations to be GARCH(1,1), it appears that the results are not so clear. For one, many of the GARCH simulated sample paths are highly irregular with extremely long tailed unconditional distributions. From these I was unable to generate a robust estimate of the unconditional distribution without drawing extremely long samples. For these I have indicated "uns." in the appropropriate spot of the table. It is interesting that these poorly behaved processes tend to coincide with the same models that proved difficult to forecast above, specifically the metals category which exhibits some of the largest bubbles of all the series. Ultimately, however, the "stable" GARCH type processes tend to generate simulated sample paths appearing visually more similar to the sample data than the non-GARCH mixed models and this is borne out when we take the average across Kullback-Leibler measures for the $KL(P, Q)$ category. In the $KL(Q, P)$ category, however, there exist a few distinct outlier values (denoted in bold text), which when omitted from the average reveal that the non-GARCH mixed model tends to slightly edge out the alternative.

9.5 Conclusions

Given these results, what does seem clear is that that the mixed causal/noncausal AR(r, s) model, while improving model fit, is unable to generate robust forcasts for many of the commodity futures price series' using the algorithm from Lanne, Luoto, and Saikkonen (2010). However, when reliable forecasts are generated, they improve upon the purely causal linear model amongst some of the series from the softs and energy sectors; for example, cocoa, coffee,

cotton, rice and lumber all benefit from the mixed causal/noncausal AR(r, s) representation, especially in the longer horizons, as do natural gas and gasoline RBOB. Interestingly, these results tend to coincide with the measures of which series were most significantly irreversible as measured by the TR test from **Table 7**. However, the disappointing mixed causal/noncausal forecasts for the livestock series as well as the superior forecasts for the two aforementioned energy series are surprising. Finally, a benefit of the mixed AR(r, s) model is that it allows us to forecast models where the purely causal forecast would otherwise be explosive.

Results from an analysis of the estimated unconditional distributions of the mixed AR(r, s) model specification suggest that it represents a better approximation to the sample data than a purely causal linear model, again when simulations are available and the model is not explosive. However, both comparison of the Kullback-Leibler measures and inspection of the simulated processes themselves from both the GARCH and non-GARCH type error term versions of the mixed AR(r, s) model suggest that this process may not be an ideal specification of the DGP of the sample data. While the GARCH type process tends to appear visually as a better match, and exhibits better in-sample model fit according to the AIC measure, its behaviour can be very eratic and ill-behaved for very long sample paths, especially when applied to the metals category. Given the results from this section, along with those from section 8, it may be interesting to consider other processes which are nonlinear in variance but take advantage of the apparent irreversibility of the commodity futures price levels data.

A Appendix: Tables and Charts

Table 20: Summary statistics - 1 of 2 - continuously compounded returns series

Series	Mean	Median	Stnd. Dev.	Skewness	Kurtosis
Soybean meal	0.011	0.000	1.656	-0.739	12.097
Soybean oil	0.009	0.000	1.543	-0.012	5.360
Soybeans	0.009	0.000	1.536	-1.237	17.737
Orange juice	0.002	0.000	1.950	0.537	16.902
Sugar	0.009	0.000	2.542	0.208	9.952
Wheat	0.013	0.000	1.752	-0.583	13.055
Cocoa	-0.009	0.000	1.947	-0.033	6.329
Coffee	-0.007	0.000	2.330	-0.290	13.336
Corn	0.013	0.000	1.596	-1.261	29.734
Cotton	0.003	0.000	1.882	-8.655	335.767
Rice	0.014	0.000	1.693	0.148	30.219
Lumber	0.012	0.000	2.002	0.338	9.098
Gold	0.026	0.000	1.226	-0.048	10.568
Silver	0.021	0.000	2.131	-0.851	48.379
Platinum	0.021	0.000	1.393	-0.515	7.518
Palladium	0.028	0.000	1.973	-0.207	9.473
Copper	0.016	0.000	1.735	-0.308	7.384
Light crude oil	0.015	0.000	2.296	-0.872	20.222
Heating oil	0.032	0.000	2.305	-1.346	23.633
Brent crude oil	0.032	0.009	2.186	-0.964	21.828
Gas oil	0.032	0.000	2.124	-1.168	20.918
Natural gas	0.012	0.000	3.387	0.234	9.900
Gasoline RBOB	0.025	0.010	2.459	-0.100	7.238
Live cattle	0.012	0.000	1.117	-1.243	15.123
Lean hogs	0.005	0.000	2.152	-1.078	70.267

* The returns have been premultiplied by 100.
Note also that the Kurtosis measure employed here is *not* an excess Kurtosis measure.

Table 21: Summary statistics - 2 of 2 - continuously compounded returns series

Series	T	Quantiles 0%	25%	50%	75%	100%	C.V.
Soybean meal	9279	-20.002	-0.779	0.000	0.826	10.483	155.217
Soybean oil	9279	-11.352	-0.857	0.000	0.842	8.247	176.207
Soybeans	9279	-23.224	-0.720	0.000	0.801	7.806	166.468
Orange juice	9279	-16.173	-0.820	0.000	0.860	23.897	1102.497
Sugar	9279	-18.218	-1.210	0.000	1.236	27.492	276.489
Wheat	9279	-23.884	-0.944	0.000	0.914	9.868	135.456
Cocoa	9279	-15.390	-1.078	0.000	1.028	12.561	-205.881
Coffee	9279	-25.655	-1.050	0.000	1.050	23.773	-338.867
Corn	9279	-32.932	-0.746	0.000	0.771	8.662	120.554
Cotton	9279	-77.653	-0.796	0.000	0.833	16.811	560.699
Rice	6308	-24.040	-0.760	0.000	0.789	27.461	121.364
Lumber	7004	-19.176	-1.121	0.000	1.083	17.328	173.106
Gold	9279	-9.909	-0.463	0.000	0.559	9.745	46.709
Silver	9279	-44.496	-0.786	0.000	0.947	41.534	101.993
Platinum	7008	-12.274	-0.626	0.000	0.752	7.622	67.511
Palladium	7008	-14.372	-0.855	0.000	0.950	15.253	69.732
Copper	6308	-11.714	-0.840	0.000	0.890	11.644	108.660
Light crude oil	7792	-40.048	-0.992	0.000	1.084	14.894	151.589
Heating oil	6943	-39.094	-1.095	0.000	1.214	13.994	72.931
Brent crude oil	6426	-38.564	-1.025	0.009	1.126	13.151	69.271
Gas oil	6159	-34.026	-1.003	0.000	1.100	13.867	67.125
Natural gas	5963	-32.769	-1.739	0.000	1.703	32.435	291.098
Gasoline RBOB	1919	-13.031	-1.206	0.010	1.396	17.476	97.901
Live cattle	9279	-13.818	-0.491	0.000	0.597	8.512	90.024
Lean hogs	7008	-45.042	-0.760	0.000	0.828	34.969	393.417

* C.V. stands for coefficient of variation, or σ/μ, and T denotes the sample size. Note that the spectral density for all series is roughly uniform, suggesting weak white noise.

Figure 13: Bilinear model

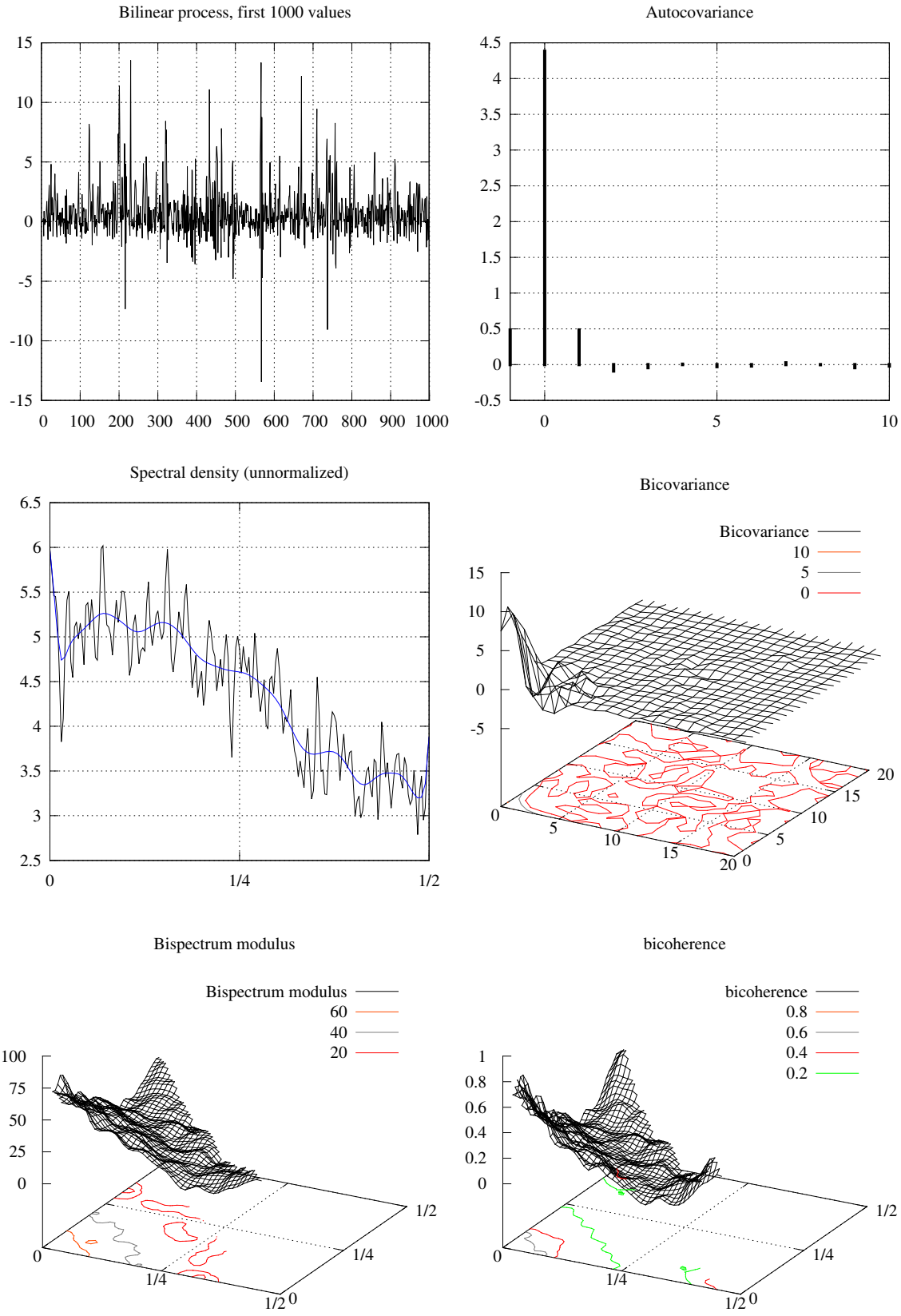


Figure 14: GARCH model

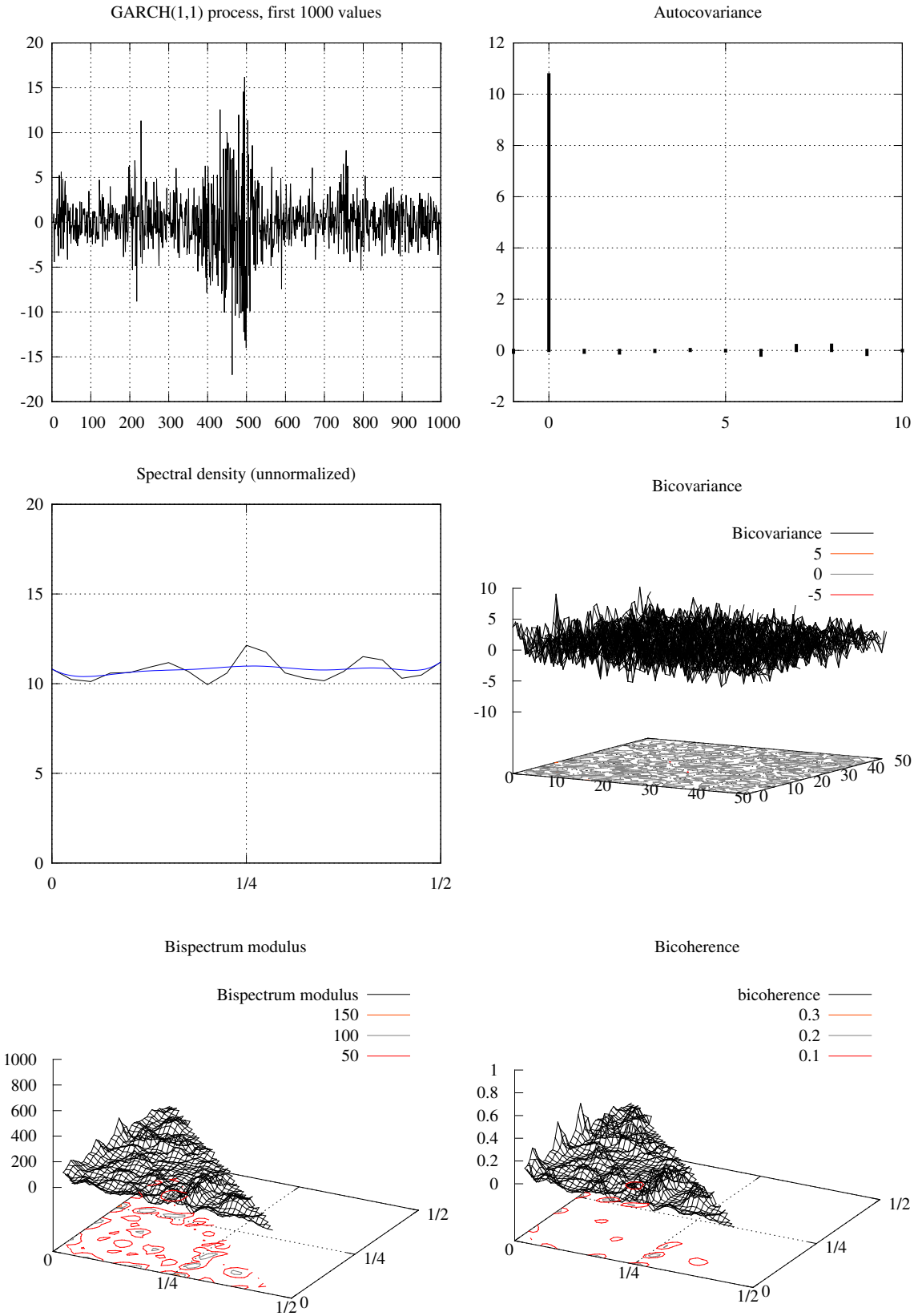


Figure 15: Bicovariance plots, 800 lags

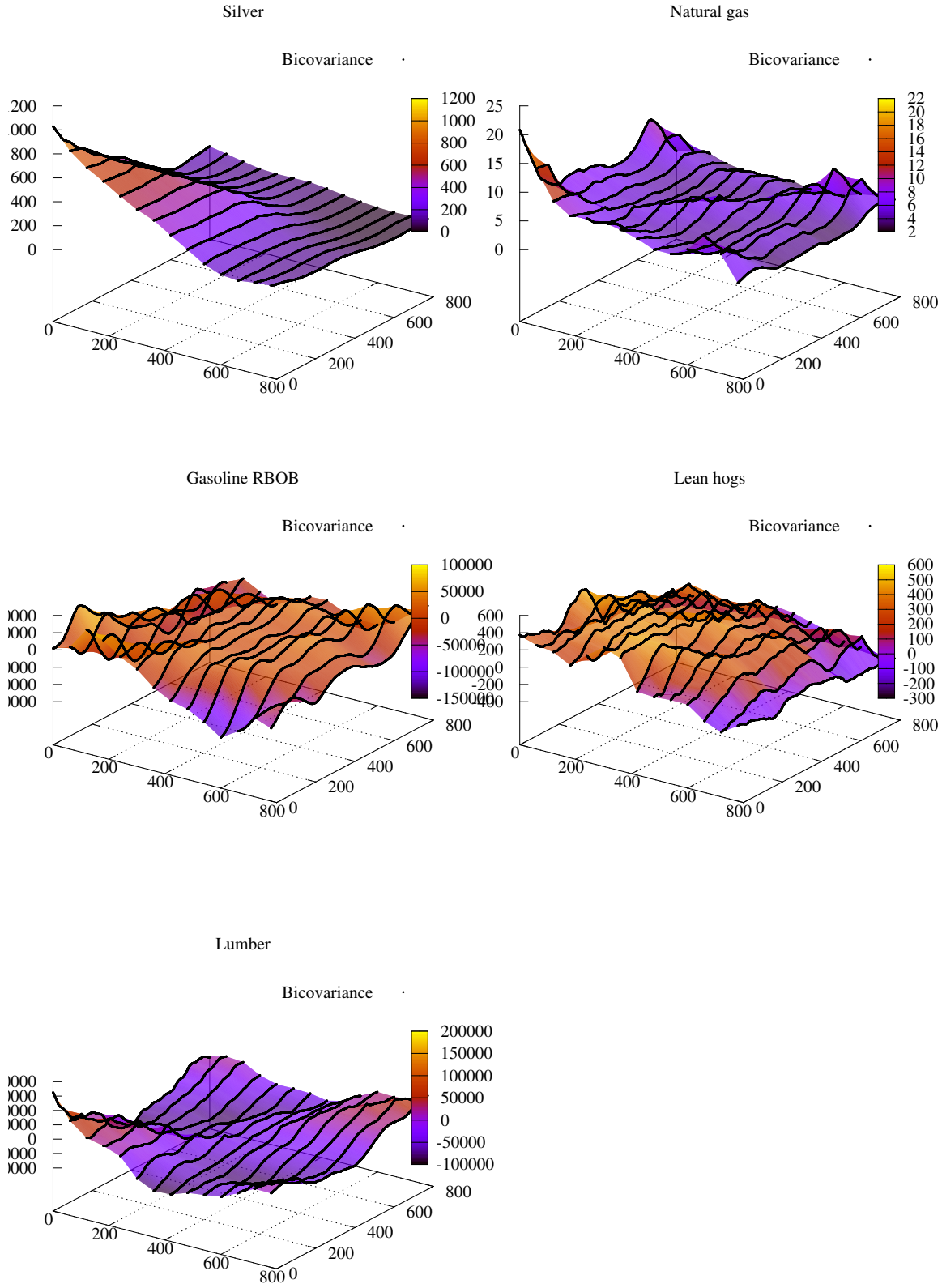


Figure 16: Bicoherence plots

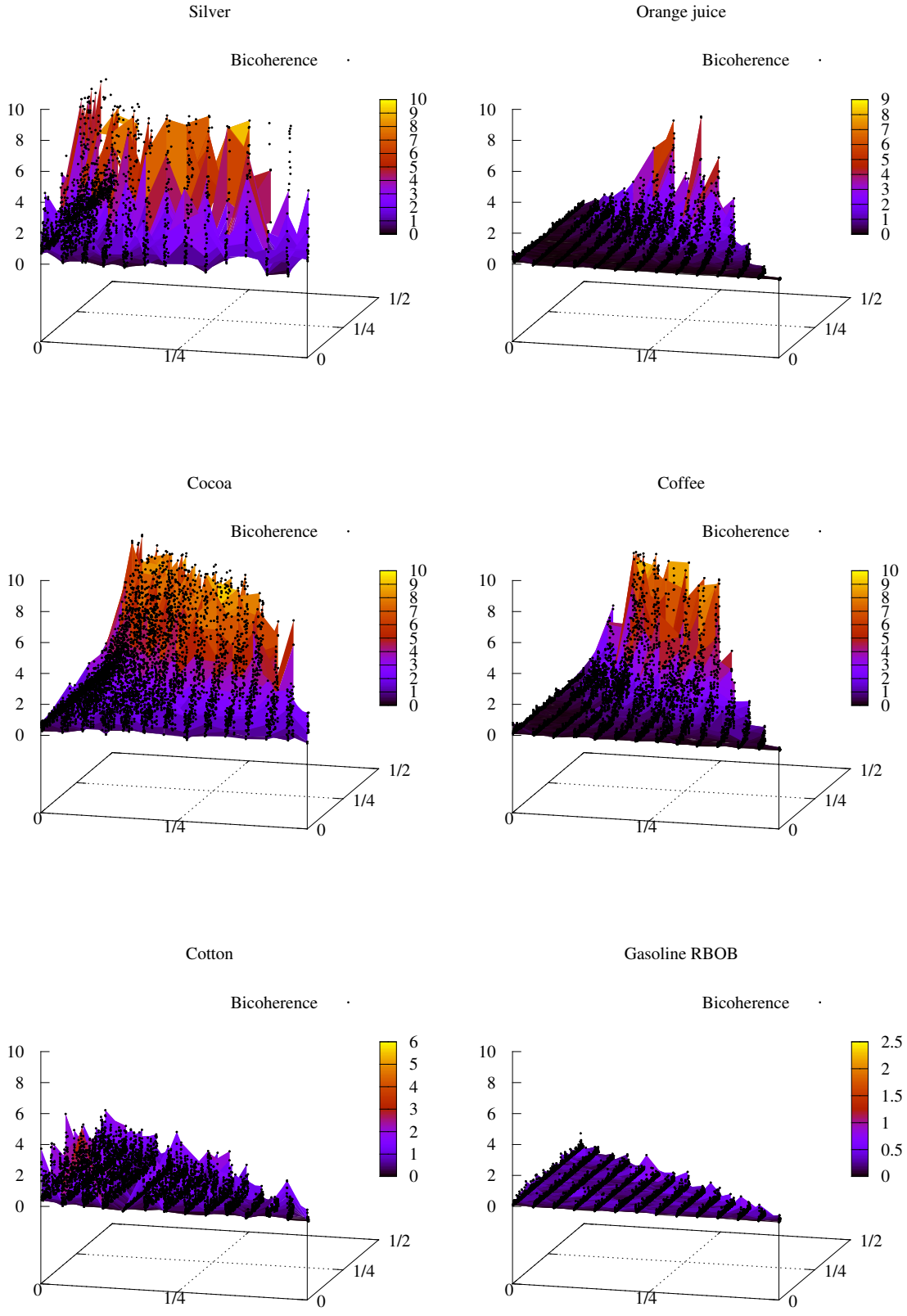


Figure 17: Bicoherence plots

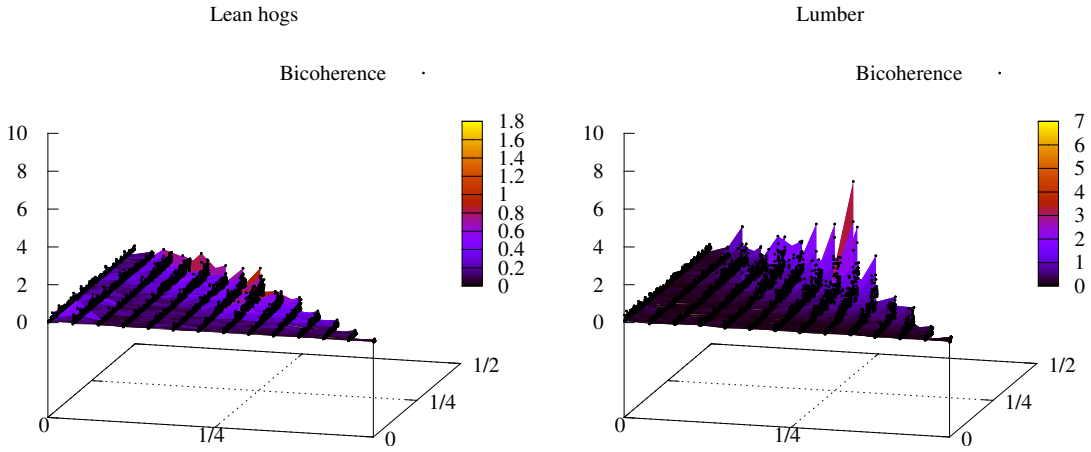


Figure 18: Hsieh (1989) statistic plots - 1 of 4

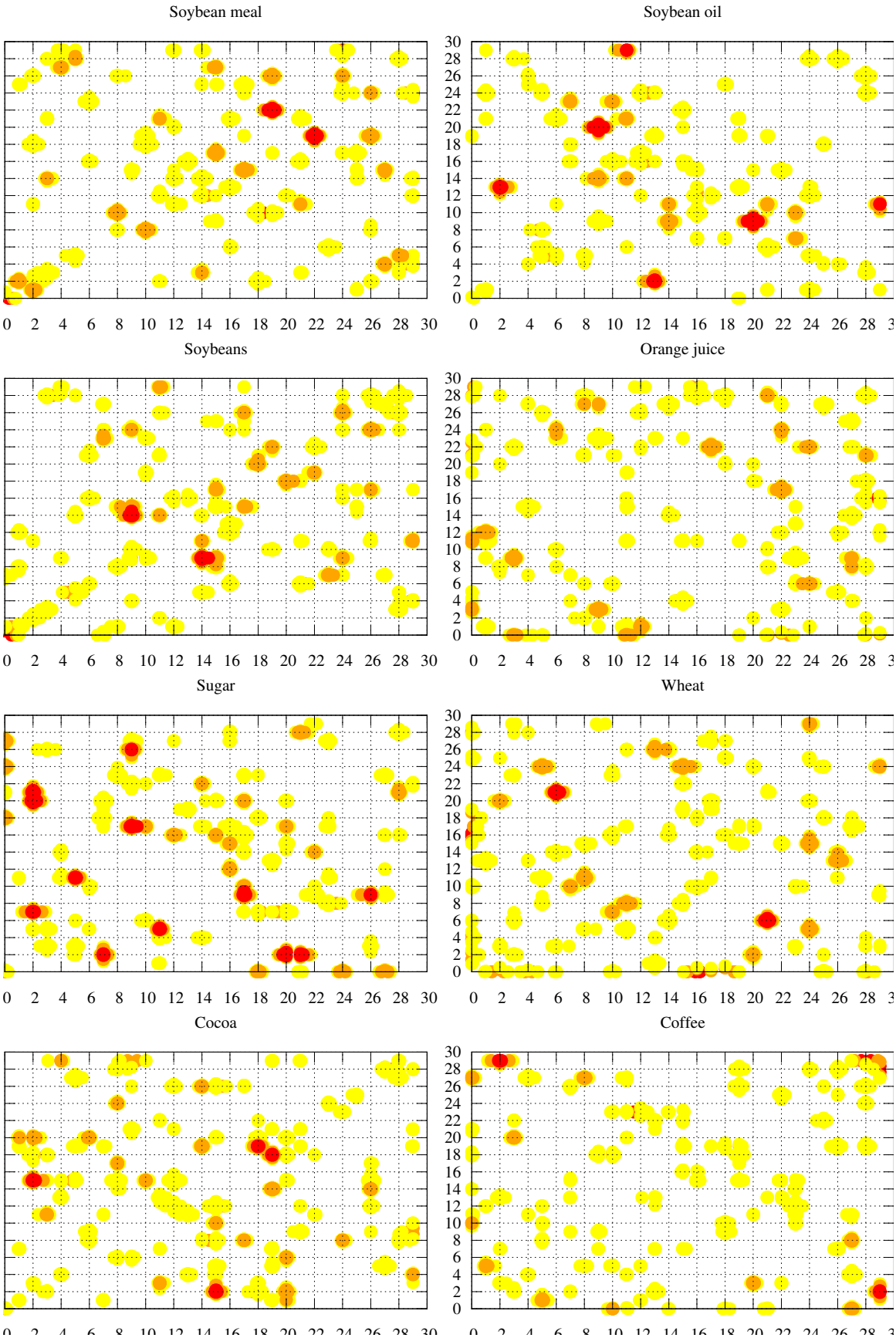


Figure 19: Hsieh (1989) statistic plots - 2 of 4

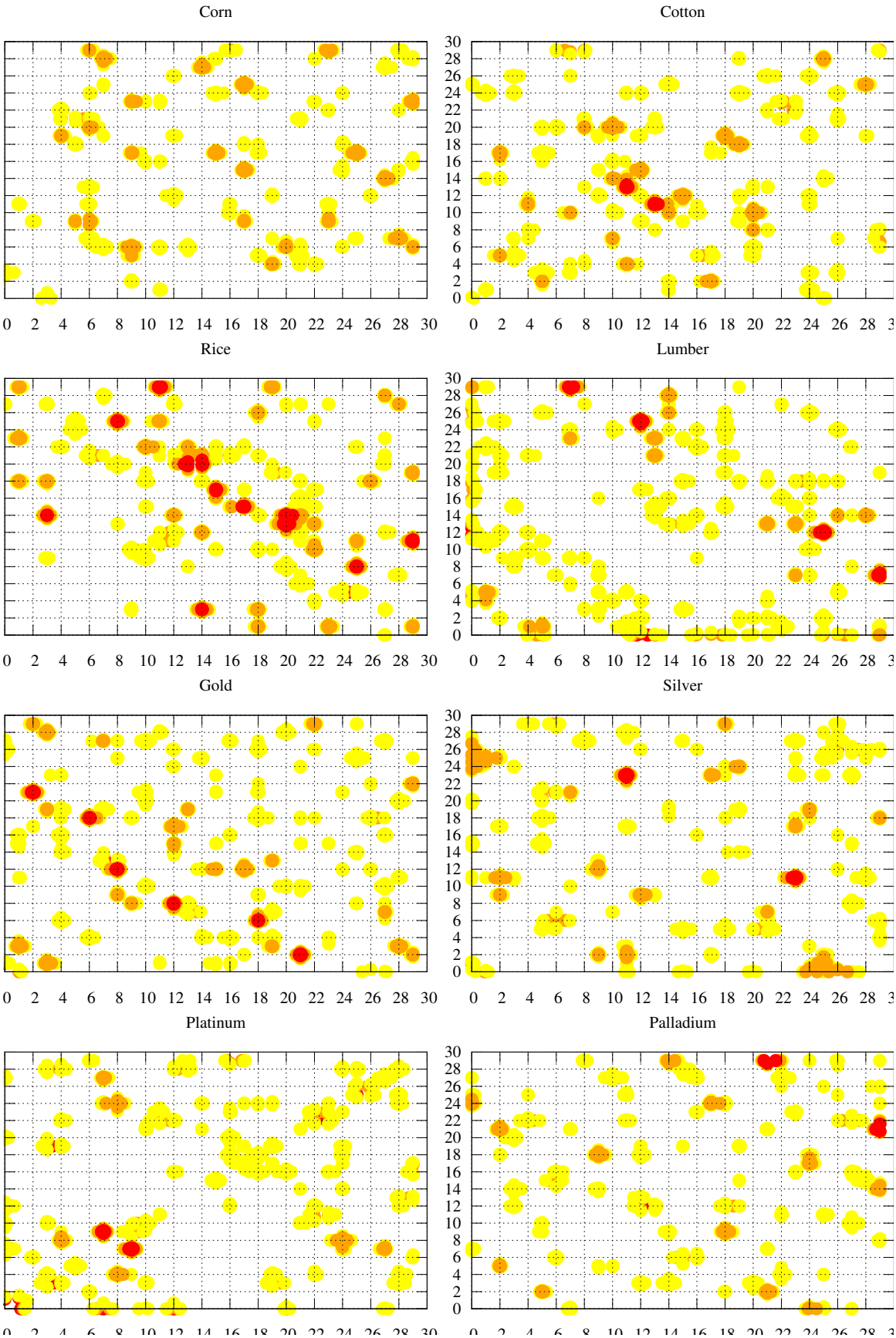


Figure 20: Hsieh (1989) statistic plots - 3 of 4

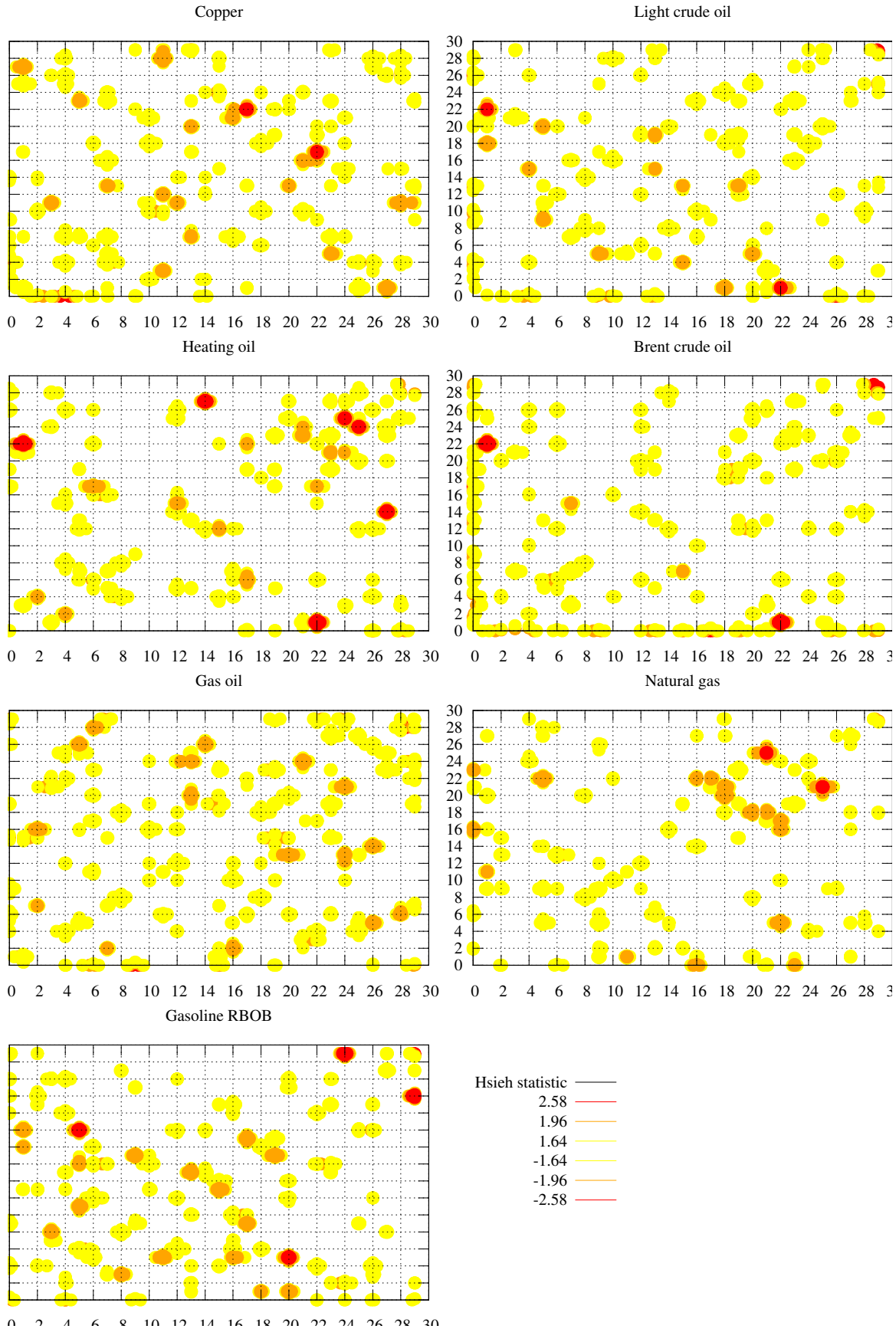


Figure 21: Hsieh (1989) statistic plots - 4 of 4

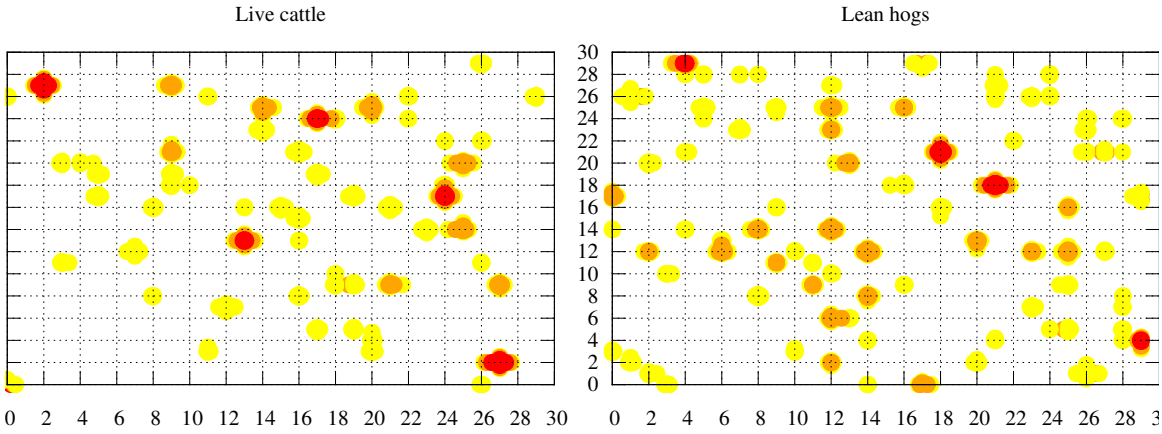


Table 22: Lag polynomial roots of the mixed and benchmark models - 1 of 3

	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 23: Lag polynomial roots of the mixed and benchmark models - 2 of 3

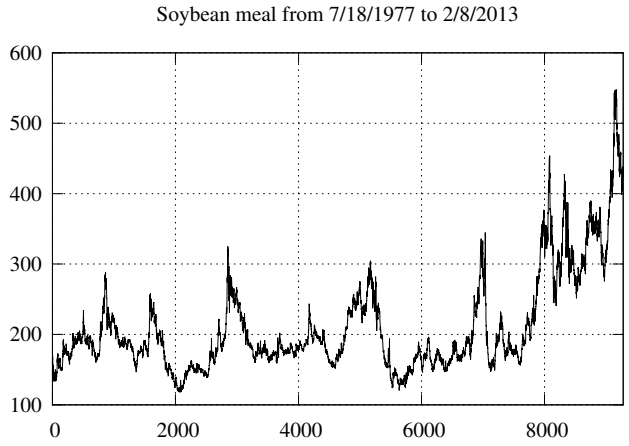
	Model	p/r,q/s	Sig.p/r	Sig.q/s	cR	cMC	ncR	ncMC	#CC
Coffee	t-dist arma	10,0	1,3		0.995	4.740			1
	skew-t mixed	10,10	1,2,5,6,10	1,2,5,6,7		1.375	1.027	1.684	5/2
						1.403	1.571	1.762	
						1.428	-1.645		
						1.430			
						1.446			
Corn	skew-t arma	2,0	1,2		1.000				0
					51.190				
	t-dist mixed	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	

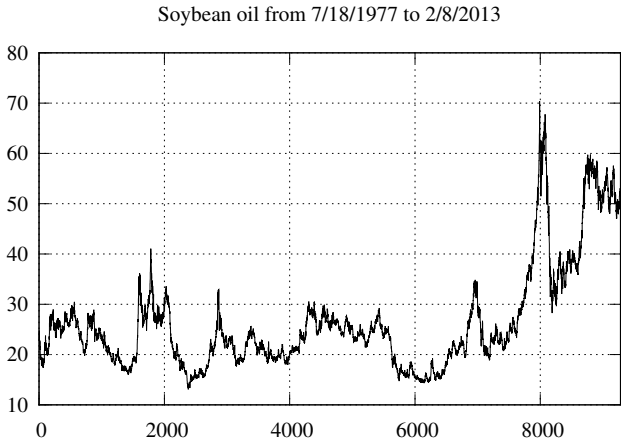
Legend:
p/r,q/s: (p,q) or (r,s) pairs for ARMA(p,q) and Mixed causal/noncausal AR(r, s) models respectively.
Sig.p/r,q/s: Significant lags at the 5% level assuming Normal distributed parameters.
cR,cMC: Causal lag polynomial; real roots and modulus of complex roots respectively.
ncR,ncMC: Noncausal lag polynomial; real roots and modulus of complex roots respectively.
#CC: Number of complex conjugate roots with the same modulus (causal/noncausal).

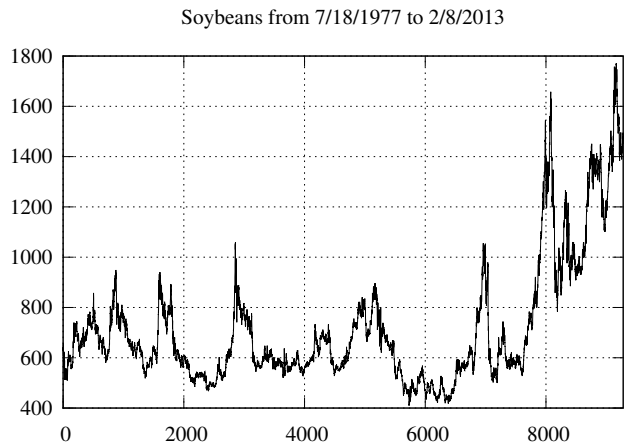
Table 24: Lag polynomial roots of the mixed and benchmark models - 3 of 3

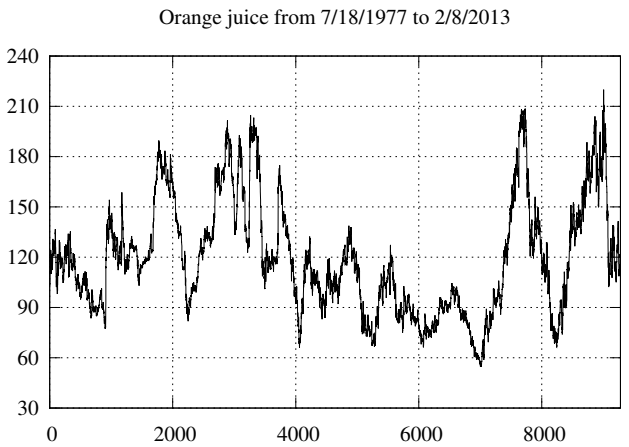
	Model	p/r,q/s	Sig.p/r	Sig.q/s	cR	cMC	ncR	ncMC	#CC
Palladium	skew-t arma	5,0	1,2,4,5		1.006	2.431			1
					-2.434				
					3.525				
	t-dist mixed	8,8	1,2-7	1,2,3,7,8	-1.618	1.621	0.989	1.547	3/3
						1.632	1.536	1.574	
						1.884		1.619	
Copper	skew-t arma	10,0	1,2,6		1.055	2.020			2
					1.696	2.101			
	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		

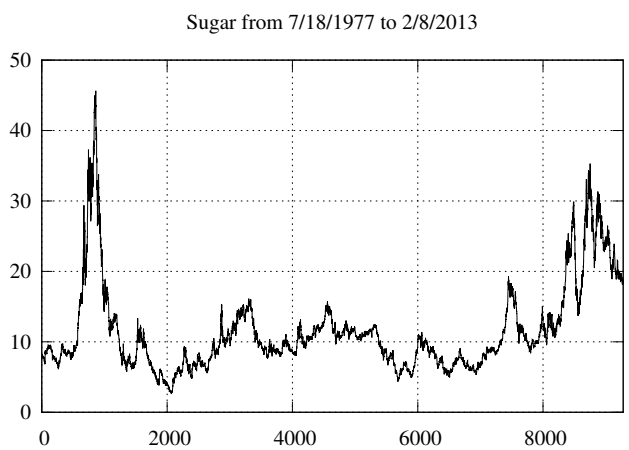
Figure 22: Plots of daily continuous contract futures price level series

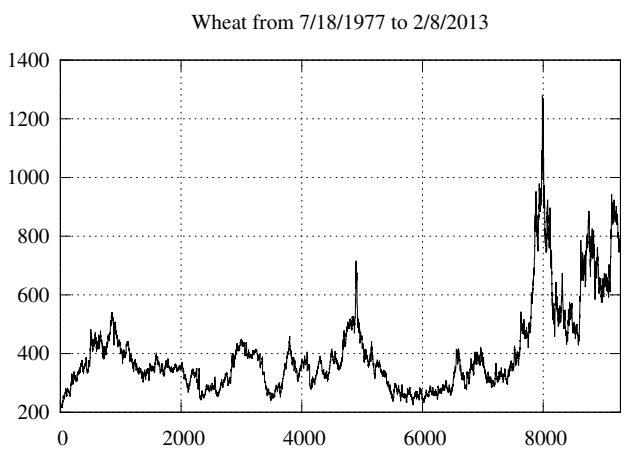


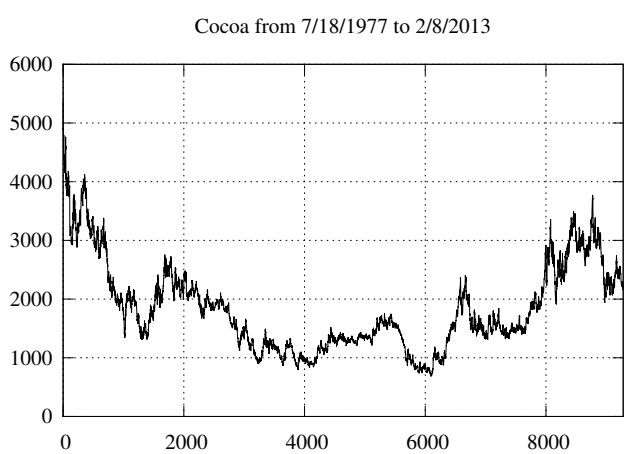












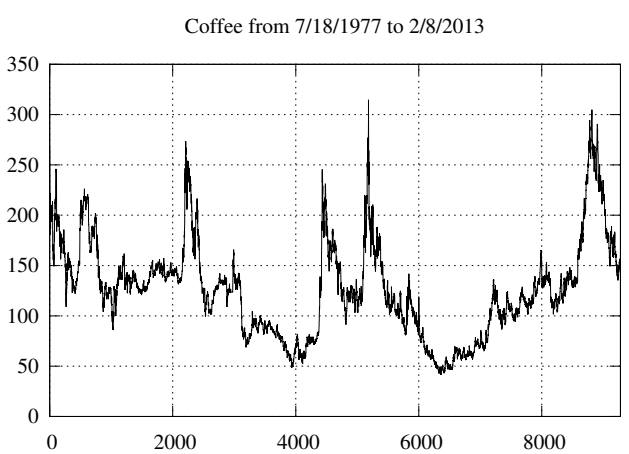


Figure 23: Plots of daily continuous contract futures price level series

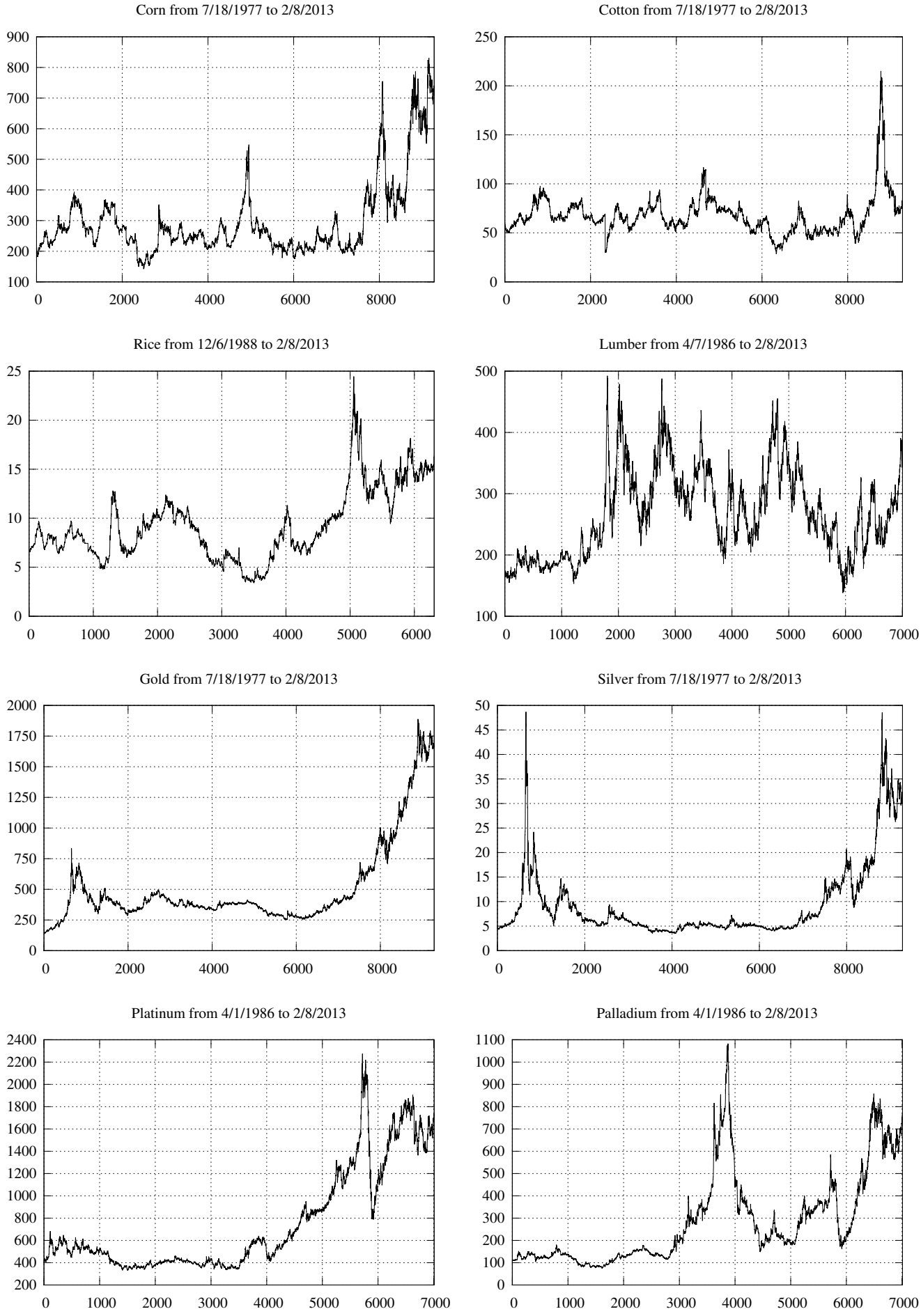


Figure 24: Plots of daily continuous contract futures price level series

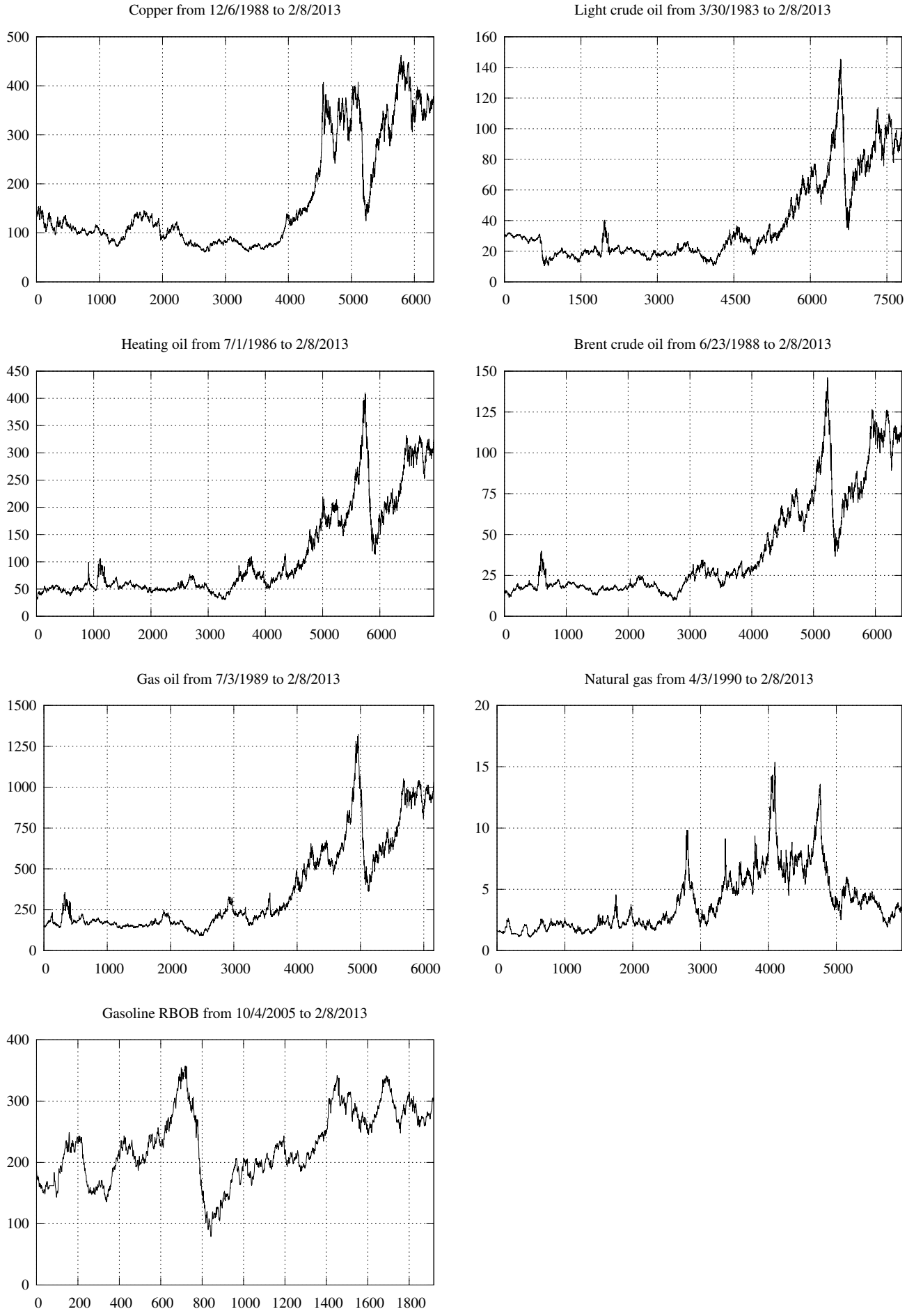


Figure 25: Plots of daily continuous contract futures price level series

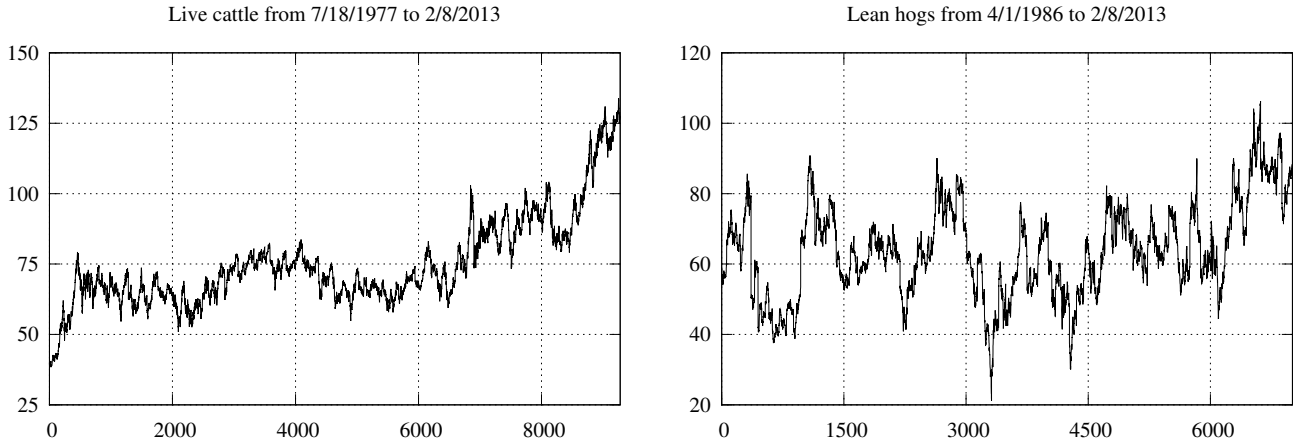


Figure 26: Histograms of daily continuous contract futures price level series, 100 bins

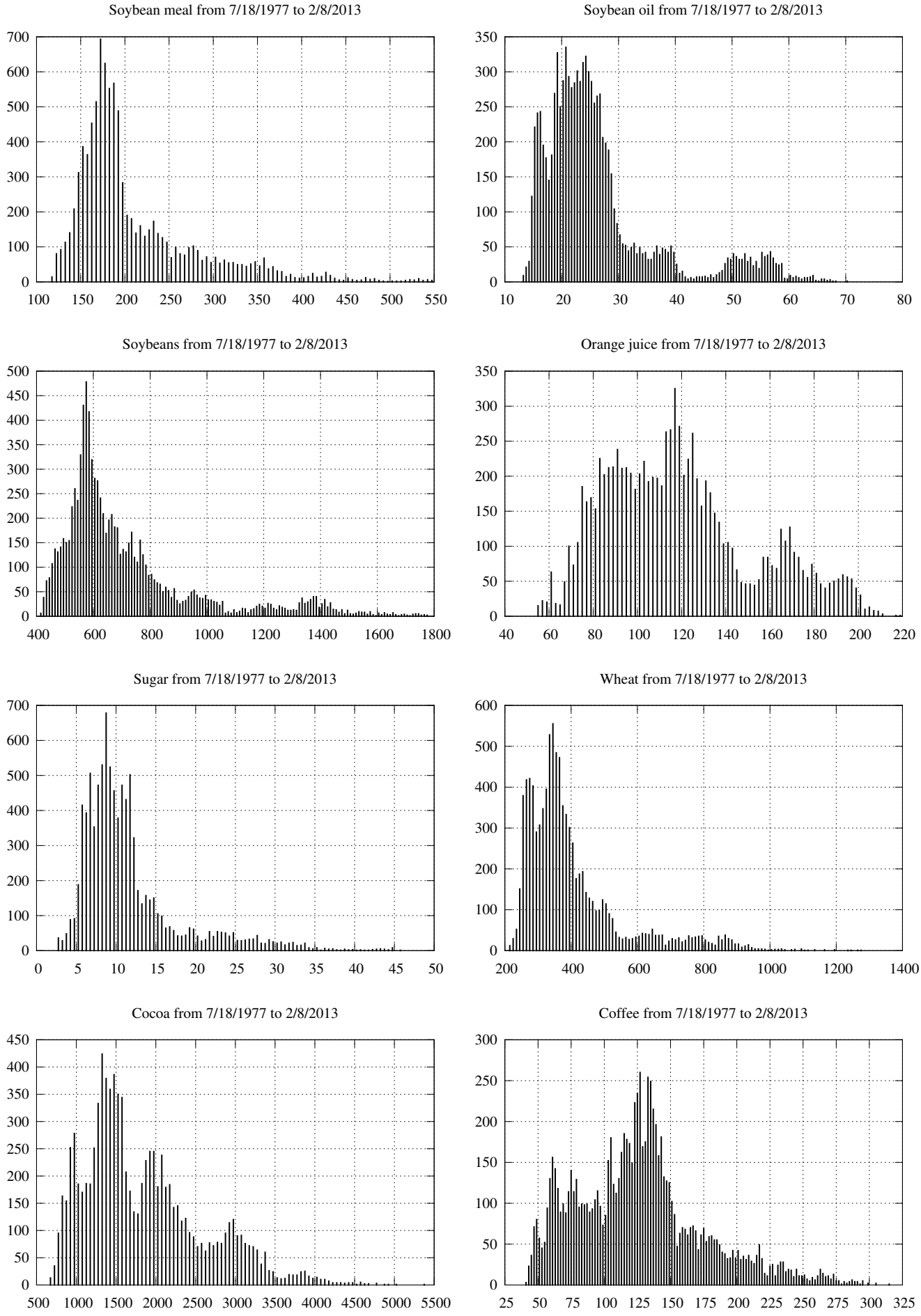


Figure 27: Histograms of daily continuous contract futures price level series, 100 bins

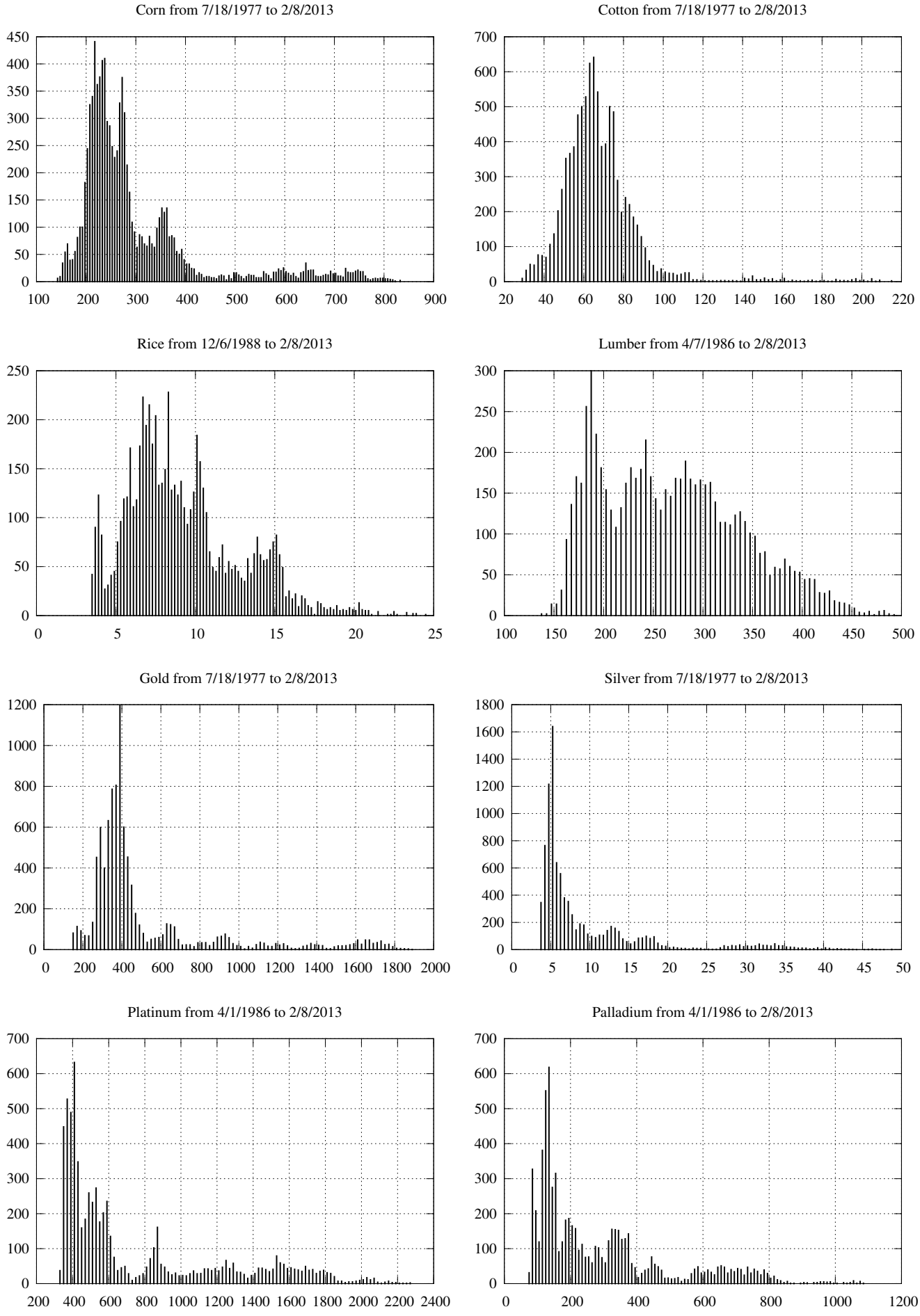
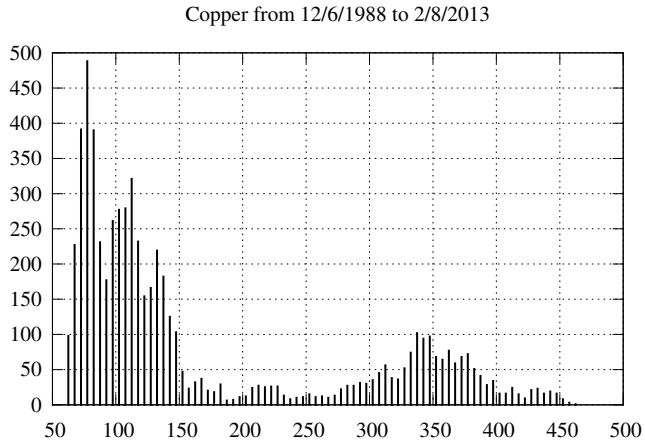
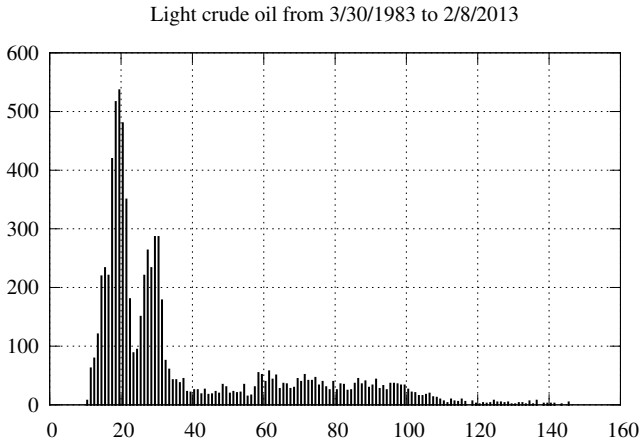
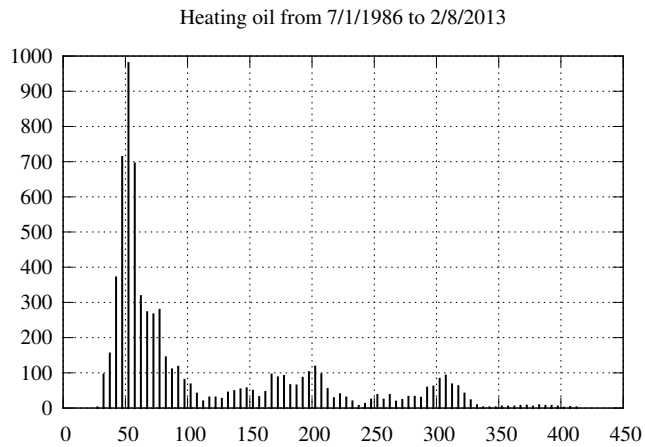
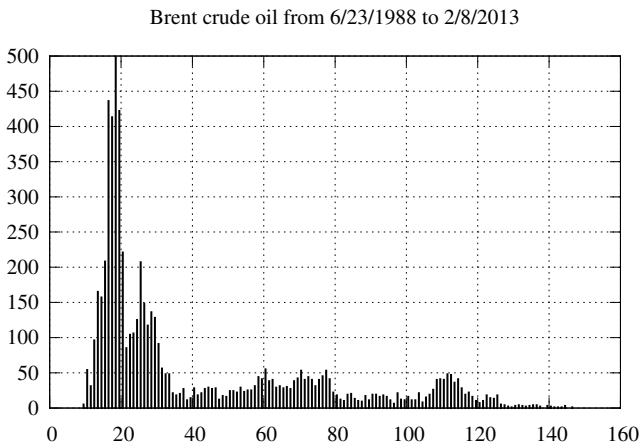


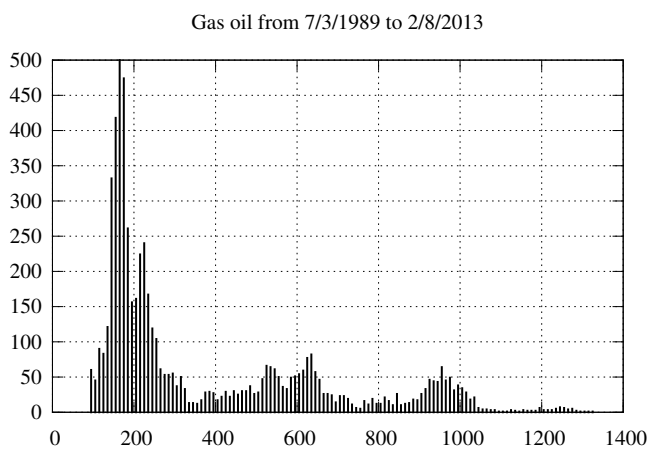
Figure 28: Histograms of daily continuous contract futures price level series, 100 bins

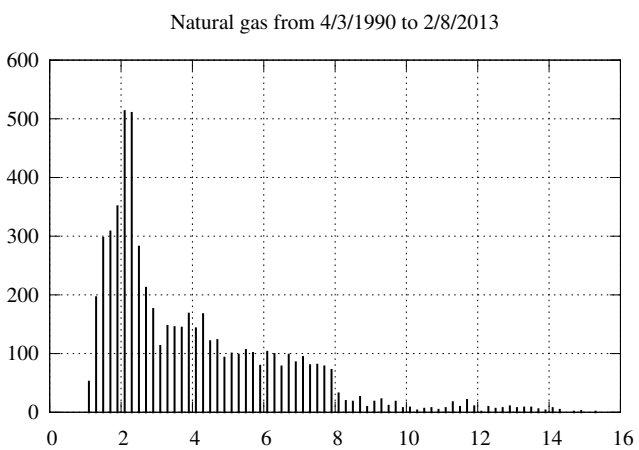












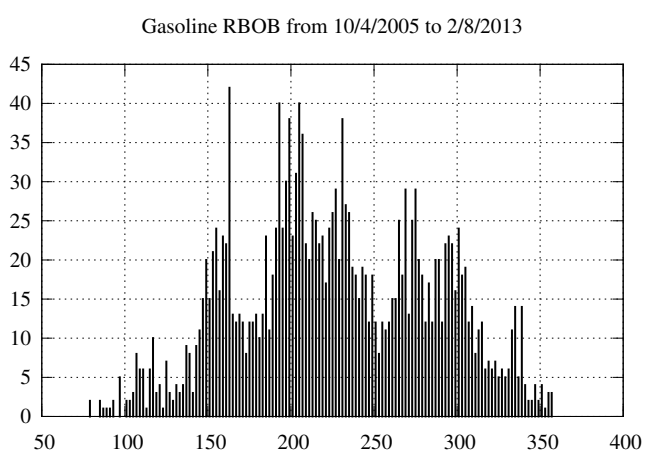
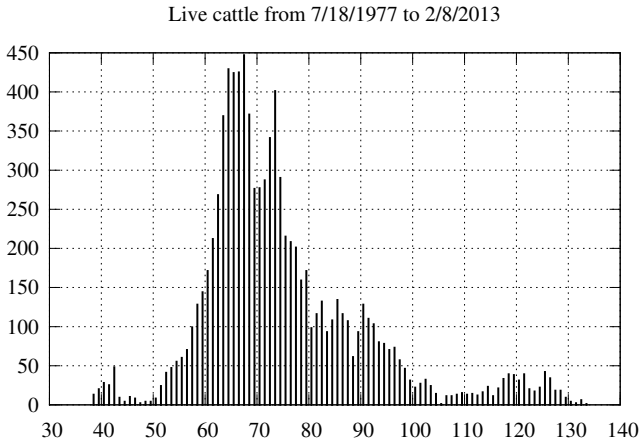
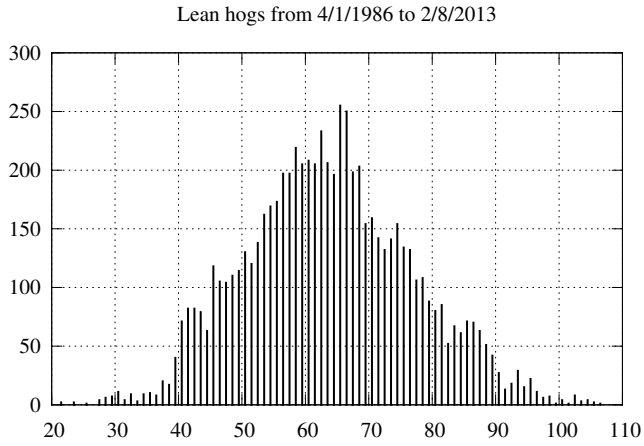


Figure 29: Histograms of daily continuous contract futures price level series, 100 bins





B References

Andrews B., David, R.A., Breidt, F.J. (2006) "Maximum likelihood estimation for all-pass time series models," *Journal of Multivariate Analysis,* 97, 1638-1659.

Barnett, W.A., Gallant, R.A., Hinich, M.J., Jungeilges, J.A., Kaplan, D.T, Jensen, M.J. (1996) "An experimental design to compare tests of nonlinearity and chaos," in *Nonlinear Dynamics and Economics: Proceedings of the Tenth International Symposium in Economic Theory and Econometrics*, eds. W.A. Barnett, A.P. Kirman, M. Salmon, 163-190, Cambridge University Press.

Black, F. (1976) "The pricing of commodity contracts," *The Journal of Financial Economics,* 3, 167-179.

Blanchard, O., Watson, M., (1982) "Bubbles, rational expectations and financial markets," Working paper No. 945. National Bureau of Economic Research.

Blank, S.C., (1991) "Chaos in futures markets? A nonlinear dynamical analysis," *The Journal of Futures Markets,* 11, 6, 711-728.

Bollerslev, T. (1986) "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics,* 31, 3, 307327.

Breidt, J., Davis, R., Lii, K., Rosenblatt, M. (1991) "Maximum likelihood estimation for non-causal autoregressive processes," *Journal of Multivariate Analysis*, 36, 175-198.

Breidt, J., Davis, R., Trindade, A. (2001) "Least absolute deviation estimation for all-pass time series models," *The Annals of Statistics,* 29, 4, 919-946.

Brillinger, D.R., Rosenblatt, M. (1967) "Asymptotic theory of estimates of k-th order spectra," in *Spectral Analysis of Time Series*, ed. B. Harris, 153-188, Wiley, New York.

Brock, W., Dechert, W.D., Scheinkman, J. (1987) "A test for independence based on the correlation dimension," working paper, Department of Economics, University of Wisconsin, Madison.

Burns, A.F., Mitchell, W.C. (1946) *Measuring Business Cycles,* Columbia University Press, New York.

Campbell, J.Y., Lo, A.W., MacKinlay, A.C. (1996) *The Econometrics of Financial Markets,* Princeton University Press, NJ.

Deaton, A., Laroque, G. (1996) "Competitive storage and commodity price dynamics," *Journal of political economy,* 104, 5, 896-923.

DeCoster, G.P., Labys, W.C., Mitchell, D.W., (1992) "Evidence of chaos in commodity futures prices," *The Journal of Futures Markets,* 12, 3, 291-305.

Dempster, A.P., Laird, N.M., Rubin, D.B. (1977) "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological),* 39, 1, 1-38.

Dusak, K. (1973) "Futures trading and investor returns: An investigation of commodity market risk premiums," *Journal of Political Economy,* 81, 1387-1406.

Eichenwald, K. (1989-12-21). "2 Hunts Fined And Banned From Trades," *New York Times online*, http://www.nytimes.com/1989/12/21/business/2-hunts-fined-and-banned-from-trades.html, Retrieved 2013-05-26.

Engle, R.F. (1982) "Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation", *Econometrica*, 50, 987-1008.

Evans, G. (1991) "Pitfalls in testing for explosive bubbles in asset prices," *The American Economic Review*, 81, 4, 922-930.

Fama, E.F., French, K.R. (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.

Frank, M., Stengos, T. (1989) "Measuring the strangeness of gold and silver rates of return," *The Review of Economic Studies,* 56, 4, 553-567.

Gettler, L. (2008-02-02). "Wake-up calls on rogue traders keep ringing, but who's answering the phone?" *The Age (Melbourne),* http://www.theage.com.au/business/wakeup-calls-on-rogue-traders-keep-ringing-but-whos-answering-the-phone-20080201-1plq.html, Retrieved 2013-

5-26.

Gordon, N. J., Salmond, D. J., Smith, A. F. M. (1993) "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEEE Proceedings F on Radar and Signal Processing,* 140, 2, 107-113.

Gourieroux, C., Jasiak, J. (2003) "Nonlinear innovations and impulse responses with application to VaR sensitivity," *Working paper, CREF 03-08.*

Gourieroux, C., Zakoian, J.M. (2012) "Explosive bubble modelling by noncausal Cauchy autoregressive process," *Working paper, CREST.*

Granger, C.W., Andersen, A.P. (1978) *An Introduction to Bilinear Time Series Models,* Vandenhoeck and Ruprecht, Gottingen.

Grassberger, P., Procaccia, I. (1983) "Measuring the strangeness of strange attractors," *Physica,* 9D, 189-208.

Hallin, M., Lefevre, C., Puri, M. (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, Sargent, T.J. (1991) "Two difficulties in interpreting vector autogressions," in *Rational Expectations Econometrics*, eds. Hansen, L.P., and Sargent, T.J., Westview Press Inc., Boulder, CO, 77-119.

Hinich, M.J. (1982) "Testing for Gaussianity and linearity of a stationary time series," *Journal of Time Series Analysis,* 3, 3.

Hinich, M.J. (1996) "Testing for dependence in the input to a linear time series model," *Journal of Nonparametric Statistics,* 6, 2-3.

Hinich, M.J. (2009) "Falsifying ARCH/GARCH models using bispectral based tests," *Communications in Statistics: Theory and Methods,* 38, 4, 529-541.

Hinich, M.J., Rothman, P. (1998) "Frequency-domain test of time reversibility," *Macroeconomic Dynamics,* 2, 72-88.

Hsieh, D.A., (1989) "Testing for nonlinear dependence in daily foreign exchange rates," *The Journal of Business*, 62, 3, 339-368.

Jones, M.C. (2010) "A skew-t distribution," in *Probability and Statistical Models with Applications,* eds. Charalambides, A., Koutras, M.V., and Balakrishnan, N., Chapman & Hall/CRC Press.

Lanne, M., Saikkonen, P. (2008) "Modeling expectations with noncausal autoregressions," *HECER Discussion paper 212.*

Lanne, M., Luoto, J., Saikkonen, P. (2010) "Optimal forecasting of noncausal autoregressive time series," *MPRA Paper 23648*, University Library of Munich, Germany.

Lanne, M., Nyberg, H., and Saarinen, E. (2011) "Forecasting U.S. macroeconomic and financial time series with noncausal and causal AR models: a comparison," *HECER Discussion paper 319.*

Ling, S., Li, D. (2008) "Asymptotic inference for a nonstationary double AR(1) model," *Biometrika,* 95, 257-263.

Lof, M. (2011) "Noncausality and asset pricing," *HECER Discussion paper 323.*

Marcellino, M., Stock, J.H., Watson, M.W. (2006) "A comparison of direct and iterated AR methods for forecasting macroeconomic time series," *Journal of Econometrics,* 135, 499-526.

McLeod, A.I., Li, W.K. (1983) "Diagnostic checking ARMA time series models using squared residuals autocorrelations," *Journal of Time Series Analysis,* 4, 269-273.

Neftci, S.N. (1984) "Are economic time series asymmetric over the business cycle," *Journal of Political Economy,* 92, 307-328.

Nelson, C.R., Plosser, C.I. (1982) "Trends and random walks in macroeconomic time series: some evidence and implications," *Journal of Monetary Economics,* 10, 139-162.

Pemberton, J., Tong, H. (1981) "A note on the distributions of nonlinear autoregressive stochastic models, *Journal of Time Series Analysis 2*, 1, 49-52.

Priestley, M.B. (1989) *Non-linear and Non-stationary Time Series Analysis,* Academic Press Ltd., London.

Ramsey J., Rothman P. (1996) "Time irreversibility and business cycle asymmetry," *Journal of*

Money and Banking, 28, 1-21.

Subba Rao, T., Gabr, M.M, (1984) "An introduction to bispectral analysis and bilinear time series models," in *Lecture Notes in Statistics,* eds. D. Brillinger et. al., Springer-Verlag.

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

Sharpe, W.F. (1964) "Capital asset prices: A theory of market equilibrium under conditions of risk," *Journal of Fiance*, 19, 3, 425-442.

Terasvirta, T. (1994) "Specification, estimation, and evaluation of smooth transition autoregressive models," *Journal of the American Statistical Association,* 89, 425, 208-218.

Tong, H., Lim, K.S. (1980) "Threshold autoregression, limit cycles, and cyclical data," *Journal of the Royal Statistical Society, Series B,* 42, 3, 245-292.

Weiss, G. (1975) "Time reversibility of linear stochastic processes," *Journal of Applied Probability,* 12, 831-836.

Working, H. (1949) "The theory of the price of storage," *American Economic Review,* 39, 1254-1262.

Yang, S.R., Brorsen, W. (1993) "Nonlinear dynamics of daily futures prices: conditional heteroskedasticity or chaos?," *The Journal of Futures Markets,* 13, 2, 175-191.