ESSAYS ON COMMODITY MARKETS

A thesis submitted in partial fulfilment of the requirements for the

Degree of

Doctor of Philosophy in Finance

in the University of Canterbury

by **Jan Koeman**

Supervised by: Prof. Jedrzej Bialkowski, University of Canterbury

Co-Supervised by: Prof. Henk Berkman, University of Auckland

Table of Contents

.

Acknowledgements

This thesis has been a fascinating journey into the world of commodities and finance.

First, I would like to thank Jedrzej Bialkowski for the opportunity to pursue a PhD at the University of Canterbury. I appreciate his insight and experience in finance research, his recommendations for suitable research topics, and his illustration of how to craft a publication quality research paper. In addition, I appreciate his efforts in helping me obtain a University of Canterbury Doctoral Scholarship at a time in my life when it was really helpful.

Second, I would like to thank Henk Berkman for being my co-supervisor and showing me how to take raw research findings and organize them into a coherent whole. Henk has a true genius for taking the research findings in a first draft of a research paper, distilling, and organizing the critical insights into a finished publication.

For Chapter 2, the data on the cocoa market and advice provided by Stephen, Matthew and Jeffrey Haws of KnowledgeCharts.com was invaluable, and impossible to obtain elsewhere.

Finally, my family – Linda, Laura, and Julia are my inspiration, and provide meaning to my work and life. I love them forever.

Essays On Commodities

General Introduction

Agricultural Commodities are a large portion of the New Zealand Economy, and a better understanding of commodity markets is useful for New Zealand market participants, especially in the dairy sector. This thesis studies two of the main functions of commodity markets, price discovery and hedging. The first chapter reports the effectiveness of hedging as a function of spot market design. The second chapter illustrates the reduced explanatory power of the Theory of Storage in recent time as contrasted with the early 1960's. The third chapter demonstrates that cutting-edge machine learning techniques are promising alternatives for learning features of commodity markets.

Hedging is one of the primary functions of commodity futures markets. In chapter one, the reasons for the success and failure of futures contracts are analyzed and the design of the spot settlement is examined in detail. The quality of the spot settlement index is characterized by several dimensions, and has a marked effect on the hedging effectiveness of the associated futures contract. The segmented dairy markets of the US and New Zealand are used to illustrate this conclusion.

The second paper is concerned with price discovery as a function of the inventory, the most important fundamental indicator in the majority of commodity markets[1]. The cocoa market is chosen as an illustration for three reasons. First, the cocoa market is one of the most idiosynchratic of commodity markets, with few links to other commodities other than coffee, which shares a similar geographic production region. Second, the cocoa market consists primarily of a single product, cocoa beans, rather than the multiple product streams of more complex commodity sectors like dairy products. It is easier to isolate the effect of inventory in the cocoa market than in markets like grain, oil, or dairy. Third, the Traditional Theory of Storage reached an apex in the work of Helmut Weymar (Weymar (1965), Weymar (1966) and Weymar (1968). After submitting his award-winning thesis[2] , and with partial initial funding from Paul Samuelson, Weymar formed Commodities Corporation in 1969. Commodities Corporation was highly profitable, and was the training ground for several of the most successful hedge fund pioneers: Paul Tudor Jones, Louis Bacon, and Bruce Kovner among others. To replicate Weymar's results, significant time was spent in researching historical archives including transcribing by hand ten years of weekly futures prices from the microfiche archives of the New York Times[3]. The model is then extended and applied to the modern period from 2009-2019, in which the model has substantially reduced explanatory power. The reduced power is attributed to the changes in the commodity markets, including the presence of large groups of momentum, and index traders. This hypothesis is examined in Chapter 3.

---

[1] Electricity is a notable counter-example, as electricity is difficult to store in any form other than reservoirs for hydro-power.

[2] Weymar's thesis was published by the MIT Press, and won the American Agricultural Economic Association's award for the best published research in 1968

[3] The New York Times Machine provides access to 150 years of papers from the famous publisher. https://timesmachine.nytimes.com/browser

Machine Learning is a new frontier for finance. Spectacular successes in speech and image recognition have encouraged enormous investment in applying similar models to understanding financial time series. The great breakthrough in machine learning occurred in 2014 in the ImageNet Large Scale Visual Recognition Challenge when AlexNet, a deep learning convolutional model created by Krishevsky, Sutskever, and Hinton (Krishevsky et al (2017)), achieved a top-five error rate of 15.1%, 10.8% better than the closest competitor. After AlexNet, a sequence of similar deep-learning architectures improved the performance to better than 95% in 2017, better than human-level performance. In 2017, IBM announced an error rate of 5.5% in speech recognition[4], also on par with human performance. The most recent breakthrough in December 2020, is in protein folding[5]. Hypothesizing from the research in Chapter 2 that the effects of momentum and index traders were having a large impact on the cocoa price, a Restricted Boltzmann Machine (RBM) is used in a recommender architecture to confirm the co-movement of groups of commodities from the commodity spectrum on the World Bank pink sheet. After verifying that co-movement is present, a more advanced Recursive Neural Network Restricted Boltzmann Machine (RNN_RBM) is used to make one-day-ahead cocoa price predictions from cocoa fundamental data and associated commodity price series. The one-day-ahead price predictions fail to outperform baseline forecasts. Possible reasons include the daily granularity being too fine, the size of the search space with

[4] Saon G, (2017). Reaching new records in speech recognition https://www.ibm.com/blogs/watson/2017/03/reaching-new-records-in-speech-recognition/
[5] Griffen A. (2020), AI solves 50-year-old science problem in 'stunning advance' that could dramatically change how we fight diseases, researchers say https://www.independent.co.uk/life-style/gadgets-and-tech/protein-folding-ai-deepmind-google-cancer-covid-b1764008.html

continuous variable inputs, and the absence of detailed worldwide inventory-ratio measurements or other relevant information. The results may also reflect the efficiency of modern commodity markets. We conclude that the RNN_RBM architecture would lend itself better to explaining rather than predicting the joint term structure of commodities. The RBM results illustrate that commodities are indeed interlinked and should be considered as a group rather than in isolation from one another.

Deputy Vice-Chancellor's Office
Postgraduate Research Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

Chapter 1 - Does the Design of Spot Markets Matter for the Success of Futures Markets? Evidence from Dairy Futures

Białkowski, J., & Koeman, J. (2018). Does the design of spot markets matter for the success of futures markets? Evidence from dairy futures. Journal of Futures Markets, 38(3), 373-389.

Please detail the nature and extent (%) of contribution by the candidate:

The candidate did all analytical and research work for the paper. Dr. Bialkowski provided research direction and commentary.

Certification by Co-authors:

If there is more than one co-author then a single co-author can sign on behalf of all

The undersigned certifies that:

- The above statement correctly reflects the nature and extent of the Doctoral candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: Prof *Jedrzej Bialkowski* Signature: J. Bialkowski Date: *31 Dec 2020*

Chapter 1

Does the Design of Spot Markets Matter for the Success of Futures Markets? Evidence from Dairy Futures

Abstract

This study provides evidence of the importance of a well-defined and functioning spot market for the success of the associated futures market. Our analysis of hedging effectiveness and hedge ratio persistence shows that none of the United States (US) spot market indices may be hedged effectively with the Chicago Mercantile Exchange nonfat dry milk futures at short hedging horizons, whereas the New Zealand (NZ) Stock Exchange whole milk powder futures contract is an effective hedge for the Global Dairy Trade spot pricing benchmark. Four important dimensions of spot market design are identified – timeliness, market-based measurement, forward-spot separation, and inclusiveness.

Citation

Białkowski, J., & Koeman, J. (2018). Does the design of spot markets matter for the success of futures markets? Evidence from dairy futures. Journal of Futures Markets, 38(3), 373-389.

1. Introduction

A number of newly introduced futures contracts have failed to attract substantial interest from market participants and their trading is characterized by low volume (Carlton, 1984; Black, 1986; Brorsen & Fofana, 2001). Several studies investigate the possible reasons behind the success or failure of exchange traded derivatives and in particular futures contracts (see Johnston and McConnell, 1989; Bialkowski & Jakubowski, 2012; Garcia et al., 2015; Till, 2014; Webb, 2015). Although past studies point out several features of futures and related cash markets that increase the chance of success, the topic is the subject of debate, and Bhardwaj, Gorton and Rouwenhorst (2015) argue that more research into the success and failure of futures contracts is needed. In this paper, we provide evidence that a previously unstudied aspect of futures markets – the underlying spot market index design – is a strong determinant of the hedging effectiveness of futures contracts and hedge ratio persistence over short to long hedging horizons.

Gray (1966) outlines three broad classes of reasons for the failure of futures contracts: poor design of the futures contract that favors either the buyer or the seller, the motivation to boycott a futures market because of the loss of pre-existing market power by either the buyer or seller, and the failure to attract speculation. In addition, Gray argues that the futures markets must serve a hedging function for commercial traders. Till (2014) and Webb (2015)[i] reinforce these considerations.

The majority of studies focus on the characteristics of the commodity, the salient features of the cash and futures market, and the aspects of the futures contract that are associated with success or failure. There is the implicit assumption that the cash market is structured to produce a single benchmark price to serve as the futures' underlying. Only a few studies, primarily in the shipping and freight markets, examine the characteristics of the cash market's underlying price index that promote success. In the case of these markets, the construction of price indices is necessary due to the wide range of product or service grades.

The contribution of this paper is threefold. First, it highlights the importance of the proper institutional design of the spot market to produce a benchmark with the necessary features to serve as the underlying for a successful futures contract. Second, the paper illustrates the trade-off between the settlement of futures contracts to a historical average and settlement to the spot price of the underlying. Third, this paper aims to serve as a source of information on the United States (US) and New Zealand (NZ) dairy futures and spot markets, a commodity market that has not received significant academic attention despite its importance and size. The analysis of the US and NZ futures contracts and their underlying spot markets illustrates that the design of spot markets strongly impacts the functioning of futures markets. In particular, the design of the spot market affects the hedging efficiency and hedge ratio persistence from short to long hedging horizons.

---

[i] Professor Robert Webb's keynote speech during the 2015 Derivatives Markets Conference in Auckland enumerated 10 characteristics that are related to the successful introduction of a futures contract: 1) price volatility in the cash market, 2) the need to hedge for commercial participants, 3) public order flow of genuine commercial (i.e. hedging) trades, 4) good contract design that does not favor the long or short side, 5) first mover advantage, 6) actively traded related futures that facilitate spread trading, 7) liquidity in comparison with existing cross-hedges, 8) low explicit trading costs (e.g. brokerage commissions), 9) speculator interest to take the long side of hedger trades, and 10) timing of the introduction of the contract.

The remainder of this paper is organized as follows. Section 2 enumerates the important dimensions of spot market design. Section 3 illustrates how the selection and design of the spot market influences the riskiness of arbitrage, hedging, and speculative activity. Section 4 provides an overview of the US and NZ dairy spot market indices, and formulates our research hypotheses. Section 5 presents the data examined in this study. Section 6 reports our methodology. Section 7 reports results, and section 8 concludes.

2. Spot Market Design Dimensions

Little academic research focuses on spot market design as a factor affecting the performance of futures markets; all studies make an implicit assumption that the spot market is structured in a competitive manner so as to produce a single cash price.[ii] This assumption is valid in the case of highly liquid underlyings with well-established mechanisms for single-price determination. The situation is different with less liquid assets traded in several locations with multiple price indices. Good examples of such assets are dairy products. In their case, the spot market was often designed shortly before a futures market was launched. In addition, government regulation has resulted in global market segmentation.

A few studies in the shipping, trucking, fishing, and forestry markets examine the design of the underlying cash market. These markets exhibit a wide range of product grades and quality, requiring the explicit construction of a single-price index. Kavussanos and Visvikis (2006) examine the maritime shipping industry and enumerate 10 characteristics that a cash market price index should exhibit – accuracy, absence of bias, familiar units, broad coverage,

[ii] Appendix 1 contains a detailed review of the factors that have been related to the success and failure of futures contracts.

frequent publication, auditability, low access cost, and the participation of major market participants. Bignell (2013) adds the ability to break down the index into separate sub-indices and the ability to update the index structure as market conditions change.

We identify four important dimensions of spot market design for producing a single-price index that can serve as an effective underlying for a futures contract: timeliness, market-based measurement, inclusiveness, and forward-spot separation. Timeliness measures the extent to which the information is current for price formation. Timeliness will be lower for indices that incorporate a range of historical information into the spot price or for indices that induce a delay between the price measurement and publication date. Hedging effectiveness is ultimately determined by the correlation between unexpected spot and futures price changes. At a point in time, the correlation between a single futures price and a range of spot prices will be lower than that between two single prices. Indices that use more current information are superior. The market-based characteristic indicates the extent to which the measure is determined by markets rather than surveys among market participants. Prices generated by financial and commodity markets are more accurate than survey prices and mandatory surveys are more accurate than voluntary surveys. Forward-spot separation indicates the separation of spot and forward market sales. A price index that mixes forward and spot sales, or that only provides spot or forward sales, will be less effective than a structure that provides both spot and forward price information. In markets for perishable commodities, it is often advantageous to have forward rather than immediate delivery. Inclusiveness assures that a significant representative portion of trades are included in the spot market index. A price index that accurately reflects the breadth of trading activity is superior for price formation to a thinly traded index. A secondary advantage of an inclusive price index is resistance to price manipulation.

The corn spot markets provide an illustrative example of effective spot market design. At any particular moment in time, it is possible to ascertain the price of corn at the nearest country elevator. The website http://www.agweb.com/markets/cash-grain-bids/ allows the entry of a zip code to immediately see the cash bids and basis levels for the closest five elevators. In addition, the daily settlement price is available from several resources, including IndexMundi.[iii] Finally, at a moment in time, it is possible to see the geographic corn basis for the entire United States.[iv] The hedging effectiveness of the corn futures markets has been reported from 74% to 80% (Sanders et al., 2003; Lien, 2008).

In our study of the US and NZ dairy markets, we provide evidence that the spot market design matters for the functioning of the associated futures market. In particular, we report higher hedging effectiveness and superior hedge ratio persistence for futures contracts with spot market indices that incorporate the above dimensions.

3. Implications of Settlement to the Average Spot Price

The setup of a spot market and method in which futures contracts are settled make hedging, speculation, and arbitrage more risky and complicated. The introduction of averaging across time into the settlement mechanism introduces distortions for speculators or participants arbitraging physical versus financial through to settlement. In addition, settlement averaging makes hedging more risky due to uncertainty about the basis.

[iii] Index Mundi, Commodity price for Maize, http://www.indexmundi.com/commodities/?commodity=corn
[iv] AG Manager, Corn basis map 6th of January 2016, http://www.agmanager.info/marketing/basis/maps/archives/2016/January/1/basismaps.asp?image=Basiscorn201601.jpg

Both US and NZ dairy futures use an average price for settlement. The average price is chosen to increase the validity of the settlement price.[v] More data points provide a more representative price for commodities that are thinly traded. A secondary possible reason for adopting an average price is an attempt to avoid manipulation of the spot price to gain from the futures settlement (Tashjian, 1995; Pirrong, 2001).

The Chicago Mercantile Exchange (CME) futures nonfat dry milk contract settles to a weighted average of the United States Department of Agriculture (USDA) Agriculture Marketing Service (AMS) National Dairy Product Sales Report (NDPSR) weekly announced prices. This index is a historical moving average, with sales included from up to 60 days prior to settlement.[vi] In New Zealand, the settlement calculation is also to the average of two biweekly auctions within the settlement month, but the convergence problem is partly mitigated by the second auction being the day before settlement and the single source of price from the Global Dairy Trade (GDT) auction platform.

Settling to a historical average increases the basis risk for hedgers, introduces risk into arbitrage trades, and requires speculators to consider the relative movement of both futures and spot prices during the hedge lifetime. First, we show that in the case of futures contracts settled to a historical average instead of the spot price at expiration, the basis is different than zero. Assume that a trade was initiated at time 0, the price of the underlying is S_0, and the price of the underlying on the spot market at time 1 is S_1. When settling to a historical average of the prices at time *0* and *1,* the settlement price is equal to $(S_0 + S_1)/2$. In this case, the basis is not 0 but $S_1 - (S_0 + S_1)/2 = (S_1 - S_0)/2$. In this model, the basis at settlement in the

[v] From a discussion with commercial traders.

[vi] Appendix 2 contains a list of dairy market acronyms

case of a futures contract with settlement to a historical average is equal to half the change in the spot price – a substantial difference from zero.

In the Chicago Mercantile Exchange nonfat dry milk (NFDM) futures market, the settlement price is a volume-weighted average price, not a simple arithmetic average as in the above example. The average is calculated from four or five separate weekly numbers. As with the above example, one would expect a non-zero basis at expiration. Cohen and Gorham (1985) note that with the feeder cattle contract, the basis risk at settlement will be non-zero, as the cash settlement price is an average over time, space, and grade (also see Kenyon et al., 1991). The non-zero basis at maturity leads to greater uncertainty in basis changes, and lower hedging effectiveness. Perversi, Feuz, and Umberger (2002) identify an unpredictable basis as the major cause of failure of the cattle stocker contract. Paul et al. (1981) argue that the failure of the Maine potato contract was attributable to a lack of convergence of the potato spot and futures prices at contract maturity.

A non-zero basis at the maturity of a contract directly affects arbitrageurs and speculators. In the case of arbitrage implied by the cost-of-carry formula, an investor is expected to make a profit equal to the absolute value of the difference between the futures price and the theoretical price if the position is kept open until maturity. Past studies show that such arbitrage is not risk free (see Kawaller, 1987; McMillan & Ülkü, 2009; Nam et al., 2010). Nevertheless, settlement to an average price increases the risk. The profit from arbitrage is equal to the difference between the futures price and the theoretical price plus the basis at contract expiration. From the perspective of a futures market speculator who bets on the direction of price movements between opening the position and the contract maturity, settlement to an average price makes the trade more complex. The potential profit or loss depends not only on the ability to predict the price at contract maturity, as with settlement to a point-in-time spot price, but also on the ability to predict prices at points taken as input for the average. For example, in the case of a simple arithmetic average calculated from the price at time 0 and 1 and a speculative short position open at time 0, the profit from speculation is higher when the spot price at time 1 is less than the spot price at 0. The speculator needs to predict the price path of the commodity rather than only the prices at contract settlement.

Hedging using futures with settlement to a historical average is more risky due to a higher basis risk. The cost of the asset or sale price for a hedged position is equal to the futures price at time 0, F_0, plus the *basis,* where the *basis* is measured at the time of closing the hedging position. As a result of the non-zero basis at maturity for futures with settlement to the average of past prices, the basis will be higher if the hedging position is closed near expiration.

To summarize, there are costs associated with settlement to the historical mean of past spot prices. Arbitrage is more risky, hedging is less certain, and speculation requires prediction of the price path rather than only the price at closing. In the US and NZ dairy markets, average value settlement results in less than optimal hedging, but due to the single source of price in the NZ spot market, the reduction in overall hedging effectiveness is expected to be substantially smaller.

4. Dairy Markets in the United States and New Zealand

In this section, we outline the structure of and highlight the differences between the dairy spot and futures markets in the United States and New Zealand. Dairy market futures are used to illustrate the importance of proper spot market design for the functioning of associated futures markets. In particular, we provide the detailed structure of each of the possible spot price indices in the context of the relevant dimensions of spot market design.

4.1. The Global Production and Processing of Milk

Over 735 billion liters of cow's milk are produced annually worldwide. The global dairy sector is a $330 billion market and significantly larger than the $100 billion worldwide coffee market. Coffee is traditionally considered the second largest commodity after oil. The $30 billion export market for dairy ingredients[vii] is approximately the same size as the coffee export market. Dairy products made from cow's milk are a dietary staple in advanced western economies, and consumption is rising quickly in China and other rapidly developing nations. The United States produces approximately 26% and New Zealand 6% of raw milk globally. NZ exports exceed 90% of milk in the form of dairy ingredients, while the US exports approximately 14%.[viii]

[vii] The Global Dairy Industry, Fonterra, http://www2.fonterra.com/our-financials/the-global-dairy-industry
[viii] US Dairy Export Council, US Export Data 2015, http://www.usdec.org/research-and-data/market-data/us-export-data

Raw milk is produced by individual farmers, and then either marketed as beverage milk or processed into a variety of longer-shelf-life commodity dairy ingredients. The processing of raw milk into beverage milk or dairy commodities is accomplished by large farmer cooperatives or independent milk processors and handlers. The beverage milk and dairy ingredients are then purchased and marketed by large multinational companies or retail chains. The main dairy ingredients produced in the United States are nonfat dry milk, cheese, whey, and butter. In New Zealand, whole milk powder is also produced in significant quantities. Approximately 70% of US and 90% of NZ raw milk production is manufactured into dairy commodities.[ix]

An important problem in dairy markets is the compensation of individual farmers for raw milk production by the processors and handlers of raw milk. This problem arises from the perishability of raw milk – large tanker trucks collect and process milk in an industrial scale process, and the farmers are paid from the sales of manufactured products. The United States and New Zealand have taken different approaches to the problem; in the former, there is substantial government involvement, while the latter favors a free market approach.

[ix] USDA, Fluid milk sales by product, 1975–2014 (millions of pounds), http://www.ers.usda.gov/data-products/dairy-data.aspx

4.2. The New Zealand Market

4.2.1. NZ Government Regulation

In contrast to the United States, the NZ government does not intervene significantly in the dairy marketplace. The largest dairy cooperative – Fonterra – has approximately a 90% market share in purchasing milk from farmers, and uses reference prices from the Global Dairy Trade auction platform.[x] The NZ Commerce Commission conducts an annual review of the methodology utilized by Fonterra to calculate the farmgate milk price ultimately paid to farmer producers. Unlike in the United States, the NZ government is not involved in the publication of reference commodity prices.

4.2.2. NZ Spot and Futures Markets

Virtually all milk collected in New Zealand is processed for export into whole milk powder, skim milk powder, cheese, butter, whey, and a few other minor dairy ingredients. The bulk of forward ingredient sales is done in the over-the-counter market, directly with Fonterra and other cooperatives. However, approximately 30% of Fonterra's output is auctioned biweekly on Global Dairy Trade,[xi] an auction platform explicitly designed for the forward selling of dairy ingredients to the worldwide marketplace.[xii] GDT also sells ingredients from several other large dairy processors, including DairyAmerica of the United States, and ARLA of

---

[x] Fonterra, Milk Price – The Facts, https://www2.fonterra.com/files/financial-docs/milk-price-methodology/Milk+Price+Questions+and+Answers+1+Aug+2011.pdf

[xi] Global Dairy Trade Profile, Fonterra, https://www.globaldairytrade.info/assets/Uploads/resources/GDT-Profile-2016.pdf

[xii] Fonterra, Submission on the Base Milk Price Calculation, https://www2.fonterra.com/files/financial-docs/industry-regulations/commerce-commission/Fonterra+Submission+on+Key+Issues+-+Review+of+Milk+Price+Calc+2012-13.pdf

Sweden. GDT was formed in 2010, and holds auctions every two weeks for six different forward contracts. Contract 1 is for delivery in the next calendar month, Contract 2 for delivery in the second calendar month, and so on. The most liquid of these contracts is the second forward delivery month (Contract 2). Contract 1 is often thinly traded. A sufficiently large volume of dairy ingredients is sold on GDT to make it the key source of benchmark prices for internationally traded dairy ingredients. Prices in a weekly survey by Agrifax of producers in New Zealand closely follow the GDT prices.

The farmgate milk price paid to farmers by Fonterra is determined from the sales of dairy ingredients on GDT, less reasonable costs. Note that the calculation of the farmgate milk price is not connected in any way with the current spot pricing of dairy ingredients – it is calculated post sales.

In contrast to the United States, the GDT spot market benchmark is a two-month forward-looking price. The GDT prices are an accurate reflection of the price of dairy ingredients in the marketplace. It is important to note that only a marketplace can give current pricing information – collecting an average of forward sales induces a delay in the spot price determination process.

The New Zealand Stock Exchange (NZX) launched whole milk powder (WMP), skim milk powder, and anhydrous milk fat futures in 2010. Open interest in whole milk powder has grown at a rate of approximately 50% per year from 2012 to 2017, the hallmark of a successful futures contract introduction. The NZX dairy futures started trading shortly after GDT became operational. The whole milk powder and skim milk powder futures settle to an average of the last two GDT Contract 2 auctions, with the last futures trading day falling on the day before the second GDT auction.

4.3. The US Market

Before discussing the US dairy market in detail, we would like to highlight one of the important features of both dairy spot and futures markets. By convention, the underlying asset for a futures contract is a price series in the spot or cash market for immediate delivery of a commodity. However, in the case of the United States, the primary spot benchmark – the AMS National Dairy Product Sales Report of nonfat dry milk survey data – includes historical information from up to two months prior and is backwards looking. In contrast, the primary NZ spot benchmark – the Global Dairy Trade Whole Milk Powder Auction Contract 2 – has elements of a forward contract.

4.3.1. US Legislation

The US milk pricing regulations are intricate in nature and arose from a perceived historical need to increase the production of beverage milk relative to manufactured milk products and to support dairy prices (Erba and Novakovic, 1995). The legislation accomplishes this objective by surveying the prices of four basic commodities – nonfat dry milk, cheese, butter, and whey on a weekly basis. The surveyed prices are then used to stipulate the ultimate payment to farmers for raw milk according to usage. There are several laws that regulate the dairy industry in the United States. The most important are the Federal Milk Marketing

Orders, the Dairy Product Price Support Program,[xiii] and the Dairy Import Tariff Rate Quotas.[xiv]

The milk marketing orders[xv] govern 10 separate geographic regions – Pacific Northwest, Arizona, Upper Midwest, Central, Southwest, Southeast, Mideast, Appalachian, Northeast, and Florida. The orders regulate dairy processors and handlers, and establish the price which farmers will be paid for raw milk. California has a similar system with northern and southern marketing areas.

According to the USDA, the marketing orders have the following objectives: "(1) assure consumers of an adequate supply of wholesome (fluid) milk at a reasonable price; (2) promote greater producer price stability and orderly marketing; and (3) provide adequate producer prices to ensure an adequate current and future Grade A milk supply" (Jesse and Cropp, 2008).

Each of the 10 marketing orders classifies milk into four different categories depending on utilization: Class I – beverage milk; Class II – milk used for soft manufactured products, including yogurt, cream, and cottage cheese; Class III – milk used for hard cheeses and cream cheese; and Class IV – milk used for dry milk products and butter. California has a similar system, but with five milk classes.[xvii]

[xiii] The DPPSP was repealed in the 2014 Farm Bill, and replaced with the Margin Protection Program (MPP), a voluntary participation program which offers insurance against low milk-feed margins.
[xiv] Randy Schnepf, Dairy Provisions in the 2014 Farm Bill, https://www.hsdl.org/?view&did=752261
[xv] USDA, Federal Milk Marketing Orders, https://www.ams.usda.gov/rules-regulations/moa/dairy
[xvii] California Department of Food and Agriculture, Milk Pricing in California, https://www.cdfa.ca.gov/dairy/pdf/Milk_Pricing_in_CA.pdf

The prices for the four classes of milk are calculated from formulas based on the surveyed prices for four commodity products manufactured from milk: nonfat dry milk, cheese, whey, and butter. The class formulas establish minimum prices for the four classes of milk[xviii] based on the surveyed prices of the four commodities, and are weighted to increase the production of beverage milk. Milk handlers and processors must pay these minimum prices into a pool that is used to compensate farmers. Farmers within a marketing order region receive uniform raw milk component prices funded by the pool payments.

Before being discontinued in 2014, the price support program set market floor prices for nonfat dry milk, butter, and cheese. The Commodity Credit Corporation bought and placed stock into inventory when prices fell before the floor levels. For example, before 2014, "through its support price program, the U.S. government agreed to buy dairy commodities at a minimum level (cwt basis) – $1.13 for block cheese, $1.10 for barrel cheese, $1.05 for butter, $.80 for non-fortified nonfat dry milk and $.81 for fortified nonfat dry milk."[xix] The Dairy Import Tariffs made the import of competitive foreign milk powder products more difficult. One effect of these combined regulations before 1990 was to cause over-production and government storage of dairy products (Erba and Novakovic, 1995). Since 1990, the market price of commodities has, in general, been greater than the minimum prices.

Shipments of the four basic commodities are surveyed weekly by the Agriculture Marketing Service, and the price average is published in the National Dairy Product Sales Report on the following Wednesday. The shipments include sales in the prior 30 days. Thus, the weekly announced price for nonfat dry milk is a backward-looking price for milk powder sold in the

---

[xviii] International Dairy Foods Association, How Raw Milk Is Priced, http://www.idfa.org/news-views/media-kits/milk/how-farm-milk-is-priced
[xix] CME Group, An Introduction to Trading Dairy Futures and Options, http://www.hedgebroker.com/documents/education/CME_IntroTradingDairyFuturesAndOptions.pdf

previous month for delivery. The class prices are based on volume-weighted averages of the weekly commodity surveys. For example, the Class IV price for May 2016, published before the 5th of June 2016, is based on the volume-weighted weekly commodity averages for May.

The purpose of the government survey and publication of weekly prices for the four basic dairy commodities is to construct class prices and ultimately determine the price that farmers will get paid for raw milk components. However, an unintended consequence of the government publication of reference settlement prices and the price support program may be to discourage the formation of suitable spot market price indices for nonfat dry milk, cheese, butter, and dry whey. Milk processors, in the absence of a timely, inclusive, and market-based index for nonfat dry milk, used the weekly backward-looking surveys as the starting point for pricing current spot sales until 2016.[xx] In 2016, milk processors shifted to pricing spot sales from last week's weekly CME average nonfat dry milk spot price.

4.3.2. US Spot Market Price Indices

Processors and handlers of milk can choose from several pricing indices for the four base products. Table 1 lists the four possible spot price indices for nonfat dry milk.

[INSERT TABLE 1 ABOUT HERE]

Spot sales priced as a differential to the weekly National Dairy Product Sales Report survey prices incorporate historical price movements from up to two months prior. Commercial sales based on last weeks average of CME spot sales are also backward looking, but to a lesser

[xx] California Dairy Information Bulletin November 2003 and MCT Dairies, Intricacies of NFDM Pricing, February 27, 2015

extent. The weekly average is used because the CME spot index is thinly traded, with less than 6% of market volume. The California index is similar in construction to the NDPSR survey.

The Dairy Market News (DMN) Surveys[xxi] are a weekly average of voluntarily provided spot sales for three geographical regions:[xxii] East, Central, and West. Dairy Market News provides a range of prices for low-, medium-, and high-heat nonfat dry milk, as well as the "mostly" price range.[xxiii] The mostly price range includes "most" commercial transactions. These indices are more timely than the NDPSR index, as the information is only one week old. However, the information is a week old, based on voluntary participation, from different geographical regions, and comes in a range with significant variance rather than a single price. Furthermore, the Dairy Market News weekly reports are descriptive in character rather than a market-based measurement of the prevailing nonfat dry milk price. By the mid-2000s, many sellers of nonfat dry milk discontinued using the midpoint of the "mostly" ranges for their pricing index in favor of the NASS[xxiv] price. Both buyers and sellers of nonfat dry milk appreciated the transparency in the NASS prices and associated volume.[xxv] The advantage of the NDPSR commodity price is the standardized reporting mechanism – all processors with production above 1 million pounds per year are required to report. However, this transparency comes with the cost of extensive historical price information being included in the definition of the nonfat dry milk price.

[xxi] The Dairy Market News is published by the USDA Agricultural Marketing Service. https://www.ams.usda.gov/market-news/dairy

[xxii] USDA AMS, Dairy Market News, Agricultural Marketing Service, Dairy Market News Guidelines, https://www.ams.usda.gov/sites/default/files/media/DMN%20Working%20Guidelines.pdf

[xxiii] USDA AMS, Dairy Market News, Nonfat Dry Milk – Central and East https://www.ams.usda.gov/mnreports/md_da650.txt

[xxiv] Until April 2012, weekly survey prices were published by the National Agricultural Statistics Service. Since April 2012, the weekly survey prices for NFDM milk have been published by the USDA Agricultural Marketing Service, Market Information Branch in the National Dairy Product Sales Report.

[xxv] MCT Dairies, Intricacies of NFDM Pricing, http://www.mctdairies.com/Compass/2015/MCT-Dairies-Compass-2015-02.pdf

The NDPSR and California weekly survey averages include forward sales up to 30 days prior, and the Dairy Market News surveys and CME prices are based on spot sales. The forward contracting order flow is opaque in the US market. In contrast, NZ-based GDT provides public timely order flow up to six months in advance.

The following cash-settled futures contracts are traded on the CME in 2017: Class III Milk Futures, Class IV Milk Futures, Nonfat Dry Milk Futures, Dry Whey Futures, Butter Futures, and Cheese Futures. All of the CME futures contracts are settled to the monthly announced class or commodity prices published by the USDA Agricultural Marketing Service.

Table 2 shows the extent to which timeliness, market-based measurement, inclusiveness, and forward-spot separation are present in the US and NZ spot indices for dairy products. From the table, it is clear that only GDT available on the NZ market has the required properties of a representative spot market. Given the nature of the spot market indices in the United States and New Zealand and considering the implication of average price settlement, we formulate the following research hypotheses. One, the hedging effectiveness of NZX dairy futures will be higher than CME nonfat dry milk futures. Two, spot market indices that are correctly designed across several dimensions will have higher hedging effectiveness and better hedge ratio persistence from short to long hedge horizons than indices with few correct dimensions. These hypotheses are examined in section 7.

[INSERT TABLE 2 ABOUT HERE]

5. Data

In this section, we present the sources of data for hedging efficiency. For the NZ market, we examine the hedging of three separate spot price indices with NZX whole milk powder futures: Agrifax whole milk powder survey prices, Global Dairy Trade biweekly auctions, and a modified GDT price series. Agrifax surveys large milk processors in New Zealand on a weekly basis. The reported survey prices are weekly averages of whole milk powder and skim milk powder contract sales for delivery in two months; the time period covers June 2013 to May 2017. Global Dairy Trade auctions are biweekly auctions for six forward contracts. The time period covers June 2013 to May 2017.

NZX dairy futures for whole milk powder and skim milk powder trade daily with 18 monthly expiration dates. We construct a near-month series changing contracts on the last contract trading day. The Agrifax survey and GDT auction prices are compared with the NZX futures prices on the same date. A synthetic GDT series – GDT-Modified – is also constructed. The GDT-Modified time series is modified by replacing each second monthly auction price with the mean of the two auction prices within the month. The GDT-Modified series tracks the average price of whole milk powder to match the underlying implied in the settlement calculation in the futures contract. In addition, the GDT-Modified series starts in June 2013, when the NZX settlement calculation was changed to the average of the two within-month auctions. Prior to that date, the settlement average was constructed from the last auction of the settlement month and the first auction of the succeeding month.

For the US market, we examine the hedging of four separate nonfat dry milk spot indices with CME nonfat dry milk futures: NDPSR survey prices, California survey prices, Dairy

Market News survey prices, and CME spot prices. CME nonfat dry milk futures trade daily for contract expirations up to 24 months. We use the near-month series from Bloomberg. CME Spot nonfat dry milk trades daily with delivery in six days. The CME spot series runs from April 2012 to May 2017.

The NDPSR weekly survey prices are compiled by the Agricultural Marketing Service of the USDA, and include all nonfat dry milk sales shipped in the previous week. The NDPSR weekly prices are announced the following Wednesday. The data series starts in April 2012, when announcements were shifted to Wednesday from Friday,[xxvi] and ends in May 2017. California weekly survey prices are compiled by the California Department of Food and Agriculture. The California prices are announced the following Friday. The series runs from April 2012 to May 2017. The Dairy Market News survey prices are also gathered by the Agricultural Marketing Service, and include spot sales for the current week. The Dairy Market News prices are published on the Thursday of the relevant week. The series runs from April 2012 to May 2017.

6. Methodology

In order to test our hypothesis, we examine the efficiency of hedging using methodology applied by Adams and Gerner (2012) and developed by Herbst et al. (1989), Ghosh (1993), and Lien (2002). First, the futures and spot prices are tested both in levels and first differences for stationary behavior. If the levels are non-stationary, but the returns are stationary, then a regression is performed of log spot prices on log future prices, and the

[xxvi] In April 2012, the reporting organization was changed from the National Agricultural Statistics Service to the USDA Agricultural Marketing Service, Market Information Branch.

residuals are tested for stationary behavior using the augmented Dickey-Fuller test (Dickey and Fuller, 1986). If the residuals are stationary, an error correction model (ECM) (Ghosh, 1993) can be estimated using OLS with the parameters from the first regression:

$$\Delta \log S_t = c + \beta \Delta \log F_t + \sum_k \gamma_k \Delta \log F_{t-k} + \sum_l \delta_l \Delta \log S_{t-l} + \lambda e_{t-1} + \varepsilon_t, \tag{1}$$

where $\Delta \log S_t, \Delta \log F_t$ are the change in log spot and futures prices; $\Delta \log F_{t-k}$ are the lagged log future price changes from the same contract; $\Delta \log S_{t-l}$ are the lagged spot price changes; γ_k, δ_l are the short-term autocorrelation coefficients; λ, e_{t-1} are the error correction coefficient and term; and ε_t are the innovations. The error correction term is calculated as

$$e_t = \log S_t - a - b \log F_t, , \tag{2}$$

where *a* and *b* are the coefficients from the original cointegration test regression between the *log(*S_t*)* and *log(*F_t*)* price-level series. The ECM also includes several lags of both the spot and future returns, thus allowing for short-term serial correlation in the future and spot return series. Hedging effectiveness is determined by the adjusted R^2 of the statistical model. In addition, hedge ratios closer to unity indicate strong co-movement between spot and futures prices.

7. Results

In this section, we examine futures hedging efficiency in the NZ and US markets. In the NZ market, the Agrifax, GDT, and modified GDT whole milk powder indices are hedged with

NZX dairy futures contracts. In the US market, the NDPSR, California, Dairy Market News, and CME spot nonfat dry milk indices are hedged using CME nonfat dry milk futures contracts.

As the hedging horizon lengthens, hedge effectiveness increases for both OLS and cointegration models (Juhl et al., 2012). However, there is substantial liquidity in the near-month contract and a stack and roll implementation is exposed to shorter contracts at the beginning and end of the hedge. For example, a hedge implemented from March 15th to April 25th would be exposed to a short-term contract from March 15th to the rollover date, and from the rollover date to April 25th. This may result in price risk if the short-term hedge correlations are low. In addition, if the shorter horizon hedge ratio estimates are substantially different from the longer horizon hedge ratio estimates, the stack and roll strategy may be less than optimal. Finally, futures markets' participants will utilize a variety of hedge intervals, and hedges should be effective both at short- and longer-term horizons.

Table 3 reports the hedging effectiveness as a function of the hedge horizon using an OLS regression model All spot market indices show an increase in hedging effectiveness as the horizon increases. The Agrifax time series reports the best effectiveness for hedging up to three weeks, and the Global Dairy Trade Modified benchmark is the most effective for time scales from four to eight weeks. The superiority of the NZ indices is reduced as the time horizon is increased.

[INSERT TABLE 3 ABOUT HERE]

Table 4 reports the hedging effectiveness as a function of the hedge horizon using the cointegration regression model. All spot market indices show an increase in hedging effectiveness as the horizon increases. The Agrifax time series reports the best effectiveness for hedging at a one-week horizon (70%). In the US market, the highest hedging effectiveness at a one-week horizon is 52% for NDPSR survey prices. Only hog (13%), cotton (32%), and silver (54%) futures have reported lower hedging effectiveness in the past (see Lien, 2008). CME futures offer an ineffective hedge for any of the price series at a one-week horizon, in contrast to the NZX futures which effectively hedge the Agrifax index. At two-week and longer horizon, the hedging effectiveness of the Agrifax, GDT-Modified, and NDPSR indices are similar (77–78%). However, the only index that provides both high hedging effectiveness and a stable estimate of the hedge ratio for both short and long horizons is the GDT-Modified benchmark.

[INSERT TABLE 4 ABOUT HERE]

The GDT-Modified time series incorporates all the relevant dimensions of spot market design, and tracks the underlying average price implied by the settlement calculation in the futures contract. The hedging effectiveness ranges from 77% at a two-week horizon to 98% at an eight-week horizon. The hedge ratios are close to 1.0 across all hedge horizons. However, short-term hedging performance would improve if the settlement calculation in the NZX futures contract were amended to settle to only the second GDT auction price. Settling to an average of the two prices distorts the incentives for both arbitrageurs and speculators. The short-term hedging effectiveness of the NZX whole milk powder futures compares favorably with that for corn (74% to 80%) (see Sanders et al., 2003; Lien, 2008).

The Agrifax hedge effectiveness ranges from 70% at a one-week horizon to 93% at an eight-week horizon, and the hedge ratio varies from .58 to .80. Although the Agrifax prices have a similar hedging effectiveness to the GDT benchmark at horizons of two-weeks or greater, the hedge ratio is not stable as the hedge horizon changes. This results from the index being survey based, and incorporating week-old information. The hedging effectiveness is still relatively high because the GDT auctions provide an effective single-source price indicator, and the Agrifax prices closely follow the GDT auctions.

The NDPSR hedge effectiveness ranges from 51% at a one-week horizon to 96% at an eight-week horizon, and the hedge ratio varies from .42 to .97. The NDPSR weekly surveys are not timely, as they incorporate historical information from up to two months prior. This results in a low hedging effectiveness at a one-week horizon and a significant change in the hedge ratio as the horizon lengthens. In addition, the forward and spot sales are mixed, as the index is based on product shipping in the prior week, not on market sales. The higher effectiveness, relative to the other US price indices primarily benefits from alignment with the settlement calculation in the CME futures contract. In addition, the higher effectiveness is supported by inclusiveness – the NDPSR surveys contain all shipments from large nonfat dry milk processors. The California surveys are similar in nature, but less inclusive and are not directly related to the settlement calculation. Thus, the California survey hedging effectiveness range is slightly lower and ranges from 51% at one week to 95% at eight weeks, and the hedge ratio varies from .37 to .78.

The Dairy Market News surveys are more timely than the NDPSR surveys, as historical information is only one week old. In addition, they provide some forward-spot separation in the form of only spot sales. However, they are not market-based or inclusive due to voluntary reporting and have separate series for geographical regions and the type of nonfat dry milk. The representative series chosen was East and Central Low and Medium Heat. Thus, the hedging effectiveness ranges from 28% at one week to 77% at eight weeks.

The lowest hedging effectiveness range of 10% at one week to 79% at eight weeks is reported for the CME spot market. This series is market based and timely. The problem stems from the lack of inclusiveness, the absence of forward pricing information, and non-alignment with the settlement calculation in the CME NFDM futures contract. Less than 6% of spot market volume is sold on the CME. Since late 2016, a substantial number of milk processors have shifted to pricing sales from last week's weekly CME NFDM average prices, but this unpublished commercial cash index is a backward looking average. It is worthwhile noting that the CME recognizes the defectiveness of their own spot market index in preferring the NDPSR historical average benchmark for the CME futures contract settlement calculation. A separate disadvantage of using the CME spot index may be the possibility of price manipulation.[xxvii]

The analysis of hedging effectiveness reveals that the NZX dairy futures contract is a better tool to use than CME nonfat dry milk futures at shorter hedge horizons and provides persistent estimates of hedge ratios. We argue that the design of the spot market contributes to the success of futures contracts as a hedging tool. Our results show that in the case of a spot market missing one or more of the elements of timeliness, market-based measurement, inclusiveness, and forward-spot separation, one would expect underperformance of the

---

[xxvii] Significant volume may be necessary to avoid price manipulation. Industry observers have raised a concern about the thin trading volume on the CME cheese spot market, in light of the CME cheese spot price being used as a reference price for cheese contracts in the industry at large. This concern was validated by a CFTC fine for the Dairy Farmers of America executives for manipulating the Class III milk futures price by trading cheese spot contracts (see US GAO, Spot Cheese Market: Market Oversight Has Increased, but Concerns Remain about Potential Manipulation, http://www.gao.gov/products/GAO-07-707).

futures market at shorter hedge horizons in terms of hedging effectiveness and a large variation between short- and long-horizon hedge ratios.

8. Conclusion

High hedging effectiveness is an indicator of the successful introduction of a futures contract. Prior studies on hedging effectiveness have focused on the design of the futures contract and other aspects of the futures and cash markets, but have neglected the underlying spot market benchmark design. This study illustrates that the design of the spot market to produce a timely, market-based, inclusive underlying with forward-spot separation is necessary for effective hedging.

In the United States, the majority of the spot pricing of dairy ingredients is published on a historical basis, leading to a multitude of different spot pricing indicators. In addition, the US forward price curve information is not published. In New Zealand, the Global Dairy Trade auction system provides all of the correct dimensions of a spot price benchmark.

Both the Chicago Mercantile Exchange and New Zealand Stock Exchange have chosen to settle dairy futures contracts to historical price averages. We illustrate that the settlement of a futures contract to a historical average introduces basis risk for hedgers and arbitrageurs, and requires speculators to predict the price path rather than only the price at maturity, making speculative trading more complex. Our results show that the New Zealand Stock Exchange dairy futures contract is a better tool for hedging at short hedge horizons, as the Global Dairy Trade spot market benchmark has all the required properties. In addition, the GDT benchmark provides persistent estimates hedge ratios from short- to long-term hedge horizons. Furthermore, the relative hedging effectiveness of each spot index is strongly related to the extent to which the particular index reflects the correct spot market design dimensions.

The above results permit us to formulate the following recommendations. First, we believe that the US nonfat dry milk spot market needs to be re-designed to provide a timely, market-based, and inclusive spot and forward price indicator. A design similar to the US corn market or the NZ Global Dairy Trade may be considered. Once a single spot benchmark has been established for the US market, the Chicago Mercantile futures contract should be changed to use the new benchmark for settlement. Finally, our results show that there would be a benefit to adjusting the settlement calculation in New Zealand to reflect only the spot price at expiration.

Appendix 1 – Success and Failure Criteria for Futures Contracts

This appendix presents a review of the academic literature on the determinants of the success or failure of futures contracts. The characteristics associated with the success or failure of futures can be broadly categorized into the properties of the underlying, the features of the cash market, the attributes of the futures contract, and the institutional setup of the futures market. Table A.1 offers a survey of relevant studies grouped by category. Criteria common to several studies are identified.

[INSERT TABLE A.1 ABOUT HERE]

Black (1986) reports that cash market size, the risk-reduction ability of the contract relative to other cross hedges, cash price volatility, and liquidity costs of a new contract in comparison with existing instruments are associated with the success or failure of futures contracts. Brorsen and Fofana (2001) consider product homogeneity (uniformity), vertical integration of the cash market, buyer concentration, and activeness of the cash market and report that an active cash market is the single characteristic that predicts whether a commodity may have a successful futures market. The study by Bekkerman and Tejeda (2013) examines the failure of the Distillers Dried Grains (DDG) contract on the Chicago Mercantile Exchange (CME). The authors identify the following factors which contribute to the success of new futures contracts:

- *Cash price variability:* A volatile cash market with substantial price uncertainty is more likely to develop a futures market than a market with little variability.

- *Size of the cash market:* The size of the cash market is an indicator of the potential revenue loss due to price volatility. Larger markets are more likely to develop futures markets to manage price risk.
- *Activeness of the cash market:* High activity and participation by different types of investors contribute to a potential higher demand for an organized futures market. In addition, a larger, more active cash market is more likely to have available and credible price information (Fortenberry and Zapata, 1997).
- *Product homogeneity:* Individual futures contracts can only hedge a single grade or quality of a commodity. If the commodity is not homogenous and interchangeable or suffers from quality variation, the hedging effectiveness will be lower for each market segment. The lack of hedging effectiveness contributes to the failure of a futures contract.
- *Product storability*: Bergfjord (2007) argues that ineffective storage may cause quality degradation and inhomogeneity. In addition, good storage infrastructure can facilitate the year-round trading of a commodity.
- *Degree of vertical integration in the market*: Vertical integration measures the extent to which the system of entities responsible for moving the product or service from producer to consumer (the supply chain) is owned by a single company. In a market with a high degree of vertical integration, most price hedging will occur within firms' structure. Consequently, activeness of the cash market and price volatility are likely to be lower.
- *Degree of market power concentration and number of market participants*: In cash markets with a high degree of concentration with only a few participants, futures markets are not expected to be successful due to the ability of firms to control the price.

- *Risk reduction through futures cross-hedging*: When effective price risk cross-hedging tools already exist, significant demand for another tool is unlikely (Black, 1986).
- *Liquidity of cross-hedge futures contracts*: Traders will compare the liquidity costs of an own-hedge futures contract to that of a competing cross-hedge product. The advantage in lower bid-ask spreads for a cross-hedge may outweigh a superior risk-reduction own-hedge capability. For example, jet fuel is a classic example of a market which is cross-hedged by a variety of oil-derived products.

Bekkerman and Tejeda (2013) conclude that the activeness of the cash market, underlying cash market volatility, product homogeneity, industry vertical integration and market power concentration, and the activeness of the futures market with which cross-hedging opportunities exist are important factors in predicting a futures market's success. Their findings are in contrast to Brorsen and Fofana (2001), who report that homogeneity, vertical integration, and buyer concentration were not significantly correlated with success or failure. Furthermore, the study by Brorsen and Fofana (2001) finds that substantial hedging activity in closely related markets is a critical determinant of a futures contract's success. For example, in the case of the DDG futures contract, low hedging activity in the related ethanol futures contract may have caused a failure of the DDG contract.

Perversi, Feuz, and Umberger (2002) provide evidence that futures markets characterized by high basis variability are less attractive for hedgers. They examine basis variability in the cattle stocker versus the cattle feeder market and conclude that one of key reasons behind the stocker contract failure was high basis risk, which discouraged producers from using the contract to hedge calves.

Recently, Till (2014) and Webb (2015) have extended the list of necessary conditions for the successful introduction of a futures contract. Both scholars highlight the importance of a commercial need for hedging and sufficient speculator interest to take the long side of hedger trades. In addition, Webb (2015) points out that the time of introduction, cost of trading, and contract design contribute to the ultimate success/failure of a new contract. Finally, public policy should not be too adverse to futures trading (see Till, 2014), otherwise futures trading is negatively affected. For example, in 1979, the Commodity Futures and Trading Commission (CFTC) banned trading in the March wheat futures contract, and in 1980, the CFTC also considered suspension of trading in silver futures to avoid price manipulation (Till, 2014).

Appendix 2 – Dairy Market Acronyms

Acronym	**Description**
AMS	Agriculture Marketing Service of the United States Department of Agriculture
CME	Chicago Mercantile Exchange
DMN	Dairy Market News
GDT	Global Dairy Trade
NASS	National Agricultural Statistics Survey
NPDSR	National Dairy Product Sales Report
NFDM	Nonfat Dry Milk
USDA	United States Department of Agriculture
WMP	Whole Milk Powder

References

Adams, Z., & Gerner, M. (2012). Cross hedging jet-fuel price exposure. *Energy Economics*, *34*(5), 1301-1309.

Bekkerman, A., & Tejeda, H. A. (2013). Revisiting the determinants of futures contracts: The curious case of distillers' dried grains. Available at SSRN 2274912.

Bergfjord, O. J. (2007). Is there a future for salmon futures? An analysis of the prospects of a potential futures market for salmon. *Aquaculture Economics & Management*, *11*(2), 113-132.

Bhardwaj, Geetesh, Gorton, Gary B., & Rouwenhorst, K. Geert. Facts and fantasies about commodity futures ten years later (June 2015). NBER Working Paper No. w21243. Available at SSRN: http://ssrn.com/abstract=2615649

Białkowski, J., & Jakubowski, J. (2012). Determinants of trading activity on the single-stock futures market: Evidence from the Eurex exchange. *Journal of Derivatives*, *19*(2), 29-47.

Bignell, A. A. S. (2013). *Characteristics of spot-market rate indexes for truckload transportation*. Doctoral dissertation, Massachusetts Institute of Technology.

Black, G. D., 1986. Success and failure of futures contracts: Theory and empirical evidence. Monograph 1986-1, Salomon Brothers Center for the Study of Financial Institutions, New York.

Brorsen, B. W., & Fofana, N. F. (2001). Success and failure of agricultural futures contracts. *Journal of Agribusiness*, *19*(2), 129-146.

Carlton, D. W. (1984). Futures markets: Their purpose, their history, their growth, their successes and failures. *Journal of Futures Markets*, *4*(3), 237-271.

Cohen, L., & Gorham, M. (1985). The projected impact of cash-settled commodity contracts on cash/futures price relationships. In *Conference on Applied Commodity Price Analysis, Forecasting and Market Risk Management* (Vol. 2).

Dickey, David A., and Wayne A. Fuller. "Distribution of the estimators for autoregressive time series with a unit root." *Journal of the American statistical association* 74.366a (1979): 427-431.

Erba, E., & Novakovic, A. M. (1995). *The Evolution of Milk Pricing and Government Intervention in Dairy Markets* (Vol. 95, No. 5). Department of Agricultural, Resource, and Managerial Economics, College of Agriculture and Life Sciences, Cornell University.

Fortenbery, T. R., & Zapata, H. O. (1997). An evaluation of price linkages between futures and cash markets for cheddar cheese. *Journal of Futures Markets*, *17*(3), 279-301.

Garcia, P., Irwin, S. H., & Smith, A. (2015). Futures market failure?. *American Journal of Agricultural Economics*, *97*(1), 40-64.

Ghosh, A. (1993). Hedging with stock index futures: Estimation and forecasting with error correction model. *Journal of Futures Markets*, *13*, 743-743.

Gray, Roger W. (1966). Why does futures trading succeed or fail: An analysis of selected commodities. Futures Trading Seminar http://www.farmdoc.illinois.edu/irwin/archive/books/Views_From_the_Trade/Section6_4.pdf

Herbst, A. F., Kare, D. D., & Caples, S. C. (1989). Hedging effectiveness and minimum risk hedge ratios in the presence of autocorrelation: Foreign currency futures. *Journal of Futures Markets*, *9*(3), 185-197.

Jesse, E., & Cropp, B. (2008). Basic milk pricing concepts for dairy farmers. *Economic Research, 180*, 200.

Johnston, E. T., & McConnell, J. J. (1989). Requiem for a market: An analysis of the rise and fall of a financial futures contract. *Review of Financial Studies, 2*(1), 1-23.

Juhl, T., Kawaller, I. G., & Koch, P. D. (2012). The effect of the hedge horizon on optimal hedge size and effectiveness when prices are cointegrated. *Journal of Futures Markets*, *32*(9), 837-876.

Kavussanos, M. G., & Visvikis, I. D. (2006). Shipping freight derivatives: A survey of recent evidence. *Maritime Policy & Management*, *33*(3), 233-255.

Kawaller, I. G. (1987). A note: Debunking the myth of the risk-free return. *Journal of Futures Markets, 7,* 327–331.

Kenyon, D., Bainbridge, B., & Ernst, R. (1991). Impact of cash settlement on feeder cattle basis. *Western Journal of Agricultural Economics*, 1991, 93-105.

Lee, Y., Ward, C. E., & Brorsen, B. W. (2012). Procurement price relationships for fed cattle and hogs: Importance of the cash market in price discovery. *Agribusiness*, *28*(2), 135-147.

Lien, Donald, and Yiu Kuen Tse. "Some recent developments in futures hedging." *Journal of Economic Surveys* 16.3 (2002): 357-396.

Lien, D. (2008). A note on estimating the benefit of a composite hedge. *Journal of Futures Markets*, *28*(7), 711-716.

McMillan, D. G., & Ülkü, N. (2009). Persistent mispricing in a recently opened emerging index futures market: Arbitrageurs invited. *Journal of Futures Markets, 29(3),* 218-243.

Nam, S. O., Kim, B. C., & Kim, H. K. (2010). An alternative approach to evaluating the agreement between financial markets. *Journal of International Financial Markets, Institutions & Money, 20(1),* 13-35.

Paul, A., Kahl, K., & Tomek, W. (1981). Performance of futures markets: The case of potatoes. U.S. Department of Agriculture, National Economics Division: Economics and Statistics Service, Technical Bulletin, No.1636, January.

Perversi, S., Feuz, D. M., & Umberger, W. J. (2002, December). Basis variability on the feeder cattle contract versus the failed stocker contract. In *2002 Western Agricultural Economics Association Annual Meetings (July 28–31, 2002: Long Beach, California)*. Western Agricultural Economics Association.

Pirrong, C. (2001). Manipulation of cash-settled futures contracts. *The Journal of Business*, *74*(2), 221-244.

Sanders, D. R., Manfredo, M. R., & Greer, T. D. (2003). Hedging spot corn: An examination of the Minneapolis Grain Exchange's cash settled corn contract. *Journal of Agribusiness*, *21*(1), 65-82.

Tashjian, E. (1995). Optimal futures contract design. *The Quarterly Review of Economics and Finance*, *35*(2), 153-162.

Till, H. F. (2014). Why some commodity (and financial) futures contracts succeed and others fail: A survey of relevant research. Available at SSRN 2573894.

Webb, R. (2015) What drives success in futures markets. Keynote speech, 2015 Derivative Markets Conference.

Table 1: Nonfat dry milk spot market price indices for the US market

Price Index	**Description**	**Estimated percentage of sales volume included in index**
USDA Agricultural Marketing Service National Dairy Product Sales Report weekly surveys	Published on Wednesday for sales shipped in the prior week. Sales occur up to 30 days before	40%
Chicago Mercantile Exchange spot market	Current prices for nonfat dry milk to be shipped within six days	6%
California weekly surveys	Published on Friday for sales shipped in the prior week. Sales occur up to 30 days before	40%
Dairy Market News surveys	Weekly average of spot sales	5%

This table lists the four major dairy spot price indices for nonfat dry milk in the United States, and the estimated percentage of sales included in each index. California produces approximately 40% of the US nonfat dry milk. Further information can be found in MCT Dairies, Intricacies of NFDM Pricing and California Dairy Information Bulletin, November 2003.

Table 2: Characteristics of the nonfat dry milk spot market price indices for the US and NZ markets

Index	**Timeliness**	**Market-Based**	**Forward-Spot Separation**	**Inclusiveness**
USDA Agricultural Marketing Service National Dairy Product Sales Report	Includes historical trades up to two months old.	No	Forward and Spot are mixed	Yes
Dairy Market News weekly	Includes historical trades from last week.	No	Only Spot	Voluntary reporting
Chicago Mercantile Exchange spot	Current	Yes	Spot	No – only 6% of market volume
Agrifax weekly	Includes historical trades from last week.	No	Only Forward	Based on survey responses
Global Dairy Trade auctions	Current	Yes	Spot and up to six months delivery visible	Yes

This table illustrates the characteristics of the existing milk powder spot market price index characteristics in the United States and New Zealand. Timeliness measures the extent to which the information is current. Market-based indicates the extent to which the measure is based on markets rather than surveys. Market-based pricing indexes allow for arbitrage in the spot market. Forward-spot separation measures the separation of the spot and forward market sales. Inclusiveness assures that a significant representative fraction of trades are included in the index..

Table 3: Hedge effectiveness as a function of horizon – OLS regression

		Model Specifications						
Spot market (S_t)		*WMP (Agrifax)*	*GDT Contract 2*	*GDT modified*	*NDPSR NFDM*	*CME NFDM*	*DMN NFDM*	*Cal NFDM*
Futures market(F_t)		*WMP (NZX)*	*WMP(NZX)*	*WMP(NZX)*	*CME NFDM*	*CME NFDM*	*CME NFDM*	*CME NFDM*
Period		*Jun 2013-May 2017*	*Jun 2013-May 2017*	*Jun 2013-May 2017*	*Apr 2012-May 2017*	*Apr 2012-May 2017*	*Apr 2012-May 2017*	*Apr 2012-May 2017*
		(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Horizon								
1 week	*β*	0.621**			0.338**	0.281**	0.155**	0.350**
		(15.38)			(7.37)	(4.04)	(3.04)	(5.16)
	Adj R^2	0.542			0.173	0.059	0.034	0.091
2 weeks	*β*	0.648**	0.496**	0.627**	0.512**	0.403**	0.421**	0.506**
		(18.78)	(6.40)	(9.57)	(12.23)	(5.66)	(8.44)	(9.38)
	Adj R^2	0.639	0.310	0.502	0.366	0.110	0.216	0.251
3 weeks	*β*	0.686**			0.711**	0.604**	0.672**	0.685**
		(21.23)			(18.88)	(8.48)	(15.13)	(15.02)
	Adj R^2	0.695			0.580	0.218	0.470	0.463
4 weeks	*β*	0.730**	0.807**	0.855**	0.851**	0.719**	0.802**	0.821**
		(23.66)	(15.53)	(23.66)	(25.78)	(10.06)	(20.17)	(20.41)
	Adj R^2	0.740	0.728	0.862	0.721	0.283	0.613	0.615
5 weeks	*β*	0.749**			0.901**	0.762**	0.862**	0.852**
		(26.88)			(30.10)	(10.94)	(24.08)	(23.87)
	Adj R^2	0.787			0.780	0.319	0.694	0.687
6 weeks	*β*	0.766**	0.841**	0.866**	0.926**	0.792**	0.865**	0.853**
		(28.77)	(16.98)	(22.79)	(32.43)	(11.87)	(25.36)	(24.80)
	Adj R^2	0.809	0.764	0.854	0.805	0.356	0.716	0.704
7 weeks	*β*	0.781**			0.944**	0.817**	0.886**	0.858**
		(30.50)			(36.71)	(13.00)	(28.56)	(26.74)
	Adj R^2	0.827			0.841	0.399	0.763	0.735
8 weeks	*β*	0.801**	0.884**	0.918**	0.974**	0.848**	0.902**	0.879**
		(32.12)	(22.29)	(34.03)	(42.72)	(14.05)	(30.26)	(28.58)
	Adj R^2	0.842	0.850	0.929	0.878	0.438	0.783	0.761

This table reports Ordinary Least Squares (OLS) regression results as a function of hedging horizon. Columns (1) to (3) are the Agrifax whole milk powder (WMP), and Global Dairy Trade (GDT) whole milk powder Auctions with New Zealand Stock Exchange whole milk powder Futures. The *GDT modified* series replaces the second auction with settlement calculation of monthly auction average. Columns (4) to (8) are the USDA AMS National Dairy Product Sales Report (NDPSR) Survey, Chicago Mercantile Exchange (CME), Dairy Market News (DMN), and California Survey prices with CME nonfat dry milk (NFDM) futures. The specifications (1)–(8) are using model

$$\Delta \log S_t = c + \beta \Delta \log F_t + \varepsilon_t$$

**, * denote statistical significance at the 5% and 10% level, respectively.

Table 4: Hedge effectiveness as a function of horizon – cointegration model

		Model Specifications						
Spot market (S_t)		*WMP (Agrifax)*	*GDT Contract 2*	*GDT modified*	*NDPSR NFDM*	*CME NFDM*	*DMN NFDM*	*Cal NFDM*
Futures market(F_t)		*WMP (NZX)*	*WMP(NZX)*	*WMP(NZX)*	*CME NFDM*	*CME NFDM*	*CME NFDM*	*CME NFDM*
Period		*Jun 2013-May 2017*	*Jun 2013-May 2017*	*Jun 2013-May 2017*	*Apr 2012-May 2017*	*Apr 2012-May 2017*	*Apr 2012-May 2017*	*Apr 2012-May 2017*
		(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Horizon								
1 week	*β*	0.582**			0.424**	0.219**	0.167**	0.365**
		(16.14)			(10.66)	(2.54)	(3.06)	(6.94)
	Adj R^2	0.697			0.516	0.105	0.282	0.515
2 weeks	*β*	0.573**	1.159**	1.021**	0.584**	0.482**	0.445**	0.361**
		(12.10)	(8.85)	(14.61)	(13.69)	(3.57)	(5.53)	(5.86)
	Adj R^2	0.776	0.610	0.773	0.769	0.296	0.450	0.651
3 weeks	*β*	0.616**			0.767**	0.782**	0.764**	0.470**
		(14.03)			(15.66)	(4.20)	(9.84)	(7.60)
	Adj R^2	0.841			0.842	0.348	0.648	0.748
4 weeks	*β*	0.722**	1.062**	1.062**	0.746**	1.353**	0.793**	0.691**
		(12.64)	(15.18)	(15.18)	(16.53)	(7.83)	(7.34)	(10.16)
	Adj R^2	0.896	0.906	0.906	0.908	0.623	0.609	0.844
5 weeks	*β*	0.811**			0.818**	1.472**	0.962**	0.809**
		(12.90)			(11.96)	(9.69)	(11.37)	(13.84)
	Adj R^2	0.906			0.866	0.762	0.853	0.924
6 weeks	*β*	0.744**	1.058**	1.019**	0.933**	1.387**	1.101**	0.552**
		(9.44)	(15.67)	(22.81)	(24.85)	(5.46)	(9.31)	(7.72)
	Adj R^2	0.888	0.941	0.967	0.962	0.571	0.765	0.864
7 weeks	*β*	0.744**			0.996**	1.579**	0.994**	0.681**
		(13.08)			(10.92)	(7.40)	(8.97)	(12.18)
	Adj R^2	0.959			0.939	0.754	0.838	0.924
8 weeks	*β*	0.796**	1.112**	1.112**	0.973**	1.500**	1.106**	0.767**
		(8.30)	(20.83)	(20.83)	(12.65)	(7.55)	(7.50)	(13.17)
	Adj R^2	0.934	0.983	0.983	0.961	0.786	0.765	0.945

This table reports cointegration regression results as a function of hedging horizon. Columns (1) to (3) are the Agrifax whole milk powder (WMP), and Global Dairy Trade (GDT) whole milk powder Auctions with New Zealand Stock Exchange whole milk powder Futures. The *GDT modified* series replaces the second auction with settlement calculation of monthly auction average. Columns (4) to (8) are the National Dairy Product Sales Report (*NDPSR*) Survey, Chicago Mercantile Exchange (CME), Dairy Market News (DMN), and California Survey prices with CME nonfat dry milk (NFDM) futures. The specifications (1)–(8) are using model

$$\Delta \log S_t = c + \beta \Delta \log F_t + \sum_k \gamma_k \Delta \log F_{t-k} + \sum_l \delta_l \Delta \log S_{t-l} + \lambda e_{t-1} + \varepsilon_t$$

**, * denote statistical significance at the 5% and 10% level, respectively.

Table A.1: Characteristics of futures markets and futures contracts associated with the success and failure of futures contracts

Category	**Characteristic**	**Studies**
Futures Market	Hedging need for commercial market participants.	Black (1986) Gray (1978) Webb (2015)
Futures Market	The attraction of speculative interest to the futures market.	Gray (1978) Till (2014) Webb (2015)
Futures Market	Favorable disposition towards the futures contract introduction by existing market power holders	Gray (1978)
Futures Market	Favorable public policy towards the contract (for example not prohibiting trading in a particular contract)	Till (2014)
Futures Contract	Good contract design that does not favor the buyer or seller.	Gray (1978) Webb (2015) Garcia, Irwin and Smith (2015)
Futures Contract	The risk reduction ability of the contract in comparison to existing futures cross-hedging contracts.	Black (1986)
Cash Market	Size and activeness of the cash market	Black (1986) Brorsen and Fofana 2001 Bekkerman and Tejeda 2013 Fortenberry and Zapata (2002)
Cash Market	Volatility of the cash price	Black 1985 Bekkerman and Tejeda 2013 Webb (2015)
Commodity	Homogeneity or uniformity of the commodity.	Bekkerman and Tejeda 2013
Cash Market	The extent of existing vertical market integration	Bekkerman and Tejeda 2013
Cash Market	Buyer concentration within the cash market.	Bekkerman and Tejeda 2013
Commodity	The storability of the commodity.	Bekkerman and Tejeda 2013
Futures Market	Superior liquidity of existing cross-hedges discourages a new own-hedge contract.	Black 1986 Brorsen and Fofana (2001) Bekkerman and Tejeda 2013 Webb (2015)
Futures Market	Significant hedger activity in closely related markets.	Bekkerman and Tejeda 2013
Futures Market	Basis Risk impedes a futures contract.	Perversi, Feuz, and Umberger (2002)
Futures Market	Public genuine commercial order flow	Webb (2015)
Futures Market	First mover advantage in the introduction of the contract.	Webb (2015)
Futures Market	Low explicit trading costs, for example broker commissions	Webb (2015)
Futures Market	Timing of the introduction of the contract	Webb (2015)

This table reports the characteristics identified in the literature that are associated with the success or failure of futures contracts. The *Category* column indicates the general area of the market the attribute refers to – the commodity, the cash market, the futures market, or the futures contract. The *Characteristic* column lists the particular characteristic of the market, contract, or commodity. The *Studies* column lists the studies in which the relevant characteristic was found to have a significant statistical relationship.

Chapter 2

The reduced explanatory power of the Traditional Theory of Storage: A comparison of the historical 1952-1963 and modern 2009-2019 cocoa spot and futures markets.

Abstract

The theory of storage explains expected commodity price movements as a function of current and expected inventory stocks-to-use levels. We report a high degree of accuracy in explaining cocoa spot price and futures market term structure movements in the historical 1952-1962 time period. Price movements in the modern 2009-2019 time period are only partially explainable by a fundamental inventory based model. The lack of accuracy is conjectured to result from several factors present in the modern but not the historical market - the presence of a large component of non-fundamental-based traders in the modern market, overt market manipulation by both producers and traders, and the modern but not historically observable intrinsic uncertainty of cocoa production.

2.1. Introduction

2.1.1 The Cocoa Market

The structure of the modern cocoa market is much the same as the cocoa market seventy years ago. Cocoa beans are grown, fermented, and dried by small scale farmers in Africa, South America, and other equatorial regions. Despite the increase in the number of global supply locations including Ecuador, Indonesia, and the Caribbean, the lions share of production still comes from the Ivory Coast and Ghana. On the consumption side, the dried beans are purchased and then ground by large multinational confectionary companies to produce ingredients for chocolate and cocoa based products. A large amount of grinding capacity formerly in consuming nations is now available on location in producing nations, but the grinding factories are still owned by large multinationals like commodity trader Cargill and confectionary giant Barry Callebut. Futures and spot trading has taken place from the 1930's in both New York and London[1].

The cocoa market is a useful market scenario in which to study the supply of storage theory for four reasons. First, the production of cocoa has a long life cycle with trees not nearing productive capacity until 6 to 8 years of age. This means the relationship between price and inventory is not affected by annual production decisions. Second, the

[1] Weymar(1965) provides general background on the historical cocoa market. The International Cocoa Organization, www.icco.org, also provides background information.

consumption of cocoa, which is mostly made into chocolate, is empirically observed to depend on lagged rather than current price. Chocolate manufacturers typically hold inventory from three to nine months and retailers only periodically change the price of a product. The dependence on lagged price lessens the activity of short-term competitive profit-seeking storers. Third, the relative perishability of cocoa beans, which are at best quality for six months to a year[2], is an additional discouragement to profit-seeking stockholding. Finally, there are no close substitutes for cocoa – chocolate can only be made from cocoa beans. These four reasons taken together tend to isolate the behaviour of the cocoa price in regard to the level of inventory, so that the inventory levels at which convenience yield and backwardation occur are clearly visible, as is the relationship between price and inventory.

The cocoa market is not unique in these respects. Several other commodity markets including cocoa, coffee, orange juice, wool, lumber, rubber, whole milk powder, and skim milk powder share the characteristics of difficulty in modifying production and short-term perishability. The conclusions from the cocoa market will likely apply also to these markets.

[2] Wong-Shing K., "Here's the Deal on Raw Cacao, AKA the Healthy Version of Chocolate", https://www.leaf.tv/articles/storage-shelf-life-of-raw-cacao-beans/

2.1.2 The Rational Expectations Competitive Storage Model

The modern "rational expectations competitive storage model" of Gustafson (1958), Gardner (1980), Williams and Wright (1991), Deaton & Laroque (1992), Cafiero (2011), and Gouel & Legrand (2017) optimizes the profit over several years (time periods) from selecting the optimal annual carryover with identical inverse demand and cost functions applying in each year(time period). Enhanced versions of the model allow a variable future production response to current inventory levels (Williams and Wright (1991)). The model is principally concerned with the optimal annual carryover of storable grains. The model is less applicable to the cocoa market where the commodity is relatively perishable, production decisions take several years to have effect, consumption is dependent on lagged rather than current price, and stock withholding from the market is unusual.

The relative simplicity of the cocoa market when compared with long-term-storable, flexible-production grains like corn allows the use of the simpler "Supply of Storage" model. We use a model based on the traditional theory of storage of Keynes (1936), Kaldor (1939), Working (1949), Brennan (1958), and Weymar (1965) to compare time periods from 1952-1963 and 2009-2019 using spot and futures data from the cocoa market. The Supply of Storage model is essentially a special case of the general "Rational Expectations Competitive Storage Model", where there is limited competitive speculative stockholding and the supply response is inelastic. Appendix A provides a history of the modern rational expectations storage model, and illustrates the consistency of the simpler Supply of Storage model when the general model is subject to the constraints of the cocoa market. One critical distinction is the omission of "working stocks" from the general model.

In addition, we study monthly data in the cocoa market as opposed to the overwhelming majority of agricultural studies that examine annual data. Peterson and Tomek (2005) provide the only monthly analysis of agricultural commodities (corn) with the related Rational Expectations Competitive Storage Model, but have difficulty reproducing the empirical annual autocorrelation visible in the market. The advantage of monthly data is that the behaviour of price as related to inventory can be studied in finer detail without being obscured by annual averaging.

2.1.3 The Traditional Theory of Storage

Keynes (1930) provides the first documented connection between supply shortages and backwardation in his observation that "If there is a shortage of supply capable of being remedied in six months but not at once, then the spot price can rise above the forward price to an extent which is only limited only by the unwillingness of the buyer to pay the higher spot price rather than postpone the date of his purchase[3]."

[3] Keynes(1930) also theorizes that in times of surplus, the forward price will rise above the current spot price, but less than the expected spot price, the concept of

Working (1932, 1933, 1948, 1949) collects detailed empirical data on wheat spot and futures prices and observes a close relationship between the degree of backwardation and the current stocks of wheat as reported in Figure 1. Backwardation, in contrast to Keynes, is redefined by Working as the now commonly accepted negative futures-spot spread rather than the forward-"expected spot" spread postulated by Keynes.

[INSERT FIGURE 1 ABOUT HERE]

Working formalizes the concept that the expected price change of a commodity was dependent on inventory levels into the traditional "Theory of Storage".

Kaldor (1939) independently develops the theory and theorizes that "stocks of all goods possess a yield measured in terms of themselves, the "convenience yield". Brennan (1958) derives supply and demand curves for storage. Weymar (1965, 1968) illustrates theoretically that the "the spread between the spot price and the price expected at some future time is a function of expected inventory behaviour over the intervening interval. Weymar also provides a new theory of short term commodity price behaviour and enhances the model with a trend-following component. Instead of academia, Weymar went on to successfully trade cocoa for Nabisco and later founded the Commodities Corporation. Commodities Corporation, in addition to commercial success in the

---

"normal backwardation". "Contango" and "Backwardation" were terms in use by traders in Liverpool in the 1850's.

commodities markets, was a pioneer in trend-following strategies and the training ground for several prominent hedge fund managers[4].

2.1.4 Possible Reasons for Reduced Explanatory Power

The Traditional Supply of Storage model exhibits greatly reduced accuracy in the modern time period of 2009-2019 when compared with 1952-1963. Some of the possible reasons are the presence of a large component of computerized trend-following or other non-fundamental based traders in the modern markets, an increase in market manipulation, the availability of options, and the availability and ambiguity of inventory information.

Trend following in the commodities markets gained major traction in the 1960s and 70s, with the publication of detailed techniques by Richard Donchian[5], the "Godfather" of modern day trend following, and the spectacular success of the "Turtle Traders", a group of non-specialist investors trained in 1983 to follow a trend following system by Richard Dennis (Covel (2007)). Trend following traders have the effect of accelerating convergence towards fundamental value when the trend is "correct", but also cause extensive overshoot when the fundamentals for the market change. Commodity index fund trading also gained interest in the 1990's leading to the Financialization of

[4] Tully S. and Carson L., *Princeton's Rich Commodity Scholars,* Time Inc., 1981, see https://www.turtletrader.com/comm-corp/
[5] Donchian R., Trend Following Methods in Commodity Price Analysis, Commodity Yearbook 1957

Commodities theme[6]. This type of trading has little correlation to trading based on cocoa market inventory fundamentals. In addition, several other strategies including machine learning and momentum are used in the commodity markets. We approximate a measure of this presence by the difference in the variance explainable by fundamental versus trend following components in our model.

Several incidences of overt market manipulation are visible in the 2010-2019 time period, including a market squeeze by Armajaro in 2010, bean smuggling and withholding by Ghana and Ivory Coast producers in 2015, and pressure by large market traders on forecast providers to provide accommodating figures. These types of manipulation appear to be present to a lesser extent in the 1950s, with only a single incident identified, a withholding from the market by Brazil in 1954.

Options began trading in the 1970s after the seminal Black-Scholes paper, and increased speculative activity in the market. This change to the markets does not impact our conclusions, but is mentioned for completeness.

The intrinsic uncertainty of production is recently observable in the conflict between multiple sources of estimates for expected production, consumption and inventory stocks from different market information services. The intrinsic uncertainty of production arises from the difficulty in knowing the production of an individual tree, which can vary from

---

[6]ETF.com, "A Brief History of Commodities Indexes", https://www.etf.com/publications/journalofindexes/joi-articles/7451-a-brief-history-of-commodities-indexes.html?nopaging=1

a few to a hundred pods from season to season. This uncertainty has not changed, but the limited information available in the 1950s may have served to provide a more "certain" information environment for trading. Despite the prediction for expected coverage levels being intrinsically inaccurate, there was only one prediction. The intrinsic uncertainty can be approximated by the divergence in analyst predictions.

The remainder of the paper is structured as follows. Section 2 reviews the Traditional Supply of Storage Theory. Section 3 describes the data sources for the 1952-1963 and 2009-2019 periods. Section 4 presents the results. In Section 4.1, the spot cocoa market analysis of Weymar (1965) is replicated (Welch, 2019). Then, in section 4.2, the historical analysis is extended to the futures market in 1952-1963. Lastly, in section 4.3, the extended spot market model is estimated on both current 2009-2019 spot and futures market data. Section 5 discusses the differences between the model accuracy in the two time periods and possible sources for the lack of accuracy in the modern period. Section 6 concludes the paper.

2.2. Supply of Storage Theory

2.2.1 The Empirical Phenomenon

The traditional theory of storage is not so much a theory as a consistent empirical pattern observed in the agricultural commodity markets, with sound reasoning as to its existence. Working (1932, 1933) was the first to carefully document the phenomenon from observations of the futures-spot price spread in grains. Figure 2 illustrates the regular pattern. Working (1948, 1949) formulates the empirical observation into the traditional “Theory of Storage”: *The expected price change of a commodity over a time interval is related to the inventory level at the beginning of the interval*.

[INSERT FIGURE 2 ABOUT HERE]

This empirical phenomenon has been confirmed by numerous other authours including Kaldor (1939), Meinken (1955), Weymar (1965), Geman (2013), and Irwin and Garcia (2015).

Four reasons have been put forward for the existence of this stylistic feature of commodity markets. First, when inventories are low there is the risk of a stockout for manufacturers where production lines can be disrupted, and lost sales can result. This risk dominates the lower section of the curve, where the expected price change can be negative but positive inventories are still held. In this situation, the instant profit from selling immediately and simultaneously locking in future inventory at a lower price is offset by the much higher costs of lost sales and production. Note that with raw commodities, the selling price of a finished good is often several times the raw material component cost[7].

[7] baked goods, steel and cars, etc.

Second, manufacturers of finished goods want to keep their pricing roughly in line with other competitors in the industry and therefore carry enough inventory to maintain flexibility. This reinforces the negative price change holding behaviour at low inventory levels, and results in a roughly linear relationship at normal inventory levels. At high inventory levels, the effect reverses as the raw commodity may reduce in price leaving the high inventory holder with surplus inventory that has to be processed at the previous higher cost.

Third, speculators will hold inventory as long as the expected price gain is greater than the cost of holding the inventory. Speculators will be active at all inventory levels. Even if there is a harvest surplus, speculators may buy and hold inventory in anticipation of a lowering of production in the following years.

Fourth, when storage facilities are taxed, stockholders will require a significant premium to continue storing the commodity. This particular regime has not received much focus as it occurs rarely in practice.

Working (1948) indicates that "The results from all lines of investigation concur in indicating that prices quoted at one time in a futures market, for two dates of delivery, stand in a relation which in general does not reflect expectations regarding events that may occur between the two delivery dates." Weymar (1965) points out the logical contradiction in this statement. If the price change between two points in time is dependant on the level of inventory at the beginning of the interval, then the inventory during the interval must not change. Otherwise, the changes over the two subintervals must add to the change over the entire interval. This will only be true if the price change is linear in inventory. Weymar generalizes the theory of storage to account for expected changes of inventory during the interval – "the spread between the spot price and the price expected at some future time is a function of expected inventory behaviour over the intervening interval."

As the time interval is shortened, the theory of storage provides an estimate for the instantaneous expected price rate of change as a differentiable function of the current inventory level. If this function is integrable, then the behaviour of price can be expressed as a function of the two endpoints. In other words, the expected price change over an interval is the integral of the differential supply of storage curve over that interval.

2.2.2 The Cocoa Market Spot Price

Weymar (1965) posits the following market model for cocoa.

Cocoa Market Economic Model

$$C_t = f_c(P_t^L) + e_c \tag{1}$$

$$H_t = f_h(P_t^L) + e_h \tag{2}$$

$$P_t^* - P_t = f_P(I_t) + e_P \tag{3}$$

$$I_t = I_{t-1} + H_t - C_t \tag{4}$$

where C_t is current consumption, P_t^L is the lagged price P_t is the current price, P_t^* is the expected price, H_t is current production, f_c , f_h , f_P are the functional forms of the relationships, and I_t is the inventory level

In the cocoa market, consumption of cocoa beans is a function of lagged price due to the inventory storage behaviour of chocolate makers, typically four to nine months[8]. The production of cocoa beans is a function of lagged price as a result of the long gestation time before cocoa trees become productive. The spread between the expected price and the current price is a function of the current inventory level – the traditional theory of storage. The last equation is an identity where the current inventory is the previous inventory plus current production and minus current consumption.

[8] For instance, Barry Callebut, one of the largest chocolate makers, maintains a four month inventory (calculated from 2018 financial statements)

Speculative storing is not common in cocoa market due to the logistics of the dependency of production on weather combined with the relative perishability of the commodity. The motivation for speculative storing is that future conditions are more favorable for selling. Since the likelihood of a poor crop in the future is equivalent to the chance of a good crop, and the annual supply will not adjust, chances are that speculatively stored cocoa beans will face the same market conditions as today as well as incurring the cost of carry. The situation is different in other markets like corn, where an oversupply will immediately cause a drop in future production due to producers adjusting their future output downwards.

2.2.3 Change of variables to Rate of Change of Price and Inventory Coverage Ratio

The original theory of Working is stated in terms of price changes and inventory levels. Weymar generalizes the theory to use percentage price changes and inventory-to-consumption ratios.

Instead of:

$$\frac{P_t^* - P_t}{h} = f_h^*(I_t) \qquad (5)$$

where P_t is the current price, P_t^* is the expected price, h is the time interval, f_h^* is the functional form of the relationship, and I_t is the inventory level the equation should be:

$$\frac{(P_t^* - P_t)/P_t}{h} = f_h^*(\frac{I_t}{\sum C_t}) = f_h^*(Y_t) \qquad (6)$$

where $\sum C_t$ is the consumption over the last 12 months.

Keeping in mind that the original theory of Working is a stylistic observation rather than a derived conclusion from a structural model, Weymar offers the following arguments for why the change of variable is a superior representation.

For the percentage change in price formulation, there are three reasons. First, speculators are interested in the return on capital rather than the price change, and the required return is a function of the amount of invested capital. Second, for hedging by dealers, a large part of the carrying costs – i.e. insurance and interest – vary with price. Third, with the assumption that finished product yields vary in accordance with raw material costs, the convenience yield will also vary with price.

For stating the Inventory as a coverage ratio there are two reasons. First, the convenience yield to processors will vary directly with the time coverage that a particular level of inventory affords. If, for example, there was a spike in consumption, the marginal convenience yield of inventory would also rise. Second, the amount of hedging and speculative activity will tend to be a function of inventory coverage rather than inventory level alone.

Finally, a significant advantage of using ratios is that dimensionless quantities are less subject to overall changes in the levels of the respective variables. An inventory coverage ratio of 50% measures the same market situation whether the price levels are USD 1500 or USD 3000.

Weymar provides empirical support for this change of variables by the high quality fit of the resultant model applied to the cocoa market. It is a question for further research whether the change of variables representation is a superior tool to the original formulation of Working. In this paper, we will follow Weymar's lead and use the formulation with percentage price changes and inventory coverage ratios.

2.2.4 Review of the Derivation of the Estimable Model for the Cocoa Price

The price change from the present until equilibrium can be decomposed into the price change from the present to a horizon time and the price change from the horizon time to equilibrium. The horizon time is chosen to be the end of the current crop year.

The traditional theory of storage does not specify the functional form of the empirical relationship, but indicates that at low inventory levels the slope is positive with a negative first derivative. We do not consider the behaviour for extreme oversupply when storage facilities become taxed as this is not relevant to the cocoa market[9]. A reasonable approximation for the price change over a small interval is the log function with the Weymar change of variables i.e.

$$\frac{d}{dh}(\ln P_t^{*h}) = f^*(Y_t^{*h}) = a + b\ln Y_t^{*h} \qquad (7)$$

where P_t^{*h} and Y_t^{*h} are the expected price and expected inventory ratio.

In discrete form, this is:

$$\ln(\frac{P_t^{*h+1}}{P_t^{*h}}) = a + b\ln Y_t^{*h} \qquad (8)$$

where P_t^{*h} and Y_t^{*h} are the expected price and expected inventory ratio at the beginning of interval h.

Taking the intervals as monthly, the price change from the present to the horizon time can be expressed as a sum of the changes in each of the months:

$$\ln(\frac{P_t^{*h_t}}{P_t}) = ah_t + b\sum_{h=0}^{h_t-1}\ln Y_t^{*h} \qquad (9)$$

---

[9] The cocoa market is sufficiently disorganized that extreme excess produce tends to be left to deteriorate, rather than forcing a price crash.

If the inventory behaviour throughout the year follows the same basic seasonal path each year (as is observed in the cocoa market), then the inventory ratio at each of the intervening months can be estimated as a function of the two endpoints. Given that the endpoints are typically either known (the current inventory ratio) or published (the expected inventory ratio before harvest at the end of the year) the price change from the present to the horizon time can be modeled. The following regression model can be used to estimate the coefficients.

$$\ln(Y_t^{*h}) = c_h + d_h \ln Y_t + e_h \ln Y_t^{*h_t} \tag{10}$$

These coefficients give the inventory ratio in the intervening months between the current time and the September horizon time. When combined with the ratios at the beginning and the end of the interval, the coefficient sums capture the behaviour of inventory throughout the interval.

Substituting into the previous equation,

$$\ln\left(\frac{P_t^{*h_t}}{P_t}\right) = b_1 h_t + b_2 \left[\sum_{h=0}^{h_t-1} c_h + \left(\sum_{h=0}^{h_t-1} d_h\right) \ln Y_t + \left(\sum_{h=0}^{h_t-1} e_h\right) \ln Y_t^{*h_t}\right] \tag{11}$$

The coefficients c_h, d_h, e_h can be estimated by running regressions on the yearly inventory data – see Table 1 for coefficients for the period 1952-1963. The mean R^2 from the individual source regressions is .95 with a standard deviation of .029. The cocoa year runs from October 1st to September 30th, so the September values, for example, give the price change from the end of September to the end of the following September. The October values from October 31st to September 30th, and so on.

[INSERT TABLE 1 ABOUT HERE]

This price change to the horizon time as a function of the inventory ratio end-points will apply in each year. Thus the behaviour of the price from the end of the year until equilibrium can be derived.

The September version of equation (11), referencing Table 1, is

$$\ln\left(\frac{P_t^{*12}}{P_t}\right) = b_1 12 + b_2 (1.2 + 5.479 \ln Y_t + 4.879 \ln Y_t^{12}) \quad (12)$$

which specifies the yearly changes in price. This equation gives the expected price change from the current end of September to the next end of September.

Since the same situation will apply in the following September,

$$\ln\left(\frac{P_t^{*(s_{n+1})}}{P_t^{*(s_n)}}\right) = b_1 12 + b_2(1.2 + 5.479\ln Y_t^{*s_n} + 4.879\ln Y_t^{*(s_{n+1})}) \quad (13)$$

Where s_n indicates the price or inventory ratio at the end of the n^{th} September.

We assume that the inventory ratio in September will approach an equilibrium value over several years in the following manner:

$$Y_t^{*s_n} = \left(\frac{Y_t^{*s_0}}{\overline{Y^{*s}}}\right)^{g^n} \overline{Y^{*s}} \quad (14)$$

where s_n is an index for the September horizon and g is the annual rate at which the September inventory level will approach the equilibrium level $\overline{Y^{*s}}$. If g =0.3, for example, the equilibrium ratio would be reached in approximately 5 years.

Taking logarithms,

$$\ln(Y_t^{*s_n}) = g^n Y_t^{*s_0} + (1 - g^n)\overline{Y^{*s}} \quad (15)$$

Substituting into equation 13 and simplifying (see Weymar (1965) pg. 183),

$$\ln\left(\frac{P_t^{*(s_n+1)}}{P_t^{*(s_n)}}\right) = b_2 g^n (5.479 + 4.879g)(\ln Y_t^{*s_0} - \ln \overline{Y^{*s}}) \tag{16}$$

The annual September to September price changes accumulated over n1 years, starting with the first horizon September S_0 are:

$$\ln\left(\frac{P_t^{*(s_n)}}{P_t^{*(s_0)}}\right) = \sum_0^{n1-1} \ln\left(\frac{P_t^{*(s_{n+1})}}{P_t^{*(s_n)}}\right) \tag{17}$$

Substituting 16 into 17, and simplifying (with the identity h_t , the September horizon, equal to s_0),

$$\ln\left(\frac{\overline{P}}{P_t^{*h_t}}\right) = b_2\left(\frac{5.479 + 4.879g}{1-g}\right)(\ln Y_t^{*h_t} - \ln \overline{Y}^{*s}) \tag{18}$$

Combining with equation (11) gives the final statistically estimable equation (19)

$$\ln\left(\frac{\overline{P}}{P_t}\right) = a + b_1 h_t + b_2 Z_t + b_3 \ln Y^{*h_t} + e_t \tag{19}$$

where

$$a = -b_3 \ln \overline{Y}^{*s} \tag{20}$$

$$b_3 = b_2 \left(\frac{5.479 + 4.879g}{1-g} \right) \tag{21}$$

$$Z_t = \sum_{h=0}^{h_t-1} c_h + \sum_{h=0}^{h_t-1} d_h \ln Y_t + \sum_{h=0}^{h_t-1} e_h ln Y^{*h_t} \tag{22}$$

and

$\overline{P}$ is the average real spot price of cocoa – an estimate for the constant equilibrium price of cocoa

P = the monthly average real spot price of cocoa

h is the horizon interval in months

Y^{*h} is the expected inventory ratio at horizon time (i.e. inventory/last years consumption)

$\overline{Y}^{*s}$ is the expected September equilibrium inventory ratio.

g = parameter indicating the speed that Y approaches equilibrium

e = error term.

This equation states that the price spread from the current spot price to the equilibrium price is a function of the month in the year, the current inventory ratio, and the expected inventory ratio at the end of the year. The coefficient sums that make up the Z coefficient result from inventory regressions that characterize the seasonal behaviour of inventory throughout the year. The coefficient b_3

is only related to the rate of approach towards equilibrium, and does not incorporate the inventory ratio estimates before the crop year end.

2.2.7 Supply of Storage with a Futures Market

The second and third terms of equation (19) give the expected spot price at each month throughout the year as a function of the current inventory ratio and the year-end inventory ratio estimate.

The price at the end of each month from the current time until the end of the crop year can be estimated in the following manner from the model of equation 19.

Equation 11 was derived for the spread from the current price to the end of the crop year, but is valid for all of the intervening months as well. Solving for the monthly price gives:

$$\ln(P_t^{*h_t}) = \ln(P_t) + b_1 h_t + b_2 \left[\sum_{h=0}^{h_t-1} c_h + \left(\sum_{h=0}^{h_t-1} d_h\right) \ln Y_t + \left(\sum_{h=0}^{h_t-1} e_h\right) \ln Y_t^{*h_t} \right] \quad (23)$$

The monthly coefficients can be estimated from the same monthly inventory regressions for each time period (September – September, October – September, …, August – September), but summed in a different manner. Instead of the inventory behaviour to the September horizon time, we are interested in the inventory levels at the intervening months. Panel 1 of Table 2 illustrates the monthly inventory regression coefficient sums that allow the calculation of the intervening inventory levels based on the endpoints for the current crop year (See Weymar (1966) Chapter 4, Appendix B for the individual regressions).

Making the assumption that the estimated spot price will be the best predictor of the futures price, Equation (23) can be used to estimate futures contract prices. As it turns out, there appears to be a roughly constant premium between the spot price and the futures price at closing in the cocoa market in 1952-1963.

The model also allows for the estimation of the futures contracts after the crop year end. The inventory ratio is anticipated to approach an equilibrium value as in equation (14), repeated here as equation 24.

$$Y_t^{*s_n} = \left(\frac{Y_t^{*s_0}}{\overline{Y^{*s}}} \right)^{g^n} \overline{Y^{*s}} \qquad (24)$$

Using the current estimated end of year inventory ratio, and letting n=1 in equation (24), the following year ending inventory can be estimated as:

$$Y_t^{*2} = \left(\frac{Y_t^{*1}}{\overline{Y^{*s}}} \right)^{g} \overline{Y^{*s}} \qquad (25)$$

Where $\overline{Y^{*s}}$ is calculated from equation 20:

$$\overline{Y^{*s}} = \exp(-\frac{a}{b_3}) \tag{26}$$

And g is calculated from equation 21:

$$g = \frac{b_3 - 5.479b_2}{b_3 + 4.879b_2} \tag{27}$$

Then, using the same procedure as with the current year but replacing the starting and ending inventory ratios with the new values and the current price with the estimated September price, we can calculate the crop futures prices from September onwards. Panel 2 of Table 2 illustrates the inventory coefficient sums for the future crop year for the historical 1952-1963 dataset. The sums for the future crop year do not change as the contracts are all estimated relative to the September of the previous crop year.

The traditional supply of storage theory as depicted in Figure 2, indicates there is a level of inventory below which stocks will be held at a loss. In other words, the term structure will shift into backwardation smoothly as inventory drops below this level. As inventory rises above this level, the carrying costs will become constant and equal to storage plus interest costs except at the high end of the curve where storage facilities are taxed.

The calculated futures prices can be compared with the actual spread between the futures contracts and the spot price. Five interesting questions can be asked regarding the historical and modern time periods:

1. How well does the model fit the spot price in the modern time period? Note that Weymar reported the spot price fit for the historical time period.
2. How well does the model fit the observed term structure in both periods?
3. Does the term structure smoothly move into backwardation or contango as a function only of expected inventory levels.
4. What are the expected inventory levels at which the term structure shifts into contango or backwardation.
5. What appears to have changed in the cocoa market between 1952-1963 and 2009-2019?

[INSERT TABLE 2 ABOUT HERE]

2.3. The Data

With the goal of allowing reproduction of the results in this paper, detailed sources for all of the data are enumerated. Matlab code and individual data files are available from the authour on request (jan.koeman@pg.canterbury.ac.nz). Data that would need to be purchased directly from the source is indicated.

2.3.1 1952-1963 Time Period

The price series for Accra beans in New York is originally sourced from the Foreign Agriculture Circulars from the United States Foreign Agriculture Service available at the Internet Archive: https://archive.org/search.php?query=cocoa%20statistics. The documents are scanned, so the price series must be transcribed by hand. An alternative paid source is the 2010 Commodity Research Bureau Yearbook with accompanying CDROM.

The deflator for the price series is the United States Bureau of Labor Statistics Wholesale Price Index (1957-59 = 100) available from the St. Louis Fed: https://fred.stlouisfed.org/series/M0448CUSM350NNBR . The base is the average price between 1957 and 1959.

The price series is deflated to factor out general movements in commodity prices from the cocoa price.

The inventory and expected inventory coverage ratios are transcribed from Weymar (1966). Figure 3 shows the Price Change, Current and Year-End estimated inventory ratios from September 1952-August 1963. Weymar (1966) provides a detailed appendix on the construction of his current and expected inventory ratios.

[INSERT FIGURE 3 ABOUT HERE]

Futures prices from 1952-1963 were manually transcribed from the New York Times archive, available by subscription at https://timesmachine.nytimes.com/browser. Four individual dates in the month were used to construct a monthly average for each contract.

2.3.2 2009-2019 Time Period

Monthly current inventory ratios are calculated from starting estimates of yearly inventory, production, and consumption (grindings) data.

Though the starting year worldwide inventory is in principle knowable (unlike the production capability of cocoa trees), the market still has a difficult time in settling on estimates. We use inventory estimates from CRA Services Ltd. (CRA Services)[10] and the International Cocoa Organization (ICCO). One curious side effect of this inefficiency that can result in some confusion is in the quotation of yearly estimates for stocks, production, and consumption. The differences in year-ending stocks are not equal to production minus consumption for the past year, but instead are quoted separately.

The cocoa year runs from October 1st to Sep 30th in the following year. Arrivals of cocoa beans at Ivorian and Ghanese ports are tracked on approximately a bi-weekly basis, though often numbers are not published. In addition, there is a fair amount of smuggling and withholding from either Ghana to the Ivory Coast or in the opposite direction

[10] Knowledge Charts, https://www.knowledgecharts.com/, is a website published by CRA Services Ltd. dedicated to information on the cocoa market. It is by far the best information resource available on the cocoa market.

depending on the farmgate price differential. The other significant producing nations are Brazil, Ecuador, Nigeria, Cameroon, and Indonesia. Detailed production data is available on a monthly or finer basis from CRA Service Ltd[11]. Only the total yearly production figures are available for several smaller producers. The monthly production pattern for Nigeria and other African countries is estimated based on the monthly production sum of Ivory Coast, Ghana, and Cameroon which have similar weather conditions. The monthly production for other South American countries is based on Brazil and Ecuador. The Asian monthly production is derived from the Indonesia patterns.

Grindings data is available regionally on a quarterly basis for the Americas, Europe, The Ivory Coast and Ghana, and Asia from CRA Services. Monthly values are estimated as one third of quarterly values. Other grindings for each region are assumed to follow the same quarterly pattern as the reported numbers.

The current inventory level is calculated by starting with the estimate of stocks at the beginning of the year, and adding production and grindings to get each monthly figure. The current inventory stocks-to-use ratio is calculated as the monthly inventory level divided by the grindings summed over the previous year. In addition to the current inventory level and current inventory ratio level, relative inventory levels and relative inventory ratio levels are calculated for the sample. The levels are relative to the level or ratio in the particular month for the entire sample from 2009-2019.

---

[11] Less accurate arrival data for the Ivory Coast and Ghana can be transcribed fromten years of monthly ICCO market commentary, as was initially done by this authour before being aware of CRA Services Ltd.

Expected cocoa production, consumption and stocks are available on a quarterly basis from the Economist Intelligence Unit and the ICCO Quarterly Bulletins. The Economist issues World Commodity Forecasts Feedstuffs and Beverages monthly by subscription. The ICCO publishes the forecast excerpt on their website (www.icco.org). From these sources both the levels and stocks-to-use ratios can be manually transcribed. The ICCO quarterly forecasts are linearly interpolated to get monthly results. Technically, the ICCO only issues three forecasts – in January, May, and September. The November report is a report of estimates for the previous crop year. Figure 4 reports the inventory ratios and deflated spot prices from 2009-2019.

[INSERT FIGURE 4 ABOUT HERE]

Seasonal Inventory ratios are compiled from monthly regressions of the cocoa inventory ratio levels and give the intermediate ratios as a function of the current and estimated year-end values (see Weymar (1966) pgs 171-176). These individual regressions are then summed to get a single figure that captures the inventory behaviour between the current point in time and the end of crop year horizon time (see constructZ.m in Appendix C for Matlab code).

Spot prices are sourced from CRA Services and from BarChart[12]. The spot price series most representative of the world is the Ivory Coast series. Series are also available for Ghana, and the cash price in New York.

Prices are deflated by the Federal Reserve Bank of St. Louis Global Price Index of All Commodities. (https://fred.stlouisfed.org/series/PALLFNFINDEXM). Robustness checks are done using the Producer Price Index for all Commodities (https://fred.stlouisfed.org/series/PPIACO), and the Consumer Price Index (https://fred.stlouisfed.org/series/CPIAUCSL).

2.4. Results

First, the model estimates obtained by Weymar for 1952-1963 are replicated and the model is extended with trend following and generalized least squares estimation. Second, the futures prices for 1952-1963 are estimated, illustrating a 93% prediction accuracy of the shift between contango and backwardation. Third, the spot price from 2009-2019 is analyzed, reporting a significant drop in model accuracy. Fourth, the 2009-2019 term structure is estimated, yielding only a 32% correlation between the model slope and the actual slope shifts. Finally, the reasons for this drop in accuracy are discussed.

2.4.1 Replication of Weymar's Results

[12] Barchart information services, www.barchart.com provides download of an extensive array of financial information by subscription.

2.4.1.1 Basic Model Results

Table 3 illustrates the result of replicating Weymar's results using model (19) and data derived from Weymar (1965):

[INSERT TABLE 3 ABOUT HERE]

The coefficient estimates are a very close match with the original estimation done by Weymar, indicating that the model has been correctly replicated.

2.4.1.2 Changing Equilibrium Price Expectations

The main problem with the basic model is the low Durbin Watson coefficient of .34, indicating serial correlation in the residuals and misspecification. The most likely source of this error is the final equilibrium price level. For the basic model, this was set equal to the average price over the interval 1952-1963. It is reasonable to assume that there are changing expectations for this equilibrium price level over time and that the equilibrium price level is influenced by the past long term trend in cocoa prices.

If we define the expected equilibrium price as a variable multiplied by the average price over the time period studied,

$$P_t^{\infty} = M_t \overline{P} \tag{28}$$

then, taking logarithms and substituting into equation (19) we obtain:

$$\ln(\frac{\overline{P}}{P_t}) = a + b_1 h_t + b_2 Z_t + b_3 \ln Y^{*h_t} - \ln M_t + e_t \tag{29}$$

$\overline{P}$ has been redefined here to be the average sample post-war cocoa price rather than the equilibrium price. The variable expected multiplier can be modeled as a function of the long term cocoa price trend. An alternative would be to assume that the multiplier was a function of past price levels, but Weymar did not find better estimates using this assumption. We follow Weymar and estimate the long-term cocoa price trend as a exponential function of a vector of twenty elements, each containing the average price change over four months.

$$M_t = \exp(b_4 \sum_{i=0}^{19} a_i \left(\frac{\Delta P}{P} \right)_i) + e_t \tag{30}$$

Where $\left(\frac{\Delta P}{P} \right)_i = \frac{1}{4} \sum_{k=4i+1}^{k=4i+4} \frac{P_{t-k} - P_{t-k-1}}{P_{t-k-1}}$.

Substituting into equation 29,

$$\ln(\frac{\overline{P}}{P_t}) = a + b_1 h_t + b_2 Z_t + b_3 \ln Y^{*h_t} + b_4 \sum_{i=0}^{19} a_i \left(\frac{\Delta P}{P}\right)_i + e_t \quad (31)$$

where a_i are the individual coefficients, and b_4 is a scaling factor to allow the sum of the a_i to equal unity. The entire trend following component covers a period of 80 months, or somewhat less than seven years.

Table 4 illustrates the result of replicating the model (31) results on data derived from Weymar (1965):

[INSERT TABLE 4 ABOUT HERE]

The differences in the trend following variant model are due to minor manual smoothing of the coefficients done by Weymar, and the use of 18 rather than 20 intervals. The a_i coefficients are reported in Appendix C. In the values obtained by Weymar, all but the last three are significant at 5%. In the reproduction, which uses 18 intervals instead of 20 due to lack of earlier price data, the last 4 coefficients are not significant.

2.4.1.3 Generalized Least Squares Regression Estimates

The R_2 increases significantly to .945 and the Durbin-Watson coefficient is significantly better at .95, but still indicates serial correlation and possible misspecification. However, the coefficient of determination is high and and the remaining misspecification is relatively minor. To remove the serial correlation in the residuals, we estimate the model using generalized least squares. An approximate value for the GLS estimates is obtained by transforming all variables using a first order autoregressive model for the residuals (see Weymar (1965) pg 213). For example, the price is transformed as:

$$P_t^T = P_t - \rho P_{t-1} \tag{32}$$

Where ρ is the first-order autocorrelation coefficient of the residuals.

Table 5 illustrates the result of incorporating the GLS transformations into model 18.1

[INSERT TABLE 5 ABOUT HERE]

With both trend-following and generalized least square transformations the Durbin Watson coefficient is 1.79 (Reproduced)/2.19 (Original) indicating that the residuals do not display serial correlation.

2.4.2 Futures Price and Estimates 1952-1963

The actual term structure for the period September 1952 to December 1962 is depicted in Figure 6[13]. The spread between the actual 1st and 5th New York cocoa contracts and the estimated values from the basic model are reported in Figure 7. The basic model reproduces the term structure with an R^2 of 84%. The correlation between the first minus fifth spread is .93. There is a tendency of the model to overestimate the spread, which is due to a persistent spot price premium over the first contract that is on average 3 cents over the period 1952-1963. This premium would be bundled into the regression constant and not accounted for in the constructed estimate of futures prices. Attributing part of the constant estimate to a spot price premium over the first contract would have the beneficial effect of raising the estimated equilibrium price ratio, which is relatively low.

[INSERT FIGURE 6 ABOUT HERE]

[INSERT FIGURE 7 ABOUT HERE]

There is an empirical spread in the current month/spot prices illustrated in Figure 8, with the average spot price generally being several cents higher than the near month contract even at expiration of the contract. An example of this is in September 1958. On the 10th, the spot price was 43.25 cents/lb and the September futures was 39.25 cents/lb. On the 23rd, the spot price of Accra beans was 42.35, and the contract expiring September 24th was 38.42 cents. On the following day, the spot price changed to 37.80 and the contract

[13] This futures data was obtained by manual transcription from the New York Times business section on microfiche. The entire collection from 1851-2002 of New York Times daily editions is available for subscription at https://timesmachine.nytimes.com/browser.

traded between 36.75 and 38.05. There is a large premium to holding the spot beans over the futures contract. The premium does not appear to vary whether the term structure is in contango or backwardation.

[INSERT FIGURE 8 ABOUT HERE]

Figure 7 illustrates that there are four periods of contango and backwardation in the historical period, and that there is a smooth switch between the two regimes that occurs when the expected inventory ratio is approximately .42 (see Figure 3).

The estimate for contracts in the new crop year are based only on approach to equilibrium reasoning, and leaves out the additional information that may be available to the market. In particular, the mid-year crop harvested in May and June, though only 10 to 15% typically of the total crop, may be indicative of the coming full year crop that starts in December. It may be possible to improve the model with a second inflection point starting around the mid-year.

Nonetheless, the estimated model provides an excellent account of the behaviour of both spot and futures prices in the 1952-1962 time period.

2.4.3 Actual and Estimated Spot Price 2009-2019

Figure 9 displays the normalized Ivory Coast spot price and the expected inventory ratios from 2009-2019. The relative current inventory ratio is calculated as the relative value of the current inventory ratio for a particular month in the time period. For example, the May 2015 relative inventory ratio is the current inventory ratio for May 2015 divided by the mean of the current inventory ratio for every month of May in the sample. The relative current inventory ratio is intended to capture whether inventory is higher or lower at a particular point in time. This is difficult to discern with the seasonal oscillation of the values.

[INSERT FIGURE 9 ABOUT HERE]

The inventory behaviour is similar in predictability to the 1952-1963 time period, and the coefficient sums are reported in Table 6. The mean R2 from the individual regressions is .97 with a standard deviation of .02. The pattern of the coefficients has changed however, with the c_h coefficients close to zero for all months.

[INSERT TABLE 6 ABOUT HERE]

Table 7 reports the values from estimating the models of equation 19 and 31 on 2009-2019 spot price data.

[INSERT TABLE 7 ABOUT HERE]

The goodness-of-fit of the model is drastically reduced in the 2009-2019 time period. The highest R2 is obtained by a combination of EIU and ICCO estimates, with significant values for the monthly changes (b1 and b2). However, the significance disappears as trend following is incorporated, and the variables are transformed for Generalized Least Square estimates. The model R^2, however, substantially increases with the incorporated equilibrium price trend multiplier. The coefficients of the equilibrium price trend multiplier display a different pattern to the 1952-1963 time period. Recent coefficients (see Appendix D) are negative for approximately three years and then turn positive. Using CPI and PPI deflators instead of the Global Price Index of All Commodities further decreases the R^2 of the model.

Futures market estimates, which rely on the significance of the structural coefficients, are poor. Only the basic model is applicable, and estimates a 32% correlation between the actual market backwardation and the estimated market backwardation.

2.5. Possible Explanations for Reduced Explanatory Power

The structure of the cocoa physical market is much the same in the time period 2009 to 2019 but the following appear to have changed: the quantity and ambiguity of inventory ratio estimates, the number of non-fundamental traders in the market, the availability of options trading, the movement of grinding capacity to production locations, the presence of market manipulation, and the general expansion of production worldwide. We examine each of these factors to identify which are likely of consequence.

2.5.1 Options Trading and Expansion in Production

First, the availability of options trading is unlikely to have a material impact except to increase the amount of speculation. Second, the expansion in production is unlikely to effect the model accuracy as the primary producers remain the same – the Ivory Coast, Ghana, and Brazil. Other additional production has tended to follow the same pattern as the primary producers. In addition, the production methods have not evolved. Cocoa is still harvested by individual small farmers with virtually no mechanization.

2.5.2 Forecast Accuracy and Information Availability

For 1952-1963, 83% of the variance in spot prices is explained by the estimate of the end-of-crop year inventory ratio whereas only 35% is explainable by either the ICCO or EIU estimates. However, the standard deviation of the modern estimates (ICCO -.025 ; EIU, .014) is substantially better than the standard deviation of the historical estimates (.046). This discrepancy might be explained by the limited information available in the earlier time period. In the 1952-1963, the primary source of information was Gill and Duffus, the leading British cocoa dealer and broker. In 2009-2019, there are several forecast providers including the EIU and the ICCO. The single-source nature of the earlier end-of-crop year estimates may have increased the prediction value to market participants. An alternative explanation may be the decrease in emphasis on fundamental values due to non-fundamental based trading.

2.5.3 Market Manipulation

Fourth, the market in the 2009-2019 period is rife with market manipulations, both by producers and traders as opposed to the 1952-1963 time period, although Weymar documents an instance of withholding by Brazil in 1954. It is possible that the market has better information in the modern period than in the historical period, but also possible that manipulations are easier to execute in a computerized environment.

In 2010, Armajaro executed a successful squeeze on the London Liffe Futures Market. In 2015, and likely at other crop harvests, Ghanese bean smugglers managed to withhold a substantial portion of the harvest that became available in the Ivory Coast in the new crop year. Lastly, in the bull run of 2015/2016 trade sources indicate strong pressure by long trading firms on information forecast providers to increase the bearishness of their estimates.

In July 2010, sixteen European cocoa trading organizations wrote to the London International Financial Futures Exchange (LIFFE) expressing their concerns about market manipulation. Armajaro had bought 250,000 tons worth of cocoa futures contracts, pushing their price to a 30-year record of £2,590 per tonne. The purchase was a very successful short squeeze on the market and resulted in LIFFE publishing a Commitment of Traders (COT) positions report for the London market. The COT had formerly only been available in the United States. The effect of this squeeze on the data is that the market was placed into heavy backwardation that was completely unrelated to underlying fundamental inventory factors. “A major concern on the London market has been the inverted futures price structure (“backwardation”), and in particular, the very large premium of the July 2010 contract (and to a lesser extent, the September contract) over the December contract. “(ICCO Monthly Review (June, 2010)).

Dummying out the 2009/2010 year and using the ICCO estimates results in an increase in the R^2 to .50, indicating that this event is affecting the model performance. Using the EIU estimates does not result in any change.

Another confusion generating factor is the behaviour of the cocoa bean producers in the Ivory Coast and Ghana. For example, in the September 2015 Monthly Review of the Cocoa Market, the ICCO notes:

“Conversely, Ghana, the world’ s second largest cocoa producing country, surprised most cocoa analysts by experiencing a sharp fall in its cocoa production, from 896,917 tonnes in the 2013/2014 season, to 735,000 tonnes following a very poor main crop, according to data released by officials in the country. The farm-gate price was increased by 22% to 6,720 cedis per tonne (US$1,759 per tonne) for the 2015/2016 season, from 5,600 cedis per tonne (US$1,465) paid in the just-ended 2014/2015 season.”

The ICCO also notes for the Ivory Coast “For the 2015/2016 season, the Government has set a farm-gate price of 1,000 CFA francs (US$ 1.67) per kilogramme, representing an 18% increase, compared with 850 CF A francs per kilogramme set for the 2014/2015 crop season.”

In the following monthly review, the ICCO notes:

“The 2015/2016 cocoa harvest started very strongly in West Africa, in contrast to the views of some analysts who had anticipated a large supply deficit for the current crop year. Indeed, news agency data estimated that total cocoa arrivals at Ivorian ports as at 1 November had reached approximately 280,000 tonnes since the start of the current cocoa season, being approximately 64,000 tonnes higher compared with the corresponding period for the previous year. In Ghana, cocoa purchases recorded by the *Ghana Cocoa Board* totalled 192,128 tonnes as at 22 October. This represented a 118% increase compared with the same period for the previous year. “

Clearly, what is happening is the Ghana bean smugglers are withholding inventory from the market in order to get a better price either in the New Year, in the Ivory Coast, or both. The effect of this behaviour is to make the seasonal inventory regressions less reliable as inventory that should be allocated to September of the current crop year is instead allocated to October in the following crop year. The amount of the beans withheld or smuggled is in the hundreds of thousands of tonnes, more than adequate to move the market.

Dummying out the 2014/2015 or the 2015/206 crop years with either the ICCO or EU estimates results in a similar or decreased R^2 for the model.

These types of market manipulations are no doubt ongoing, and may serve to confound the values of fundamental variables on which predictions are made.

2.5.4 Non-Fundamental Based Trading

Finally, and probably most importantly, the 2009-2019 market appears to have a significant component of non-fundamental based trading. These traders include trend followers like the Turtle Traders, momentum traders, Commodity Index Funds, and several other types.

An approximate measure of this trading is the variance explainable by our fundamental based model versus our model incorporating equilibrium price trend following. The model with trend following is designed to incorporate the changes in equilibrium price estimations based on past price trends, but will also capture other types of trading that are using historical price changes as input. The difference in R^2 values for the models incorporating trend following versus the basic models are 11% in the 1952-1963 time period, but 58% for the 2009-2019 time period.

Anecdotal evidence supports the prevalence of technical trading in the cocoa market. After receiving his doctorate, Weymar went on to trade cocoa successfully at Nabisco using his fundamental-based model. Subsequently, Weymar raised five million in equity capital (including 150,000 from Paul Samuelson) and formed Commodities Corporation. To his surprise, Weymar discovered that after two years of trading, the fundamental model was not making economically valuable price predictions. The firm then designed and applied a hybrid trend-following system designed by Paul Vannerson which is still in use today at Goldman Sachs. From subsequent historical results, it appears that Commodities Corporation performed the strongest when trend following and fundamental models suggested the same trading strategy.

In addition, Burghardt (2010) compares the performance of a subset of trend following Commodity Trading Advisors with the overall performance of the Newedge CTA Index, and found a correlation of .97 over the period 2000-2009. This result indicates that virtually all CTA based trading, a significant component of general commodity trading, utilized trend following rather than fundamental analysis.

2.6. Conclusion

We have compared the performance of a Supply of Storage based commodity price and inventory model in the cocoa markets of 1952-1963 and 2009-2019. The performance of the model is exceptional in the earlier period, but poor in the more recent time frame.

Several possible sources of the model error are investigated including market manipulation, information forecast quality, and the relative degree of non-fundamental based trading. There is some evidence that manipulation is making prices unrelated to fundamentals, but the most likely source of the discrepancy appears to be extensive trading based on technical rather than fundamental factors. Market commentary points to significant difficulty for hedge funds operating in the commodity space, with two-thirds of funds closing their doors in the period 2012-2018[14] which may be related with the plethora of technical trading strategies.

The comparison of the modern and historical periods raises an interesting question about the effectiveness of the cocoa market in pricing cocoa beans. The price in the modern period is efficient in the sense of being less predictable, but at the same time the market price is less related to the fundamental inventory ratio and inventory ratio estimates. Are the technical traders assisting or confounding price discovery in the cocoa marketplace?

[14] Schaefer K., 2018, Why Commodity Hedge Fund Managers Are Disappearing

References

Alquist, R., & Kilian, L. (2010). What do we learn from the price of crude oil futures?. Journal of Applied econometrics, 25(4), 539-573.

Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. The Journal of Finance, 68(3), 929-985.

Banks, F. E. (1994). Oil stocks and oil prices. OPEC review, 18(2), 173-184.

Bellman, R. (1966). Dynamic programming. Science, 153(3731), 34-37.

Bobenrieth, E., Wright, B., & Zeng, D. (2013). Stocks-to-use ratios and prices as indicators of vulnerability to spikes in global cereal markets. Agricultural Economics, 44(s1), 43-52.

Brennan, M. J. (1958). The supply of storage. The American Economic Review, 48(1), 50-72.

Brennan, D., Williams, J., & Wright, B. D. (1997). Convenience yield without the convenience: a spatial–temporal interpretation of storage under backwardation. The Economic Journal, 107(443), 1009-1022.

Burghardt, G. (2010). Measuring the impact of trend following in the CTA space. Opalesque interview.

Cafiero, C., & Wright, B. D. (2011). The empirical relevance of the competitive storage model. Journal of Econometrics, 162(1), 44-54.

Cartea, A., & Figueroa, M. G. (2005). Pricing in electricity markets: a mean reverting jump diffusion model with seasonality. Applied Mathematical Finance, 12(4), 313-335.

Carter, C. A., Rausser, G. C., & Smith, A. (2011). Commodity booms and busts.

Covel, M. (2007). The Complete TurtleTrader: The Legend, the Lessons, the Results. Collins.

Seamon G. (2010), CME Group, Understanding Wheat Futures Convergence , https://www.cmegroup.com/trading/agricultural/files/AC-422_WheatResearchPaper_FINAL_SR.PDF

Deaton, A., & Laroque, G. (1992). On the behaviour of commodity prices. The review of economic studies, 59(1), 1-23.

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

Deaton, A. (2011). Puzzles and paradoxes: a life in applied economics. Eminent Economists II: Their Life and Work Philosophies.

Erb, C. B., & Harvey, C. R. (2006). The strategic and tactical value of commodity futures. Financial Analysts Journal, 62(2), 69-97.

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

Fisher, F. M., & Cootner, P. H. (1971). An econometric model of the world copper industry.

Geman, H., & Smith, W. O. (2013). Theory of storage, inventory and volatility in the LME base metals. Resources Policy, 38(1), 18-28.

Gorton, G. B., Hayashi, F., & Rouwenhorst, K. G. (2012). The fundamentals of commodity futures returns. Review of Finance, 17(1), 35-105.

Joseph, K., Irwin, S. H., & Garcia, P. (2015). Commodity storage under backwardation: Does the working curve still work?. Applied Economic Perspectives and Policy, 38(1), 152-173.

Gardner, B. L. (1980). Optimal stockpiling of grain. Lexington Books..

Geman, H., & Smith, W. O. (2013). Theory of storage, inventory and volatility in the LME base metals. Resources Policy, 38(1), 18-28.

Gibson, R., & Schwartz, E. S. (1990). Stochastic convenience yield and the pricing of oil contingent claims. The Journal of Finance, 45(3), 959-976.

Gilbert, C. L. (2016). The dynamics of the world cocoa price. The economics of chocolate, 307.

Gilbert, C.L. (2014). "Stockholding, Investment and Commodity Price Dynamics: The World Cocoa Market", cocoa model estimates for "The dynamics of the world cocoa price". (Unpublished) Model_paper.pdf

Gouel, C., & Legrand, N. (2017). Estimating the competitive storage model with trending commodity prices. Journal of Applied Econometrics, 32(4), 744-763.

Gustafson, R. L. (1958). Carryover levels for grains: a method for determining amounts that are optimal under specified conditions (No. 1178). US Department of Agriculture.

Kaldor, N. (1939, 1976). Speculation and economic stability. In The Economics of Futures Trading (pp. 111-123). Palgrave Macmillan, London.

Keynes, J. M. (1936, 2018). The general theory of employment, interest, and money. Springer.

Knight, F. H. (2012). Risk, uncertainty and profit. Courier Corporation.

Legrand, N. (2019). The empirical merit of structural explanations of commodity price volatility: Review and perspectives. Journal of Economic Surveys, 33(2), 639-664.

Lucia, J. J., & Schwartz, E. S. (2002). Electricity prices and power derivatives: Evidence from the nordic power exchange. Review of derivatives research, 5(1), 5-50.

Meinken, K. W. (1955). The demand and price structure for wheat (No. 1136). US Dept. of Agriculture.

Miranda, M., and X. Rui. (1996). An Empirical Reassessment of the Nonlinear Rational Expectations Commodity Storage Model. Unpublished Paper, The Ohio State University

Miranda, M. J. (1998). Numerical strategies for solving the nonlinear rational expectations commodity market model. Comput. Econ. 11, 71–87.

Miranda, M. J., & Fackler, P. L. (2004). Applied computational economics and finance. MIT press.

Muth, J. F. (1961). Rational expectations and the theory of price movements. Econometrica 29, 315–335.

Peterson, H. H., & Tomek, W. G. (2005). How much of commodity price behavior can a rational expectations storage model explain?. Agricultural Economics, 33(3), 289-303.

Pindyck, R.S. (1994). Inventories and the Short-Run Dynamics of Commodity Prices. The RAND J. of Econ. 25 (1), 141-159.

Pindyck, R. S. (2001). The dynamics of commodity spot and futures markets: a primer. The energy journal, 1-29.

Routledge, B. R., Seppi, D. J., & Spatt, C. S. (2000). Equilibrium forward curves for commodities. The Journal of Finance, 55(3), 1297-1338.

Schwartz, E. S. (1997). The stochastic behavior of commodity prices: Implications for valuation and hedging. The journal of finance, 52(3), 923-973.

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

Tang, K., & Xiong, W. (2012). Index investment and the financialization of commodities. Financial Analysts Journal, 68(6), 54-74.

Verleger, P. K. (1982). The determinants of official OPEC crude prices. The Review of Economics and Statistics, 177-183.

Welch, I. (2019). Reproducing, Extending, Updating, Replicating, Reexamining, and Reconciling. Critical Finance Review, 8(1-2), 301-304.

Weymar, F. H. (1965). The dynamics of the world cocoa market (Doctoral dissertation, Massachusetts Institute of Technology).

Weymar, F. H. (1966). The supply of storage revisited. The American Economic Review, 56(5), 1226-1234.

Weymar, F. H. (1968). The dynamics of the world cocoa market. The dynamics of the world cocoa market.MIT Press ISBN: 9780262230285

Williams, J. B. (1936). Speculation and the carryover. The Quarterly Journal of Economics, 50(3), 436-455.

Williams, J. C., & Wright, B. D. (2005). Storage and commodity markets. Cambridge university press.

Working, H. (1927). Forecasting the Price of Wheat. J. of Farm Econ. 9 (3), 273–287.

Working, H. (1933). Price Relations between July and September Wheat Futures at Chicago Since 1885. Wheat Studies of the Food Research. Inst. 9 (6). 34

Working, H. (1934). Price Relations Between May and New-Crop Wheat Futures at Chicago Since 1885. Wheat Stud. of the Food Res. Inst. 10 (5).

Working, H. (1948). Theory of the Inverse Carrying Charge in Futures Markets. J. of Farm Econ. 30 (1), 1-28.

Working, H. (1949). 'The Theory of Price of Storage'. The Am. Econ. Rev. 39 (6), 1254–1262.

Wright, B. D. (2011). The economics of grain price volatility. Applied Economic Perspectives and Policy, 33(1), 32-58.

Appendix A – The modern rational expectations competitive storage (MRECS) model

A.1 Model History

Williams (1936) provides the first analysis of the optimal annual carryover in agricultural commodities and illustrates a pre-computer, graphical method of finding the optimal amount of grain to store from harvest to harvest given a surplus, and an estimate of future production.

Gustafson (1952) revisits the problem and illustrates how the carryover problem for grains, formulated as a Bellman Equation, can be solved using dynamic programming. The solution consistes of optimal storage rules that determine the current year carryover as a function of the carry-in from the previous year and current production. Gustafson illustrates that there was a single optimal solution that involved the same rule being applied each year.

Gardner (1979) improves on Gustafson by allowing for a flexible supply response to price. Gustafson justifies the omission of this condition by the historical observation that variability in yield per acre amounted for the vast majority of the variation in yearly output. Williams and Wright (1991) provide a detailed review of Gardner, and source code routines that implement the dynamic programming problem. In addition Williams and Wright examine the implications of public interventions including floor price and public storage schemes.

Deaton and Laroque (1992, 1995, 1996), noting the lack of empirical testing of the theory, provide a detailed test with 13 different commodities. Their conclusion is that the theory is not capable of explaining all of the empirical evidence, in particular the high annual autocorrelation visible empirically in commodity prices.

Miranda and Rui (1999, unpublished) illustrate that combining the modern rational expectations competitive storage model with a traditional supply of storage formulation can explain the observed autocorellation of commodity prices. The conclusions are drawn from using price data alone, as high quality inventory data was not obtainable. They suggest that more research into the microstructure foundation of the behaviour of inventory at low levels is needed.

Peterson and Tomek (2005) create a monthly simulation of the US corn market based on Deaton and Laroque, and find that despite including an explicit convenience yield, the high levels of autocorellation in the corn market were not reproducible. This is an interesting finding in light of the Miranda (1999) result and possibly results from estimating the model using inventory rather than price data.

Cafiero (2011) re-estimates the D&L model with a finer grid and some minor adjustments to the slopes of the inverse demand function. The re-estimation accurately reproduces the autocorellation of the empirical commodity prices. However, Gouel and Legrand (2017) report that the Cafiero modifications result in the model always remaining in the non-stock–out state, questioning the usefulness of a two-state model in which one state is never realized. Gouel and Legrand instead estimate the model after detrending the commodity price series and find autocorellations in line with empirical measurements.

A.2 The MRECS Model under the constraints of the Cocoa Market.

Since production decisions take several years to implement in the cocoa market, the MRECS model can be represented without an elastic supply response as (Gardner(1979) pp 15):

$$\prod{}^{*}(X_1 + I_0) = \max E\left\{\sum_{t=1}^{T}[P_t * (X_t + I_{t-1} - I_t) - G(I_t)](\frac{1}{(1+r)^t})\right\} \quad \text{(A.1)}$$

where $\prod^{*}$ is the optimal profit, X_t is the random production in year t, I_{t-1} are beginning stocks (carry-in), I_t are carryover stocks at the end of the year, P_t is the market clearing price for the year, G is the cost function for storage, and r is the discount rate.

The optimization problem can be stated as follows: given a starting carry-in inventory I_0, and production X_1 in year 1, what is the optimal carryover each year from the current year to the horizon year T that maximizes profit. Gardner states the model in more general terms, with a value function instead of revenue and a welfare function for profit, but otherwise the specification above is the same.

This model excludes "working stocks" from the definition of inventory. Inventory is only composed of speculative or discretionary stocks. Thus, a "stockout" in this model is not really a complete stockout, as plant and machinery can still be kept running with working stocks. This creates some ambiguity as to the size of non-speculative stocks, as it assumes that manufacturers have control of inventory when inventory is run down to the working stock level. If that were not the case, and speculators instead had cornered the market, then manufacturers would be obliged to obtain the stock necessary to keep machinery running on the open market. This would have the effect of increasing the portion of inventory that behaved in a different manner to purely speculative stocks. In addition, stocks at lower aggregate inventory levels function in certain market situations to provide "cost coverage", making their behaviour non-speculative. There is strong empirical support for the non-speculative behaviour of inventory at low levels as there has never been a stockout of corn in recorded history in the United States (though the stocks were very small in 1934 and 1936)[15]. If stocks at low levels were purely speculative, then all stock would be consumed, and the industry would start over the next year with zero stocks.

[15] From Peterson and Tomek (2005)

The Supply of Storage model in the cocoa market states that the expected price change over a small interval where inventory does not change is a function of the inventory level at the start of the interval. A reasonable approximation is:

$$\ln\left(\frac{P_t^{*h+1}}{P_t^{*h}}\right) = a + b\ln I_t^{*h} \qquad \text{(A.2)}$$

where P_t^{*h} and I_t^{*h} are the expected price and expected inventory at the beginning of interval h.

The SOS model avoids the artificial distinction between "working" and "speculative" stocks, and is designed to allow stockholding at a loss – the "convenience" yield, which arises both from insurance against plant shutdown and insurance against cost coverage (see Weymar (1965) pp, 105-110).

The MRECS model was designed to apply to annual carryover, but the time interval can be changed to monthly intervals. Using monthly time intervals, the MRECS model will maximize, over several time periods, the discounted revenue minus the total storage cost. This implies that more inventory will be stored when the discounted expected price in the next period is greater than the current price. In other words, the expected price change in the following period must be greater if the starting level of inventory is greater. The latter statement is consistent with the SOS model when expected price change is positive and speculative storing is minimal as in the cocoa market. When the expected price change is negative, all speculative stocks are consumed and we are in the regime of "working stocks", which is not covered by the MRECS model. In the cocoa market, which is often in backwardation, there is an advantage to using the SOS model which allows for both regimes.

The general SOS model implies an increase in storage costs when storage capacity is taxed. Oversupply in the grains market can result in lack of convergence between cash and futures prices in the wheat market (Seamon (2010)). The MRECS model can accommodate oversupply costs with an appropriate storage cost function G. In the cocoa market, overproduction has less of an impact as the storage industry is not sufficiently organized to register the marginal effects of taxed storage capacity. For example, in 2017, when an unexpectedly large crop materialized, the excess beans sat in trucks and became rotten[16]. In the functional form chosen in A.2, there is no provision for increased costs when storage capacity is taxed.

[16] Gro Intelligence (2017)," Cocoa Prices Turn Bitter in Ivory Coast", https://www.gro-intelligence.com/insights/articles/ivory-coast-cocoa-prices

Appendix B. How to Estimate the Traditional Model from Futures Prices Only (Optional Appendix)

An alternative model can estimate the same coefficients in the spot price model but using only the futures data. This assumes that the futures price is a good estimate of the spot price or differs by a roughly constant premium from the expected spot price.

Equation 19 can be estimated directly for contracts up to the September horizon:

$$\ln(\frac{P_t^{*h_t}}{P_t}) = a + b_1 h_t + b_2\left[\sum_{h=0}^{h_t-1} c_h + (\sum_{h=0}^{h_t-1} d_h)\ln Y_t + (\sum_{h=0}^{h_t-1} e_h)\ln Y_t^{*h_t}\right] + e_t \qquad (24)$$

For the contracts after the horizon month, we can use the return to equilibrium argument with a single year.

Taking logs of equation 21,

$$\ln(Y_t^{*2}) = g\ln Y_t^{*1} + (1-g)\ln\overline{Y^{*s}} \qquad (25)$$

Substituting into equation 24 we obtain:

$$\ln(\frac{P_t^{*h_t}}{P_t}) = b_2(1-g)\overline{Y^{*s}} + b_1 h_t + b_2\left[\sum_{h=0}^{h_t-1} c_h + (\sum_{h=0}^{h_t-1} d_h)Y_t^{1^*}\right] + b_2 g\left[(\sum_{h=0}^{h_t-1} e_h)\ln Y_t^{1^*}\right] + e_t \quad (26)$$

These equations can be estimated with the futures data from 1952-1963 to get an alternate estimate of the coefficients b_1, b_2, g, and $\overline{Y^{*s}}$. Embedded in the constant term would be any risk premium (which is observed in the spot to current month spread).

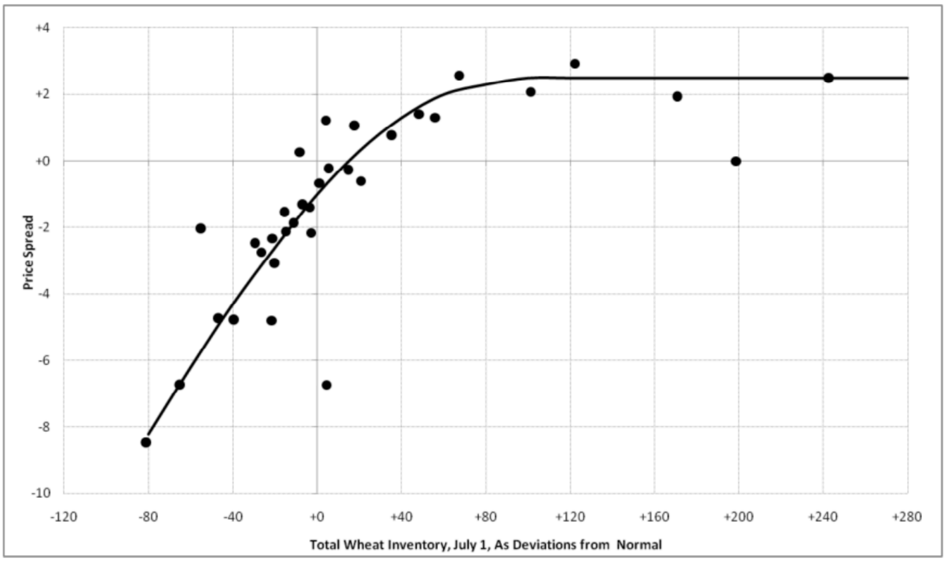
Appendix C. Equilibrium Price Change Multiplier coefficient estimates 1953-1962.

[INSERT TABLE C1 ABOUT HERE]

Appendix D. Equilibrium Price Change Multiplier coefficient estimates 2009-2019.

[INSERT TABLE C2 ABOUT HERE]

Figure 1 – Supply of Storage Curve (Working,1933)



This figure depicts the relationship between the July-September spread and wheat stocks on July 1st in the 1930's.

Figure 2 – Supply of Storage (Weymar,1966)

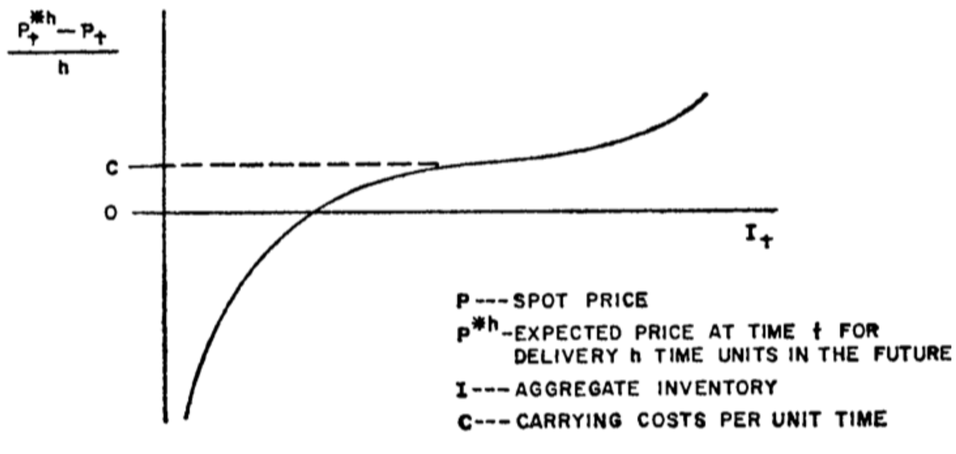
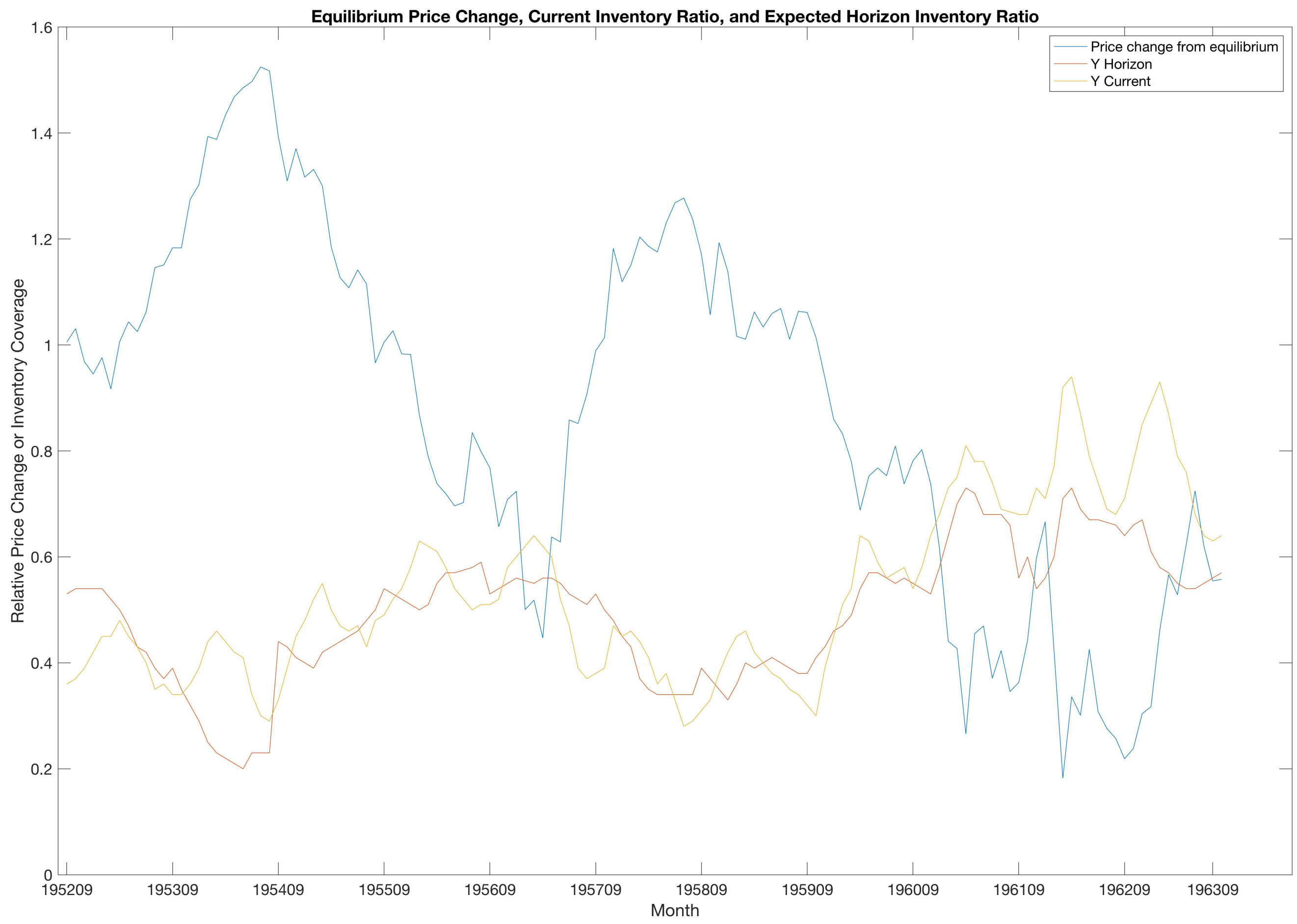


FIGURE 1. SUPPLY OF STORAGE CURVE

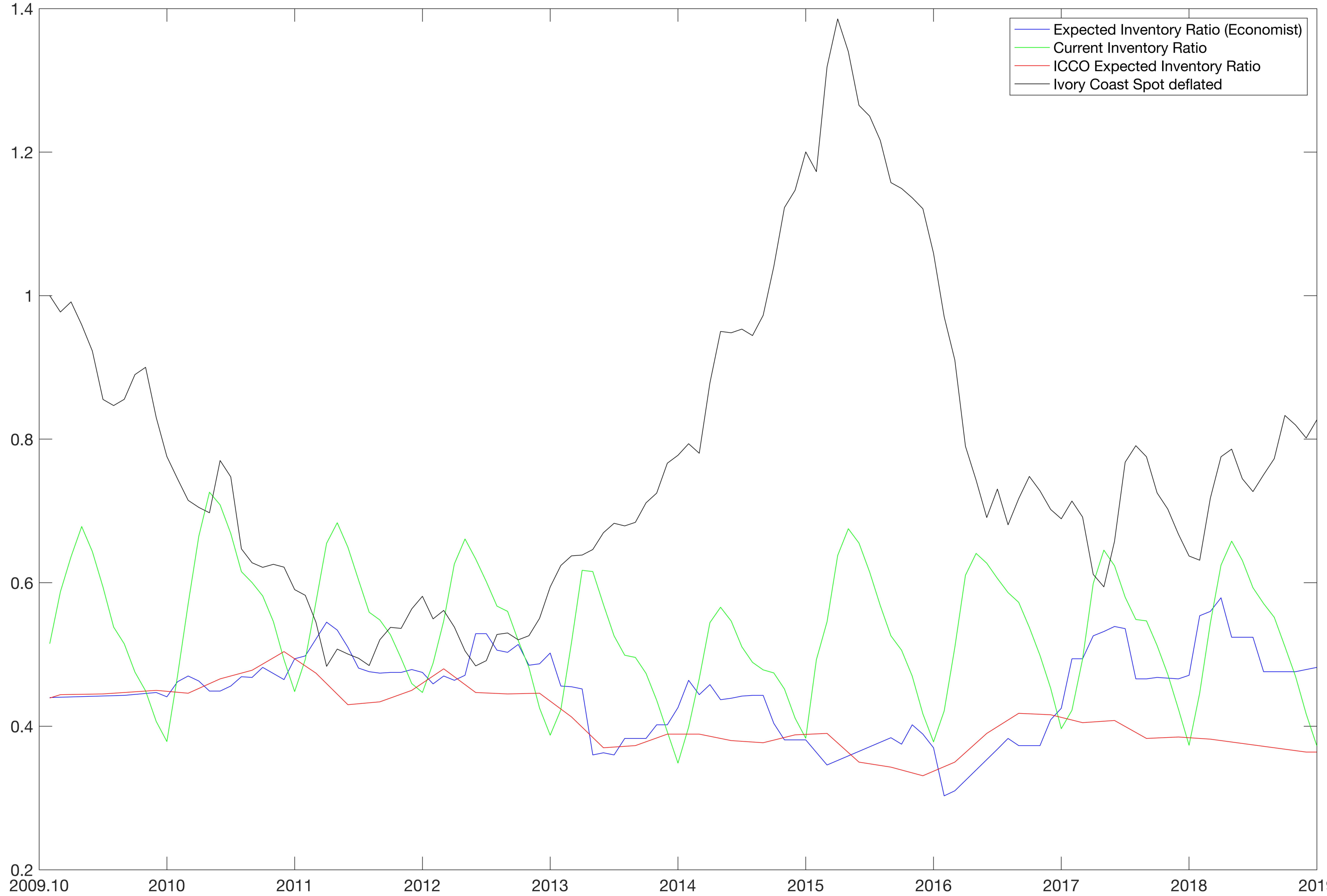
The figure reproduced above is from Weymar (1966).

Figure 3 – Price Change, Current and Year-End estimated inventory ratios from September 1952-August 1963



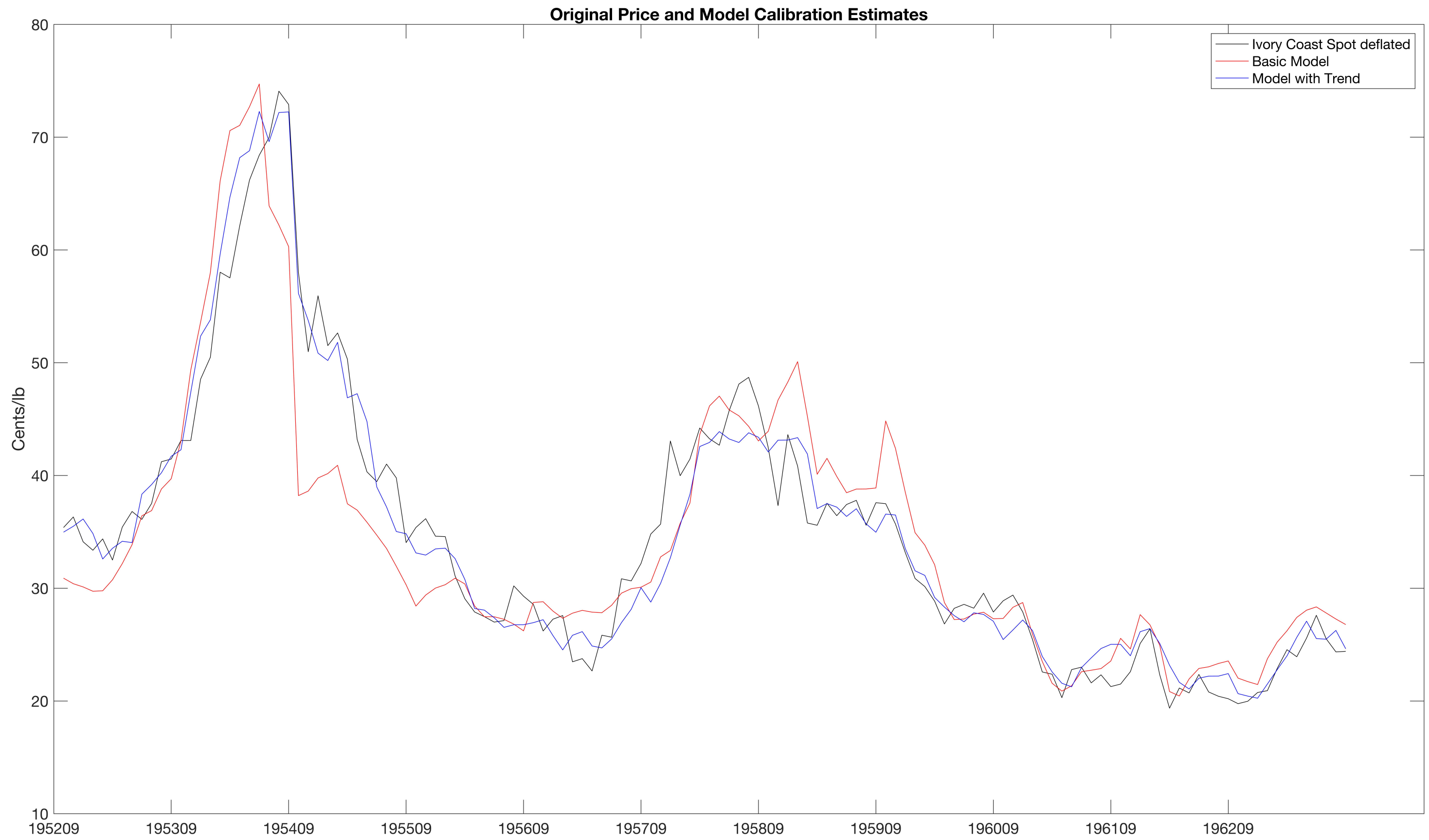
The figure is reproduced from data in Weymar (1966). The change from the equilibrium price is depicted along with the current inventory ratio and the expected horizon inventory ratio. There is a strong inverse correlation between the inventory ratios and the price change.

Figure 4 – Price Change, Current and Year-End estimated inventory ratios from October 2009-September 2019



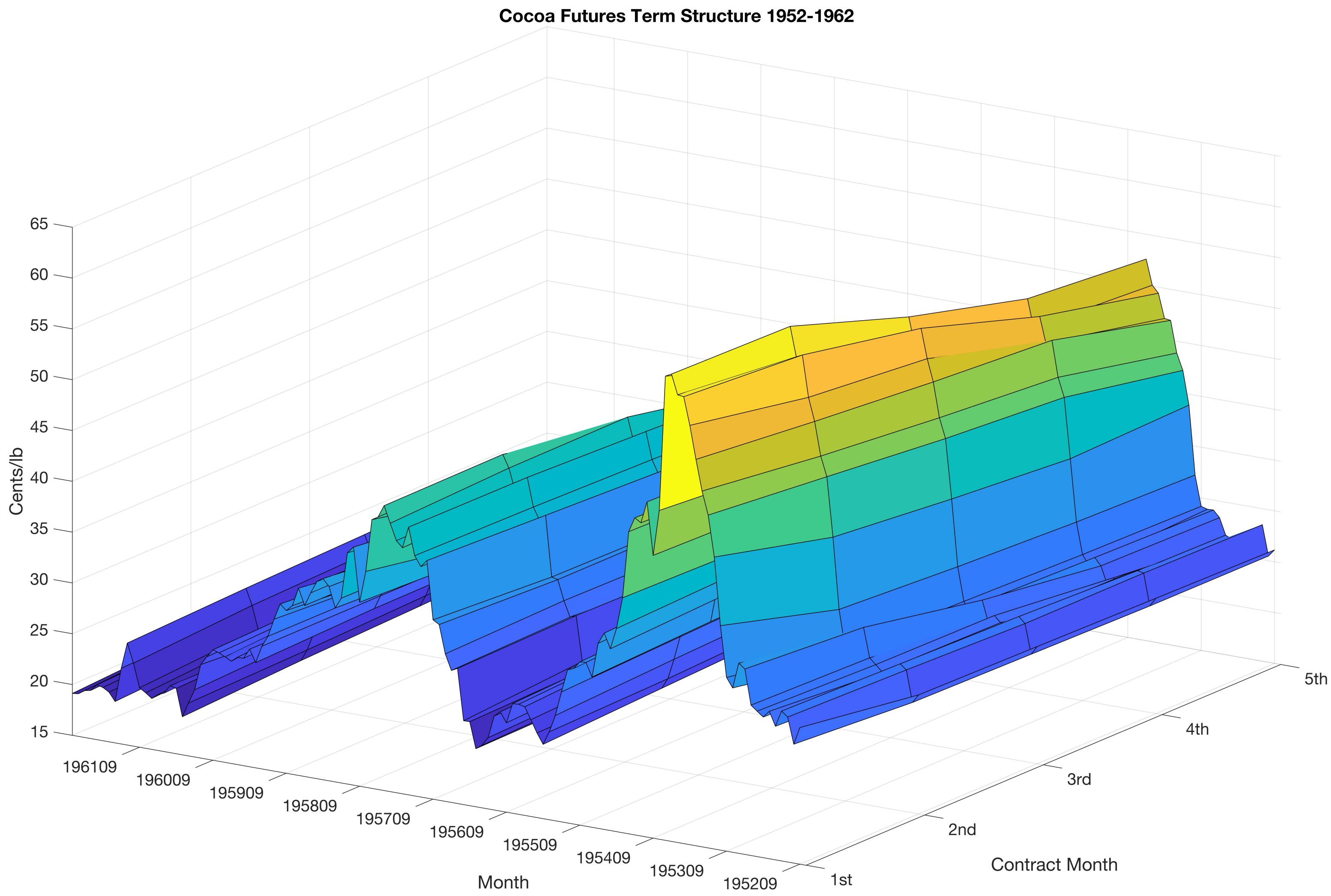
This figure reports the deflated Ivory Coast spot price from 2009 to 2019 along with the inventory coverage ratios estimated by the Economist Intelligence Unit and the International Cocoa Organization. The current inventory ratio is calculated from data sources from CRA Services Ltd. There is an inverse correlation between the inventory ratios and the price change, but significant disagreement in the estimated inventory ratios.

Figure 5 – Cocoa Real Spot Price and Model Estimates September 1959 to October 1963



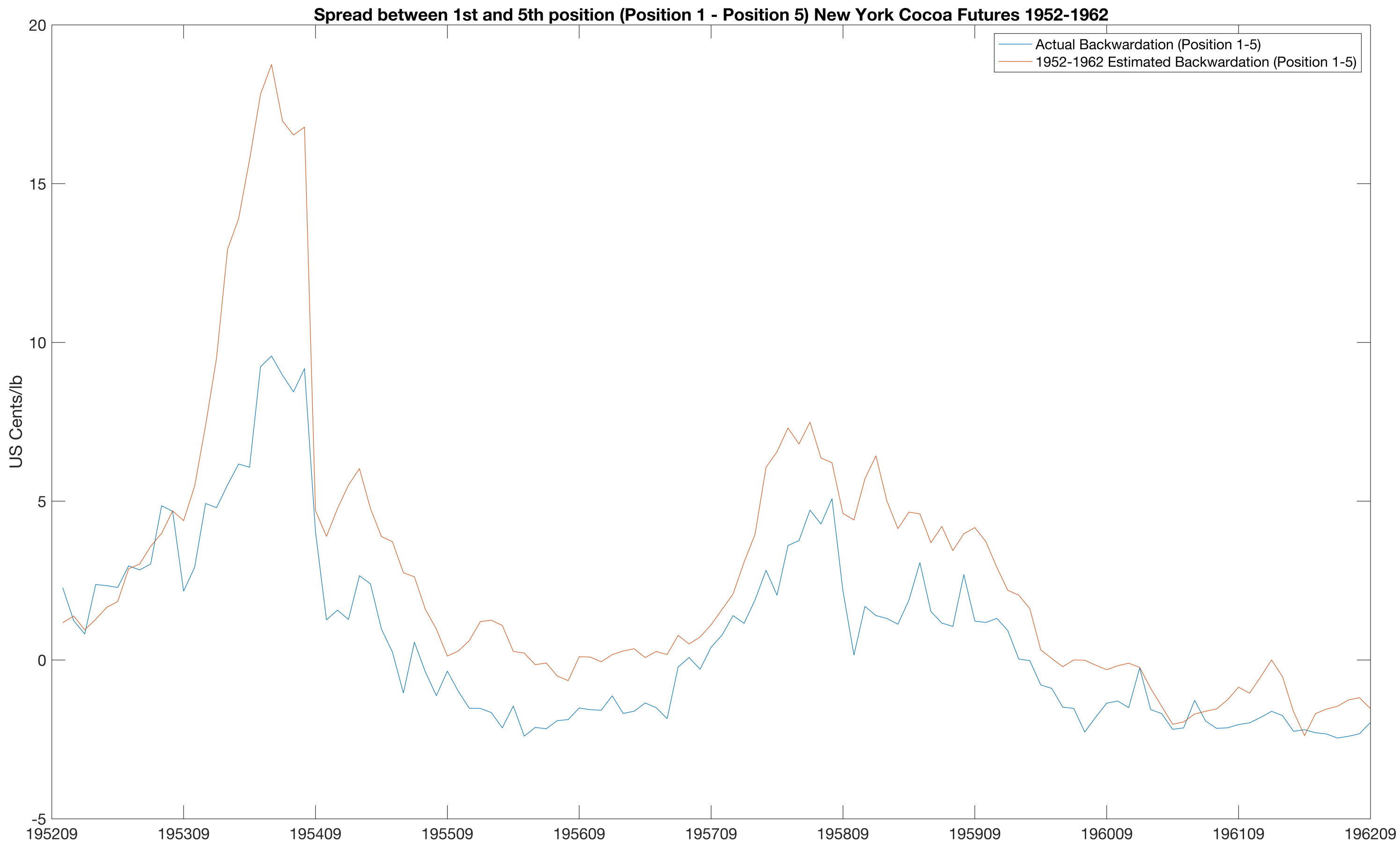
This figure illustrates the cocoa real deflated price, the basic model calibration estimate and the trend following estimate from September 1952 to October 1963.

Figure 6 – Futures Term Structure September 1959 to September 1962



This figure illustrates the actual cocoa term structure on a monthly basis from September 1952 to September 1962.

Figure 7 – Futures Term Structure Spread September 1952 to December 1962

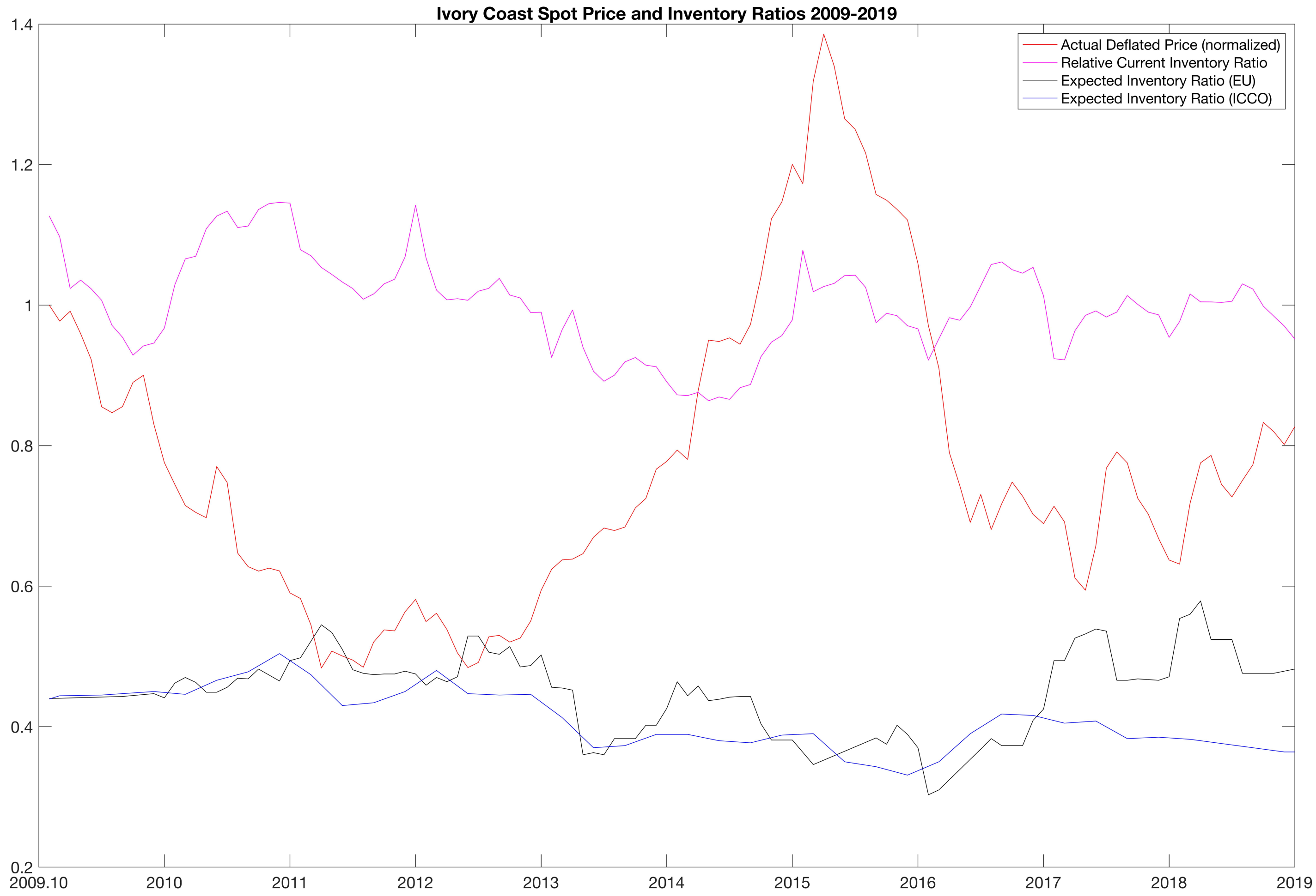


This figure reports the spread between the first and fifth New York cocoa futures contracts on a monthly basis from September 1952 to September 1959.

Figure 8 – Spot and Futures Closing Price September 1952 to December 1962

This figure illustrates the spot accra deflated price and the deflated futures closing price from September 1952 to December 1962.

Figure 9 – Ivory Coast Spot Price (normalized) and Inventory Ratios 2009-2019



This figure illustrates normalized Ivory Coast spot price, the Economist Intelligence Unit expected inventory ratio, the International Cocoa Organization expected inventory ratio, and the relative current inventory ratio over the period 2009-2019.

Table 1 – Seasonal Sums of Inventory Ratio Coefficients from the month in question until the September horizon 1952-1963.

	$\sum c_h$	$\sum d_h$	$\sum e_h$
September	1.2	5.479	4.879
October	0.783	4.189	5.951
November	0.595	3.583	5.797
December	0.578	3.368	5.228
January	0.305	3.352	4.515
February	0.16	2.616	4.231
March	-0.223	3.706	2.336
April	-0.174	2.914	2.052
May	-0.038	2.194	1.91
June	-0.025	2.416	0.709
July	0	1.603	0.397
August	0	1	0

This table shows the seasonal regression coefficients generated by regressing the monthly inventory on the beginning and estimated ending values for each month:

$$\ln(Y^{*h}) = c_h + d_h \ln Y_t + e_h \ln Y_t^{*h_t} + e_t$$

The sums, when combined with the current inventory ratio and the inventory ratio at the end of the horizon interval, capture the inventory behaviour over the full interval. The mean R^2 from the individual regressions is .95 with a standard deviation of .029.

Table 2 – Seasonal Regression Coefficient Sums 1952-1963

Panel 1 – Current Crop Year

		h	DECC	DECD	DECE	h	MARC	MARD	MARE	h	MAYC	MAYD	MAYE	h	JULC	JULD	JULE	h	SEPC	SEPD	SEPE
1952	9	3	0.19	2.85	0.14	6.00	0.70	4.66	1.05	8.00	1.12	5.27	2.34	10.00	1.25	5.59	3.85	12.00	1.20	5.48	5.91
1952	10	2	0.06	1.90	0.04	5.00	0.43	3.70	0.88	7.00	0.75	4.14	2.27	9.00	0.83	4.35	3.85	11.00	0.78	4.19	5.95
1952	11	1	0.00	1.00	0.00	4.00	0.26	2.97	0.79	6.00	0.56	3.44	2.18	8.00	0.63	3.68	3.74	10.00	0.60	3.58	5.80
1952	12					3.00	0.19	2.46	0.43	5.00	0.47	3.09	1.70	7.00	0.53	3.42	3.20	9.00	0.51	3.37	5.23
1953	1					2.00	0.06	1.85	0.14	4.00	0.29	2.79	1.19	6.00	0.32	3.32	2.54	8.00	0.30	3.35	4.52
1953	2					1.00	0.00	1.00	0.00	3.00	0.17	2.15	0.85	5.00	0.17	2.73	2.16	7.00	0.15	2.62	4.23
1953	3									2.00	-0.02	1.94	0.08	4.00	0.01	3.29	0.69	6.00	-0.04	3.71	2.34
1953	4									1.00	0.00	1.00	0.00	3.00	-0.11	2.39	0.52	5.00	-0.18	2.92	2.06
1953	5													2.00	-0.01	1.63	0.38	4.00	-0.04	2.19	1.91
1953	6													1.00	0.00	1.00	0.00	3.00	-0.03	2.42	0.71
1953	7																	2.00	0.00	1.60	0.40
1953	8																	1.00	0.00	1.00	0.00

Panel 2 – Future Crop Year

		h	DECC	DECD	DECE	h	MARC	MARD	MARE	h	MAYC	MAYD	MAYE	h	JULC	JULD	JULE	h	SEPC	SEPD	SEPE
1952	9	3	0.19	2.85	0.14	6.00	0.70	4.66	1.05	8.00	1.12	5.27	2.34	10.00	1.25	5.59	3.85	12.00	1.20	5.48	5.91
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
1953	8	3	0.19	2.85	0.14	6.00	0.70	4.66	1.05	8.00	1.12	5.27	2.34	10.00	1.25	5.59	3.85	12.00	1.20	5.48	5.91

This table depicts the sum of the inventory regression coefficients that allow the estimation of the inventory in the intervening months for the period 1952-1963. The values in Panel 2 are constant, as the starting point is always the end of the current crop year.

Table 3 – Simple Model Replication Results

	a	b_1	b_2	b_3	R^2	F	DW
Weymar's Values (Original)	.761**	.016**	.0368**	.866**	.843	229	.3459
With Correction (Reproduction)	.7482**	.0188**	.0391**	.8548**	.8467	236	.34

These are the results from estimating the following model:

$$\ln(\frac{\overline{P}}{P_t}) = a + b_1 h_t + b_2 Z_t + b_3 \ln Y^{*h_t} + e_t$$

The reproduced results are in close agreement with the original values obtained by Weymar. Differences are attributed to minor errors in transcribing the data from the thesis.

**indicates significance at the 5% level

Table 4 – Trend Following Model Replication Results

	a	b_1	b_2	b_3	R^2	F	DW
With trend following (Original)	.440**	.0175**	.0289**	.455**	.948	86	1.05
With trend following (Reproduction)	.4844**	0.0181**	0.0305**	0.5043**	.945	91.6	.95

These are the results from estimating the following model with an equilibrium price multiplier based on the long term cocoa price trend:

$$\ln\left(\frac{\overline{P}}{P_t}\right) = a + b_1 h_t + b_2 Z_t + b_3 \ln Y^{*h_t} + b_4 \sum_{i=0}^{19} a_i \left(\frac{\Delta P}{P}\right)_i + e_t$$

See Appendix C for the a_i estimates.

**indicates significance at the 5% level

Table 5 – Full Model Replication Results

	a	b_1	b_2	b_3	R^2	F	DW
With Generalized Least Squares and trend (Original)	.265**	.0092**	.0209**	.524**	.85	183	2.19
With Generalized Least Squares and trend (Reproduction)	.52**	.0114**	.0232**	.5462**	.87	34.5	1.79

These are the results from estimating the following model with all variables transformed by a first order autoregressive model for the residuals (see eqn 32):

:

$$\ln(\frac{\overline{P}}{P_t}) = a + b_1 h_t + b_2 Z_t + b_3 \ln Y^{*h_t} + b_4 \sum_{i=0}^{19} a_i \left(\frac{\Delta P}{P}\right)_i + e_t$$

**indicates significance at the 5% level

Table 6 – Seasonal Sums of Inventory Ratio Coefficients from the month in question until the September horizon 2009-2019.

	$\sum c_h$	$\sum d_h$	$\sum e_h$
September	-0.011	2.350	5.895
October	0.014	3.417	4.371
November	0.011	3.869	3.808
December	0.021	5.429	2.980
January	-0.032	4.981	2.913
February	-0.044	4.393	2.521
March	-0.040	3.667	2.139
April	-0.011	3.383	1.547
May	-0.005	2.696	1.279
June	-0.005	2.420	0.669
July	-0.002	1.762	0.296
August	0.000	1.000	0.000

This table shows the seasonal regression coefficients generated by regressing the monthly inventory on the beginning and estimated ending values for each month:

$$\ln(Y^{*h}) = c_h + d_h \ln Y_t + e_h \ln Y_t^{*h_t} + e_t$$

The sums, when combined with the current inventory ratio and the inventory ratio at the end of the horizon interval, capture the inventory behaviour over the full interval. The mean R2 from the individual regressions is .97 with a standard deviation of .02

Table 7 – Model Calibration Results

	a	b_1	b_2	b_3	R^2	F	DW
ICCO values	-3.88 **	0.0135	0.0297	1.0406 **	.348	20.64	.1229
EIU values	-6.677 **	-0.0426	. -0.0712	1.8053 **	.37	22.7	.0954
ICCO and EU	-3.8268 **	0.0433 **	0.0814 **	1.0326 **	. 3727	22.97	.14
With trend following	.7869**	0.004	0. 0044	-0.2456 **	.94	78.8	1.2
With Generalized Least Squares and trend	.1696	.0003	.0001	-0.0771	.88	34	1.67

These are the results from estimating the following model on 2009-2019 deflated Ivory Coast cocoa spot price data:

$$\ln(\frac{\overline{P}}{P_t}) = a + b_1 h_t + b_2 Z_t + b_3 \ln Y^{*h_t} + b_4 \sum_{i=0}^{19} a_i \left(\frac{\Delta P}{P} \right)_i + e_t$$

The values with trend following incorporate the b4 term. The values with Generalized Least Squares are obtained by transforming the variables using the first order autoregressive coefficient on the residuals. Estimates from the International Cocoa Organization (ICCO) and the Economist Intelligence Unit (EIU) are contrasted. See Appendix D for the a_i estimates.

**indicates significance at the 5% level

Table C1 – Coefficients of price changes

Period	Original	Reproduced
1	0.056**	0.046**
2	0.068**	0.062**
3	0.072**	0.088**
4	0.076**	0.085**
5	0.078**	0.085**
6	0.058**	0.070**
7	0.065**	0.071**
8	0.076**	0.078**
9	0.07**	0.067**
10	0.052**	0.046**
11	0.068**	0.069**
12	0.062**	0.076**
13	0.053**	0.053**
14	0.041**	0.043**
15	0.03**	0.023
16	0.031**	0.023
17	0.019**	0.004
18	0.011	0.011
19	0.008	
20	0.004	

This table reports the coefficients of the price changes over 20 four month time periods in the original model, and 18 in the reproduced model. The coefficients are normalized to sum to unity using a scaling factor.

** indicates significance at the 5% level

Table C2 – Coefficients of price changes

Period	Value
1	-0.138**
2	-0.092**
3	-0.058**
4	-0.069**
5	-0.073**
6	-0.049**
7	-0.057**
8	-0.051**
9	0.007**
10	0.088**
11	0.153**
12	0.226**
13	0.277**
14	0.255**
15	0.222**
16	0.172**
17	0.103**
18	0.084**

This table reports the coefficients of the price changes over 18 four month time periods. The coefficients are normalized to sum to unity using a scaling factor.

** indicates significance at the 5% level

Chapter 3

The performance of machine learning in modeling commodity price co-movement and predicting cocoa prices.

Abstract

In Chapter 2 of this thesis, we report that cocoa price movements in the recent time period 2009-2019 are no longer explainable by a fundamental, inventory-ratio based model that worked extremely well in calculating commodity prices during the period 1952-1962. This reduced explanatory power is conjectured to result from the trading activity of large groups of momentum and index traders, whose presence is indicated in the modern time period by the Commitments of Traders report of the Commodity Futures Trading Commission. These groups trade multiple commodities simultaneously, the momentum traders going long and short different groups of commodities, and the index traders rebalancing their portfolios to maintain target percentages of each commodity. The effect of these multi-commodity trades is hypothesized to contribute to the reduced explanatory power of the Weymar model of Chapter 2, by making prices less dependent on fundamental measurements like inventory ratios and resulting in complex multi-commodity price change behaviour. Machine learning was chosen to investigate this hypothesis because of its ability to build multivariate, non-linear probability distributions, and also for machine learning's spectacular success in other fields including recommender systems, speech

recognition, and image recognition. In particular the Recursive Boltzmann Machine (RBM) architecture has been used to create recommender systems that won the Netflix prize for the largest improvement over the internal Netflix recommender system. The RBM architecture is used in this paper to illustrate the co-movement of commodity prices, supporting the assertion that momentum and index traders are affecting price movements. A more advanced architecture, the Recurrent Neural Network Recursive Boltzmann Machine (RNN_RBM) is used to attempt a very difficult problem in finance, predicting tomorrow's price. This attempt is not successful for several possible reasons including the daily granularity being too fine, the large size of the search space with continuous variable inputs, and the absence of detailed worldwide inventory-ratio measurements or other relevant information. The results may also reflect the efficiency of modern commodity markets. Future work with the RNN_RBM might focus on the less difficult problem of modeling financial structures with a natural, contemporaneous sequence such as the sequential contracts in the term structures of commodity prices. Finally, we provide an illustration of fractional differencing, a new machine learning technique from Lopez de Prado with extensive potential applications in Finance. Due to the difficulty in implementing neural network systems, full source code is provided as is common in the machine learning literature.

3.1. Introduction

In my second paper for this thesis, I illustrate that a model based on inventory ratios for the cocoa market has substantially reduced explanatory power in the period 2009-2019, when compared with the period 1952-1963. One major difference is that it is unlikely that the earlier historical period had index traders or complex multi-commodity momentum strategies due to the computational difficulty.

Recent studies have illustrated that a large flow of index funds has entered into commodity markets since 2000 (Henderson (2015), Tang and Xiong (2012) , Main et al (2018)). In addition, the returns of a large component of traders, Commodity Trading Advisors, are closely related to the returns from momentum strategies (Bhardwaj et al (2014), Bollen et al (2019)). These heterogeneous traders have been accused of drawing prices away from fundamental measurements like inventory and causing price bubbles (Shu (2009)). However, there is substantial debate (Irwin and Sanders (2011))

Traders in the commodity markets can now be categorized into three groups: commercial (dealers and processors), non-commercial (speculators, including momentum traders), and index traders. The Supplemental Commitment of Traders report published since 2006 provides a breakdown into these groups for US Commodity open interest. In 2020, the open interest totals of each group are of the same magnitude (i.e. (1:2:1 for Non-Commercial:Commercial:Index)), indicating that each group has a large impact on price expectations and changes. In other words, 25% of trading is Non-Commercial Momentum Trading, 50% is industry trading, and 25% is index trading. There are subtle interactions between the groups. Kang et al (2019) report that commercial traders provide liquidity for impatient speculative traders.

Fundamental traders will tend to base their trading on fundamental quantities like inventory ratios. However, as illustrated by Adhikari et al (2018) there is a statistical relationship between the inventories of closely related products and the price changes of a given commodity. For instance, soybean traders pay attention to the inventories of corn and wheat, as the same land can be used to produce the three grains.

Momentum traders are heterogeneous in regard to time frame (ranging from HFT to annual), but typically will short a selection of poorly performing commodities and go long a selection of strongly performing commodities. This means that the trading decisions are made over information on several commodities instead of a single commodity.

Index traders will rebalance their funds to maintain a particular allocation of market value to each component. In addition, the flow of capital into or out of the funds will cause a trading decision across a large number of commodities simultaneously. Index traders in 2006 tended to be substantially net long, but in 2020, the ratio of long to short positions is close to 2:1 (CFTC 2020 Commitment of Traders Supplemental Report).

Thus, the information universe that traders are using is not limited to data on a single commodity. Fundamental traders will use inventory of closely related commodities. Momentum traders will use returns on perhaps up to 10 commodities over varying time frames. Index traders may use price changes on the entire universe of approximately 40 trading commodities. To analyze this situation, it is necessary to model a multivariate probability distribution for commodity prices, with possible non-linear interactions between groups of commodity prices. This is possible with

Machine Learning, in particular Recursive Boltzmann Machine architectures, but difficult with traditional linear regression techniques.

Recently, an intense interest has been placed on using Machine Learning in Finance, due to the spectacular success of ML in other fields. In particular, RBM architectures won the Netflix prize, and are used throughout the industry by large on-line marketplaces such as Amazon or Ebay. The phenomenal jump in speech recognition accuracy, demonstrated by products such as Amazon's Alexa, are driven by advanced Recurrent Neural Networks. Alpha-Go, a reinforcement-based learning neural network, recently defeated 18-time world champion Lee Sedol. The most recent success is in protein folding by AlphaFold, a network from Deep Mind, the inventors of Alpha Go[1]. See Appendix A for a discussion of the advantages and disadvantages of Machine Learning in comparison with traditional econometric tools.

In single time series forecasting, after comparing the post-sample accuracy of popular ML methods including multi-layer perceptrons and recurrent neural networks with eight traditional statistical ones including ARIMA and Comb methods, Makridikakis et al (2018) finds that the former are dominated by the latter across both accuracy measures used and for all forecasting horizons examined. However, Makridikakis only examines the single time series case.

In a multi-factor environment, Li (2019) reports that when both text features and financial features are utilized, more accurate crude oil price forecasts are possible.

---

[1] AlphaFold, using AI for scientific discovery, https://www.deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery

Sirignano (2019) using a Deep Learning system indicates “The universal model --- trained on data from all stocks --- outperforms, in terms of out-of-sample prediction accuracy, asset-specific linear and nonlinear models trained on time series of any given stock”. Zhao et al (2017) uses a deep network model, stacked denoising encoders, to model the nonlinear and complex relationships of oil price with its factors. Using 198 exogenous variables, Zhao et al report more accurate crude oil price forecasts. Ghodussi et al. (2019) provide a review of machine learning in finance and determine that support vector machines, artificial neural networks, and genetic algorithms are among the most popular techniques used in energy economics papers. Shah et al (2019) in a review also indicate that “Application of machine learning techniques and other algorithms for stock price analysis and forecasting is an area that shows great promise” From the literature, Machine learning appears to particularly effective when the non-linear relationships between numerous time series are taken into account.

In this paper, we choose the Restricted Boltzmann Machine (RBM) architecture to analyze commodity prices and the Recurrent Neural Network Restricted Boltzmann Machine (RNN_RBM) to forecast the cocoa price. The RBM and RNN_RBM models are selected because of the evidence in the literature that ML approaches are capable of forecasting security prices, the effectiveness and widespread use of the RBM architecture in recommender systems, the usage of the RBM as a feature extractor in classification systems and because of the relative simplicity and transparency of the RBM.

Several studies have incorporated the RBM as part of a larger classification system. Assis et al (2018), using a RBM as a feature extractor and a support vector machine (SVM) as a classifier, illustrate that results in identifying stock market trends are better with the combined architecture than with a SVM alone. Liang (2017) et al also find improved performance using the features extracted by the RBM in conjunction with SVM, random-forest, or logistic regression classifiers.

A recommender architecture is useful because the system can be trained, for instance, by showing movie ratings from a large number of users. From the ratings, the network builds a set of hidden feature detectors. Then, when a new user is presented to the system with an incomplete set of ratings, the system will fill in the missing ratings thus making recommendations. In a similar manner, an RBM architecture can be trained on price changes from a large selection of the commodity universe and then queried to find out whether commodities tend to change in price at the same time.

In this paper, an RBM is utilized to analyze a binary matrix encoding the co-movement of changes in monthly commodity prices. Each absolute commodity price change greater than a certain percentage amount is encoded as a 1, and the lesser price changes as zero. Cross-validation of the RBM results indicate that the groups of commodity prices tend to change together.

The RNN_RBM combines a Recurrent Neural Network with a RBM to allow sequencing of multivariate time series, and one-step-ahead predictions. In essence, the RBM recognizes features in multiple time series, and the RNN learns the probability of each feature following or preceding another feature. This architecture has been successfully used in areas diverse as polyphonic music transcription (Boulanger-Lewandowski (2012), forecasting cellular data traffic (Bäärnhielm, A. (2017)), and predicting human motion (Taylor et al (2007)). The RNN_RBM has also been used to predict predict the trend of stock prices by focusing on news events with long-term effects (Yoshihara et al (2014). Zhang (2015) reports that the RNN_RBM effectively incorporates 10-K report information into trend predictions but does not investigate the transaction cost, slippage, and borrowing costs needed to implement a real world trading strategy.

We use a RNN_RBM to estimate the continuous multivariate distribution of cocoa prices, cocoa inventory, coffee prices, coffee inventory, and series representative of the information set of momentum and index traders. The multivariate distribution is then used to make one-day-ahead predictions for cocoa prices. This is a very difficult problem and the results are not superior to a baseline prediction of tomorrow's price as today's price, which may simply be a reflection of the efficiency of modern commodity markets. A number of potential issues with the model are highlighted but we suggest that the RNN_RBM would be better utilized in studying the term structure of multiple commodities, a problem that has a natural sequence in the order of the futures contracts, and does not require predicting the future.

3.2. Neural Network Theory

3.2.1 The Intuition behind Neural Networks

One of the simplest useful forms of a neural network is a two neuron perceptron that transforms input X into output Y with the following equation:

$$y = \sigma(b_2 + w_2\sigma(b_1 + w_1 x)) \qquad (1)$$

The input x is multiplied by a weight w_1 and added to a bias b_1. This amount is fed through a non-linear activation function σ, multiplied by weight w_2 ,added to bias b_2 and fed through the same activation function. With a set of training examples consisting of a loss function, a set of inputs, and a set of correct outputs, the network can be iteratively trained by adjusting the weights for each separate input. The adjustments to the weights are controlled by a small learning rate. A typical loss function is the squared error of the correct value minus the output. First, the partial derivatives of the loss function to each of the weights and biases are determined. In simple networks these can be calculated in closed form and in more complicated networks the partial derivatives can be approximated numerically. Then, for each input in the training set, each of the weights and biases are adjusted by a small amount to reduce the squared error. This structure allows the system to arbitrarily partition the input space. Yadav(2016) provides a detailed example of partitioning two concentric circles of data points, a task impossible with linear classification systems.

A Recursive Boltzmann Machine (Hinton (2010) utilizes neurons in a two layer structure. The hidden layer is composed of feature detectors, and each hidden neuron is associated with the probability that a particular input matches a particular feature. In the context of movie recommendations, a RBM recommender system is trained by feeding the network a large number of movie ratings from a large number of users.

The hidden nodes or neurons serve as feature extractors, in a similar manner to how Principal Components Analysis selects a smaller number of components to approximate a vector space (see Bu et al (2015)). These hidden nodes are effectively trained to represent abstract categories like Science Fiction, Adventure, or Crime Drama. After the system is trained, a partial new set of ratings from a new user are input into the system, and the RBM reconstructs an estimated complete set of ratings from its internal probability distribution. The new ratings for the previously unrated movies are sorted and become the recommendations for the user.

The Recurrent Neural Network Recursive Boltzmann Machine (RNN_RBM) (Boulanger-Lewandowski (2012)) combines a RBM with another recurrent network structure that allows sequencing of inputs and features over time. In the context of movie recommendations, the RNN_RBM would be able learn the changes in user preferences over time. For instance, the RNN_RBM might deduce that a teenager that liked Adventure movies would prefer Crime drama as an adult. The RNN_RBM adds an additional group of hidden units to the RBM that allows for the learning of sequential information.

Training of the RBM is done by establishing training and holdout sets, and testing is done by cross-validation. In this paper, the training set for the RBM is randomly generated as 90% of the data with the holdout set comprising the remaining 10%. For each split of the data into training and test sets, the network is fed batches of elements of the training set for a number of epochs, and the psuedo-likelihood of the holdout set is compared with the training set. This procedure is done ten times and the likelihood values are averaged to get a final result.

For the RNN_RBM, the sequential nature of the data must be maintained and the data is split into ten approximately equal sized batches. The RNN_RBM is trained on the first nine batches and the remaining batch is used to compare the accuracy of predictions. Prediction accuracy is calculated as the sum of the squared errors between the actual and predicted values over the test set.

3.2.2 A Simple Feedforward Network

The simplest neural network is a two layer feedforward network that is trained by backpropagation of errors calculated by gradient descent, depicted in Figure 1.

[INSERT FIGURE 1 ABOUT HERE]

The hidden layer is fully connected to the input layer, and the output node is fully connected to the hidden layer. Input nodes receive data directly. Hidden nodes weight and sum each input, add in a bias, and adjust the total by a non-linear, differentiable activation function such as the sigmoid function. A non-linear activation function is necessary or the network is only capable of simple linear discrimination. The cost of a set of inputs to the network is a monotonic differentiable function of the difference between the actual output and the desired output. The network is trained by calculating the gradient of the output with respect to each weight in the network, and then adjusting the weight in the direction that reduces the cost function(Hecht-Nielsen

(1992), Rumelhart et al. (1995)). This relatively straightforward system has been shown capable of universal function approximation (Hornick (1991), Kidger (2020)).

3.2.3 The Restricted Boltzmann Machine (RBM)

The Restricted Boltzmann Machine (RBM) is a two layer energy based network that learns in an unsupervised manner (see Figure 2) (Smolensky (1986), Hinton (2002)).

[INSERT FIGURE 2 ABOUT HERE]

Each visible and hidden unit has a bias, and the visible and hidden layers are completely connected. However, there are no intra-layer connections between hidden or visible unit. The network will converge to an energy minimum by Gibbs sampling. First the visible units are set with a particular configuration (the input), the hidden unit probabilities are calculated and the hidden units are sampled from the probabilities. Then the reverse process of setting the visible units from the units is done. This process is repeated until equilibrium, and the output is the converged value of the reconstructed visible units (although the hidden units can also be used as outputs for subsequent layers). Explicitly, for a binary RBM, the hidden probabilities are calculated as

$$p(h_j = 1) = \sigma(b_j + \sum_i v_i w_{ij}) \quad (2)$$

where sigma is the sigmoid distribution, h_j are the hidden units, v_i are the visible units, a_i are the hidden biases, b_j are the hidden biases, and w_{ij} are the symmetric weights.

Then, the hidden states are sampled from this probability distribution. The reverse is done to calculate the visible probabilities:

$$p_i = p(v_i = 1) = \sigma(a_i + \sum_j h_j w_{ij}) \qquad (3)$$

In the case of continuous inputs, the input time series are normalized to zero mean and unit variance, and the visible probabilities are used in place of sampling. In other word, the hidden binary units are still sampled, but the sampling step is omitted for the visible units (Hinton (2010)).

Hopfield (1982), who uses the terminology Harmonium to describe an RBM, specifies the energy of a joint distribution of hidden and visible states as:

$$E(\mathbf{v}, \mathbf{h}) = -\sum_{i \in \text{visible}} a_i v_i - \sum_{j \in \text{hidden}} b_j h_j - \sum_{i,j} v_i h_j w_{ij} \qquad (4)$$

where v_i are the visible unit values, h_j are the hidden unit values, a_i are the visible biases, b_j are the hidden biases, and w_{ij} are the symmetric weights between the visible and hidden units.

The Boltzmann distribution from statistical mechanics states that the probability of a particular state is related to the energy of that state by the following:

$$p_i \propto e^{-\frac{\varepsilon_i}{kT}} \qquad (5)$$

where pi is the probability of being in a particular state, e_i is the energy of that state, k is the Boltzmann constant, and T is the temperature.

Reasoning by analogy, the probability of a given joint distribution of a Hopfield Harmonium is :

$$p(\mathbf{v}, \mathbf{h}) = \frac{1}{Z}\, e^{-E(\mathbf{v}, \mathbf{h})} \qquad (6)$$

where v is the visible vector, h is the hidden vector, E is the energy function, and Z is the partition function or the normalization constant.

The normalization constant is intractable to estimate as it requires the calculation of the energy in every possible configuration. The marginal probability of a visible(hidden) vector is derived by summing over the hidden (visible) states.

The derivative of the log probability of a visible vector can be derived as (Hinton (2002), Hinton (2010):

$$\frac{\partial \log p(\mathbf{v})}{\partial w_{ij}} = \langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model} \qquad (7)$$

leading to a training rule of:

$$\Delta w_{ij} = \epsilon(\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model}) \qquad (8)$$

where epsilon is a learning rate, and the expectations <> are taken with regard to the data and then the model.

The problem with this learning rule is that to determine the expectation of the joint distribution under the model, Gibbs sampling has to be run for numerous steps, making the process computationally intractable. Hinton (2002) in one of the most important contributions to the field of Neural Networks, discovered that a far simpler procedure, Contrastive Divergence (CD), would be an adequate approximation of the precise learning rule. Contrastive Divergence substitutes only a single Gibbs sampling step for the normally lengthy equilibrium calculation. It is not completely clear what gradient CD is following (Sutskever and Hinton (2010)). In Hinton's words (Hinton (2010), *"Nevertheless, it works well enough to achieve success in many important applications"*. Indeed, RBM architectures were among the top performing entrants in the Netflix prize competition[2].

[2] The Netflix prize competition awarded one million dollars to an algorithm that exceeded the Netflix internal recommendation engine in accuracy by 10% .https://en.wikipedia.org/wiki/Netflix_Prize

With binary units, the RBM learns the distribution of binary patterns in the data. With continuous units, the RBM learns the joint distribution of the visible units. In this paper, we are interested in approximating the joint distribution of scaled, continuous commodity financial time series.

A reference implementation of a Bernoulli RBM is in the sklearn Python library, sklearn.neural_network.BernoulliRBM.

The RBM has achieved notable success in recommender systems. In this architecture, the RBM is trained on a set of binary inputs such as movie ratings by users. The trained RBM is then fed a new user without ratings for several movies and run through a 1-step Gibbs sampling chain. The probability distribution learned by the RBM will then "recommend" ratings for the new movies not yet seen by the user. This type of architecture will be used in this paper to study the coincidence of changes to commodity prices.

3.2.4 The Recurrent Neural Network (RNN)

The standard feedforward networks and RBMs cannot model changes over time. Rumelhart (1986) lays the groundwork conceptually for unfolding a single feedforward network into multiple layers in order to process sequences. Werbos (1990) devised the Backpropagation Through Time (BPTT) algorithm that is used to train Recurrent Neural Networks (RNNs). A typical RNN is illustrated in Figure 3.

[INSERT FIGURE 3 ABOUT HERE]

It is not necessary that a sequence be related to time, the sequence only needs to have an order. The RNN consists of a simple feedforward network replicated for each step of the sequence. In other words, there are input nodes, hidden nodes, and output nodes for each sequential step. The critical innovation in a RNN is the transfer of the hidden state to the next node group. The hidden state at each step is calculated as a combination of the input and the hidden state from the previous step. The initial "hidden" input to the first time step can be learned as a parameter of the network. Each node group shares the same matrices (W and U in the figure)

The BPTT algorithm calculates the gradient of the cost function with respect to the output for the last step in the sequence, and then backpropagates the gradients through each time step in reverse. Detailed algorithms in Python for implementing RNNs can be found in Behane (2018) and Weidman (2019)

RNNs underly many of the recent spectacular successes of neural networks, including the speech recognition performance of products like Alexa from Amazon or Siri from Apple.

3.2.5 The Recurrent Neural Network Restricted Boltzmann Machine (RNN_RBM)

The Recurrent Neural Network Restricted Boltzmann machine (Boulanger-Lewandowski et al. (2012)) combines the RBM and RNN into a single network as detailed in Figure 4.

[INSERT FIGURE 4 ABOUT HERE]

The $u^{(i)}$, W_{uu}, and W_{vu} are the hidden nodes and weight matrices of the RNN component of the RNN_RBM. The initial RNN unit $u^{(0)}$ is a parameter to the network and can be learned.

A sequence of input $v^{(t)}$ is presented to the network, one element at a time. The RNN hidden units are calculated as

$$u_t = \sigma(b_u + v_t W_{vu} + u_{t-1} W_{uu}) \tag{9}$$

where b_u are the hidden unit biases.

The same weight matrix is used for the RBMs at each time step, but biases for the RBMs at each time step are calculated from the visible inputs:

$$\begin{aligned} b_h^t &= b_h + u_{t-1} W_{uh} \\ b_v^t &= b_v + u_{t-1} W_{uv} \end{aligned} \tag{10}$$

Each RBM runs a separate Gibbs chain of 15 to 25 cycles to determine the reconstructed output for that time step. The changes to W and the initial biases (b_h and b_v) of each RBM are calculated and summed using Contrastive Divergence. The other parameters, W_{vu}, W_{uv}, W_{uh}, W_{uu}, b_u are adjusted by BPTT. Explicit formulas for the changes are provided by Boulanger-Lewandowski et al. (2012). In the source code for the RNN_RBM (see Appendix A), the weight adjustments are automatically calculated by Theano[3].

The RNN_RBM, after being trained, can generate one-step ahead predictions by initializing the biases of each conditional RBM with the previous days observations, and running a 25-step Gibbs chain until convergence. The visible units are initially set to zero.

In theory, the conditional RBMs at each time step learn the multivariate distribution of the inputs, and in the simplest case, the RNN component learns the transition probabilities between the different time samples of the learned multivariate distribution. However, the RNN is capable of learning longer term dependencies between the elements in the input sequence.

Complete source code for the RNN_RBM utilized in this paper is provided in Appendix A. The RNN_RBM is implemented in Python and uses the Theano deep learning library. Theano provides automatic differentiation capability to calculate and adjust the weights and biases of the network. In addition, Theano allows for execution of the compiled network on a graphical processing unit like the Nvidia GTX 1080, which is can dramatically speed up training time for systems with hundreds of hidden units.

---

3 The Theano machine learning library was developed by the MILA lab at the University of Toronto under Yoshua Bengio. It is available at deeplearning.net, and also on github.

3.2.6 Basic Validation of the RBM and RNN_RBM

A critical step in using the RBM or RNN_RBM is deriving simpler test sets with known outputs to validate that the network is functioning properly. It is not possible simply to monitor the weights or the outputs from complex inputs. These test datasets are not normally included in published papers, but due to their importance in establishing baseline functionality are included here. (Surprisingly, some implementations of the RNN_RBM found on github and the web fail to pass these basic tests)

For the RBM, the simplest training and test set is a sequence of binary input vectors with only a single unit on. For example, with six visible units, the set of inputs would be:

100000

010000

001000

000010

000001.

This sequence would be repeated several hundred times, and fed into the RBM. The correct output for the RBM is then simply the input. If 100000 is fed into the network, then one cycle of Gibbs sampling must produce the same output.

For the RNN_RBM, which detects the transition probabilities over time, the test sequence is similar:

110000

011000

001100

000110

000011

In a similar manner, the RNN_RBM is trained on this sequence repeated numerous times. Then, in the generation phase, the RNN_RBM must produce the identical output.

For continuous time series normalized to zero mean and unit variance, an appropriate test dataset is a sequence of sine waves with different periods.

3.2.7 Fractional Differentiation of Time Series

It is well known that financial time series are typically not stationary, hetero-skedastic, and drawn from a non-normal distribution. The non-stationary aspect can lead to spurious regression results and incorrect inference. The standard method of dealing with unit roots is either to look for co-integration or to integer differentiate the time series. Lopez de Prado (2018) suggests a third, intriguing approach, fractional differentiation. This approach was first applied by Hosking (1981) to ARIMA series, but has not received much focus to date. Using a binomial series expansion of the backshift operation, Lopez De Prado illustrates that it is possible to differentiate a series by only partially subtracting previous values. In other words, the fractional constant can range from 0 to 1, with 1 being full integer differentiation. The advantage of this method is that some of the long-term memory of the series is preserved with fractional differentiation whereas with integer differentiation all memory beyond one step is lost. Lopez de Prado speculates that this may be why the efficient market hypothesis has gained such stature in finance. It is possible to use the Augmented Dickey-Fuller (ADF test) on an iteratively fractionally differentiated time series to find the amount of differentiation necessary to remove a unit root. In this paper, we use the FracDiff library or https://github.com/simaki/fracdiff, inspired by Lopez de Prado (2018). Data preparation source code is provided in Appendix B.

3.3. Data

Commodity Futures prices and open interest for cocoa, coffee, and sugar are kindly made publically available (on acceptance of license terms) by the Intercontinental Exchange (ICE) at

https://www.theice.com/FuturesUSReportCenter.shtml

This data spans the first five contracts with daily frequency from January 2000 to the present.

Monthly pricing data from the 1960s to the present, the “commodity pink sheet”, is obtainable from the World Bank at https://www.worldbank.org/en/research/commodity-markets.

End of Month warehouse stocks for both cocoa and coffee at licensed US warehouses are also downloadable from ICE.

3.4. Results

3.4.2 Change Analysis of Monthly Commodity Prices

First, a Restricted Boltzmann Machine (RBM) was used to analyze the binary-encoded monthly changes in commodity prices for co-movement. Monthly price changes for forty-seven commodities from the World Bank Pink Sheet were analyzed. Only a single commodity was included from each commodity group. For instance, Tea Columbo was selected from the Tea Group which also includes Tea Mombasa, and Tea Kolkata. The intention is to detect co-movement of changes in disparate commodities.

The commodity price change was encoded in the following manner. If the absolute value of the price change since the previous month was greater than 4%, then the commodity was assigned a one, otherwise a zero. A sample of the input data from 2010 to 2013 is depicted in Figure 5. The entire dataset runs from January 1960 to June 2020. A RBM was trained ten times for 100 epochs with a random withholding of ten percent of the data as a training set. The psuedo-likelihood of a data sample is an approximation of the likelihood function which measures the goodness of fit of the

RBM to a statistical sample. The average pseudo-likelihood of the test data set (mean -26.57) was consistently closer to zero than the pseudo-likelihood of the RBM before training on the training data(mean -30.32), indicating that the fitted RBM learned co-movement of groups of commodity prices in the training data and extrapolated to the test data set. Source code for this system is included in Appendix C.

[INSERT FIGURE 5 ABOUT HERE]

3.4.2 Cocoa One-Step-Ahead Price Prediction

With the RBM results indicate that there is co-movement in commodity prices, we now attempt the (significantly) more difficult problem of predicting the one-day ahead cocoa price with the RNN_RBM.

Six data series are selected from the futures data and two derivative series are created from the commodity pink sheet. The selected series are Nearby Cocoa Close, Total Cocoa Open Interest, Cocoa End of Month Warehouse Stocks, Nearby Coffee Close, Coffee Total Open Interest, Coffee End of Month Warehouse Stocks. The End of Month stocks are linearly interpolated to get daily values.

Coffee is chosen because of the overlap of coffee and cocoa production regions around the globe. The weather in the tropical band where cocoa and coffee are grown will affect both crops, and there is substitution between the crops dependent on which crop is more profitable to farm in a particular year. The Total Open Interest was identified by Weymar (1968) as an important component of a fundamental based model. The End of Month Warehouse stocks were chosen as representative of the current inventory of cocoa or coffee. These series are not exhaustive, but are indicative of the information used by fundamental cocoa traders.

The first derivative series is the previous month return of cocoa relative to the return on the following commodities: Sugar World; Crude oil, Brent; Coal, Australian; Natural gas, US; Coffee, Arabica; Tea, Colombo; Palm oil; Soybeans; Maize; Rice, Thai 5% ; Wheat, US SRW; Orange; Beef; Cotton, A Index; Rubber, SGP/MYS; Aluminum; Copper; Lead; Tin; Nickel; Zinc; Gold; Platinum; and Silver. These are commodities with liquid futures contracts representative of the information universe of momentum and index traders. The relative return of cocoa to the mean return would indicate the amount of rebalancing to be done by index traders.

The second derivative series is the index position of last month's cocoa return relative to the last month returns of the commodities listed above. This index is related to the long/short mix chosen by momentum traders who will typically go long the group of top performing commodities and short the group of worst performing commodities.

All the series are scaled to a zero mean and unit variance. The data is split into two sets for training and one-step-ahead prediction.

Table 1 illustrates the fractional differencing results. The Cocoa Nearby Close, Cocoa Stocks are stationary in levels, and the Cocoa Relative Return and Cocoa Position are stationary by construction. The Cocoa Totals Open Interest and the Coffee Nearby Close require only seven-tenths differentiation. The Coffee Totals Open Interest requires eight tenths, and the Coffee Stocks require full integer differentiation. The advantage of using fractionally differentiated series is that the RNN_RBM does not need to learn long term non-stationary trends in the input series.

[INSERT TABLE 1 ABOUT HERE]

Figure 6 reports heatmaps of the correlations between the level series and the fractionally differentiated series. The fractionally differentiated series appear to maintain long run price information. For example, the level correlation of the Coffee Nearby Close and the Cocoa Nearby Close is 0.596, and the fractional correlation with the 0.7 differentiated Cocoa Nearby Close is .46. The correlation between the Cocoa Nearby Close and the fully differentiated Coffee Stocks changes by a much larger percentage amount from -0.51 to -0.19. The heatmaps also confirm significant correlations between the time series, including a 0.6 correlation between Coffee Open Interest and Cocoa Open Interest, and support the choice of time series for analysis.

[INSERT FIGURE 6 ABOUT HERE]

The RNN_RBM is trained for 200 epochs with a batch size of 100, 50 hidden units and a learning rate of .01. The results from one–step-ahead prediction for both level and fractionally differenced data are depicted in Table 2. The residual sum of squares (RSS) is calculated as the sum of the squared difference between the actual value and the prediction. The RSS for the model is compared with the RSS for the baseline prediction of using today's price as tomorrow's prediction. Kolmogorov-Smirnov

(KS) two-sample p-values are reported for the predictions and the actual values. The KS test fails to reject the null hypothesis that the predictions and the actual values are drawn from the same distribution for both the level and fractionally differenced data. Figure 7 is a graphical representation of the input and predicted time series. The predicted series incorporates a large amount of noise relative to the actual values.

[INSERT TABLE 2 ABOUT HERE]

These results indicate that the RNN_RBM is not particularly effective in this particular application of predicting one-day ahead cocoa prices and is unable to use the observed correlations to improve on a baseline prediction of tomorrow's price as today's price. Despite the co-movement of commodity prices detected by the RBM, the RNN_RBM was not able to learn the time-dependency of the hidden features. There are four likely reasons for this lack of success.

First, the co-movement in commodity prices is on monthly rather than daily data. A further analysis would use monthly measurements as input to the RNN_RBM, though this would be complicated by the shortness of the monthly data series over approximately ten years. One hundred and twenty data-points would be considered a small amount of data to train a machine learning system. If the time horizon were increased from one day, then the impact from such factors as the market's liquidity and transactions costs should also be considered.

Second, the RNN_RBM is using continuous rather than binary data which is known to complicate the learning process (Hinton (2010)). The input space is significantly larger with continuous than binary variables. One possibility would be to encode the changes in commodity prices in a similar manner to the RBM data, and attempt to sequence the changes with the RNN_RBM.

Third, there is a great deal of information that is not input to the model, and might produce superior results if utilized. For instance, the inventory ratio calculated in Chapter 2 of this thesis uses substantially more raw information than the warehouse stocks and the total open interest.

Finally, predicting future prices is one of the most difficult problems in finance due to the efficiency of modern markets and it is not surprising that the RNN_RBM is unable to easily accomplish this task.

A more fruitful application of the RNN_RBM for future research would be to analyze the multi commodity term structure. Each commodity has several contracts with a natural sequential order, and the RNN_RBM could be used to predict the value of the final contract for a particular commodity like cocoa based on the structures of other related commodities. This system would then be useful in explaining the convenience yield of a single commodity in terms of the convenience yields of related commodities.

[INSERT FIGURE 7 ABOUT HERE]

3.5. Conclusions and Further Work

The reduced explanatory power of the fundamental inventory-ratio based model of Chapter 2 was hypothesized to result from the presence of index and momentum traders in the cocoa markets. Using machine learning, this paper investigates two aspects of this hypothesis, the co-movement of commodity prices and the prediction of future cocoa prices using fundamental information and information from related commodities.

The Restricted Boltzmann machine detects the co-movement of commodity prices from the World Bank Pink Sheet. This architecture indicates that commodity prices move together in groups on a monthly basis, supporting the hypothesis that cocoa price movements are partially explainable by the movements in prices of other commodities.

One-day-ahead price prediction of the cocoa price using a Recurrent Neural Network Restricted Boltzmann machine is not successful for four identifiable reasons. One, the time frame is too short and monthly price changes would be better utilized. Second, continuous input data representations may make the search space too large, and binary representations might be used instead. Third, additional information including inventory to consumption ratios might be necessary as input to the model. Finally, price prediction is one of the most difficult problems in finance and the difficulty of prediction attests to the general efficiency of markets.

We recommend that further research with the RNN_RBM on commodity prices focus on two areas, sequencing binary encoded price changes, and exploring the joint commodity term structure. Binary encoded price changes would significantly restrict the search space for the RNN_RBM, and the joint commodity term structure has a natural sequence from the near-month to the final contract.

Finally, we report the use of a new technique from Lopez de Prado, fractional differencing, on producing stationary time series without removing all long term price movement information. Several of the commodity time series analyzed require only partial differentiation including the Cocoa Totals Open Interest and Coffee Nearby Close.

References

Adhikari, R., & Putnam, K. J. (2019). Comovement in the commodity futures markets: An analysis of the energy, grains, and livestock sectors. Journal of Commodity Markets, 100090.

Assis, C. A., Pereira, A. C., Carrano, E. G., Ramos, R., & Dias, W. (2018). Restricted Boltzmann machines for the prediction of trends in financial time series. In 2018 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

Bäärnhielm, A. (2017). Multiple time-series forecasting on mobile network data using an RNN-RBM model.

Bhardwaj, G., Gorton, G. B., & Rouwenhorst, K. G. (2014). Fooling some of the people all of the time: The inefficient performance and persistence of commodity trading advisors. The Review of Financial Studies, 27(11), 3099-3132.

Berhane F. (2021) Building your Recurrent Neural Network - Step by Step. https://datascience-enthusiast.com/DL/Building_a_Recurrent_Neural_Network-Step_by_Step_v1.html

Bollen, N. P., Hutchinson, M. C., & O'Brien, J. (2019). When It Pays to Follow the Crowd: Strategy Conformity and CTA Performance. Available at SSRN 3481243.

Boulanger-Lewandowski, N., Bengio, Y., & Vincent, P. (2012). Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription. arXiv preprint arXiv:1206.6392.

Bu, Y., Zhao, G., Luo, A. L., Pan, J., & Chen, Y. (2015). Restricted Boltzmann machine: a non-linear substitute for PCA in spectral processing. Astronomy & Astrophysics, 576, A96.

Castelvecchi, D. (2016). Can we open the black box of AI?. *Nature News*, *538*(7623), 20.

De Prado, M. L. (2018). Advances in financial machine learning. John Wiley & Sons.

Ghoddusi, H., Creamer, G. G., & Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. Energy Economics, 81, 709-727.

Hecht-Nielsen, R. (1992). Theory of the backpropagation neural network. In Neural networks for perception (pp. 65-93). Academic Press.

Henderson, B. J., Pearson, N. D., & Wang, L. (2015). New evidence on the financialization of commodity markets. The Review of Financial Studies, 28(5), 1285-1311.

Hinton, G. E. (2002). Training products of experts by minimizing contrastive divergence. Neural Computation, 14(8):1711–1800.

Hinton, G. E. (2010). A Practical Guide to Training Restricted Boltzmann Machines , https://www.cs.toronto.edu/~hinton/absps/guideTR.pdf

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(4), e1312.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences, 79:2554–2558.

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. Neural networks, 4(2), 251-257.

Irwin, S. H., & Sanders, D. R. (2011). Index funds, financialization, and commodity futures markets. Applied Economic Perspectives and Policy, 33(1), 1-31.

Kang, J., & Kwon, K. Y. (2019). How about selling commodity futures losers?. Journal of Futures Markets, 39(12), 1489-1514.

Kang, W., Rouwenhorst, K. G., & Tang, K. (2019). A tale of two premiums: The role of hedgers and speculators in commodity futures markets. The Journal of Finance.

Kidger, P., & Lyons, T. (2020). Universal approximation with deep narrow networks. In Conference on Learning Theory (pp. 2306-2327).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90.

Lago, J., De Ridder, F., & De Schutter, B. (2018). Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. Applied Energy, 221, 386-405.

Li, X., Shang, W., & Wang, S. (2019). Text-based crude oil price forecasting: A deep learning approach. International Journal of Forecasting, 35(4), 1548-1560.

Main, S., Irwin, S. H., Sanders, D. R., & Smith, A. (2018). Financialization and the returns to commodity investments. Journal of Commodity Markets, 10, 22-28.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. PloS one, 13(3), e0194889.

Liang, Q., Rong, W., Zhang, J., Liu, J., & Xiong, Z. (2017). Restricted Boltzmann machine based stock market trend prediction. In 2017 International Joint Conference on Neural Networks (IJCNN) (pp. 1380-1387). IEEE.

Loyola-Gonzalez, O. (2019). Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view. *IEEE Access*, *7*, 154096-154113.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. nature, 323(6088), 533-536.

Rumelhart, D. E., Durbin, R., Golden, R., & Chauvin, Y. (1995). Backpropagation: The basic theory. Backpropagation: Theory, architectures and applications, 1-34.

Shah, D., Isah, H., & Zulkernine, F. (2019). Stock market analysis: A review and taxonomy of prediction techniques. International Journal of Financial Studies, 7(2), 26.

Shu, T. (2009). Does positive-feedback trading by institutions contribute to stock return momentum?. In AFA 2007 Chicago meetings paper.

Sirignano, J., & Cont, R. (2019). Universal features of price formation in financial markets: perspectives from deep learning. Quantitative Finance, 19(9), 1449-1459.

Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. In Rumelhart, D. E. and McClelland, J. L., editors, Parallel Distributed Processing, volume 1, chapter 6, pages 194–281. MIT Press, Cambridge.

Sutskever, I. and Hinton, G.E. (2007). Learning multilevel distributed representations for high-dimensional sequences. In AISTATS, pp. 544–551, 2007.

Tang, K., & Xiong, W. (2012). Index investment and the financialization of commodities. Financial Analysts Journal, 68(6), 54-74.

Taylor, G. W., Hinton, G. E., & Roweis, S. T. (2007). Modeling human motion using binary latent variables. In Advances in neural information processing systems (pp. 1345-1352).

Weidman, S. (2019). Deep Learning from Scratch, O'Reilly Media Inc.

Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. Proceedings of the IEEE, 78(10), 1550-1560.

Yadav V. (2016). How neural networks learn nonlinear functions and classify linearly non-separable data? https://medium.com/@vivek.yadav/how-neural-networks-learn-nonlinear-functions-and-classify-linearly-non-separable-data-22328e7e5be1

Yoshihara, A., Fujikawa, K., Seki, K., & Uehara, K. (2014). Predicting stock market trends by recurrent deep neural networks. In Pacific rim international conference on artificial intelligence (pp. 759-769). Springer, Cham.

Zhang, Y. (2015). Using financial reports to predict stock market trends with machine learning techniques (Doctoral dissertation, University of Oxford).

Zhao, Y., Li, J., & Yu, L. (2017). A deep learning ensemble approach for crude oil price forecasting. Energy Economics, 66, 9-16.

```python
#Appendix A

# Jan Koeman
# Gaussian RNN_RBM with one-step ahead prediction
# University of Canterbury, Christchurch, New Zealand
# Department of Economics and Finance
# jan.koeman@pg.canterbury.ac.nz

# Based on a tutorial provided by
# Nicolas Boulanger-Lewandowski
# University of Montreal (2012)
# RNN-RBM deep learning tutorial
# More information at http://deeplearning.net/tutorial/rnnrbm.html

# Note that Theano will not run under the latest Nvidia GPUs like the RTX-2080
# A recommended configuration is Ubuntu 16 with the GTX-1080
# such as the Dell 5810 Precision Workstation.
# Follow the ananconda installation of Theano, and install the CUDA 8.0
# libraries from the Nvidia website. Make sure the Theano path variable
# for the CUDA library is set correctly.
# A GPU is not important unless the number of hidden units is in the hundreds



from __future__ import print_function

import glob
import os
import sys
import pickle
import midi_manipulation2
import matplotlib.pyplot as plt
from theano.printing import Print
import pandas as pd
from scipy.stats import ks_2samp

import numpy
try:
    import pylab
except ImportError:
    print ("pylab isn't available. If you use its functionality, it will crash.")
    print("It can be installed with 'pip install -q Pillow'")

# from midi.utils import midiread
# from midi.utils import midiwrite
import theano
import theano.tensor as T
from theano.sandbox.rng_mrg import MRG_RandomStreams as RandomStreams

#Don't use a python long as this don't work on 32 bits computers.
numpy.random.seed(0xbeef)
rng = RandomStreams(seed=numpy.random.randint(1 << 30))
theano.config.warn.subtensor_merge_bug = False


def build_rbm_gaussian(v, W, bv, bh, k):
    '''Construct a k-step Gibbs chain starting at v for an RBM.

    v : Theano vector or matrix
        If a matrix, multiple chains will be run in parallel (batch).
    W : Theano matrix
        Weight matrix of the RBM.
    bv : Theano vector
```

```
        Visible bias vector of the RBM.
    bh : Theano vector
        Hidden bias vector of the RBM.
    k : scalar or Theano scalar
        Length of the Gibbs chain.

    Return a (v_sample, cost, monitor, updates) tuple:

    v_sample : Theano vector or matrix with the same shape as `v`
        Corresponds to the generated sample(s).
    cost : Theano scalar
        Expression whose gradient with respect to W, bv, bh is the CD-k
        approximation to the log-likelihood of `v` (training example) under the
        RBM. The cost is averaged in the batch case.
    monitor: Theano scalar
        Pseudo log-likelihood (also averaged in the batch case).
    updates: dictionary of Theano variable -> Theano variable
        The `updates` object returned by scan.'''

    def gibbs_step_gaussian(v):
        mean_h = T.nnet.sigmoid(T.dot(v, W) + bh)
        h = rng.binomial(size=mean_h.shape, n=1, p=mean_h,
                         dtype=theano.config.floatX)
        val = T.dot(h, W.T) + bv
        mean_v = val
        # mean_v = rng.normal(size=v.shape, avg=(T.dot(h, W.T) + bv), std=1.0, dtype=theano.config.floatX)
        # v = rng.binomial(size=mean_v.shape, n=1, p=mean_v, dtype=theano.config.floatX)
        v = mean_v
        return mean_v, v



    chain, updates = theano.scan(lambda v: gibbs_step_gaussian(v)[1], outputs_info=[v],
                                 n_steps=k)
    v_sample = chain[-1]

    mean_v = gibbs_step_gaussian(v_sample)[0]
    # print("v, meanv", v, mean_v)
    monitor = T.xlogx.xlogy0(v, mean_v) + T.xlogx.xlogy0(1 - v, 1 - mean_v)
    monitor = monitor.sum() / v.shape[0]

    def free_energy(v):
         return -(v * bv).sum() - T.log(1 + T.exp(T.dot(v, W) + bh)).sum()

    def free_energy_bak(x, W, bv, bh):
        # The function computes the free energy of a visible vector.
        hidden_term = T.sum(T.log(1. + T.exp(T.dot(x / 1, W) + bh)), axis=1)
        vbias_term = T.sum(T.square(T.sub(x, bv)), axis=1) / ((1 ** 2) * 2)
        return -hidden_term + vbias_term


    cost = (free_energy(v) - free_energy(v_sample)) / v.shape[0]

    return v_sample, cost, cost, updates

def build_rbm(v, W, bv, bh, k):
    '''Construct a k-step Gibbs chain starting at v for an RBM.

    v : Theano vector or matrix
        If a matrix, multiple chains will be run in parallel (batch).
```

```
    W : Theano matrix
        Weight matrix of the RBM.
    bv : Theano vector
        Visible bias vector of the RBM.
    bh : Theano vector
        Hidden bias vector of the RBM.
    k : scalar or Theano scalar
        Length of the Gibbs chain.

    Return a (v_sample, cost, monitor, updates) tuple:

    v_sample : Theano vector or matrix with the same shape as `v`
        Corresponds to the generated sample(s).
    cost : Theano scalar
        Expression whose gradient with respect to W, bv, bh is the CD-k
        approximation to the log-likelihood of `v` (training example) under the
        RBM. The cost is averaged in the batch case.
    monitor: Theano scalar
        Pseudo log-likelihood (also averaged in the batch case).
    updates: dictionary of Theano variable -> Theano variable
        The `updates` object returned by scan.'''

    # print_op = theano.printing.Print('V')
    # printed_v = print_op(v)
    # f = theano.function([v], printed_v)
    # r = f(v)

    def gibbs_step(v):
        mean_h = T.nnet.sigmoid(T.dot(v, W) + bh)
        h = rng.binomial(size=mean_h.shape, n=1, p=mean_h,
                         dtype=theano.config.floatX)
        mean_v = T.nnet.sigmoid(T.dot(h, W.T) + bv)
        v = rng.binomial(size=mean_v.shape, n=1, p=mean_v,
                         dtype=theano.config.floatX)
        return mean_v, v

    chain, updates = theano.scan(lambda v: gibbs_step(v)[1], outputs_info=[v],
                                 n_steps=k)
    v_sample = chain[-1]

    mean_v = gibbs_step(v_sample)[0]
    monitor = T.xlogx.xlogy0(v, mean_v) + T.xlogx.xlogy0(1 - v, 1 - mean_v)
    monitor = monitor.sum() / v.shape[0]

    def free_energy(v):
        return -(v * bv).sum() - T.log(1 + T.exp(T.dot(v, W) + bh)).sum()
    cost = (free_energy(v) - free_energy(v_sample)) / v.shape[0]

    return v_sample, cost, monitor, updates



def shared_normal(num_rows, num_cols, scale=1):
    '''Initialize a matrix shared variable with normally distributed
    elements.'''
    return theano.shared(numpy.random.normal(
        scale=scale, size=(num_rows, num_cols)).astype(theano.config.floatX))


def shared_zeros(*shape):
    '''Initialize a vector shared variable with zero elements.'''
    return theano.shared(numpy.zeros(shape, dtype=theano.config.floatX))
```

```python

def build_rnnrbm(n_visible, n_hidden, n_hidden_recurrent, gaussian=False):
    '''Construct a symbolic RNN-RBM and initialize parameters.

    n_visible : integer
        Number of visible units.
    n_hidden : integer
        Number of hidden units of the conditional RBMs.
    n_hidden_recurrent : integer
        Number of hidden units of the RNN.

    Return a (v, v_sample, cost, monitor, params, updates_train, v_t,
    updates_generate) tuple:

    v : Theano matrix
        Symbolic variable holding an input sequence (used during training)
    v_sample : Theano matrix
        Symbolic variable holding the negative particles for CD log-likelihood
        gradient estimation (used during training)
    cost : Theano scalar
        Expression whose gradient (considering v_sample constant) corresponds
        to the LL gradient of the RNN-RBM (used during training)
    monitor : Theano scalar
        Frame-level pseudo-likelihood (useful for monitoring during training)
    params : tuple of Theano shared variables
        The parameters of the model to be optimized during training.
    updates_train : dictionary of Theano variable -> Theano variable
        Update object that should be passed to theano.function when compiling
        the training function.
    v_t : Theano matrix
        Symbolic variable holding a generated sequence (used during sampling)
    updates_generate : dictionary of Theano variable -> Theano variable
        Update object that should be passed to theano.function when compiling
        the generation function.'''

    W = shared_normal(n_visible, n_hidden, 0.01)
    bv = shared_zeros(n_visible)
    bh = shared_zeros(n_hidden)
    Wuh = shared_normal(n_hidden_recurrent, n_hidden, 0.0001)
    Wuv = shared_normal(n_hidden_recurrent, n_visible, 0.0001)
    Wvu = shared_normal(n_visible, n_hidden_recurrent, 0.0001)
    Wuu = shared_normal(n_hidden_recurrent, n_hidden_recurrent, 0.0001)
    bu = shared_zeros(n_hidden_recurrent)
    # v_test = shared_zeros(n_visible)
    # predict_flag = shared_zeros(1)

    params = W, bv, bh, Wuh, Wuv, Wvu, Wuu, bu  # learned parameters as shared
                                                # variables

    v = T.matrix()  # a training sequence
    u0 = T.zeros((n_hidden_recurrent,))  # initial value for the RNN hidden
                                         # units

    # If `v_t` is given, deterministic recurrence to compute the variable
    # biases bv_t, bh_t at each time step. If `v_t` is None, same recurrence
    # but with a separate Gibbs chain at each time step to sample (generate)
    # from the RNN-RBM. The resulting sample v_t is returned in order to be
    # passed down to the sequence history.
    def recurrence(v_t, u_tm1):
        bv_t = bv + T.dot(u_tm1, Wuv)
        bh_t = bh + T.dot(u_tm1, Wuh)
        generate = v_t is None

```

```
        if generate:
            if gaussian:
                v_t, _, _, updates = build_rbm_gaussian(T.zeros((n_visible,)), W, bv_t, bh_t, k=25)
            else:
                v_t, _, _, updates = build_rbm(T.zeros((n_visible,)), W, bv_t, bh_t, k=25)

        u_t = T.tanh(bu + T.dot(v_t, Wvu) + T.dot(u_tm1, Wuu))
        return ([v_t, u_t], updates) if generate else [u_t, bv_t, bh_t]

    def recurrence_predict(v_t, u_tm1):
        bv_t = bv + T.dot(u_tm1, Wuv)
        bh_t = bh + T.dot(u_tm1, Wuh)

        # u_tm1 = theano.printing.Print('u_tm1')(u_tm1)
        # v_t = theano.printing.Print('v_t')(v_t)

        if gaussian:
                vp_t, _, _, updates = build_rbm_gaussian(T.zeros((n_visible,)), W, bv_t, bh_t, k=25)
        else:
                vp_t, _, _, updates = build_rbm(T.zeros((n_visible,)), W, bv_t, bh_t, k=25)
        # vp_t = theano.printing.Print('v_t prediction')(vp_t)

        u_t = T.tanh(bu + T.dot(v_t, Wvu) + T.dot(u_tm1, Wuu))
        # u_t = theano.printing.Print('u_t')(u_t)
        return ([vp_t, u_t], updates)

    # For training, the deterministic recurrence is used to compute all the
    # {bv_t, bh_t, 1 ⩽ t ⩽ T} given v. Conditional RBMs can then be trained
    # in batches using those parameters.

    (u_t, bv_t, bh_t), updates_train = theano.scan(
        lambda v_t, u_tm1, *_: recurrence(v_t, u_tm1),
        sequences=v, outputs_info=[u0, None, None], non_sequences=params)

    if gaussian:
        v_sample, cost, monitor, updates_rbm = build_rbm_gaussian(v, W, bv_t[:], bh_t[:],
                                                        k=15)
    else:
        v_sample, cost, monitor, updates_rbm = build_rbm(v, W, bv_t[:], bh_t[:],
                                                         k=15)
    updates_train.update(updates_rbm)


    # DO NOT ENABLE BOTH GENERATION AND PREDICTION FUNCTIONS OR THEANO WILL NOT COMPILE!
    # USE ONE OR THE OTHER.

    # symbolic loop for sequence generation
    # (v_t, u_t), updates_generate = theano.scan(
    #     lambda u_tm1, *_: recurrence(None, u_tm1),
    #     outputs_info=[None, u0], non_sequences=params, n_steps=600)


    (v_t, u_t), updates_predict = theano.scan(
        lambda v_t1, u_tm1, *_: recurrence_predict(v_t1, u_tm1), sequences=v[:,:], #full sequence used
        outputs_info=[None, u0], non_sequences=params)

```

```
    updates_generate = None


    return (v, v_sample, cost, monitor, params, updates_train, v_t,
            updates_generate, updates_predict)


class RnnRbm:
    '''Simple class to train an RNN-RBM from MIDI files and to generate sample
    sequences.'''

    def __init__(
        self,
        n_visible=6,
        n_hidden=30,
        n_hidden_recurrent=20,
        lr=0.01,
        r=(24, 102),
        dt=0.3,
        gaussian=False
    ):
        '''Constructs and compiles Theano functions for training and sequence
        generation.

        n_hidden : integer
            Number of hidden units of the conditional RBMs.
        n_hidden_recurrent : integer
            Number of hidden units of the RNN.
        lr : float
            Learning rate
        r : (integer, integer) tuple
            Specifies the pitch range of the piano-roll in MIDI note numbers,
            including r[0] but not r[1], such that r[1]-r[0] is the number of
            visible units of the RBM at a given time step. The default (21,
            109) corresponds to the full range of piano (88 notes).
        dt : float
            Sampling period when converting the MIDI files into piano-rolls, or
            equivalently the time difference between consecutive time steps.'''

        self.r = r
        self.dt = dt
        # (v, v_sample, cost, monitor, params, updates_train, v_t,
        #     updates_generate) = build_rnnrbm(
        #         r[1] - r[0],
        #         n_hidden,
        #         n_hidden_recurrent
        #     )

        (v, v_sample, cost, monitor, params, updates_train, v_t,
            updates_generate, updates_predict) = build_rnnrbm(
                n_visible,
                n_hidden,
                n_hidden_recurrent,
                gaussian
            )

        gradient = T.grad(cost, params, consider_constant=[v_sample])
        updates_train.update(
            ((p, p - lr * g) for p, g in zip(params, gradient))
        )
        self.train_function = theano.function(
            [v],
            monitor,
```

```
            updates=updates_train
        )
        # self.generate_function = theano.function(
        #     [],
        #     v_t,
        #     updates=updates_generate
        # )
        self.predict_function = theano.function(
            [v],
            v_t,
            updates=updates_predict
        )



    def train(self, files, batch_size=100, num_epochs=200):
        '''Train the RNN-RBM via stochastic gradient descent (SGD) using MIDI
        files converted to piano-rolls.

        files : list of strings
            List of MIDI files that will be loaded as piano-rolls for training.
        batch_size : integer
            Training sequences will be split into subsequences of at most this
            size before applying the SGD updates.
        num_epochs : integer
            Number of epochs (pass over the training set) performed. The user
            can safely interrupt training with Ctrl+C at any time.'''

        # assert len(files) > 0, 'Training set is empty!' \
        #                        ' (did you download the data files?)'
        # dataset = [midiread(f, self.r,
        #                      self.dt).piano_roll.astype(theano.config.floatX)
        #             for f in files]
        # dataset = [midi_manipulation2.r(f, self.r,
        #           self.dt).piano_roll.astype(theano.config.floatX)
        #  for f in files]
        # strPickledData = "/Users/jankoeman/dlFromScratch/Data/PickledFraction.pkl"
        strPickledData = "/Users/jankoeman/dlFromScratch/Data/PickledSeries.pkl"
        # strPickledData = "/Users/jankoeman/dlFromScratch/Data/PickledCount.pkl"
        try:
            f = open(strPickledData, 'rb')
            dataset = pickle.load(f)
            f.close()
        except OSError as error:
            print("Could not open/read file:", strPickledData)
            print(error)
            sys.exit()



        dataset_size = len(dataset)
        test_dataset = dataset[:dataset_size-1]
        train_dataset = dataset[dataset_size-1]

        try:
            for epoch in range(num_epochs):
                numpy.random.shuffle(test_dataset)
                costs = []
                # print("Training", epoch)
                for s, sequence in enumerate(dataset):
                    sequence = sequence.astype(numpy.float)
                    for i in range(0, len(sequence), batch_size):
                        cost = self.train_function(sequence[i:i + batch_size])
```

```
                        # if not(numpy.isnan(cost)):
                        mincost = cost.min()
                        if not(numpy.isnan(mincost)):
                            costs.append(mincost)

                print('Epoch %i/%i' % (epoch + 1, num_epochs))
                print(numpy.mean(costs))
                sys.stdout.flush()

        except KeyboardInterrupt:
            print('Interrupted by user.')

        td = train_dataset.astype(numpy.float)
        self.predict(td)


    def generate(self, filename, show=True):
        '''Generate a sample sequence, plot the resulting piano-roll and save
        it as a MIDI file.

        # filename : string
        #     A MIDI file will be created at this location.
        # show : boolean
        #     If True, a piano-roll of the generated sequence will be shown.'''
        #
        pass

        # piano_roll = self.generate_function()
        # # midi_manipulation2.write_song(filename,piano_roll)
        # # midiwrite(filename, piano_roll, self.r, self.dt)
        # print("Piano Roll", piano_roll)
        # if show:
        #
        #     lstColumns = ['COCOA NEARBY CLOSE','COCOA TOTALS OPEN INTEREST','Cocoa Stocks',
        #                   'Coffee Stocks','Sugar Stocks','SUGAR 11 NEARBY CLOSE','SUGAR 11 TOTALS OPEN INTEREST',
        #                   'COFFEE NEARBY CLOSE', 'COFFEE TOTALS OPEN INTEREST', 'Crude oil', 'Brent  Coal, Australian']
        #     dfPianoRoll = pd.DataFrame(piano_roll[:,0:8])
        #     dfPianoRoll.columns = lstColumns[0:8]
        #     dfPianoRoll.plot(subplots=True)

    def predict(self, test_dataset, filename=None, show=True):
        '''Generate a sample sequence, plot the resulting piano-roll and save
        it as a MIDI file.

        filename : string
            A MIDI file will be created at this location.
        show : boolean
            If True, a piano-roll of the generated sequence will be shown.'''

        predictions = self.predict_function(test_dataset)
        # midi_manipulation2.write_song(filename,piano_roll)
        # midiwrite(filename, piano_roll, self.r, self.dt)
        # print("Piano Roll", predictions)
        if show:
            # extent = (0, self.dt * len(piano_roll)) + self.r
            # pylab.figure()
            # pylab.imshow(piano_roll.T, origin='lower', aspect='auto',
            #              interpolation='nearest', cmap=pylab.cm.gray_r,
            #              extent=extent)
            # pylab.xlabel('time (s)')
```

```
            # pylab.ylabel('MIDI note number')
            # pylab.title('generated piano-roll')
            # lstColumns = ['COCOA NEARBY CLOSE','COCOA TOTALS OPEN INTEREST','Cocoa Stocks',
            #                'Coffee Stocks','Sugar Stocks','SUGAR 11 NEARBY CLOSE','SUGAR 11 TOTALS OPEN INTEREST',
            #                'COFFEE NEARBY CLOSE', 'COFFEE TOTALS OPEN INTEREST', 'Crude oil', 'Brent  Coal, Australian']
            nColumns = 8
            strMainCol = 'COCOA NEARBY CLOSE'
            # strMainCol = 'Sine Wave 1'
            lstColumns = [strMainCol,'COCOA TOTALS OPEN INTEREST','Cocoa Stocks',
                          'Coffee Stocks',
                          'COFFEE NEARBY CLOSE', 'COFFEE TOTALS OPEN INTEREST', 'Relative Return', 'Position']
            # lstColumns = [strMainCol,'Sine Wave 2','Sine Wave 1C',
            #                'Sine Wave 1C',
            #                'Sine Wave 1C', 'Sine Wave 1C', 'Sine Wave 1C', 'Sine Wave 1C']
            dfPianoRoll = pd.DataFrame(predictions[:,0:nColumns])
            dfPianoRoll.columns = lstColumns[0:nColumns]
            dfPianoRoll.plot(subplots=True)
            print("predictions size", dfPianoRoll.shape)

            # lstColumns = ['COCOA NEARBY CLOSE','COCOA TOTALS OPEN INTEREST','Cocoa Stocks',
            #                'Coffee Stocks',
            #                'COFFEE NEARBY CLOSE', 'COFFEE TOTALS OPEN INTEREST', 'Relative Return', 'Position']
            dfInputs = pd.DataFrame(test_dataset[:,0:nColumns])
            dfInputs.columns = lstColumns[0:nColumns]
            dfInputs.plot(subplots=True)
            print("Data:")
            print(dfInputs.head(5))
            print("Data shape", dfInputs.shape)

            differenceSeries = dfPianoRoll[strMainCol]
            inputSeries1 = dfInputs[strMainCol].to_numpy()
            predSeries = dfPianoRoll[strMainCol].to_numpy()
            sizeSeries1 = inputSeries1.shape[0]
            differenceSeries2 = predSeries[0:sizeSeries1-1] - inputSeries1[1:sizeSeries1]
            print ("Sum of squares predictions:", sum(differenceSeries2**2))
            print('Kolmogorov-Smirnoff values')
            print(ks_2samp(predSeries, inputSeries1))


            dfCocoaInputSeries = dfInputs[strMainCol]
            sizeSeries = dfCocoaInputSeries.shape[0]
            dfPrevValues = dfCocoaInputSeries[0:sizeSeries-1].to_numpy()
            dfPredictions = dfCocoaInputSeries[1:sizeSeries].to_numpy()
            differenceNaive = dfPredictions-dfPrevValues
            print("baseline sum of squares:", sum(differenceNaive**2))




def test_rnnrbm(batch_size=100, num_epochs=200):
    model = RnnRbm(8,50,50,.01,gaussian=True)
    cwd = os.path.dirname(os.path.abspath(__file__))
    re = os.path.join(os.path.split(cwd)[0],
                      'data', 'Nottingham', 'train', '*.mid')
    model.train(glob.glob(re),
```

```
                  batch_size=batch_size, num_epochs=num_epochs)
    return model

if __name__ == '__main__':
    kPrecision = 5
    numpy.set_printoptions(precision=kPrecision)
    pd.set_option("display.precision", kPrecision)
    theano.config.openmp = False
    model = test_rnnrbm()
    # model.generate('sample1.mid')
    #model.predict('testpredict')
    # model.generate('sample2.mid')
    pylab.show()
```

```python
# Appendix B


# Jan Koeman
# Data Preparation code for analysis with
# Gaussian RNN_RBM with one-step ahead prediction
# University of Canterbury, Christchurch, New Zealand
# Department of Economics and Finance
# jan.koeman@pg.canterbury.ac.nz

# The raw data series are publicly available on the web as described in
# my thesis Essays on Commodities, Chapter 3.




from pandas import read_csv
from datetime import datetime
from matplotlib import pyplot
import pandas as pd
import pickle
import sys
import numpy as np
from fracdiff import fdiff
from statsmodels.tsa.stattools import adfuller
import os
import seaborn as sns


def parser(x):
    return pd.to_datetime(x)


strFileName = "/Users/jankoeman/Documents/Main/_PHD2/_NeuralNetwork Data/Sugar Ending Stocks 3.csv"
dfSugarStocks = read_csv(strFileName, header=0, parse_dates=[0], index_col=0, squeeze=True, date_parser=parser, delimiter=";")
print("Sugar")
#print(dfSugarStocks.head())
#df['Date'] = pd.to_datetime(df['Date'])
#print(dfSugarStocks.head())
upSugarStocks = dfSugarStocks.resample('D').interpolate()
print(upSugarStocks.head(10))

strFileName = "/Users/jankoeman/Documents/Main/_PHD2/_NeuralNetwork Data/Cocoa Stocks 2.csv"
dfCocoaStocks = read_csv(strFileName, header=0, parse_dates=[0], index_col=0, squeeze=True, date_parser=parser, delimiter=";")
print("Cocoa")
#print(dfCocoaStocks.head())
#df['Date'] = pd.to_datetime(df['Date'])
#print(dfCocoaStocks.head())
upCocoaStocks = dfCocoaStocks.resample('D').interpolate()
print(upCocoaStocks.head(10))

dfSugar = pd.DataFrame(upSugarStocks)
dfCocoa = pd.DataFrame(upCocoaStocks)
dfBoth = dfSugar.join(dfCocoa, how="inner")
print("Joined")
print(dfBoth.head(10))
print("Shape:", dfBoth.shape)

strFileName = "/Users/jankoeman/Documents/Main/_PHD2/_NeuralNetwork Data/Coffee
```

```
Stocks 2.csv"
dfCoffeeStocks = read_csv(strFileName, header=0, parse_dates=[0], index_col=0,
squeeze=True, date_parser=parser, delimiter=";")
print("Coffee")
#print(dfCocoaStocks.head())
#df['Date'] = pd.to_datetime(df['Date'])
#print(dfCocoaStocks.head())
upCoffeeStocks = dfCoffeeStocks.resample('D').interpolate()
print(upCoffeeStocks.head(10))
dfCoffee = pd.DataFrame(upCoffeeStocks)

dfCocoaCoffeeSugar = dfBoth.join(dfCoffee, how="inner")
print(dfCocoaCoffeeSugar.tail(10))
print("Shape:", dfCocoaCoffeeSugar.shape)

strFileName = "/Users/jankoeman/Documents/Main/_PHD2/_NeuralNetwork Data/World Bank
10.csv"
dfWorldBankPrices = read_csv(strFileName, header=0, parse_dates=[0], index_col=0,
squeeze=True, date_parser=parser, delimiter=";")
print("World Bank")
print(dfWorldBankPrices.head())
#df['Date'] = pd.to_datetime(df['Date'])
#print(dfCocoaStocks.head())

print(dfWorldBankPrices[dfWorldBankPrices.index.duplicated()])
#exit(0)

# upWorldBank = dfWorldBankPrices.resample('D').interpolate()
# print(upWorldBank.head(10))
# dfWorld = pd.DataFrame(upWorldBank)

dfWorldReturns = dfWorldBankPrices.pct_change()
print(dfWorldReturns.head(10))
dropWorldBankReturns = dfWorldReturns.drop([parser('2000-01-15')])
print(dropWorldBankReturns.head(10))


upWorldBankReturns = dropWorldBankReturns.resample('D').fillna("ffill")

print(upWorldBankReturns.head(10))
dfWorld = pd.DataFrame(upWorldBankReturns)
dfJustMean = pd.DataFrame(dfWorld.mean(axis=1))
print("Just mean")
print(dfJustMean.head(3))
dfMeanIndex= pd.DataFrame(dfWorld['Cocoa']/dfWorld.mean(axis=1))
dfMeanIndex.columns = [ 'Relative Return']
print(dfMeanIndex.head())

positionArray = np.zeros((upWorldBankReturns.shape[0],1))
matrixReturns = upWorldBankReturns.to_numpy()
for i in range(matrixReturns.shape[0]):
    sortArray = matrixReturns[i]
    cocoaValue = sortArray[0]
    sortArray = np.sort(sortArray)
    positionCocoa = np.where(sortArray==cocoaValue)
    if len(positionCocoa[0]) > 1:
        pass
        # print("i, value, positions, len", i, cocoaValue, positionCocoa, len(
positionCocoa[0]))
        # print(sortArray)
    positionArray[i] = positionCocoa[0][0]

# print(positionArray[:200])
```

```

dfMeanIndex['Cocoa Index'] = positionArray

print(dfMeanIndex.head(10))




#cocoa term structure
strFileName = "/Users/jankoeman/Documents/Main/_PHD2/_NeuralNetwork Data/Cocoa Futures 4.csv"
dfCocoaFut = read_csv(strFileName, header=0, parse_dates=[0], index_col=0, squeeze=True, date_parser=parser, delimiter=";")
print("Cocoa Futures")
print(dfCocoaFut.head())
#df['Date'] = pd.to_datetime(df['Date'])
#print(dfCocoaStocks.head())

#coffee term structure
strFileName = "/Users/jankoeman/Documents/Main/_PHD2/_NeuralNetwork Data/Coffee Futures 3.csv"
dfCoffeeFut = read_csv(strFileName, header=0, parse_dates=[0], index_col=0, squeeze=True, date_parser=parser, delimiter=";")
print("Coffee Futures")
print(dfCoffeeFut.head())
#df['Date'] = pd.to_datetime(df['Date'])
#print(dfCocoaStocks.head())

#sugar term structure
strFileName = "/Users/jankoeman/Documents/Main/_PHD2/_NeuralNetwork Data/Sugar 11 2.csv"
dfSugarFut = read_csv(strFileName, header=0, parse_dates=[0], index_col=0, squeeze=True, date_parser=parser, delimiter=";")
print("Sugar Futures")
print(dfSugarFut.head())
#df['Date'] = pd.to_datetime(df['Date'])
#print(dfCocoaStocks.head())

dfSugarJ = pd.DataFrame(upSugarStocks)
dfCocoaJ = pd.DataFrame(upCocoaStocks)
dfCoffeeJ = pd.DataFrame(upCoffeeStocks)
dfWorldJ = pd.DataFrame(upWorldBankReturns)
# dfSugarFutJ = dfSugarFut[[ 'SUGAR 11 NEARBY CLOSE', 'SUGAR 11 NEARBY OPEN INTEREST','SUGAR 11 2ND CLOSE',
#                            'SUGAR 11 2ND OPEN INTEREST','SUGAR 11 3RD CLOSE','SUGAR 11 3RD OPEN INTEREST',
#                            'SUGAR 11 4TH CLOSE', 'SUGAR 11 4TH OPEN INTEREST', 'SUGAR 11 5TH CLOSE', 'SUGAR 11 5TH OPEN INTEREST' ]]
dfSugarFutJ = dfSugarFut[[ 'SUGAR 11 NEARBY CLOSE', 'SUGAR 11 TOTALS OPEN INTEREST' ]]

print(dfSugarFutJ.head())

# dfCoffeeFutJ = dfCoffeeFut[[ 'COFFEE NEARBY CLOSE', 'COFFEE NEARBY OPEN INTEREST','COFFEE 2ND CLOSE',
#                            'COFFEE 2ND OPEN INTEREST','COFFEE 3RD CLOSE','COFFEE 3RD OPEN INTEREST',
#                            'COFFEE 4TH CLOSE', 'COFFEE 4TH OPEN INTEREST', 'COFFEE 5TH CLOSE', 'COFFEE 5TH OPEN INTEREST' ]]
dfCoffeeFutJ = dfCoffeeFut[[ 'COFFEE NEARBY CLOSE', 'COFFEE TOTALS OPEN INTEREST']]

# dfCocoaFutJ = dfCocoaFut[[ 'COCOA NEARBY CLOSE', 'COCOA NEARBY OPEN INTEREST','COCOA 2ND CLOSE',
```

```
#                            'COCOA 2ND OPEN INTEREST','COCOA 3RD CLOSE','COCOA 3RD OPEN INTEREST',
#                            'COCOA 4TH CLOSE', 'COCOA 4TH OPEN INTEREST', 'COCOA 5TH CLOSE', 'COCOA 5TH OPEN INTEREST' ]]

dfCocoaFutJ = dfCocoaFut[[ 'COCOA NEARBY CLOSE', 'COCOA TOTALS OPEN INTEREST' ]]

print("All Data Series Joined")
# dfAll = dfCocoaFutJ.join([dfCocoaJ, dfCoffeeJ, dfSugarJ, dfSugarFutJ, dfCoffeeFutJ, dfWorldJ],how="inner")
dfAll = dfCocoaFutJ.join([dfCocoaJ, dfCoffeeJ, dfCoffeeFutJ, dfMeanIndex],how="inner")
print(dfAll.head())

strOutputPath = "/Users/jankoeman/Documents/Main/_PHD2/_NeuralNetwork Data/All Series.csv"
dfAll.to_csv(strOutputPath, sep=";")

#add the mean of all the data series



from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import StandardScaler

dfReplaced = dfAll.replace(to_replace = ',', value ='')
dfAllFloat = dfReplaced.astype(float)
matAll = dfAllFloat.to_numpy()
scaler = StandardScaler()
scaledData = scaler.fit_transform(matAll)
print("scaled data")
print(scaledData.shape)
print(scaledData)

lstColumnNames = list(dfAll.columns.values)

dfScaled = pd.DataFrame(scaledData)
dfScaled.columns = lstColumnNames
print(dfScaled.head())
strOutputPath = "/Users/jankoeman/Documents/Main/_PHD2/_NeuralNetwork Data/All Series Scaled.csv"
dfScaled.to_csv(strOutputPath, sep=";")

lstColumns = ['COCOA NEARBY CLOSE','COCOA TOTALS OPEN INTEREST','Cocoa Stocks',
                            'Coffee Stocks',
                            'COFFEE NEARBY CLOSE', 'COFFEE TOTALS OPEN INTEREST', 'Cocoa Relative Return', 'Cocoa Position']
lstFractions = []
matDifferenced = np.zeros(matAll.shape)
for i in range(matAll.shape[1]):
    for j in range(0,11):
        if j == 0:
            colDiff = matAll[:,i]
        else:
            diff = j/10
            col = matAll[:,i]
            colDiff = fdiff(col, diff, axis=0)
        dftest = adfuller(colDiff, autolag='AIC')
        if dftest[1] < 0.05:
            # print("Column Number", i)
            diff = j/10
            # print("Fraction", diff)
            # print("Test statistic = {:.3f}".format(dftest[0]))
```

```
            # print("P-value = {:.3f}".format(dftest[1]))
            if diff == 1.0:
                matDifferenced[1:,i] = colDiff #one item lost
            else:
                matDifferenced[:,i] = colDiff
            lstFractions.append([lstColumns[i],j/10,dftest[0], dftest[1]])
            break

pd.set_option("display.precision", 3)
dfFractions = pd.DataFrame((lstFractions))
dfFractions.columns = ['Series', 'Fraction','ADF Test Stat', 'p-value']
print(dfFractions.head(10))
strOutputPath = "/Users/jankoeman/Documents/Main/_PHD2/_NeuralNetwork Data/Fraction Diff.csv"
dfFractions.to_csv(strOutputPath, sep=";")

scaledFracData = scaler.fit_transform(matDifferenced)
#save in pickledFraction.pkl


# pyplot.figure()
dfPlot = dfScaled.iloc[:600, 0:11]
dfPlot.plot(subplots=True, legend=True)
# pyplot.plot(scaledData[:,:9])
pyplot.show()

strPickledData = "/Users/jankoeman/dlFromScratch/Data/PickledSeries.pkl"
lstSections = []
for i in range(0, scaledData.shape[0], 463):
    matSequence = scaledData[i:i + 463,0:9]
    lstSections.append(matSequence)

#print("Training Data", lstSections)
try:
    f = open(strPickledData, 'wb')
    pickle.dump(lstSections, f)
    f.close()
except OSError as error:
    print("Could not open/read file:", strPickledData)
    print(error)
    sys.exit()

strPickledData = "/Users/jankoeman/dlFromScratch/Data/PickledFraction.pkl"
lstSections = []
for i in range(0, scaledFracData.shape[0], 463):
    matSequence = scaledFracData[i:i + 463, 0:9]
    lstSections.append(matSequence)

# print("Training Data", lstSections)
try:
    f = open(strPickledData, 'wb')
    pickle.dump(lstSections, f)
    f.close()
except OSError as error:
    print("Could not open/read file:", strPickledData)
    print(error)
    sys.exit()

dfLevel = pd.DataFrame(scaledData)
lstCols = ['Cocoa Nearby Close', 'Cocoa Open Interest', 'Cocoa Stocks', 'Coffee Stocks', 'Coffee Nearby Close', 'Coffee Open Interest', 'Relative Return', 'Relative Position']
dfLevel.columns = lstCols
```

```python
print("Level Correllations")
print(dfLevel.corr())

fig = pyplot.figure()
ax1 = fig.add_subplot(211)
ax2 = fig.add_subplot(212)

dfCorrLevel = pd.DataFrame(dfLevel.corr())
dfCorrLevel.columns = lstCols
dfCorrLevel.index = lstCols
sns.heatmap(dfCorrLevel,cmap='RdGy_r',linewidths=.1,xticklabels=1, yticklabels=1, linecolor='black', ax=ax1)
# pyplot.show()
ax1.set(xticklabels=[])
ax1.tick_params(bottom=False)
ax1.set_title('Level Correlations')

dfFraction = pd.DataFrame(scaledFracData)
print("Fractional Correllations")
print(dfFraction.corr())
dfCorrFrac = pd.DataFrame(dfFraction.corr())
dfCorrFrac.columns = lstCols
dfCorrFrac.index = lstCols
sns.heatmap(dfCorrFrac,cmap='RdGy_r',linewidths=.1,xticklabels=1, yticklabels=1, linecolor='black', ax=ax2)
ax2.set_title('Fractional Correlations')
pyplot.show()
```

```python
# Appendix C


# Jan Koeman
# Data Preparation code for analysis with
# Standalone sci-kit-learn RBM
# University of Canterbury, Christchurch, New Zealand
# Department of Economics and Finance
# jan.koeman@pg.canterbury.ac.nz

# The raw data series are publicly available on the web as described in
# my thesis Essays on Commodities, Chapter 3.



from pandas import read_csv
from datetime import datetime
from matplotlib import pyplot
import pandas as pd
import pickle
import sys
import numpy as np
from fracdiff import fdiff
from statsmodels.tsa.stattools import adfuller
import os
import csv
import sys
import midi_manipulation2
from tqdm import tqdm
import pickle
import jkRnnRBM as jk

import rbmdlfs as dlfs

import matplotlib.pyplot as plt

from scipy.ndimage import convolve
from sklearn import linear_model, datasets, metrics
from sklearn.model_selection import train_test_split
from sklearn.neural_network import BernoulliRBM
from sklearn.pipeline import Pipeline
from sklearn.base import clone

import seaborn as sns

def parser(x):
    return pd.to_datetime(x)




strFileName = "/Users/jankoeman/Documents/Main/_PHD2/_NeuralNetwork Data/World Bank Binary 06.csv"
# dfWorldBinary = read_csv(strFileName, header=0, parse_dates=[0], index_col=0, squeeze=True, date_parser=parser, delimiter=";")
dfWorldBinary = read_csv(strFileName, header=0, index_col=0, squeeze=True, delimiter=";")
print("World in Binary")
print(dfWorldBinary.head())
print(dfWorldBinary.tail())
print("Shape:", dfWorldBinary.shape)
#df['Date'] = pd.to_datetime(df['Date'])
#print(dfCocoaStocks.head())
```

```python
# upCoffeeStocks = dfCoffeeStocks.resample('D').interpolate()
# print(upCoffeeStocks.head(10))
# dfCoffee = pd.DataFrame(upCoffeeStocks)
print("Percent Change")
dfWorldReturns = dfWorldBinary.pct_change()
print(dfWorldReturns.head(10))
# dropWorldBankReturns = dfWorldReturns.drop([parser('2000-01-15')])

dfDropped = dfWorldReturns.drop(index='1960M01')
print("1st row dropped")
print(dfDropped.head(10))
print(dfDropped.tail(10))

strOutputPath = "/Users/jankoeman/Documents/Main/_PHD2/_NeuralNetwork Data/_World Binary Source.csv"
dfDropped.to_csv(strOutputPath, sep=";")


dfBinary = dfDropped.apply(lambda x: [0 if np.abs(y) <= 0.03 else 1 for y in x])
print("In Binary")
print(dfBinary.head(10))
print(dfBinary.tail(10))
binaryColumns = dfBinary.columns

X = dfBinary.to_numpy()

x_train, x_test = train_test_split(X,test_size=100)
model = BernoulliRBM(n_components=30, n_iter=100, verbose=True)
# print("pseudoLL before training", np.mean(model.score_samples(x_test)))
model.fit(x_train)
# print("SklearnW.shape", model.components_.T.shape)
# print("SklearnW\n", model.components_.T)
# print(model.intercept_hidden_)

print("pseudoLL trained", np.mean(model.score_samples(x_test)))

# lstFractions = []
# lstColumns = dfBinary.columns
# matDifferenced = np.zeros(X.shape)
# for i in range(X.shape[1]):
#     for j in range(0,11):
#         if j == 0:
#             colDiff = X[:,i]
#         else:
#             diff = j/10
#             col = X[:,i]
#             colDiff = fdiff(col, diff, axis=0)
#         dftest = adfuller(colDiff, autolag='AIC')
#         if dftest[1] < 0.05:
#             # print("Column Number", i)
#             diff = j/10
#             # print("Fraction", diff)
#             # print("Test statistic = {:.3f}".format(dftest[0]))
#             # print("P-value = {:.3f}".format(dftest[1]))
#             if diff == 1.0:
#                 matDifferenced[1:,i] = colDiff #one item lost
#             else:
#                 matDifferenced[:,i] = colDiff
#             lstFractions.append([lstColumns[i],j/10,dftest[0], dftest[1]])
#             break

# Y = matDifferenced
# model = BernoulliRBM(n_components=40, n_iter=100, verbose=True)
```

```
# model.fit(Y[0:600,:])
# print("pseudoLL Y", np.mean(model.score_samples(Y[601:,:])))



sample = np.zeros(47)
sample[38:44] = 1
# sample[39]= 0
# sample[40] = 0
# sample[3] = 0
reconstructed = model.gibbs(sample)
print("Sample", sample)
print("Reconstructed", reconstructed*1)

# reconstructed2 = model.gibbs(reconstructed)
# # print("Sample", sample)
# print("Reconstructed2", reconstructed2*1)
recon = reconstructed*1
# for i in range(0,47):
#     print(binaryColumns[i], int(sample[i]),recon[i])
# for i in range(0,47):
#     print(binaryColumns[i], recon[i])


reconstructions = np.zeros((47,47))
for i in range(0,47):
    sample = np.zeros(47)
    sample[i] = 1
    reconstructed = model.gibbs(sample)
    # print("Sample", sample)
    # print("Reconstructed", reconstructed*1)
    reconstructions[i] = reconstructed*1

print("results shape", reconstructions.shape)
print("results", reconstructions)
dfRecon = pd.DataFrame(reconstructions)
dfRecon.columns = binaryColumns

strOutputPath = "/Users/jankoeman/Documents/Main/_PHD2/_NeuralNetwork Data/_World Binary.csv"
dfRecon.to_csv(strOutputPath, sep=";")

dfRecon.insert(1, 'Commodity', binaryColumns, True)
dfHeatmap = dfRecon.set_index('Commodity')
print("heatmap df")
print(dfHeatmap.head())

# sns.heatmap(dfHeatmap, cmap="YlGnBu",linewidths=.1,xticklabels=1, yticklabels=1, linecolor='black')
dfGraph = dfBinary.iloc[600:647, :]
# sns.heatmap(dfGraph, cmap="YlGnBu",linewidths=.1,xticklabels=1, yticklabels=1, linecolor='black')
# pyplot.show()
```

Appendix D: The advantages and disadvantages of Machine Learning in comparison with traditional econometric tools

Machine Learning methods, in comparison with more traditional econometric models, are quite often considered to be black box tools. Following is a short summary of the advantages and disadvantages of ML as opposed to traditional econometric tools. For more information, consult Castelvecchi (2016), Loyola-Gonzalez (2019), or Holzinger et al. (2019).

DISADVANTAGES

1. A significant amount of data is required.
2. ML systems are very sensitive to changes in input parameters. There is an art to training ML systems and the training skills can take time to acquire.
3. Expert level programming expertise is required to modify or debug basic system structures.
4. Since the system is a black box, it is difficult to understand how it works by examining the interior weights. This makes it necessary to devise a system of training inputs where the output of the system is known, which may take a significant amount of time.

ADVANTAGES

1. ML models can accept non-stationary series as input and can learn non-linear associations to generate output.

2. A Restricted Boltzmann Machine has similar capabilities to Principal Components Analysis (PCA). A Restricted Neural Network RBM can simultaneously learn input associations and sequence the associations through time. This is not possible with PCA or econometric tools.
3. The RNN_RBM and RBM learn the joint probability distribution of several variables instead of the single distribution of one variable as a function of the others. In addition, the RNN_RBM is capable of learning the evolution of the joint probability distribution over time.
4. The RNN_RBM is a generative model. In other words, once trained a RNN_RBM can generate output that is drawn from the probability distribution learned from the input.
5. The RBM architecture is unsupervised, where it is not necessary to know the correct output for each set of inputs. The system learns the patterns in the data simply by being shown the possible range of inputs.

Table 1 – Fractional Differentiation Results on Commodity Time Series.

Series	Fraction	ADF Stat	p-value
Cocoa Nearby Close	0	-2.98	0.04
Cocoa Totals Open Interest	0.7	-3.69	0.00
Cocoa Stocks	0	-4.61	0.00
Coffee Stocks	1	-4.94	0.00
Coffee Nearby Close	0.7	-3.35	0.01
Coffee Totals Open Interest	0.8	-4.20	0.00
Cocoa Relative Return	0	-9.07	0.00
Cocoa Position	0	-8.26	0.00

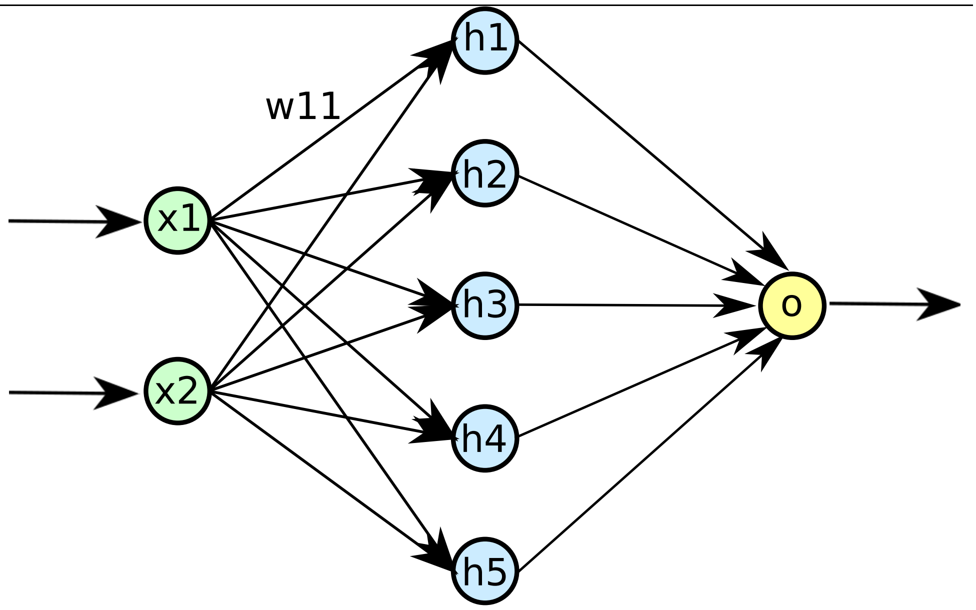
This table depicts the results from iteratively fractionally differencing the listed commodity time series using the method of Lopez de Prado (2018)

Table 2 – RNN-RBM prediction results.

Description	RSS	Baseline RSS	Kolmogorov-Smirnov p-value
Level Data	3.66	.65	.89
Fractional Differenced Data	2.98	.65	.96

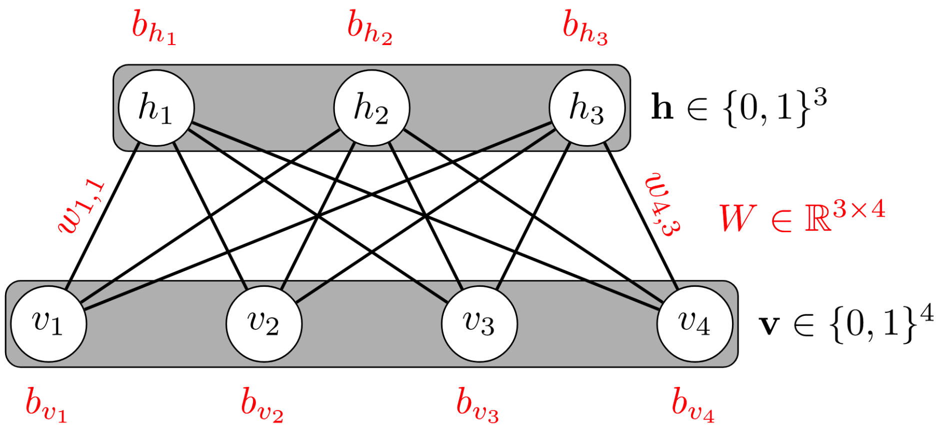
This table illustrates the results from one-step-ahead predictions on hold out date after training the RNN_RBM for 100 epochs. The baseline prediction uses today's price as tomorrow's estimate.

Figure 1 – Simple Feedforward Network



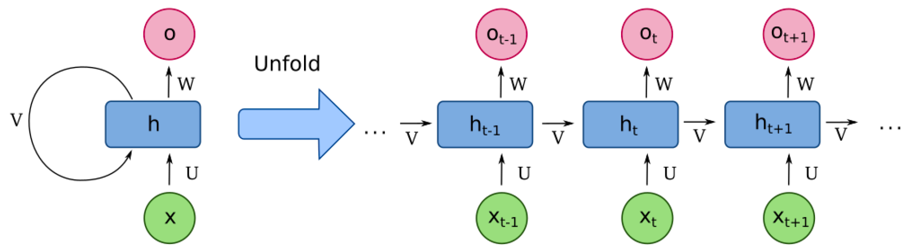
This figure depicts the simplest type of feedforward neural network where the inputs x_i are fed through the hidden nodes to the output node.

Figure 2 – Restricted Boltzmann Machine (RBM)



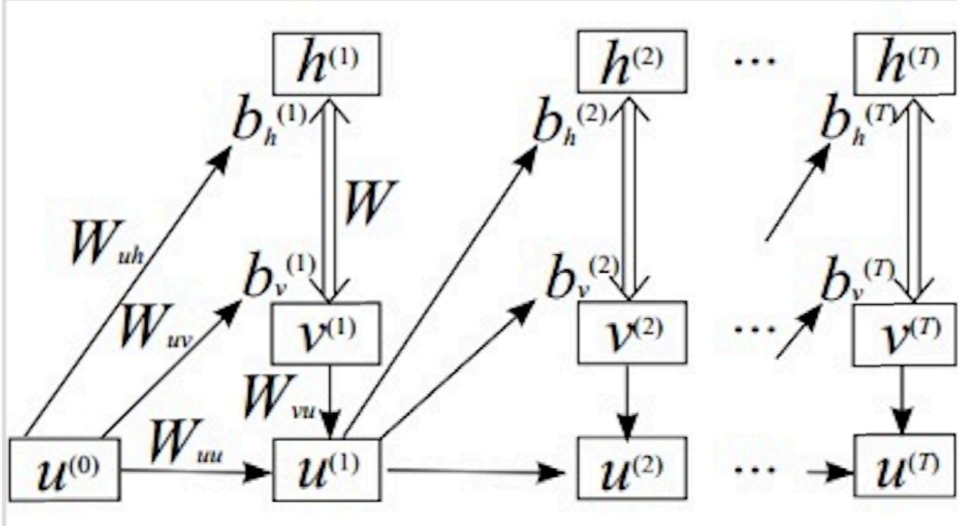
This figure depicts a Restricted Boltzmann Machine (RBM).

Figure 3 – Recurrent Neural Network (RNN)



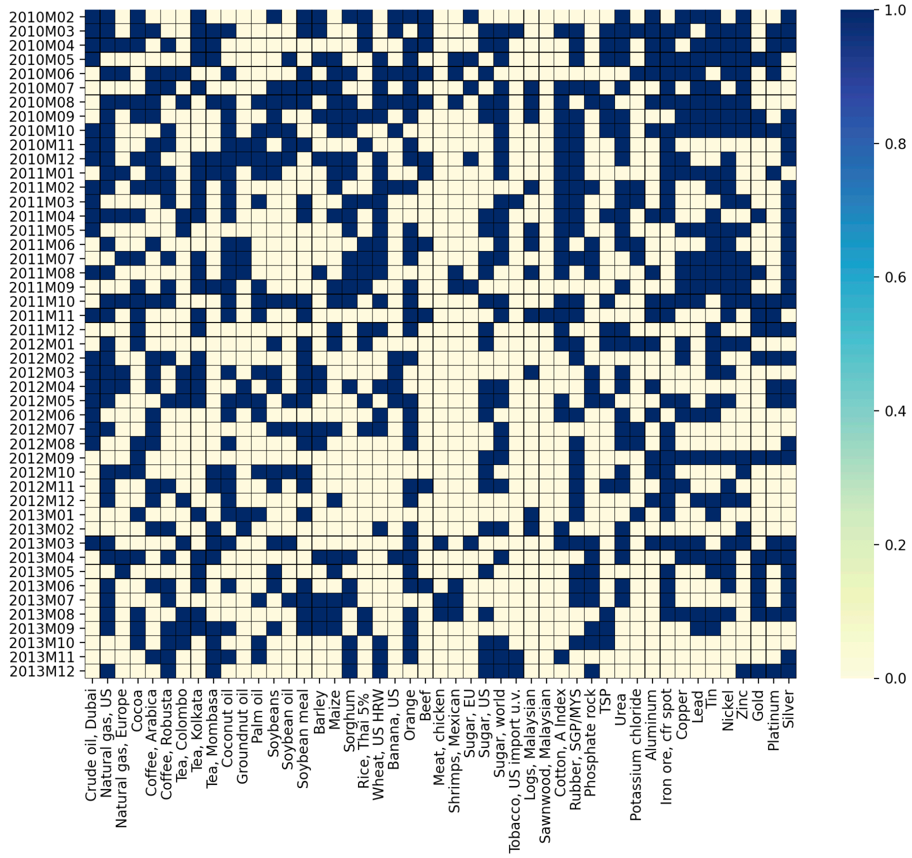
This figure depicts a Recurrent Neural Network (RNN). X_t is a sequence that is fed to each timestep in turn. There are as many (x_t, h_t, o_t) node groups as there are steps in the sequence.

Figure 4 – The Recurrent Neural Network Restricted Boltzmann machine (RNN_RBM)



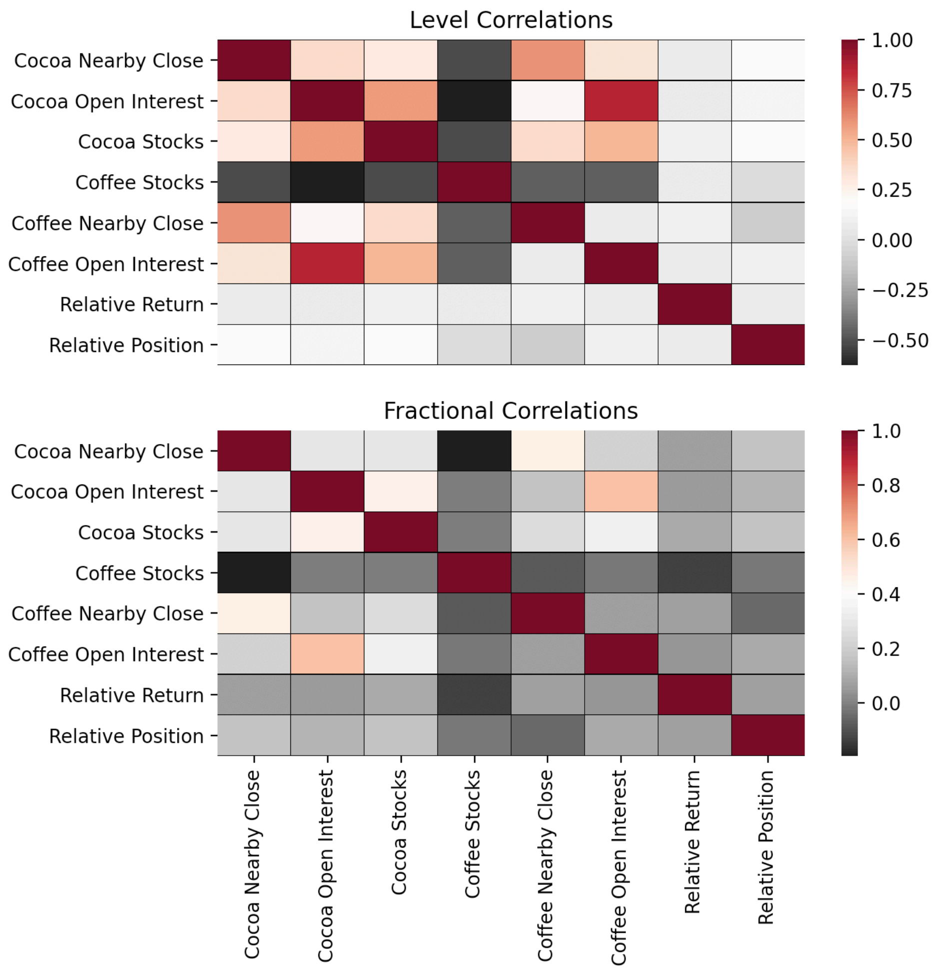
This figure depicts a Recurrent Neural Network Restricted Boltzmann machine (RNN_RBM). At each time step, the biases for a separate conditional RBM are initialized from the RNN. (Boulanger-Lewandowski et al. (2012))

Figure 5 – Monthly Commodity Changes Sample Data



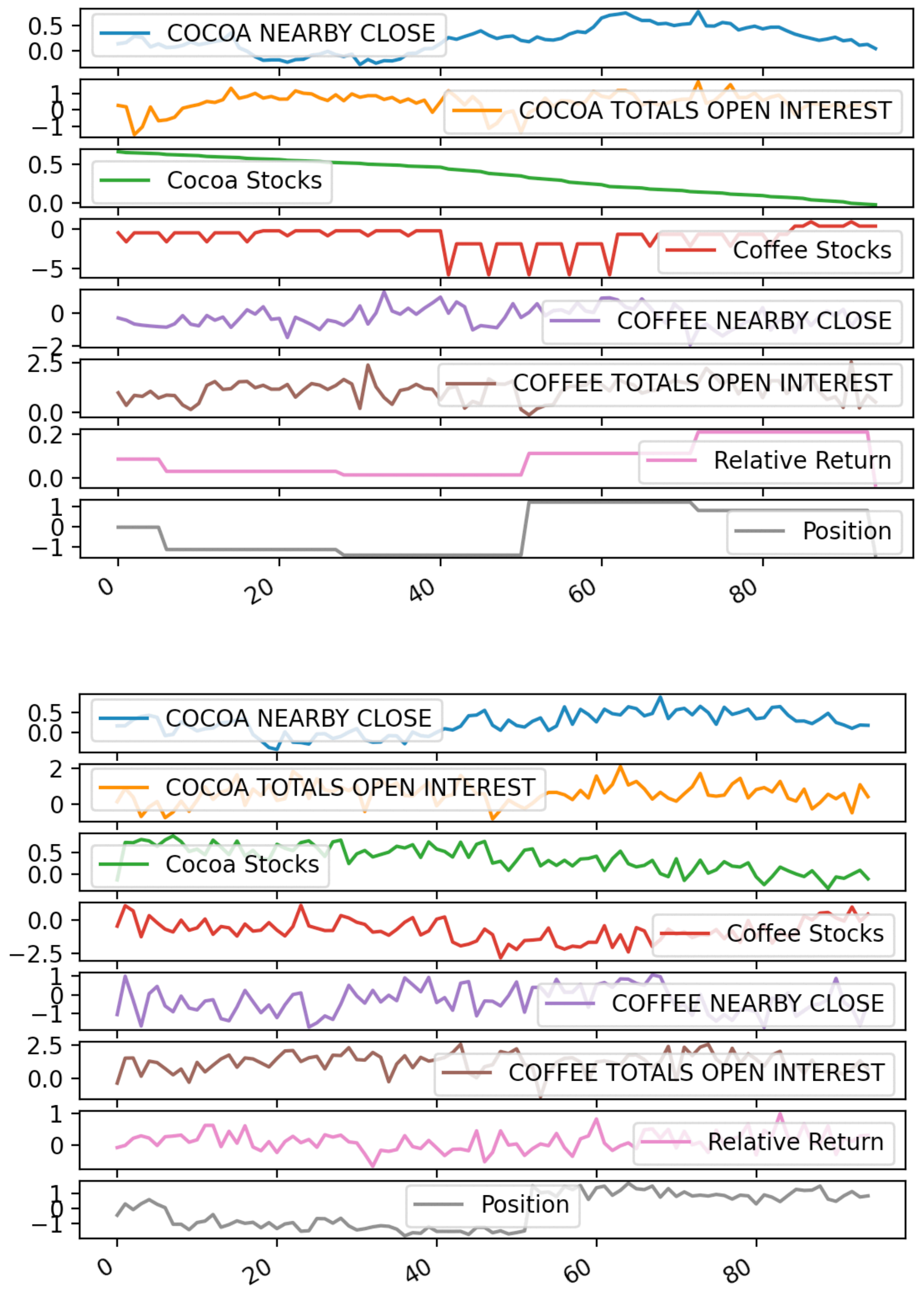
This figure reports a sample of monthly commodity price changes from the World Bank pink sheet from the period February 2010 to December 2013. The commodities are listed along the x-axis. If the absolute monthly percentage change is greater than .04, then the matrix value is set to 1 and appears as black, otherwise it is set to zero and appears as yellow. The complete

Figure 6 –Heatmaps of cocoa related time series.



This figure illustrates heatmaps of the level correllations and the fractionally differenced correlations of the cocoa and related time series input to the RNN_RBM. The fractional correlations indicate that some long term price information is retained.

Figure 7 – RNN_RBM cocoa and related time series actual values versus predictions



This figure illustrates the fractionally differenced commodity time series actual values in the top panel, and the predictions made by the RNN_RBM in the bottom panel. The RNN_RBM was trained for 200 epochs with 50 hidden units and a learning rate of .01. Ideally, the bottom panel would be a close match to the top panel but he predictions are only a noisy estimate of the actual values.

General Conclusions

Government regulation of commodities markets can be a double-edged sword. Whilst regulation in the United States may protect (and enrich) dairy farmers from price fluctuations, the effect on the proper hedging role of futures markets can be dramatic. Chapter one of this thesis illustrates that the correct design of the Global Dairy Trade spot market index in New Zealand results in substantially better hedging effectiveness that the corresponding indices constructed in the US.

Modern commodities markets are substantially more complicated than commodity markets from the last century. The advent of high-speed data collection and analysis has lead to the presence of completely new trader categories including momentum and index traders. Chapter two illustrates that a carefully crafted inventory and consumption-based fundamental model , of the cocoa market, the apex of the Traditional Theory of Storage, no longer can adequately explain price movements in modern markets.

Machine learning systems that have achieved spectacular success in speech and image pattern recognition tasks are starting to find application in finance, despite a high degree of technical complexity. Paper three of this thesis applies two state-of-the-art neural network models to continuous and binary financial time series. A Restricted Boltzmann Machine (RBM) is utilized to characterize the contemporaneous changes in commodity prices across the commodity universe and a Recursive Neural Network Restricted Boltzmann Machine (RNN_RBM) is used to analyze fundamental, momentum, and index time series inputs related to the cocoa market. Imperfections in time series including non-stationarity are dealt with by fractional differencing, a new technique invented by Lopez de Prado. The RNN_RBM predictions do not better simple baseline predictions but the RBM analysis indicates that prices of groups of commodities do move together. The automatic construction of time-dependent multivariate probability distributions appears to hold great promise for automatic feature detection and modelling in finance. Several other machine learning applications, using either the RNN_RBM or RBM could be explored as extensions to our research. In particular, the RNN_RBM could be used to explore the co-movement of commodity convenience yields, and to explain the multi-commodity term structure.