THREE ESSAYS ON QUANTITATIVE ANALYSIS IN COMMODITY MARKETS

BY

JIARUI LI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Agricultural and Applied Economics
in the Graduate College of the
University of Illinois Urbana-Champaign, 2022

Urbana, Illinois

Doctoral Committee:

Professor Scott H. Irwin, Chair and Director of Research
Professor Gary Donald Schnitkey
Professor Teresa Serra
Associate Professor Xiaoli Etienne, University of Idaho
Dr. Todd Hubbs, United States Department of Agriculture

ABSTRACT

This dissertation consists of three essays that apply quantitative analysis in commodity markets. The first paper essay studies the pass-through impacts from the Low Carbon Fuel Standard (LCFS) program to wholesale fuels provided in California. The goal of LCFS is to reduce carbon intensity of transportation fuels provided in California. LCFS mandates an overall carbon intensity reduction for fuels in gasoline pool and diesel pool by 10% by 2020 through the tradable LCFS credit system. To evaluate if LCFS effectively discourages the consumption of traditional fuels, we estimate the long-run equilibrium and short-run dynamics of the pass-through from the LCFS credit prices to wholesale gasoline and wholesale diesel prices in California from 2016 to early 2020. Our pass-through models control for fundamental time-constant fixed effects and time-varying seasonal patterns in wholesale fuel prices. Wholesale gasoline fuels have quick and complete pass-through over the full sample period, suggesting fuel suppliers can pass the full LCFS credit costs to downstream buyers about 4 business days; for wholesale diesel fuels, they have incomplete long-run pass-through, and over 15 business days, they can only recoup 64% of the LCFS credit costs.

The second essay examines the forecasting accuracy of a batch of yield forecasting models that directly transform the ordinal crop condition ratings to the numeric condition index along with a recently developed model introduced by Begueria and Maneta in 2020 that applies the cumulative link mixed model to transform the condition ratings to the continuous condition index. We conduct the out-of-sample yield forecasts recursively for corn and soybean from 2000 through 2020 for all models. We measure the forecasting errors of this group of models and find throughout the growing season, the average root-mean-square-percentage-error (RMSPE) is about 5% for corn

and 6% for soybean. Our findings suggest this group of models that use crop conditions data provide accurate yield forecasts. Next, we compare the model developed by Begueria and Maneta (BM model) with its four competing yield forecasting models that have already been widely applied by industry practitioners. Single-horizon model forecasting comparison tests like modified Diebold Mariano test and Model Confidence Set test fail to show that BM model significantly outperforms its competitors for each week throughout the growing season. We also conduct the multi-horizon forecasting comparison test. The results from the average Superior Predictive Ability test show that despite BM model has more complex model specification, throughout the growing season, it does not provide superior out-of-sample yield forecasts than its competitors.

The third essay applies a recently developed cross-quantilogram (CQ) test to examine the impact of Commodity Index Traders (CIT) positions on returns in four agricultural futures markets from 2004 – 2019. Most previous studies reject the basic tenet of the Masters Hypothesis that financial index investments have pressured agricultural futures prices upward. However, the impact of this investment may be more complicated and nuanced than can be detected by the relatively simple linear Granger causality tests used in many previous studies. We conduct three linear causality tests to provide the baseline about the relationship between CIT positions and futures returns. Test results fail to reject the null of no causality in most of the cases across the different tests, measures of position pressure, or the sample period considered. Next, we apply the CQ test of directional predictability in the tails of the distributions of the CIT positions and price movements. Consistent with the standard linear causality tests, we find no evidence that supports the Master Hypothesis.

ACKNOWLEDGMENTS

Foremost, I want to thank my academic advisor Dr. Scott Irwin for always being kind, patient and supportive. I think I am very lucky to have the chance to learn from him and work with him. He opened my eyes to the world of commodity markets and price analysis; he taught me how to construct interesting research questions; and he also showed me how to think critically and provide answers to these questions. His continuous encouragements and positive feedbacks helped me finish this dissertation and let me always be passionate about conducting quantitative analysis in these areas.

Besides my advisor, I am thankful to my committee members: Dr. Xiaoli Etienne, Dr, Todd Hubbs, Dr. Teresa Serra, and Dr. Gary Schnitkey. Their careful reviews and insightful suggestions significantly bring my work to a higher level. I am also grateful to Dr. Philip Garcia, Dr. Michel Robe, and Dr. Mindy Mallory for giving me valuable suggestions for my study. I would like to thank Hongxia Jiao for her help in collecting and organizing data for my research projects.

I also want to thank the wonderful staff of the Department of Agricultural and Consumer Economics, especially Dung Quach-Wisdom, Carli J. Miller, Pam Splittstoesser, Melissa Warmbier, Donna Stites, and Ian Chong for giving me a lot of administrative supports and for making the paperwork so easy and efficient. I want to thank Dr. Nick Paulson for offering us graduate students a lot of valuable opportunities. I am also thankful to my peers, Xinyue He, Siyu Bian ,Kun Peng, Zhepeng Hu, Yijia Li, Tengjiao Chen, Anabelle Couleau, Quanbiao Shang, Pedro Tremacoldi-Rossi, Chi Ta for your help and support throughout my graduate life.

Finally, I want to thank my friends and family members in China and the U.S.. I am thankful to my friendship with Chang Lu, Xiaoxi Xue, Xiaofei Yang, Kaiyi Chen, Liqing Li,

Shihang Wang, Kehui Zhang, Jiayi Yang, Yingyu Liu, Yujun Zou, Yaning Zhou, Mengyuan Wang. Thank you for giving me confidence and your full supports. I am thankful to my parents, my Mom Jia Ji and my Dad Honglin Li. Thank you for always encouraging me to think positively and have faith in myself. I want to thank my grandparents, Yinfa Ji and Shihui Xu for your infinite love. Special thanks to my husband Yichen Zhang. Thank you for being my rock and bringing me happiness.

To my parents Jia Ji and Honglin Li

TABLE OF CONTENTS

CHAPTER 1:

INTRODUCTION

The first essay of this dissertation focusses on the prices of wholesale petroleum fuels and how they are affected by market-based policy that aims at reducing carbon emissions in transportation sectors. For example, in California, the Low Carbon Fuel Standard (LCFS) designs a transaction mechanism based on the carbon intensity of regulated fuels that levies tax on traditional fuels and rewards credits to renewable fuels. In a competitive market, upstream obligated parties have the incentives to fully pass-through the LCFS credit costs and cause the inflation of prices. LCFS reaches its goal by increasing the prices of these less-clean fuels. We build wholesale gasoline and diesel price spreads between Los Angeles and other bulk markets outside of California to measure the pass-through impacts from LCFS. To validate the design of the pass-through framework, we build the price spreads between two markets that are free from the obligation of LCFS and then measure if they respond to the variations in LCFS credit prices as the placebo analysis. We find that there are no pass-through impacts from LCFS. Pooling three wholesale gasoline spreads between Los Angeles and the other three bulk markets, we find complete and fast pass-through, suggesting within a few business days, fuel suppliers can fully pass-through LCFS credit costs. In wholesale diesel market, we find incomplete pass-through in both long-run and short-run, suggesting fuel suppliers have to undertake a certain portion of the credit costs. This study estimates the magnitude and speed of the pass-through rate of LCFS credit prices. Our results show the effectiveness of LCFS and we further present the different market structures for wholesale gasoline market and wholesale diesel market.

The second essay of this dissertation focuses on the fundamentals of agricultural spot markets. Crop yield forecasts have been an essential component of supply, demand, and price forecasting. We focus on a group of yield forecasting models that use the crop conditions data to provide yield forecasts. Among these models, a recently developed model by Begueria and Maneta (2020) has sophisticated specifications, and they claim it outperforms other simpler models that use the same crop condition information provided by USDA. We conduct forecast competitions between the model developed by Begueria and Maneta (2020) and other simpler crop condition models developed by Irwin and Good (2017a,b) and Bain and Fortenbery (2017) using recursively out-of-sample yield forecasts. The data for the study consists of weekly state and national crop condition ratings from 1986 through 2020 for corn and soybean. Statistical forecast comparison tests do not provide any supporting evidence to show the model proposed by Begueria and Maneta (2020) outperforms simpler models.

The third essay of this dissertation focuses on agricultural futures market, and we revisit the Master Hypothesis that blames the buying pressure from index traders drive up the commodity futures prices. Most previous studies apply the linear Granger causality tests find evidence against the Master Hypothesis. However, as the impact of financial index investment in agricultural futures markets is more complicated and nuanced than can be detected by relatively simple linear frameworks, our study applies the recently developed cross-quantilogram test to examine the relationship between index investment and futures prices in the tails of the data. We use weekly index traders' positions and futures returns from January 6, 2004 through December 31, 2019 for Chicago Board of Trade (CBOT) corn, wheat, soybeans, and Kansas City Board of Trade (KCBOT) wheat. Similar to the linear tests, we find very little evidence of a directional relationship in the extremes of the distributions.

CHAPTER 2:

THE PASS-THROUGH OF THE LOW CARBON FUEL STANDARD CREDIT PRICES TO WHOLESALE FUELS PRICES IN CALIFORNIA

2.1 Introduction

The California Low Carbon Fuel Standard (LCFS) was created in 2007 to enhance the regulations on in-state petroleum carbon emissions. The program aims at reducing carbon pollution and dependencies on non-renewable fuels in the transportation sector besides incentivizing the development of alternative fuels and alternative fuel vehicles. This program is fuel-neutral as it encourages the use of biofuels, electricity, and hydrogen fuels; and it has been a key tool applied to reduce emissions of greenhouse gas (GHG), nitrogen oxides, and particulate matter. In 2011, LCFS provided the annual Carbon Intensity (CI) Compliance Schedule for all transportation fuels supplied in California. CI measures the "life-cycle" emissions of each fuel. The goal of LCFS is to reduce the average carbon intensity (CI) of the California transportation fuel pool by at least 10% by 2020, based on crude oil's CI level in 2010 (LCFS, 2015). In 2018, LCFS amendments were voted and passed with a 2030 target of 20% CI reduction from the 2010 CI levels (CARB, 2018).

Fossil fuels have higher CI ratings than their obligated LCFS standards. They generate deficits in proportion to the amount that are above the standard CI. On the other hand, CI ratings of renewable fuels are lower than their assigned LCFS standards. They generate credits in proportion to the amount that are below the standard CI. To meet the mandates of LCFS CI standards, fossil fuel suppliers can either purchase the LCFS credits from fuel providers who hold surplus credits, or they can implement more renewable fuel content in their final products to offset the deficits. With no outside money invested in the LCFS credit market, this policy is revenue

neutral by building upon the bilateral "over the counter" transaction system between obligated parties. The value of LCFS credits reflect the cost difference between the more expensive renewable fuels and the more economic traditional fuels. Moreover, in the market, it sends out signals of tax and subsidy to drivers in California and affect their choice of fuels. Through the credit transfer system, LCFS simultaneously subsidizes cleaner fuels and taxes fossil fuels. The economic effectiveness of LCFS is realized through its credit trading system at different stages of the transportation fuel supply chain. With the hypothesis that obligated parties have incentives to pass-through the additional LCFS credit costs to their downstream buyers, as traditional fuels get more expensive through the supply chain, LCFS eventually motivates the consumption of cleaner fuels in California.

The pass-through model is a powerful tool to investigate if tax policies impact markets and consumer welfares. Alfred Marshall (2009) defines "pass-through" as the "diffusion throughout the community of economic changes which primarily affect some particular branch of production or consumption". In energy markets, "economic changes" refer to the context where fuel prices become more expensive as the result of more stringent energy policies that usually levy taxes on carbon emissions. It also refers to the additional input costs borne by suppliers who has the incentives to pass them down to downstream buyers through the supply chain (Stolper, 2016). The pass-through from cost changes to fuel prices has been widely discussed in the energy industry (Chesnes, 2016; Lade and Bushnell, 2016). For example, Borenstein, Cameron, and Gilbert (1997), Bachmeier and Griffin (2003), and Lewis (2011) studied the price transmissions from crude oil to retail gasoline and the asymmetric responses from retail gasoline prices to the increase and decrease of crude oil prices.

The tax incidence model in a perfectly competitive fuel market suggests that a pass-through rate is bounded between 0% – 100%. Pass-through rate is defined as $\rho = \frac{1}{1+\frac{\varepsilon_D}{\varepsilon_S}}$, where ε_D and ε_S are the elasticities of demand and supply, respectively (Weyl and Fabinger, 2013). Figure 2.1 illustrate in a perfectly competitive market, how tax costs are split between sellers and buyers. When demand and supply are relatively elastic, as shown in Figure 2.1(a), pass-through rate $\rho \in (0,1)$, tax costs are undertaken proportionately by buyers and sellers. The pass-through rate is 100% when supply is perfectly elastic ($\varepsilon_S \to \infty$), as shown in Figure 2.1(b) or when demand is perfectly inelastic ($\varepsilon_D \to 0$), as shown in Figure 2.1 (c). The two extreme cases show buyers shoulder the complete tax burden. Evidence of pass-through in competitive markets was found in RINs credit market where wholesale fuels can fully pass-through the mandating credit costs to fuel prices (Knittel, Meiselman, and Stock, 2017). In the U.S. manufacturing industries, Ganapati, Shapiro, and Walker (2016) showed that suppliers in the plant-level units can pass down 70% of energy price-driven changes in input costs to consumers. When demand function is sufficiently convex, or suppliers have market power at a certain level, pass-through rate ρ is greater than 100% and known as overfull pass-through (Stolper, 2016). Evidence of the overfull pass-through was found in Spain that 24% of gas stations have estimated pass-through rates greater than 100% (Stolper, 2016). Burkhardt (2016) found that during 2013–2014, when the RINs credit price increased due to unexpected shocks, in the gasoline market, the costs were borne 16 times more by consumers than producers. Similarly in the diesel market, the extra costs were borne 1.28 times more by consumers than producers.

The U.S. Renewable Fuel Standard (RFS) implemented nationwide in 2005 has a similar goal but different regulatory framework from LCFS. RFS mandates that fuels for transportation use must satisfy volumetric requirements through different categories of biofuels. These

requirements are met through Renewable Identification Numbers (RINs), which are generated with the production or import of renewable fuels, and they are tradable credits after renewable fuels are blended with traditional fuels. Under RFS, the RINs transaction mechanism provides subsidy to biofuels producers by levying extra taxes from conventional fuels producers. Studies observed that variations in wholesale fuel price spreads and net RINs obligations are linked. The co-movement of fuel price spreads[1] and RINs prices suggests that transportation fuel prices in the U.S. are responsive to the incidence of RFS. Evidence showed that fuel providers can fully pass-through the RINs costs to fuel prices quickly and sufficiently, but not immediately (Burkholder, 2015; Knittel, Meiselman, and Stock, 2017).

To build the pass-through framework, we measure how price difference between fuels provided in California and other states respond to the movement of LCFS credit prices. The strategy of using price spreads follows the empirical design proposed by Knittel, Meiselman, and Stock (2017) for the following reasons. First, RFS and LCFS share similar mechanisms of credit generations, and they both focus on stimulating the consumption of cleaner fuels. It is reasonable to expect that LCFS credit costs are likely to be passed down to wholesale fuels prices in California. Second, fuel price spreads between wholesale fuels in California and other states remove price impacts from the RINs and other time-constant fixed effects. Third, because similar wholesale fuels in other states are free of the LCFS obligations, fuel price spreads reflect the extra LCFS credit costs carried by California fuel providers and should respond to the variations in LCFS credit prices. To collect vigorous spot market prices, we choose Los Angeles wholesale petroleum market and three representative markets across the nation: New York, Chicago, and Gulf Coast as they are three recognized bulk petroleum markets where wholesale fuels prices are

[1] Price spreads are defined as price differences for the same fuel between two different markets.

determined and large amount of fuels transactions take place (Pouliot, Smith, and Stock, 2017). Before we estimate the pass-through impacts from LCFS credit prices to wholesale fuels in California, we conduct a placebo analysis that estimates the long-run pass-through from LCFS credit prices to price spreads between markets outside of California. The placebo analysis of having zero impact from the fake treatment groups further validates our framework of using price spreads and LCFS credit costs to estimate the pass-through from LCFS credits to wholesale fuels provided in California. We provide long-run and short-run estimations of pass-through rates of LCFS credit obligations to wholesale fuel prices. The empirical analysis allows us to estimate the magnitude and speed of the pass-through from LCFS credit prices to wholesale petroleum fuels prices in California.

This paper contributes to the literature by extending the pass-through model applied in the RINs market to a subnational LCFS credits market in California. Some studies show that LCFS and RFS are two mutually reinforcing policies (Whistance, Thompson, and Meyer, 2017) because together they promote the use of cleaner fuels especially fuels with high ethanol-content. Huang et al. (2013) provide evidence that in California, having both programs in effective can achieve a higher reduction in GHG emissions and lead to a cut down on first generation biofuels (corn ethanol and sugarcane ethanol) and an increase in second generation biofuels (cellulosic ethanol and Biomass to Liquid (BtL)). Our study provides a different angle to show besides RFS, whether the incidence of LCFS in California successfully encourages the consumption of renewable energies and ultimately reaches the goal of reducing GHG emission.

2.2 Background

LCFS defines regulated parties as the initial fuel producers and importers of designated transportation fuels provided in California. Deficits are generated by fossil fuels, mainly gasoline

and diesel blendstocks. Credits are generated by alternative fuels, for example, renewable fuels (biofuels, biogas), hydrogen, and electricity. LCFS proposes two separate standards for gasoline and its substitutes (gasoline pool), and diesel and its substitutes (diesel pool). Each individual standard is based on the CI rating of the fuel, and this setup prevents the incentives of shifting from one fossil fuel to another (Yeh et al., 2016). Table 2.1 provides the LCFS Carbon Intensity (CI) compliance schedule and the average CI ratings of CARBOB and ULSD for both gasoline pool and diesel pool respectively. In mid-2015, CARB started the process of readopting LCFS and establishing a more stringent compliance target with 10% CI reduction from a 2010 baseline by 2020. Starting January 1, 2016, LCFS was officially readopted and set up an annual 2% CI reduction for both gasoline pool and diesel pool.

LCFS uses the equations below to calculate credits or deficits for each fuel in gasoline or diesel pools:

$$Credit_{i,t}^{XD}\left(Deficit_{i,t}^{XD}\right) = \left(CI_{Standard,t}^{XD} - \frac{CI_{i,t}^{XD}}{EER_{i,t}^{XD}}\right) \cdot E_{displace}^{XD} \cdot \frac{1MTCO_2e}{10^6 gCO_2e}, \tag{2.1}$$

and

$$E_{displace}^{XD} = E_{i,t} \cdot EER_{i,t}^{XD}, \tag{2.2}$$

where XD identify whether a fuel (i) is in gasoline pool or diesel pool; $CI_{Standard,t}^{XD}$ (gCO_2e/MJ) is the carbon intensity standard of fuel i under XD in year t; $CI_{i,t}^{XD}$ (gCO_2e/MJ) is the carbon intensity of fuel i in year t; $E_{i,t}$ $(MJ/gallon)$ is the fuel energy density; $EER_{i,t}^{XD}$ is the Energy Economy Ratio of fuel i, as an adjustment for the energy efficiency compared to gasoline or diesel.

For example, energy density of CARBOB is 119.53 $MJ/gallon$.Table 2.1 shows that in 2017 the standard CI of gasoline pool is 95.02 ($MJ/gallon$), and the CI score of CARBOB is 99.78 ($MJ/gallon$). Applying the Equations (2.1) and (2.2), we can calculate that one gallon of CARBOB generates $|(95.02 - 99.78)| \times 119.53 \times 10^{-6} = 5.69 \times 10^{-4}(MT)$ deficits. For

ULSD, the energy density is 134.47 ($MJ/gallon$). Table 2.1 shows in 2017 the standard CI of diesel pool is 98.44 (gCO_2e/MJ) and the CI score for ULSD is 102.01 (gCO_2e/MJ). Using Equations (2.1) and (2.2), we can calculate that burning one gallon of ULSD would generate $|(98.44 - 102.01)| \times 134.47 \times 10^{-6} = 4.8 \times 10^{-4}(MT)$ deficits. To convert LCFS deficits from $\$/MT$ to $\$/Gallon$, we assume LCFS credit price is transacted at $\$200/MT$, for gasoline, the credit price is $|(95.02 - 99.78)| \times 119.53 \times 10^{-6} \times 200 = 0.114$ ($\$/gallon$), for ULSD, the credit price for diesel is $|(98.44 - 102.01)| \times 134.47 \times 10^{-6} \times 200 = 0.098$ ($\$/gallon$).

LCFS offers multiple options for deficit generators to meet their annual obligations. They can directly purchase credits from alternative fuels providers, or they can transfer obligations to downstream blenders with the ownership of fuels. These options under this mechanism increases fuel costs by the amount of LCFS credit prices. For this study, we focus on fuels provided publicly in the wholesale markets. A simplified illustration of LCFS credits and deficits generation and obligation transfer mechanisms on the gasoline supply chain is presented in Figure 2.2. In the first case, a simplified credit transfer mechanism is shown in Figure 2.2 (a), where fuels initial providers are the designated regulated parties, and downstream blenders are free of compliance obligations. Here we use wholesale gasoline (CARBOB) and ethanol as an example to illustrate the pass through of LCFS credit prices to wholesale fuel prices, considering wholesale diesel (ULSD) and diesel pool have a very similar credit transfer mechanism. LCFS credit price is determined by the demand and supply in the market. For example, now we assume credits are transacted between regulated parties at $\$a/gallon$. For each gallon of ethanol, producers are awarded by CARB at $\$a/gallon$. For each gallon of CARBOB, refiners and importers are required to retire a gallon of credit at $\$a/gallon$ to CARB to demonstrate their compliance. This can be achieved through directly purchasing credits from ethanol producers. Purchasing credits brings extra cost to gasoline

providers. To recoup it, when they sell their fuels to downstream blenders, they charge a higher gasoline price at $\$(x+a)/gallon$, assuming the net cost of gasoline is $\$x/gallon$. When ethanol producers sell their fuels to blenders because there is no LCFS credit value remained in ethanol at this stage, assuming price for ethanol is $\$y/gallon$, now without the LCFS value in ethanol, price becomes $\$(y-a)/gallon$. In this scenario, the resulting blended fuel costs are $\$(x+y)/gallon$. In the second case, initial regulated parties choose to transfer their compliance obligations to fuel blenders, along with the transfer of their fuels' ownerships. A simplified credit transfer system is shown in Figure 2.2 (b) using wholesale gasoline and ethanol as an example. After acquiring the ownership of CARBOB and ethanol, fuel blenders are considered as regulated parties under LCFS. They are required to retire credits to CARB for blends' gasoline components at $\$a/gallon$ and they are awarded with credits for blends' ethanol components at $\$a/gallon$. In this transaction, fuel blenders receive a gallon of wholesale gasoline from CARBOB refiners and imports for $\$x/gallon$, and a gallon of ethanol with $\$a/gallon$ LCFS credit value from ethanol fuel producers for $\$y/gallon$. In this scenario, the resulting blended fuel costs are $\$(x+y)/gallon$.

LCFS created the Credit Clearance Market system that relaxes the situation when the market is short of credits. Through the process, a regulated party can choose to purchase its "pro-rata shares of credit"[2] when it fails to retire the requested credits by the end of that year (LCFS, 2015). The LCFS credits are directly exchanged between credit providers and deficit holders "over the counter". The total amount of available credits determines the price. With credit price being positive, LCFS provides a cash transfer mechanism that directly brings the obligation payments to cleaner fuels producers as an incentive for more environment friendly fuels. Yeh et al. (2016) point

[2] The calculation of the "pro-rata shares" of a regulated party (A) is defined as $(\text{A's deficit}/\text{total deficit}) \times \min(\text{pledge credits}, \text{total deficits})$.

out that the current positive credit prices indicate a binding and effective program, and it would continue to have an expected positive marginal compliance with the development of the program.

2.3 Data and Descriptive Statistics

2.3.1 Data and construction of variables

Our data consists of daily spot prices of wholesale gasoline and diesel from Los Angeles, New York, Chicago, and Gulf Coast markets, and LCFS credit prices for gasoline pool and diesel pool respectively. They are all collected from the Ethanol & Biodiesel Information Service Report provided by OPIS. The full sample period covers the U.S. business days from January 4, 2016, to March 12, 2020, with 1050 observations. Wholesale gasoline and diesel in Los Angeles are provided in the form of California Reformulated Blendstock for Oxygenate Blending (CARBOB) and Ultra Low Sulfur Diesel (ULSD), respectively. In New York, Chicago, and Gulf Coast markets, wholesale gasoline and diesel are supplied in the form of Reformulated Blendstock for Oxygenate Blending (RBOB) and ULSD, respectively.

Pass-through models measure the impacts from LCFS credit prices to wholesale fuel prices in California. Since wholesale fuels provided outside of California are free of LCFS obligations, we assume the price spreads between Los Angeles and other wholesale fuel markets are positive and they would respond to variations in LCFS credit prices. The estimations of pass-through rates are based on the following assumptions: (1) RFS obligations have the same price impacts on wholesale fuels provided across the nation. (2) Wholesale price spreads remove similar time-constant fixed effects on individual wholesale price. The spreads are mainly affected by variations in LCFS credit prices[3]. (3) Wholesale fuels provided in Los Angeles were more expensive than

[3] RFS has a pull on the availability of renewable fuels to California. As alternative fuels in California enjoy subsidies from both RFS and LCFS, they have incentives to expand their supplies and produce more of the LCFS credits, which would affect the price of renewable fuels and the LCFS credit prices. However, the impacts from the

other bulk markets before the implementation of LCFS, which are mainly caused by local supply shocks and the more stringent fuel quality required for fossil fuels in California by CARB. We assume neither of them directly relate to price fluctuations in the LCFS credit prices.

Based on the above assumptions, fuel spreads are the price differences between fuels supplied in LA and three wholesale fuel markets outside of California: Los Angeles (LA), Chicago (CHI), and Gulf Coast (GC) as presented in the equations below:

Wholesale gasoline price spreads:

- LA–NY RBOB spread = LA RBOB − NY RBOB
- LA–CHI RBOB spread = LA RBOB − CHI RBOB
- LA–GC RBOB spread = LA RBOB − GC RBOB

Wholesale diesel price spreads:

- LA–NY diesel spread = LA ULSD − NY ULSD
- LA–CHI diesel spread = LA ULSD − CHI ULSD
- LA–GC diesel spread = LA ULSD − GC ULSD

2.3.2 Descriptive statistics and time series plots

Table 2.2 provides summary statistics for six wholesale fuel spreads and their corresponding LCFS credit prices over the full sample period. Figure 2.4 plots the LCFS credit prices for both gasoline pool and diesel pool. Figure 2.5 and Figure 2.6 plot wholesale gasoline spreads and diesel spreads with the corresponding LCFS credit prices.

RINs are limited, and this is because the credit-generating systems of two policies have different emphases. LCFS is a fuel-neutral program, and it is designed as a performance standard that provides LCFS credit benefits to alternative fuels based on their CI ratings. RFS promotes the blending of cleaner fuels by requesting specific volumetric targets for different categories of biofuels. However, this leaves an open question for the future development of this paper.

Figure 2.4 presents the development of two LCFS credit prices from 2014 to 2020, including LCFS credit prices before and after the readoption of program in the beginning of 2016. Before 2016, LCFS credit prices are close to zero and are lack of volatilities. After 2016, when LCFS with more rigorous standards was readopted, prices started to increase by each year. We observe a few structural breaks in credit prices. They usually occur at the beginning of each year from 2016 – 2020. They are likely caused by the transaction mechanism of LCFS credits, which requires obligated parties to offset their deficits by the end of year, and each year the CI standards for gasoline pool and diesel pool continue to decline causing greater CI gap between the CI of traditional petroleum fuels and the assigned standard CI. The jumps in the beginning of each year reflect demand shocks for LCFS credits as traditional fuels providers expect the increasing demand in the following years and they have the incentives to purchase more credits that can meet their program mandates. Figure 2.4 shows two credit prices have similar volatility, and from Table 2.2 it presents two credit prices have similar standard deviations of $0.06/gallon. And yet LCFS credit prices for gasoline pool are slightly greater than that of diesel pool, reflecting the fact that in California credit demand in gasoline pool are higher than the credit demand in diesel pool.

Table 2.2 shows that, on average, all spreads are positive. For three wholesale gasoline spreads, they are over 10 cents/gallon; and for three wholesale diesel spreads, they are smaller than 10 cents/gallon. Positive fuel price spreads suggest during the full sample period, wholesale fuel gasoline and diesel are more expensive in Los Angeles market than that in other bulk markets. Figure 2.5 and Figure 2.6 show the development of three gasoline spreads and three diesel spreads from 2014 – 2020. We can observe that before 2016, spreads fluctuated around zero, and since early 2016, spreads have been positive and showed the tendency of moving along with LCFS credit prices. We also observe the oscillation patterns in spreads. As spreads remove time-constant fixed

effects, the remaining components are time-varying fixed effects on fuel prices. The patterns of spreads suggest daily seasonality, so before we run pass-through analysis, we need to specify seasonal components in the regressions.

2.4 Time Series Analysis

2.4.1 Empirical strategy for pass-through estimations

The pass-through model reflects the impacts from LCFS credit prices to wholesale fuel prices in California. For example, if wholesale gasoline and diesel markets are perfectly competitive and we relax the assumption on the elasticities of demand and supply, we would assume the estimated pass-through rate is between 0 and 1. When pass-through rate is 1 it indicates the wholesale fuels markets are perfectly competitive and it further implicates either the elasticity of supply is perfectly elastic, or the elasticity of demand is perfectly inelastic. It might seem plausible to directly apply the price level of two series to estimate the pass-through impacts on wholesale fuel prices from LCFS credit prices, but this would only work theoretically when two series have similar levels of price variations. In practice, factors like supply and demand shocks and transportation costs would magnify fluctuations in fuel prices.

Having wholesale fuel prices and LCFS credit prices, we follow the approach proposed by Knittel, Meiselman, and Stock (2017) to remove time-constant fixed effects and aviod omitted variable bias. Spreads are the price difference between wholesale fuels supplied in Los Angeles and other three bulk markets, where fuels in Los Angeles are considered as the treatment group and fuels in other three markets are considered as the control group.

For the long-run equilibrium of pass-through model, the LCFS credit prices and price spreads of wholesale fuels are defined:

$$Spread_{i,t}^{j} = \alpha_i^j + \beta_i^j LCFS_t^j + u_{i,t}^j, \tag{2.3}$$

where $Spread_{i,t}^{j}$ refers the price differences between Los Angeles and other three markets (i) of fuel j (diesel or gasoline) on day t, $u_{i,t}^{j}$ refers to other factors that contribute to the price variations in $Spread_{i,t}^{j}$, β_i^j is the pass-through coefficient for $Spread_{i,t}^{j}$ which captures the impact from LCFS credit prices to wholesale fuel prices in Los Angeles. For example, suppose the estimated pass-through rate is 1, when we have \$1/gallon increase in the LCFS credit prices, $Spread_{i,t}^{j}$ would increase by \$1/gallon. Because fuels sales in other states are free of LCFS credit obligations, ceteris paribus, we would further expect fuel price in LA increase by \$1/gallon. Therefore, our pass-through model is equivalent to directly estimating the relationship between fuel prices in LA and the LCFS credit prices, but with a more comprehensive control over time-constant fixed effects on individual wholesale fuel price.

This approach fits the pass-through relationship well for the following reasons. First, fuel spreads fully capture the magnitude the LCFS obligations. At the same time, because both wholesale gasoline and diesel are under the national RINs obligations, the spreads can remove price fluctuations caused by RFS. Second, spreads take away some common factors that fluctuate fuel prices in different markets, for example, demand and supply shocks, crude oil prices, crack spread, and transportation costs. Third, because fuels provided in different markets are physically comparable, this strategy reassures that the spreads do not contain varying input costs from fuels' different chemical components.

2.4.2 Validation of pass-through model

To validate the framework of estimating the impact from LCFS credit prices to wholesale fuel price in Los Angeles, we conduct the placebo analysis. We build control group and fake treatment group that consist of price spreads between wholesale fuels provided in Chicago, New York, and

Gulf Coast. As fuels transacted in these places are not affected by LCFS, therefore, these price spreads should not respond to the variations of LCFS credit prices.

The placebo analysis has the following steps: (1) build pairwise price spreads from three wholesale markets outside of California for gasoline and diesel. For each fuel, there are three combinations of two markets without repetition. (2) As the pass-through model does not distinguish different wholesale markets, the analysis is built on pooling observations over: three gasoline spreads, three diesel spreads, and six gasoline and diesel spreads. (3) We run OLS regressions where dependent variable is price spread and independent variables are LCFS credit prices and seasonal components. (4) The impact from LCFS credit to price spreads is estimated using Newey-West heteroskedasticity-and autocorrelation-consistent (HAC) standard errors with 30 lags.

Placebo analysis results are summarized in Table 2.3. For pooled wholesale gasoline spreads, pooled diesel spreads, and pooled spreads of two fuels, the impact from LCFS credit prices on price spreads between markets outside of California are not significantly different from zero. To account for time-varying seasonal components in price spreads, we first estimate seasonal components using full sample period from 2016 – early 2020. We also estimate seasonality of spreads using sample period from 2013 – 2015 where LCFS was not fully adopted in California. This approach removes any potential impacts from LCFS credit prices on the normal seasonal pattern of spreads. Results of placebo analysis suggest that fuels provided outside of California are not affected by LCFS program. And they further validate the design and results of pass-through models by using fuel price spreads between Los Angeles and other wholesale markets and LCFS credit prices.

2.4.3 Identification problem

One factor that might cause biased pass-through estimation is other clean energy programs promoted in California. One example is the Cap-and-Trade program that requires regulated parties to turn in allowances for their emissions. The allowances are obtained quarterly through auctions organized by CARB. The collected funds are deposited at CARB and will later be used to support the development of cleaner transportation. The Cap-and-Trade allowances and LCFS credits increase obligated parties' marginal costs, but they are distinctively different in terms of the level and volatility of credit prices as shown in Figure 2.2. From 2016 to 2018, the LCFS credit prices increased from $80/MT to over $180/MT with relatively high price variations, while the Cap-and-Trade allowances stayed at around $14/MT. The behaviors of two price series indicate that the LCFS credit prices have downward impacts on the Cap-and-Trade allowances, as LCFS has more stringent GHG emissions targets than Cap-and-Trade. When a regulated party meet its LCFS compliance, it can automatically satisfy the compliance of Cap-and-Trade. Therefore, due to the low-price levels and lack of variations of Cap-and-Trade allowances, we assume it does not cause any additional marginal input costs to fuel providers in California.

In energy markets, there are some other factors that also impact fuel prices. For these factors that cannot be removed by taking price differences of petroleum fuels, they usually have time-varying characteristics, for example, changing demand, supply, and income. The application of dynamic OLS that includes leads and lags of the first difference of LCFS credit prices can remove the endogeneity between independent variable and omitted variables in error term, especially for the pass-through regression where we have a I(0)/I(1) system between spreads and LCFS credit prices (Stock and Watson, 1993).

2.4.4 Seasonality specifications

The movement of wholesale fuel spreads suggest their variations are perceived as the combination of a deterministic seasonal component caused by market fundamentals and the response to LCFS credit prices. Figures 2.4 and 2.5 show that the seasonal patterns are available in all spreads. The inclusion of seasonal components in pass-through regressions avoid omitted variable bias.

Our study consists of daily prices of the LCFS credit and wholesale fuels spreads. To smooth out the daily variations and avoid the price jumps from week to week, we set up a seasonality model that has a parameterized linear combination of trigonometric functions with daily seasonal frequencies (Hannan, Terrell, and Tuckwell, 1970; Knittel, Meiselman, and Stock, 2017). The daily indicator variables use four sine and cosine functions applying on calendar days to reflect weekly peak and trough without breaking the continuity of daily prices, as shown in equation (1.4)[4]. Including eight harmonic frequencies in our seasonality model fits well for daily patterns, as they capture variation periods from annually (k=1) up to quarterly (when k=4).

$$Spread_{i,t}^{j} = \mu_i + \sum_{k=1}^{4} \gamma_{i,c,k}^{j} \cos\left(\frac{2\pi tk}{366}\right) + \sum_{k=1}^{4} \gamma_{i,s,k}^{j} \sin\left(\frac{2\pi tk}{366}\right) + v_{i,t}^{j},^{5} \quad (2.4)$$

To validate the specification of seasonality and to reduce any potential impacts from LCFS credit prices to the normal seasonal pattern of spreads, we apply data before 2016 to test if all the seasonal coefficients are significant. Before 2016, as LCFS credit prices are not binding, and

[4] Other specifications of this model are applied for robustness check, for example, we extend to including 10 and 12 indicator variables. Results are similar, and they have the same level of statistical significance. Therefore, having four sine and four cosine seasonality model would be sufficient and parsimonious for each spread. Including eight harmonic frequencies in our seasonality model fits well for daily patterns, as they capture variation periods from annually (k=1) up to quarterly (when k=4).

[5] We also provide an alternative approximation for the seasonality model. As there is no observation for non-business days, there are only about 260 daily prices for each year. In this model, we index time only on business days and adjust seasonal frequencies to 260 as a robustness check for our base seasonality model. This model is specified as:

$$Spread_{i,t}^{j} = \mu_i + \sum_{k=1}^{4} \gamma_{i,c,k}^{j} \cos\left(\frac{2\pi tk}{260}\right) + \sum_{k=1}^{4} \gamma_{i,s,k}^{j} \sin\left(\frac{2\pi tk}{260}\right) + v_{i,t}^{j}$$

F-tests conclude that all the spreads have seasonal factors at 1% significance level.

credits transactions were relatively small, we assume there is no impacts from LCFS credit prices to spreads normal seasonality. We estimate coefficients in Equation (2.4) with observations from 2013–2015. We conduct an F-test on all seasonality coefficients. The test statistics and R^2 of our seasonal model are summarized in Table 2.2. F-tests conclude that all the spreads have seasonal factors at 1% significance level. R^2 indicates that for each spread, daily seasonal indicators explain a considerable portion of variations in spreads.

2.4.5 Time series properties

Before we run time series regression models on fuel spreads and the LCFS credit prices, the first thing we check is each series order of integration. When two series have different degrees of integration, their relationship tends to be spurious, and the least square estimates could be inconsistent. Figure 2.5 and Figure 2.6 present the price movement of six spreads and two LCFS credit prices. We use Augmented Dickey Fuller (ADF) test and Dicky-Fuller Generalized Least Squares (DF-GLS) test for unit root test. Test results are summarized in Table 2.5 under ADF and DF-GLS columns. ADF test statistics suggest that at 1% significance level, all the six spreads are integrated of order zero, i.e., being stationary, and two LCFS credit price series fail to reject the null hypothesis of being nonstationary. DF-GLS tests show four out of six spreads are stationary and two credit prices are nonstationary. Altogether, for spreads 10 out of 12 cases suggest they are stationary; and for LCFS credit prices, none of unit root tests suggest they are stationary.

Without accounting for structural breaks, unit root test tends to provide biased conclusions about whether a series is stationary. During full sample period, there are a few structural breaks in both LCFS credit prices. They usually appear at the beginning of each year, and throughout the year, credit prices show a mean-reverting pattern. One reason behind is that LCFS tightens the standard CI for regulated fuels year by year, and the program requires obligated parties to offset

the deficits by each year, which cause demand shocks in the market that appear in the beginning of each year. However, as credits can be stored in the bank and would not expire, and LCFS incorporates more renewable energy fuels that build larger credit supply, the lagging effect from supply side slowly takes away the demand shocks. We find in Figure 2.5 and Figure 2.6, at the beginning of each year, there is no clear evidence to show structural breaks in spreads. Therefore, we apply the dummy variables to control for price jumps in these periods and the model is defined:

$$Spread_{i,t}^{j} = \alpha_i^j + \beta_i^j LCFS_t^j + u_{i,t}^j + \alpha_{i,1}^j D_{2017} + \alpha_{i,2}^j D_{2018} + \alpha_{i,3}^j D_{2019} + \alpha_{i,4}^j D_{2020}$$
$$+\beta_{i,1}^j D_{2017} \times LCFS_t^j + \beta_{i,2}^j D_{2018} \times LCFS_t^j + \beta_{i,3}^j D_{2019} \times LCFS_t^j + \beta_{i,4}^j D_{2030} \times LCFS_t^j \quad (2.5)$$

Table 2.4 presents estimated coefficients before the slope dummy variables. We assume the relationship between LCFS credit prices and spreads should be constant, except for a few days where the relationship is changed due to demand shocks. Test results suggest that all the slope dummy variables are not significantly different from zero, indicating they cannot change the overall impact from LCFS credit prices to spreads.

2.4.6 Long-run pass-through estimations and results

With the assumption that the LCFS credit prices are exogenous to spreads, we specify the pass-through model for two wholesale fuels as below:

$$Spread_{i,t}^{gas} = \alpha_i^{gas} + \beta_i^{gas} LCFS_t^{gas} + w_{i,t}^{gas} + u_{i,t}^{gas}, \quad (2.6)$$

and,

$$Spread_{j,t}^{diesel} = \alpha_j^{diesel} + \beta_j^{diesel} LCFS_t^{diesel} + w_{j,t}^{diesel} + u_{j,t}^{diesel}, \quad (2.7)$$

where β_i^{gas} and β_j^{diesel} are two pass-through coefficients, $w_{i,t}$ and $w_{j,t}$ refer to the 8 seasonal indicator variables as specified in equation (2.4), and we apply Newey-West HAC standard errors with 10 lags.

We apply long-run pass-through models over the full sample period from 2016 to the early 2020 for the following reasons. (1) Before 2016 LCFS was not fully developed. Although LCFS credits were available for transactions between obligated parties, the number of transactions were limited, and the general credit price was relatively low. Figure 2.4 shows during this period from 2014 – 2015, LCFS credit prices are lack of volatilities, therefore variations in two credit prices were too small to be reflected in the corresponding spreads. (2) The updated LCFS policy was implemented at the beginning of the 2016. Since then, the credit prices became relatively stable and reflected a relatively stable demand and supply for the credits.

Knittel, Meiselman, and Stock (2017) pointed out that the pass-through theory does not distinguish between the spreads derived from different wholesale markets, therefore we can assume the same pass-through rates for different spreads. Moreover, as individual pass-through estimate has large standard error, pooling price spreads improves the precision of estimates, and it averages out the idiosyncratic time-constant fixed effects for price spreads.

Table 2.5 summarizes the estimated long-run pass-through coefficients for six individual spread and the pooled regressions for two groups: three gasoline spreads and three diesel spreads with different seasonality specifications. For wholesale gasoline, the estimated pass-through coefficients for individual spread and the pooling of three spreads are not significantly different from 1, indicating wholesale gasoline providers in LA have complete pass-through and downstream buyers fully undertake the LCFS credit costs. For wholesale diesel, three individual spreads and the pooling of three spreads show pass-through rate is significantly different from one and is about 0.5. This indicates fuel providers can only pass down half of their LCFS credit costs. Similar results are provided by the dynamic OLS that include five leads and lags of the first differences of LCFS credit prices as additional regressors. Dynamic OLS corrects for endogeneity

in the regression and provides efficient estimator when variables in equation (2.6) and equation (2.7) have unit roots and are cointegrated.

The findings of complete pass-through in California wholesale gasoline market suggests that over the long-run, downstream buyers of wholesale fuels undertake the full LCFS credit costs. As traditional fuels are more expensive under LCFS, it further shows the effectiveness of the program in terms of discoursing the consumption of less-clean fuels. However, in wholesale diesel market, it shows an incomplete pass-through of LCFS credit prices to wholesale fuel prices, which indicates that fuel providers can only pass down half of LCFS credit costs. The incomplete pass-through implies LCFS is not as effective in wholesale diesel market as it is in wholesale gasoline market, as the extra costs from LCFS are not fully passed down to downstream buyers, and the incentives of increasing the consumption of cleaner fuels are discounted. This can be explained by the underlying elasticities of demand and supply, where both demand and supply should be elastic. In California, gasoline and diesel are the largest and the second largest used fuels in the state. It is reasonable to assume the demand elasticity of wholesale diesel is more elastic than gasoline. Moreover, as the total consumption of wholesale diesel is smaller than gasoline, diesel suppliers does not have the flexibility of passing LCFS credit costs to their buyers over the sample period. Supporting evidence is presented in Figure 2.4 where LCFS credit price for diesel pool is lower than gasoline pool, which implies the demand for credits is lower and the demand for wholesale diesel fuels is lower.

2.4.7 Short-run pass-through estimations and results

In this section, we discuss the pass-through dynamics between the LCFS credit prices and wholesale fuel price spreads by applying the structural vector autoregressions (SVARs) model.

The SVAR model estimates spreads' dynamic responses to unexpected shocks in the LCFS credit prices with a mathematical specification below:

$$Y_t = \Pi_0 + \sum_{k=1}^{p} \Phi_k Y_{t-k} + \Gamma\Psi_t + u_t, \quad (2.8)$$

where $Y_t' = (LCFS_t^j, Spread_t^{j,i})$. $LCFS_t^j$ are credit prices for diesel or gasoline on day t. $Spread_t^{j,i}$ refers to the individual wholesale spread on day *t*. Ψ_t includes eight seasonal indicator variables for each spread. u_t is the VAR innovation. The impulse response function of the short-run model provides dynamic pass-through estimates from the LCFS credit price to wholesale fuels, especially measures how wholesale fuels in California respond to an unexpected shock in LCFS credit prices. We assume LCFS credit prices are exogenous and uncorrelated to any structural shocks of fuel price spreads. Therefore, in the SVAR specification, the unexpected credit price shock is placed as the first element of u_t, and the dynamic response of Y_{2t} to u_{1t} estimates the response of the spread to the credit price shocks. The choice of lags $k = 2$ in the autoregression is determined by the Akaike information criterion (AIC). AIC shows that five out of six spreads can be specified with 2 lags in the autoregression. The bivariate SVAR short-run pass-through dynamics over the full sample period are summarized in Table 2.6.

We also apply the pooled SVAR model to estimate the dynamics of short-run pass-through. Theoretically, pass-through theory does not distinguish different markets and the response of individual spread to an unexpected shock in the same LCFS credit price should not be different from each other (Knittel, Meiselman, and Stock, 2017). Pooling three spreads for wholesale gasoline and wholesale diesel can also average out the idiosyncratic price variations in individual spread. For the pooled model, we extend the response vector to $Y_t' = (LCFS_t^j, Spread_t^{i,1}, Spread_t^{i,2}, Spread_t^{i,3})$. The restriction lies upon the matrix Φ_k. For example, the equation for $Spread_t^{i,1}$ should have the following restrictions: (1) coefficient before other

spreads (i.e. $Spread_{t-k}^{i,2}$ and $Spread_{t-k}^{i,3}$) should all be zero; (2) coefficient before $LCFS_{t-k}^{j}$ are the same for all spread equations; (3) own-lag coefficient should be the same for all spread equations. Estimation results over the full sample period are summarized in Table 2.7.

The bivariate short-run pass-through estimates suggest the following conclusions. First, the estimates for gasoline show a pattern that after 15 days, all spreads show the evidence of full pass-through, and for the first few days, cumulative pass-through rates have large standard errors. Second, for diesel, three spreads have very different pass-through patterns. For the LA–NY spread, over the first 15 days, LCFS has negative pass-through rate. For the LA–CHI spread, pass-through rates are all below 1, however, the t-test shows that they are not significantly different from 1, suggesting the complete cumulative short-run pass-through. For the LA–GC spread, after 13 days, the cumulative pass-through rates become positive, and yet they are significantly smaller than 1, indicating the incomplete dynamic pass-through. The evidence in wholesale diesel markets is not consistent for all three spreads. This is because individual spread has unique behavior over the full sample period, and the volatility of LCFS credit prices for diesel is lower than gasoline. Therefore, the measure of how spreads respond to the shocks in LCFS credit prices over a short period of time, i.e., we use 15 days to illustrate the reaction, are likely to be less accurate.

To avoid the inconsistent pass-through estimates from individual spread with the bivariate SVAR model, we also report the dynamic pass-through rates from the pooled SVAR model that increases the accuracy of estimators. For wholesale gasoline, we observe that after 4 days, fuel suppliers can achieve complete pass-through. The estimated pass-through rates become more precise as they have lower standard errors till after 15 days, the estimated pass-through rate is about 1.22 with the lowest standard error of 0.39. For wholesale diesel, the pass-through rates in the early days are all negative and they become positive after 6 days when there is an unexpected

shock in LCFS credit prices. After 15 days, the short-run pass-through rate is about 0.64, which is significantly smaller than 1, suggesting incomplete pass-through. The findings in short-run pass-through present a similar pattern that for gasoline, it has full pass-through, whereas for diesel, it has incomplete pass-through. The contrasting findings are caused by more elastic demand for wholesale diesel and the lower credit prices for diesel pool. Therefore, the obligated parties of wholesale diesel do not have the ability and incentives to fully pass down LCFS credit costs to their buyers.

2.5 Conclusions

In conclusion, our study estimates the long-run pass-through equilibrium and the short-run pass-through dynamics from LCFS credit prices to California wholesale gasoline and diesel prices. Our discussions focus on wholesale petroleum prices and LCFS credit prices from 2016 to the early 2020, during which LCFS was readopted, and the variations of credit prices are large enough for the estimations of pass-through. For wholesale gasoline, pooled long-run estimated coefficient is 1.06 with a Newey-West HAC standard error of 0.2, suggesting in wholesale gasoline markets, fuel providers and fully recoup their LCFS obligations by passing the extra credit costs to their downstream buyers. For wholesale diesel, pooled estimated coefficient is about 0.5, with a Newey-West HAC standard error of 0.1, showing that suppliers of wholesale diesel can only pass down half of the credit costs to their downstream buyers. For short-run pass-through, in wholesale gasoline markets, obligated parties have complete pass-through after 4 business days, but in wholesale diesel markets, in the first few days, pass-through rates are negative and after 15 days, they can only recoup 60% of the LCFS credit costs.

The evidence of the complete LCFS pass-through rates in wholesale gasoline market for both long-run and short-run shows eventually customers pay for the LCFS credit prices and shows

the effectiveness of the program by discouraging the consumption of traditional fuels. The evidence of incomplete pass-through in wholesale diesel for both long-run and short-run shows obligated parties can only pass down a portion of LCFS credit costs to their buyers. The more elastic demand for diesel, the lower credit prices, and the lack of volatility for diesel pool make obligated parties do not have the capacity and incentives to pass down LCFS credit costs.

Some other factors can also explain the partial pass-through rates. For example, LCFS has the back-loaded credit system: deficits and credits are matched and cleared up by the end of each fiscal year. Due to this mechanism, daily variations in prices cannot accurately reflect the fundamentals of the fuel markets. For large refineries, they usually provide fuels with the lowest possible prices. Because of the high productivity, these refineries can price-match their external competitors and recover themselves from the reduced markups, i.e., price difference between the selling price and the cost. For fuels providers, they face a situation where the change in their markups is greater than the additional LCFS credit costs. Even though they add the full LCFS credit costs to their prices, the prices would still be lower than their internal competitors. For small and more vulnerable refineries, they are less productive, so cannot reduce the markup and they have to add the full LCFS credit costs to the selling prices. Between the above two extreme cases, we have refineries that are in the middle. The prices we have for wholesale fuels are at an average level that neutralizes each fuel provider's pricing decisions. Therefore, it is reasonable to observe estimated pass-through rates that are between 0 and 1. Future studies may wish to apply the pass-through rates and include key variables that affect fuel prices for the estimate of demand and supply elasticities. As spreads are highly volatile, it would be increasing to estimate pass-through at the volatility level.

2.6 Tables and Figures

Table 2.1: Current Carbon Intensity of wholesale gasoline and diesel and LCFS Carbon Intensity compliance schedule for gasoline pool and diesel pool

Year	CARBOB (gCO_2e/MJ)	Gasoline Standard (gCO_2e/MJ)	ULSD (gCO_2e/MJ)	Diesel Standard (gCO_2e/MJ)
2016	99.78	96.5	102.01	99.97
2017	99.78	95.02	102.01	98.44
2018	99.78	93.55	102.01	96.91
2019	100.82	93.23	100.45	94.17
2020	100.82	91.98	100.45	92.92

Notes:
(1) Gasoline Standard is applied to gasoline and fuels used as a substitute for gasoline.
(2) Diesel Standard is applied to diesel and fuels used as a substitute for diesel.
(3) In 2015, California readopted the LCFS with an updating CI modeling system. The average CI standard of each fuel for years from 2016 to 2018 reflects reductions from the revised base year (2010).
(4) Staring in 2016, LCFS set up a reduction of 2% in CI for both diesel and gasoline fuel pools each year from 2010 baseline levels, with the target of achieving 10 percent by 2020. From 2013-2015, the annual reduction mandates were 1% for both fuel pools.
(5) Since 2019, to smooth and strengthen the CI reduction through 2030, the amended regulation changes the annual reduction to 1.5%, from the 5% total reduction in 2018 to the 20% total reduction in 2030.
(6) The current CIs of two fuels are provided in Market Data Used in Price Assessments, OPIS's California LCFS Carbon Intensity Calculations. Available at https://www.opisnet.com/about/pdf/OPIS-RenewableFuels-RINCredits.pdf.
(7) The target CI of two fuels are available in Low Carbon Fuel Standard regulation. Final Regulation Or-der.Title17, California Code of Regulations. PP. 32-33.

Table 2.2: Summary statistics: daily fuel spreads and LCFS credit prices

	Summary Statistics and Unit Root Tests: Jan 4, 2016 - Mar 12, 2020						Seasonal Components Jan 4, 2013 - Dec 31, 2015	
	Mean	SD	Min	Max	DF-GLS	ADF	F-Stat	R2
Fuel spreads:								
LA-NY RBOB	0.147	0.184	-0.314	1.227	-1.560	-5.077^^^	11.400^^^	0.156
LA-CHI RBOB	0.104	0.223	-1.096	1.145	-1.900^	-5.563^^^	16.440^^^	0.211
LA-GC RBOB	0.168	0.186	-0.310	1.265	-1.608	-5.185^^^	10.560^^^	0.147
LA-NY ULSD	0.047	0.065	-0.108	0.382	-4.381^^^	-4.582^^^	35.420^^^	0.365
LA-CHI ULSD	0.071	0.090	-0.155	0.398	-4.484^^^	-4.631^^^	17.980^^^	0.226
LA-GC ULSD	0.092	0.066	-0.044	0.425	-4.455^^^	-4.517^^^	24.780^^^	0.287
LCFS Credit Prices:								
Gasoline Pool	0.104	0.062	0.022	0.230	-1.033	-2.593	-	-
Diesel Pool	0.086	0.060	0.015	0.221	-1.100	-2.192	-	-

Notes:
(1) We conduct a joint test on the seasonal variables in a regression of individual wholesale fuel spreads on a constant and the seasonal components using data from January 4, 2013 to December 31, 2015, before LCFS was readopted by California.
(2) We use F-statistic to test on the joint significance of the seasonal components with the null hypothesis that they have no power in explaining the variation in wholesale fuel spreads. R^2 of the seasonal components are presented in the last column.
(3) To test the null hypothesis that wholesale fuel spreads and LCFS credit prices have a unit root, we use the DF-GLS test and ADF test. For fuel spreads, both tests have specifications of intercept, no linear time trend, maximum lags of six, and lags chosen by modified AIC. For LCFS credit prices, two tests have the same specification as the one for spreads, except that we include linear time trend.
(4) ^^^, ^^, ^ indicates significance level at 1%, 5%, 10%, respectively.

Table 2.3: Placebo analysis of the long-run regression between price spreads outside of California and LCFS credit prices

	RBOB	ULSD
Full Sample	0.192	-0.154
(Seasonal 2016 - 2020)	(0.117)	(0.099)
Full sample	0.0834	-0.2365
(Seasonal 2013 - 2015)	(0.174)	(0.171)

Notes:
(1) Table presents t-test statistic and Newey-West standard error with 30 lags in the parenthesis.
(2) For the first set of tests, we use sample from 2016 – 2020 to estimate the seasonal components. For the second set of tests, we use sample from 2013 – 2015, when LCFS program was not officially introduced to the public, to estimate the seasonal components.
(3) The null hypothesis of test statistic is that the coefficient is equal to zero.

Table 2.4: Dummy variables for potential structural breaks in LCFS credit prices

	LA-NY RBOB	LA-CHI RBOB	LA-GC RBOB	LA-NY ULSD	LA-CHI ULSD	LA-GC ULSD
D_2017	49.796	60.054	2.499	-18.627	-24.649	-14.516
	(85.845)	(112.806)	(4.849)	(40.058)	(57.914)	(39.724)
D_2018	0.693	1.539	0.000	0.729	-0.606	-0.306
	(17.802)	(23.393)	(1.810)	(10.806)	(15.623)	(10.716)
D_2019	145.366	160.971	13.288	78.167	170.747	71.947
	(273.311)	(359.149)	(47.546)	(119.901)	(173.344)	(118.900)
D_2020	6.625	27.641	-1.420	-2.192	-1.680	5.241
	(50.399)	(66.227)	(11.184)	(20.733)	(29.974)	(20.560)

Notes:
(1) Table presents t-test statistic and Newey-West standard error (in parenthesis) with 10 lags.
(2) The null hypothesis of the t-test is that the coefficient is equal to zero.

Table 2.5: Level regressions and cointegration statistics

Regression Coefficients (SEs)	Individual Wholesale Spreads						Pooled Wholesale Spreads	
	LA-NY RBOB	LA-CHI RBOB	LA-GC RBOB	LA-NY ULSD	LA-CHI ULSD	LA-GC ULSD	RBOB	ULSD
(1) OLS, full sample, no seasonals	1.052	1.05	1.306	0.240***	0.697**	0.428***	1.136	0.455***
	(0.272)	(0.318)	(0.269)	(0.117)	(0.132)	(0.119)	(0.274)	(0.113)
(2) OLS, full sample, seasonals	1.03	0.895	1.256	0.305***	0.720**	0.486***	1.06	0.504***
	(0.214)	(0.248)	(0.212)	(0.098)	(0.124)	(0.098)	(0.214)	(0.098)
(3) OLS, full sample, augmented seasonals	1.031	0.894	1.254	0.302***	0.716**	0.483***	1.06	0.500***
	(0.215)	(0.248)	(0.211)	(0.093)	(0.115)	(0.093)	(0.213)	(0.091)
(4) DOLS, full sample, seasonals	1.112	0.965	1.339	0.310***	0.690**	0.484***	1.139	0.495***
	(0.218)	(0.255)	(0.215)	(0.102)	(0.127)	(0.103)	(0.216)	(0.101)
Engle-Granger ADF cointegration test	-5.593^^^	-5.938^^^	-5.957^^^	-5.066^^^	-6.065^^^	-5.344^^^		

Notes:
(1) Full sample are daily observations from January 4, 2016 to January 4, 2016 to March 12, 2020.
(2) Seasonally adjusted uses pre-2016 fuel spreads to estimate seasonal components and then subtract predicted seasonal components from the full sample spreads.
(3) Newey-West standard errors with 10 lags are provided in the parenthesis below estimated coefficients.
(4) The null hypothesis of level regressions is that estimated coefficients are zero, i.e., there is no pass-through from LCFS obligation costs to fuel prices.
(5) DOLS regressions include five leads and five lags of the first difference of LCFS credit prices.
(6) The null hypothesis of Engle-Granger ADF cointegration test is that the wholesale spreads and their corresponding LCFS credit prices have no cointegration.
(7) ***, **, * indicates significance level at 1%, 5%, 10%, respectively. The null hypothesis is that the coefficient is equal to 1.
(8) ^^^, ^^, ^ indicates significance level at 1%, 5%, 10%, respectively. The null hypothesis of cointegration test is that two series are not correlated.

Table 2.6: Bivariate VARs for individual wholesale spreads: cumulative structural impulse response functions

	LA-NY RBOB		LA-CHI RBOB		LA-GC RBOB		LA-NY ULSD		LA-CHI ULSD		LA-GC ULSD	
Lag	Coef.	(SE)	Coef.	(SE)	Coef.	(SE)	Coef.	(SE)	Coef.	(SE)	Coef.	(SE)
0	-0.207*	(0.719)	-0.833	(1.120)	-0.178	(0.744)	-0.130***	(0.181)	0.363**	(0.248)	-0.106***	(0.185)
1	-0.866*	(1.052)	-2.341**	(1.590)	-0.955*	(1.074)	-0.201***	(0.268)	0.325*	(0.372)	-0.188***	(0.273)
2	-1.048	(1.278)	-2.612*	(1.886)	-1.024	(1.304)	-0.476***	(0.333)	0.600	(0.470)	-0.473***	(0.336)
3	0.386	(1.460)	-0.830	(2.123)	-0.042	(1.497)	-0.546***	(0.382)	0.540	(0.540)	-0.501***	(0.385)
4	0.573	(1.414)	-0.569	(1.975)	0.110	(1.449)	-0.509***	(0.371)	0.561	(0.530)	-0.448***	(0.372)
5	0.684	(1.322)	-0.327	(1.782)	0.281	(1.365)	-0.452***	(0.351)	0.544	(0.502)	-0.380***	(0.350)
6	0.819	(1.217)	-0.008	(1.588)	0.478	(1.262)	-0.397***	(0.327)	0.552	(0.464)	-0.318***	(0.325)
7	0.879	(1.093)	0.173	(1.366)	0.598	(1.129)	-0.348***	(0.307)	0.555	(0.436)	-0.261***	(0.304)
8	0.913	(0.973)	0.311	(1.164)	0.696	(0.999)	-0.301***	(0.288)	0.559	(0.411)	-0.207***	(0.285)
9	0.938	(0.863)	0.433	(0.990)	0.781	(0.881)	-0.258***	(0.270)	0.563	(0.388)	-0.157***	(0.267)
10	0.956	(0.766)	0.526	(0.840)	0.850	(0.774)	-0.217***	(0.255)	0.566	(0.367)	-0.110***	(0.251)
11	0.970	(0.680)	0.601	(0.715)	0.908	(0.680)	-0.179***	(0.241)	0.569	(0.348)	-0.067***	(0.236)
12	0.981	(0.605)	0.662	(0.614)	0.957	(0.599)	-0.144***	(0.228)	0.572	(0.330)	-0.027***	(0.224)
13	0.990	(0.542)	0.712	(0.533)	0.999	(0.531)	-0.111***	(0.217)	0.575	(0.314)	0.010***	(0.212)
14	0.997	(0.488)	0.753	(0.470)	1.034	(0.474)	-0.080***	(0.206)	0.578	(0.300)	0.044***	(0.202)
15	1.004	(0.443)	0.787	(0.422)	1.064	(0.428)	-0.051***	(0.197)	0.580	(0.287)	0.076***	(0.193)

Notes:
(1) We use full sample period to apply the structural VAR analysis, from January 4, 2016 to March 12, 2020.
(2) Estimates are impose responses from the structural VAR model with standard errors in the parenthesis below.
(3) For structural VAR model, we include 2-day lags and control for seasonal components.
(4) ***, **, * indicates whether the estimated coefficients are significantly different from 1 at 1%, 5%, 10% level respectively.

Table 2.7: Pooled VARs for wholesale spreads: cumulative structural impulse response functions

	RBOB		ULSD	
Lag	Coefficient	(SE)	Coefficient	(SE)
0	-0.168	(0.715)	-0.080***	(0.181)
1	-0.856*	(0.988)	-0.134***	(0.255)
2	-0.936*	(1.165)	-0.364***	(0.311)
3	0.412	(1.295)	-0.406***	(0.351)
4	0.715	(1.394)	-0.223***	(0.382)
5	0.779	(1.281)	-0.170***	(0.352)
6	0.867	(1.169)	-0.119***	(0.323)
7	1.010	(1.060)	-0.089***	(0.294)
8	1.068	(0.956)	-0.065***	(0.272)
9	1.103	(0.837)	-0.038***	(0.255)
10	1.131	(0.734)	-0.012***	(0.240)
11	1.159	(0.643)	0.013***	(0.226)
12	1.180	(0.566)	0.035***	(0.213)
13	1.196	(0.498)	0.055***	(0.201)
14	1.210	(0.441)	0.074***	(0.190)
15	1.220	(0.394)	0.091***	(0.181)

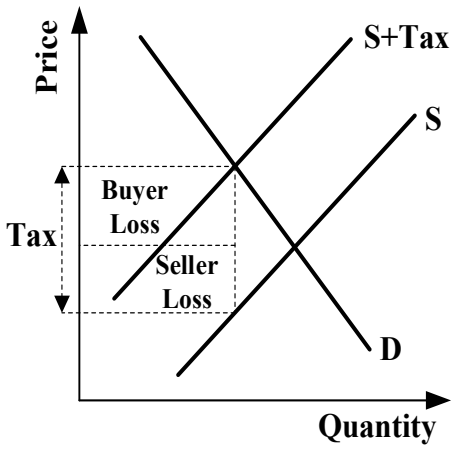
Note:

(1) We use full sample period to apply the structural VAR analysis, from January 4, 2016 to January 4, 2016 to March 12, 2020.

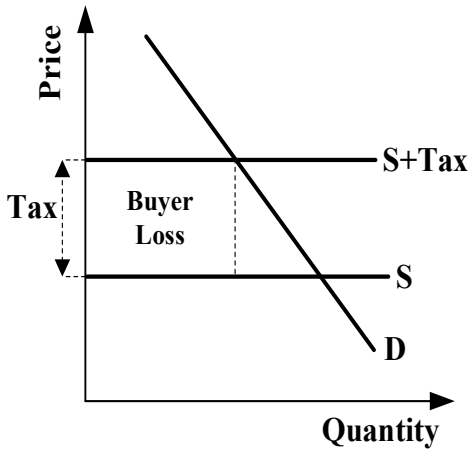
(2) Estimates are impulse responses from the structural VAR model with standard errors in the parenthesis below.

(3) In the structural VAR model, we include 2-day lags ($p = 2$) and seasonal components for spreads in the model specification.

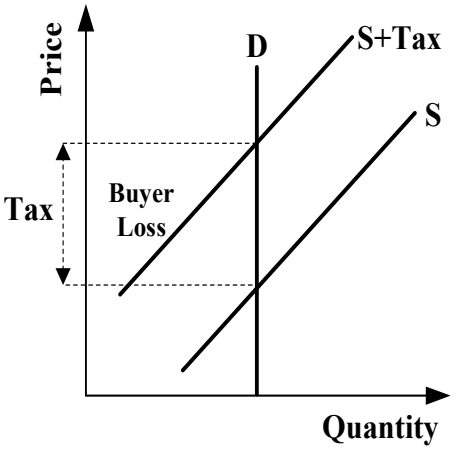
(4) ***, **, * indicates whether the estimated coefficients are significantly different from 1 at 1%, 5%, 10% level respectively.



(a) Elastic demand and supply

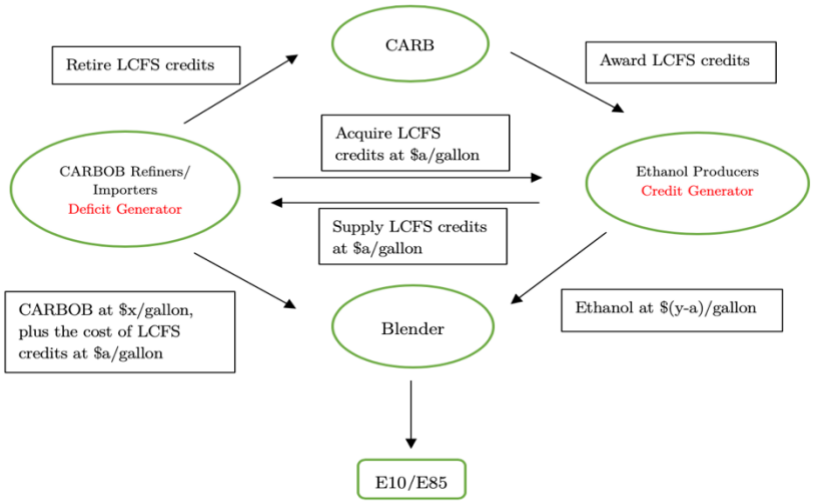


(b) Perfectly elastic supply

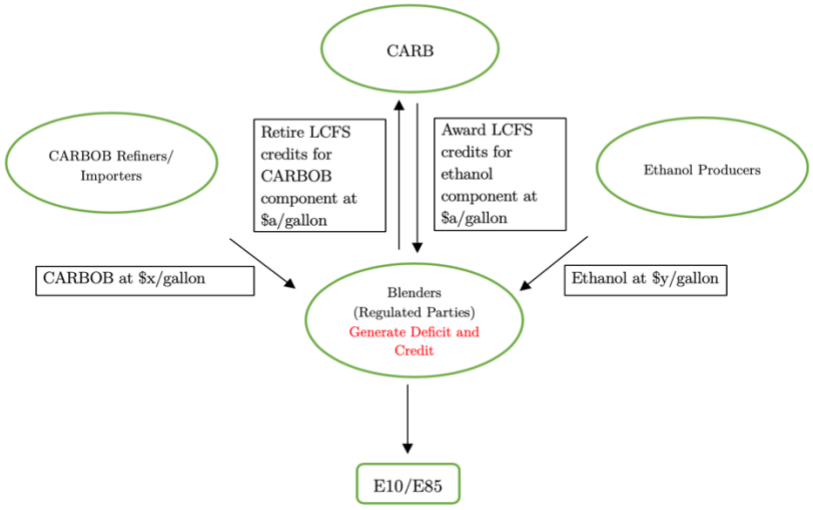


(c) Perfectly inelastic demand

Figure 2.1: Illustration of tax incidence under perfect competition



(a) Scenario 1: Fuel providers are regulated parties



(b) Scenario 2: Fuel providers transfer their obligations to blenders

Figure 2.2: Simplified LCFS credits and deficits generation and obligation transfer mechanisms on the gasoline supply chain (ethanol only)

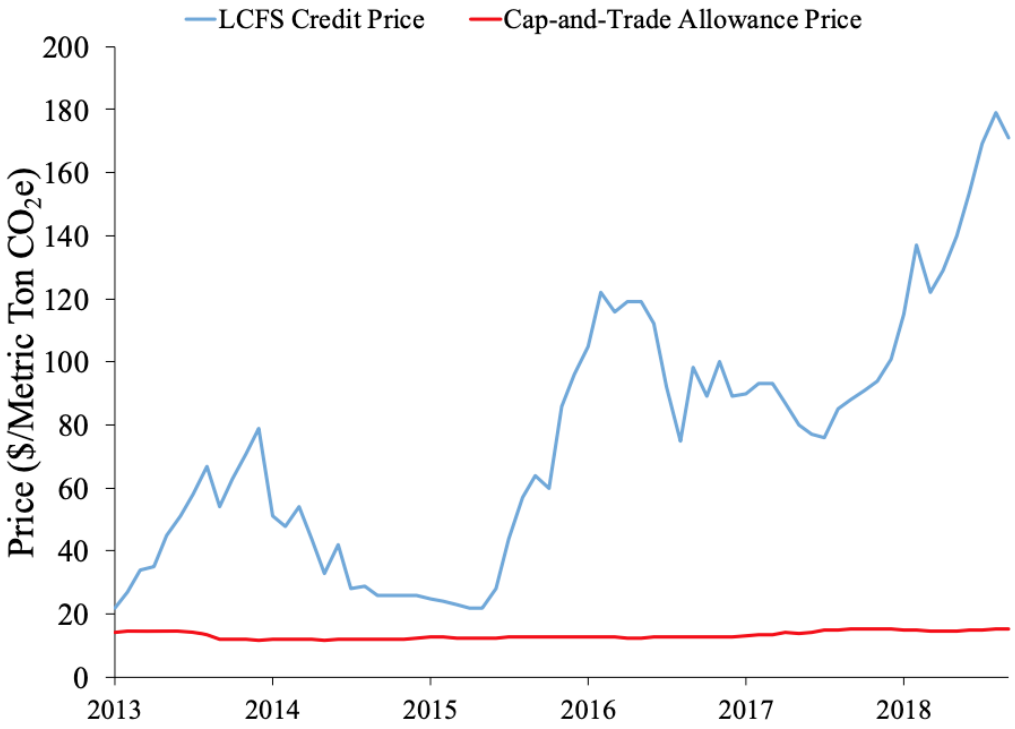


Figure 2.3: LCFS Credit prices and GHG Cap-and-Trade allowances prices from 2013 – 2019

Notes: CARB publishes the LCFS credit prices monthly, and CARB publishes Cap-and-Trade allowances quarterly. Source: Plot is derived from https://scholar.harvard.edu/_les/stavins/_les/dp92 schatzki-stavins.pdf

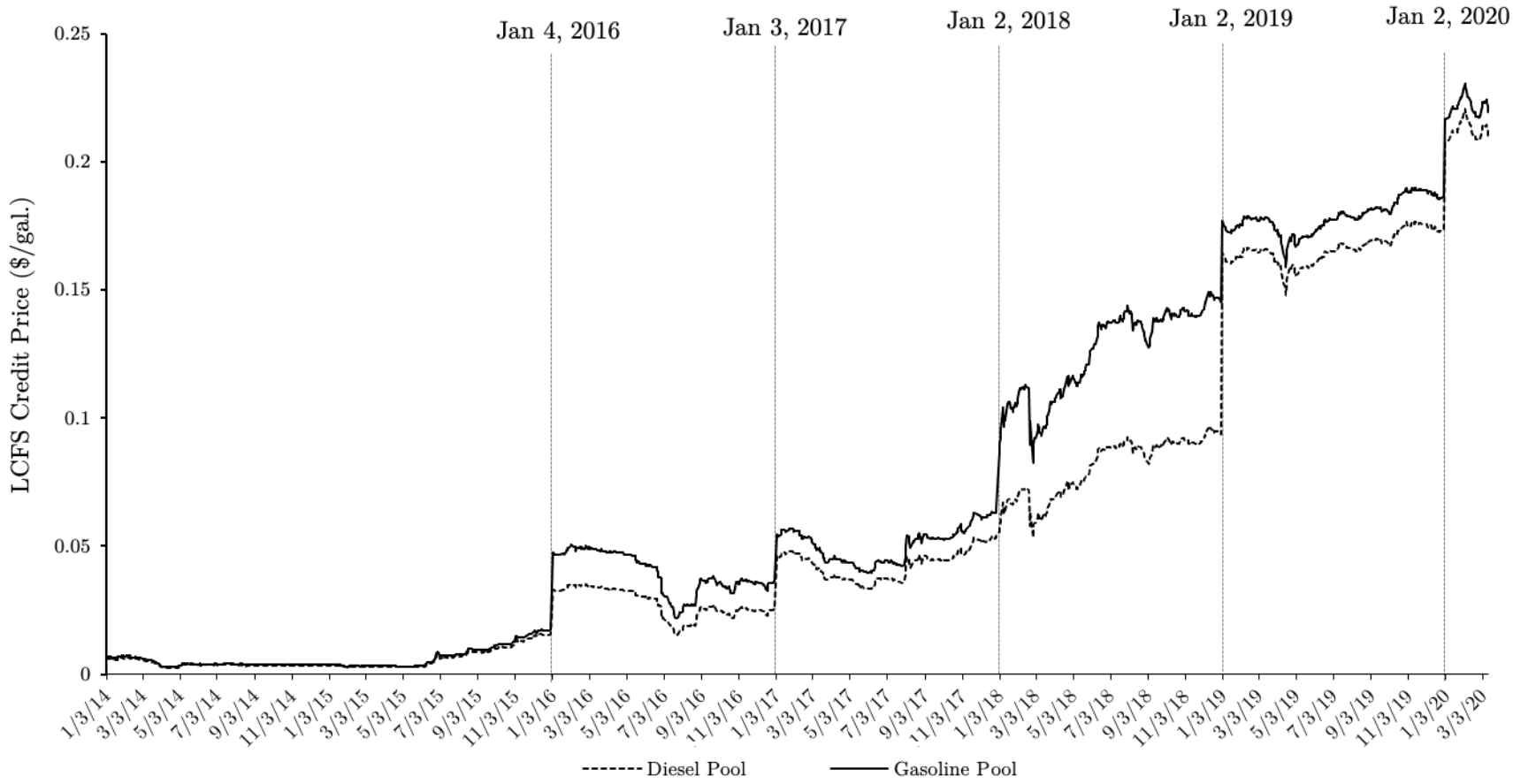
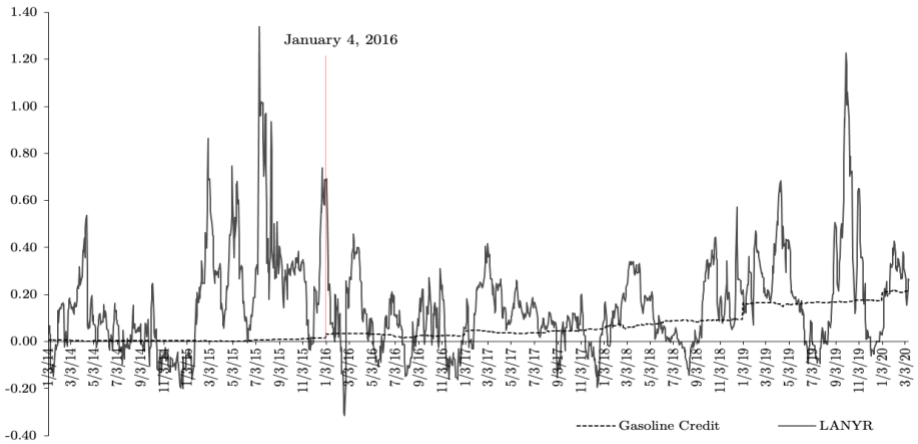
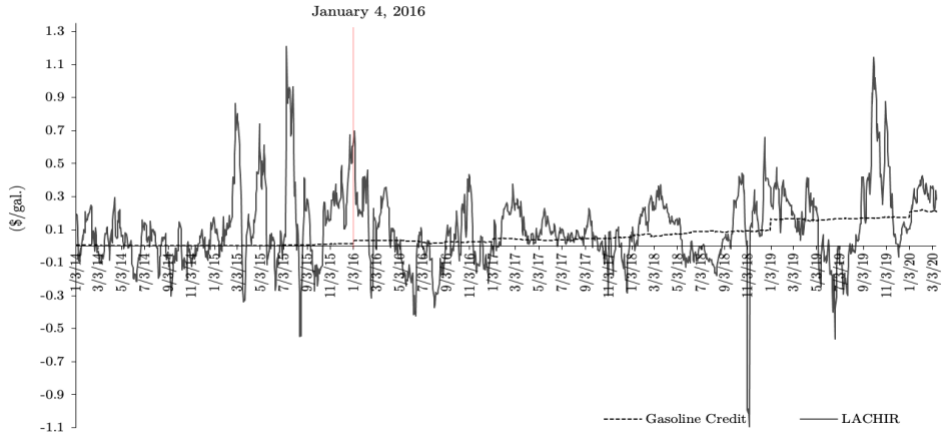


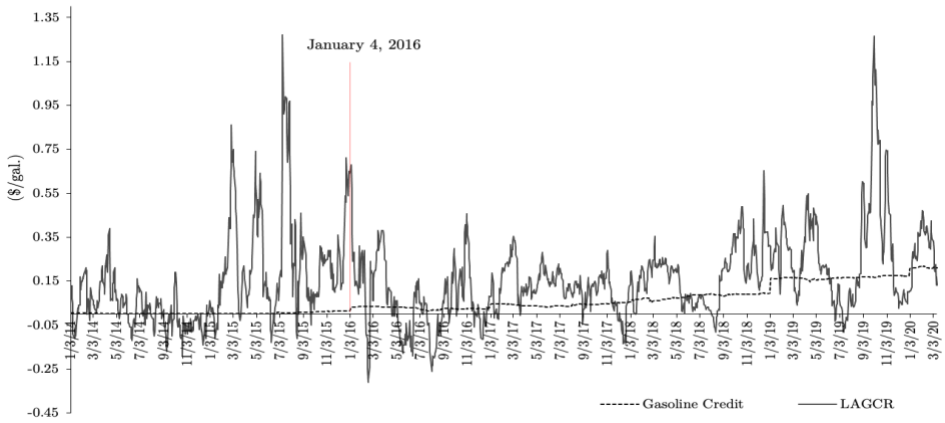
Figure 2.4: LCFS Credit prices for LCFS gasoline pool and diesel pool from 2014 – 2020



(a) LA–NY RBOB spreads and LCFS credit price for gasoline pool

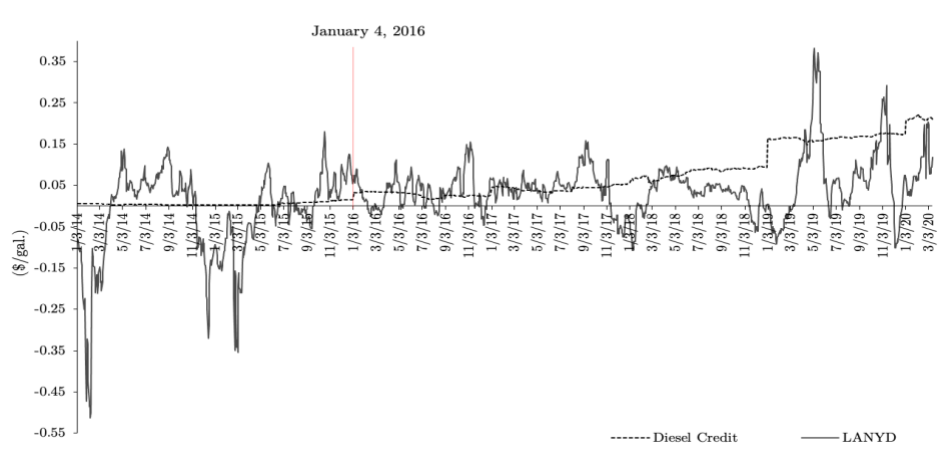


(b) LA–CHI RBOB spreads and LCFS credit price for gasoline pool

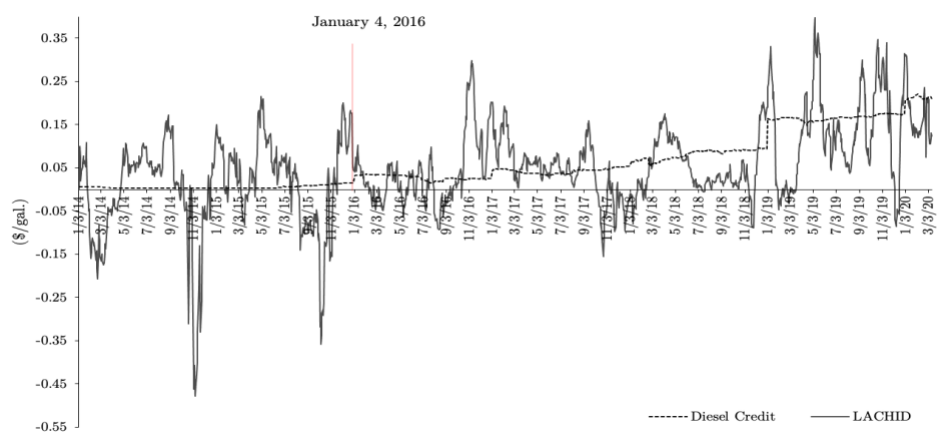


(c) LA–GC RBOB spreads and LCFS credit price for gasoline pool

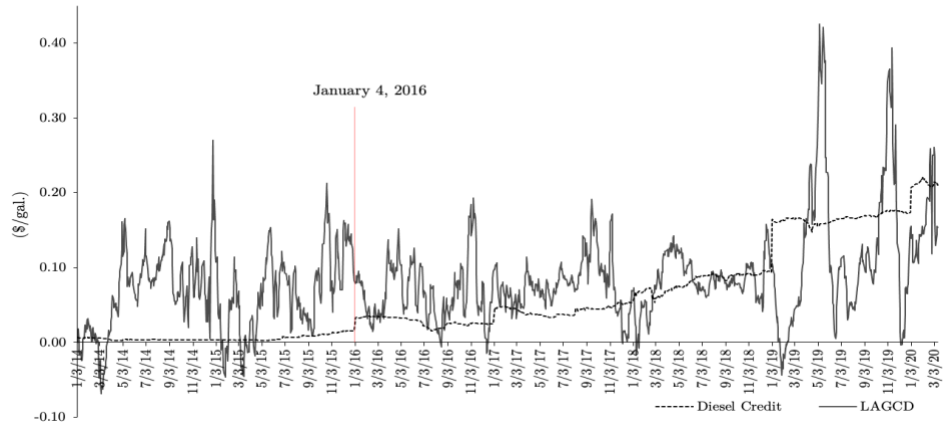
Figure 2.5: Wholesale gasoline fuel spreads and LCFS credit price for gasoline pool from 2014 – 2020



(a) LA–NY ULSD spreads and LCFS credit price for diesel pool



(b) LA–CHI ULSD spreads and LCFS credit price for diesel pool



(c) LA–GC ULSD spreads and LCFS credit price for diesel pool

Figure 2.6: Wholesale diesel fuel spreads and LCFS credit price for diesel pool from 2014 – 2020

CHAPTER 3:

DOES COMPLEXITY PAY? FORECASTING CORN AND SOYBEAN YIELDS USING CROP CONDITION RATINGS

3.1. Introduction

Accurate forecasts of crop yield are highly valuable from several perspectives. From a market perspective, yield forecasts are an essential component of supply, demand, and price forecasting. From a policy perspective, yield forecasts are important to governments around the world to assess drought impacts and food insecurity. In addition, these forecasts are crucial for farmers and agribusiness firms in developing marketing and risk management plans.

Given the importance of crop yield forecasts, it is no surprise that there is a very large literature on the relationship between weather, technology, and crop yields dating back to the early 1900s (Tannura, Irwin, and Good, 2008). Broadly speaking, this literature shows that summer precipitation and air temperature directly influence yield potential, along with other factors including soil quality, planting date, disease, insects, and technological improvements from seed genetics, fertilizers, and producer management techniques.

Kaufmann and Snell (1997) and others observe that the various methods of forecasting crop yields can be categorized into three groups. The first group consists of crop simulation models that directly assess the effects of weather and soil properties on plant physiology. While such models have a strong foundation in biological theory and experimental data, they are nonetheless highly complex and difficult to generalize to aggregate areas such as crop reporting districts or states (e.g., Walker 1989). The second group consists of multiple regression models that estimate the relationship of weather and technology to crop yields. Regression models are relatively simple to specify and estimate for aggregate areas, an advantage when forecasting, but aggregation of

variables across time and space can harm accuracy (Shaw, 1964). The third group consists of models based on remote-sensing data collected from earth-orbiting satellites. By far the most popular approach is to convert remote-sensing data into a vegetative index (NDVI) and correlate this to yield (e.g., Sakamoto, Gitelson, and Arkebauer, 2013). While there has been a great deal of work along these lines, the advantage in forecasting crop yields has not been proven convincingly to date.

There is still another approach to crop yield forecasting that is widely used by market analysts in both the private and public sectors. The U.S. Department of Agriculture (USDA) publishes weekly condition ratings for important crops during the growing season. The condition ratings reflect the subjective judgment of nearly 4,000 observers about crop yield prospects and are reported as the percentage of a crop rated in five mutually exclusive and exhaustive categories: very poor, poor, fair, good, and excellent. A popular approach is to use the sum of good and excellent condition ratings to build a simple condition index and relate this to trend-adjusted crop yields. Several representative articles applying this approach to forecasting U.S average corn and soybean yields can be found at the *farmdoc daily* website (Irwin and Good, 2017a,b; Irwin and Hubbs, 2018a,b,c,d).

Despite the widespread use of crop condition ratings to forecast crop yields in the private and public organizations, there are only a few studies in the academic literature that investigate condition-based forecasts. The general idea behind these studies is to transform the ordinal condition ratings to a numeric condition index, and then construct a time-series model between yields and the condition index. For example, Kruse and Smith (1994) developed a weighting system that estimates a changing yield weight for each crop condition class in the growing season for corn and soybean. By multiplying each crop condition ratings by its yield weights, they

computed an average in-sample yield estimate at state-level. Fackler and Norwood (1999) built a similar state-level yield forecasting model for corn, cotton, soybean, and spring wheat with estimated yield weight that is unchanging throughout the growing season for each crop condition level. They showed that for each condition class, the product of estimated yield weight and condition ratings reflects its average yields. Bain and Fortenbery (2017) used fixed weights to construct a condition index in a yield forecasting model for wheat. Their condition index is based on a straightforward system where for the lowest very poor condition is assigned a weight of 0, and as the condition increase by one level, the corresponding weight will increase by 0.25 until it reaches the highest excellent condition with a weight of 1.

Most recently, Begueria and Maneta (2020) developed a sophisticated two-stage yield forecasting model based on crop condition ratings for corn, cotton, soybean, and winter wheat at the state level. They argue that spatial and temporal differences in crop condition information should be directly modeled before making yield forecasts. Hence, the authors developed a cumulative link mixed model to transform raw condition data to a continuous and almost normal-distributed crop condition index. After removing space and time effects, they argue that maximum information can be extracted from crop condition ratings, which offers a better possibility of providing unbiased and accurate yield forecasts. Begueria and Maneta (BM) purport that their modeling approach achieves large improvements in accuracy over simpler condition-based forecasts, such as Irwin and Good (2017a,b).

The improvements in forecast accuracy reported by BM are interesting for two reasons. First, the finding that a complex model beats simpler models in terms of forecast accuracy runs counter to a large body of literature on forecasting. Armstrong (2001, p. 693) summarizes the evidence as "…showing that while some complexity may improve accuracy, seldom does one need

highly complex methods. In some studies, complexity harmed accuracy." The results provided by BM may represent an important exception to this general result. Second, the forecast results in Begueria and Maneta model (BM model, hereinafter) are based on a cross-validation procedure that leaves out one observation at a time and forecasts the "missing observation" regardless of its ordering in time. This procedure is only applied to the second stage of the estimation as well. This approach is quite different from the recursive out-of-sample procedures that are standard in the time-series forecasting literature. Third, BM did not compute forecast error statistics for simpler models using the same data set as in their study, but, rather, relied on forecast statistics reported in the original articles.

The purpose of this study is to conduct a forecast competition between BM model and simpler crop condition models in forecasting U.S. average corn and soybean yields. Specifically, we compare the forecast accuracy of BM model to Irwin and Good (2017a,b) and Bain and Fortenbery (2017) models. The data for the study consists of weekly state and national crop condition ratings from 1986 through 2020 for corn and soybean. To evaluate the predictability of all yield forecasting models, we use data from 2000 through 2020 as the out-of-sample period. We first recursively estimate all yield forecasting models and provide true out-of-sample yield forecasts. Next, we apply the modified Diebold-Mariano test to conduct a weekly pair-wise comparison between BM model and its competitors. Test results suggest that BM model does not have a systematic superior predictability than other more straightforward yield forecasting models. We also apply the Model Confidence Set tests to select the best individual yield forecasting models. Moreover, we add the composite forecasts as the arithmetic average of the five individual yield forecasts in the set of models. Test results for individual yield forecasting models suggest early in the growing season, Irwin and Good Bias Adjustment model is the best

model, and by the end of growing season model BM model and Bain and Fortenbery model are the two models selected out with the best yield forecasting performance. When we include composite forecasts from Equal Weighted model, test results show composite forecasts provide the most accurate yield predictions. However, test statistics are not significant, indicating all best models fail to outperform their competitors. Last, we apply the multi-horizon average Super Predictive Ability (aSPA) test developed by Quaedvlieg (2021) to compare BM model with its four competitors across the entire growing season. Again, test results indicate that BM model fails to provide more accurate yield forecasts than the competing and simpler forecasting models.

3.2. Data

3.2.1 Crop condition ratings

From roughly late April until the end of November each growing season, USDA weekly Crop Progress reports provide progress and condition ratings for corn and soybean in 18 major producing states. The reports are published on the first business day of the week after 4:00 pm Eastern time. Estimates in the report are based on non-probability subjective surveys conducted by nearly 4,000 local crop observers, who are drawn from the ranks of extension agents, USDA Farm Service Agency (FSA) staff, elevator managers, and other agricultural professionals. Each local observer follows the standard definitions and guidelines provided by the USDA to conduct assessments of crops in their local area. Data are reported on the progress of producer activities (e.g., planting and harvesting), various phenological stages of development (e.g., emergence, flowering), and crop condition ratings. It is important to emphasize that weekly observations are entirely subjective and the result of visual field observations, direct conversations with farmers, and expert local knowledge. For this reason, the data collection process for USDA Crop Progress reports can be described as a system of "people as crop sensors." Finally, state-level estimates are

based on weighting of local observer estimates, usually at the county level, and national-level estimates are based on weighting of each state's planted acreage estimate from the previous year (NASS, 2021; Irwin and Good, 2017a).

The data released in the weekly Crop Progress report are followed closely by grain market participants. For example, Lehecka (2014) notes that these reports are among the most requested publications distributed by the USDA between monthly Crop Production and World Agricultural Supply and Demand Estimates (WASDE) reports. Using event study methods, Lehecka shows the strongest corn and soybean futures market reactions are found in July and August, when weather conditions are most critical for crop development. He also finds that market reactions have increased over time.

Lehecka's work shows that Crop Progress reports have substantial informational value to participants in the grain futures markets. As discussed above, this is especially true during the heart of the summer growing season for corn and soybean. It is during these months that crop condition ratings take center stage. The ratings are reported in five exhaustive categories as follows (NASS, 2021):

Very Poor – Extreme degree of loss to yield potential, complete or near crop failure. Pastures provide very little or no feed considering the time of year. Supplemental feeding is required to maintain livestock condition.

Poor – Heavy degree of loss to yield potential which can be caused by excess soil moisture, drought, disease, etc. Pastures are providing only marginal feed for the current time of year. Some supplemental feeding is required to maintain livestock condition.

Fair – Less than normal crop condition. Yield loss is a possibility but the extent is unknown. Pastures are providing generally adequate feed but still less than normal for the time of year.

Good – Yield prospects are normal. Moisture levels are adequate and disease, insect damage, and weed pressures are minor. Pastures are providing adequate feed supplies for the current time of year.

Excellent – Yield prospects are above normal. Crops are experiencing little or no stress. Disease, insect damage, and weed pressures are insignificant. Pastures are supplying feed in excess of what is normally expected at the current time of year.

The ratings for a given crop in each condition category are expressed as a percentage, reflecting the proportion of the crop rated in a particular category. Since the categories are exhaustive, the percentages in the five categories sum to 100.

We collected all weekly condition ratings for corn and soybean at the state and national level starting in 1986, when the program was established, through 2020. For each year, the coverage of weeks in the growing season is not the same because ratings do not begin until a substantial part of the crop has emerged and do not end until most of the crop is mature. Since dates for emergence and maturity vary from year-to-year, the beginning and ending dates for condition ratings also vary. To have a consistent evaluation period for all competing models, we use weeks 23 – 39 for corn and weeks 25 – 39 for soybean to evaluate the weekly yield forecasts. The ranges of these weeks roughly correspond to early June to late September for corn and late June to late September for soybean. Corn and soybean ratings are available for all years during the sample period for these weeks and for all but a few of the 18-states included for each crop in the Crop Progress report.

3.2.2 Harvested acres

BM model provides weekly yield forecasts at the state-level for the 18 major-producing states included in the Crop Progress report for corn and soybean due to the design of their modelling framework. We are interested in yield forecasts at the national level because this is a key determinant of market prices rather than yield in any individual state. To compare all competing models at the national level, we developed a straightforward method of converting a set of state-level forecasts to one national level forecast. Specifically, we use the ratio of weighted-average yields of 18 states to the national yields. Once the state-level yield forecasts are available, forecasts of national yields can be easily calculated using the estimated ratio. For these 18 states, each individual state has different productivity for corn and soybean. We use the proportion of individual state's harvested acres out of the total harvested acres of 18 states to estimate the yield weight for each state. Each year for each state, we use previous five-year moving-average yield weight as a forecast for current year's yield weight. For the ratio of weighted sum of state-level yields to the final estimates of national yields, we apply a similar previous five-year moving-average procedures to acquire a forecast for the current year's state-to-national yield ratio.

Since five-year moving-average procedures are applied to harvested acres, and the first year we use the crop condition ratings for yield forecasts is 1986, we collected harvested acres for each state from 1976 – 2020. The harvested acres data are obtained from NASS Quick Stats, and they are published in the Acreage reports released by NASS each year by the end of June. The Acreage report produces the revised harvested acres for the previous year and forecasted harvested area for the current year. The timing of the Acreage report roughly lines up with the beginning of the forecast window each year.

3.2.3 Annual yield estimates

All weekly yield forecasts are compared to the final yield estimates published at the NASS Quick Stats website. As only one five-year moving-average procedure is applied to yields, we collected yield data from 1981 through 2020.

3.3. Yield Forecasting Models

3.3.1 Yield forecasting cycles

The goal of all yield forecasting models in this study is to provide early yield projections when weekly condition ratings are available for corn and soybean. Figure 3.1 uses corn to illustrate a typical forecast cycle. Each year of our sample, the first yield prediction starts in week 23. The yield forecasts for week 23 are obtained using crop condition ratings published in this week. Importantly, all the forecast models are estimated recursively using samples that end before a given forecast week. The out-of-sample period is 2000 through 2020 and forecasts for corn are made for week 23 – week 39 in each year and for soybean for week 25 – week 39. To evaluate the performance of yield forecasting models, we compare the weekly forecasts with final yield estimates published in the USDA's Crop Production Annual Summary report that is released in January after the growing season.

3.3.2 Begueria and Maneta model

Begueria and Maneta (BM) model (2021) is the most technically sophisticated model considered in this forecast competition. They argue that spatial and temporal differences in crop condition information should be directly modeled before making yield forecasts. Hence, BM developed a cumulative link mixed model (CLMM) to transform raw condition data to a continuous and almost normal-distributed crop condition index (CCI). After removing space and time effects, they argue

that maximum information can be extracted from crop condition ratings, which offers a better possibility of providing unbiased and accurate yield forecasts.

In formal terms, the first step of BM model is to estimate the CLMM using a probit link function to connect ordinal response with numeric factors. The CLMM is specified as:

$$probit\big(P(Y_i \leq j|s,y,w)\big) = \theta_j + \beta_y y + \beta_w w + v_s + v_{y,s} y + v_{w,s} w + \epsilon_{si}, \quad (3.1)$$

where $probit\big(P(Y_i \leq j|s,y,w)\big)$ is the probability that the ith report's condition ratings are no greater than category j, and $j\epsilon[1,4]$ since there are five condition categories; s, y and w are state year and week in report i, respectively; and θ_j is a threshold parameter which remains constant and determines the range of the response variable in a certain category j. There are two fixed effects in the model: a long-term (year) effect and a temporal (week) effect. Three random effect components are included: state, the interaction between state and year, and the interaction between state and week. The error term ϵ_{si} is the unbiased CCI that is specific for each state and is free of any long-term or temporal time effects.

In the second stage of their modeling process, BM develop a mixed model, where the fixed effects are the long-term (year) effects and CCI effects and the random effect is conditional on each state including the intercept two slopes with the interactions from year and CCI. This model provides weekly yield forecasts for each state and is specified as:

$$\mu_i(s) = \beta_0 + \beta_y y_i + \beta_c CCI_i + v(s) + v_y(s) y_i + v_c(s) CCI_i + \epsilon_i \quad (3.2)$$

where $\mu_i(s)$ is the expected yield at state s and time i, y_i is the transformed year index at time i, CCI_i is the crop condition index at time i, β_0 is the global intercept, β_y is the long-term year effects and β_c is the CCI effect (they are both fixed effects and have the same effects on all the states). The BM model treats state as the random components, meaning for different states, they have different temporal effects and CCI effects.

Figure 3.2 uses corn as an example to illustrate how BM model recursively provide out-of-sample weekly yield forecasts. Yield forecasts for week 23 in 2000 were estimated with the following steps: i) the CLMM model is estimated using crop condition ratings from the first published Crop Progress report in 1986 to the most recent report published in week 23 of 2000. With the updated model, we can transform and update the ordinal crop condition ratings for all the weeks till week 23, 2000. Second, we can estimate the mixed model using the updated CCI and other variables in week 23 from 1986 to 1999. Third, updated CCI and year index of week 23, 2000 were entered in the mixed model and we can obtain a yield projection for week 23, 2000. Following these steps, as we move forward in the growing season, we can have weekly updates of yield forecasts.

We also present BM yield forecasts for corn and soybean at the national level. We calculate the weighted average yields of 18 states where the weight for each state is the proportion of harvested acres. To transform the weighted average yields of 18 states to the national yields, we apply the ratio of weighted average yields of 18 states to the national yields. Yield forecasts at national level are available from 2000 to 2020.

3.3.3 Irwin and Good model

The design of Irwin and Good model (Irwin and Good, 2017a) makes it applicable for both state-level and national-level yield forecasts. At the national level, Irwin and Good model (IG National model, hereinafter) is specified as below:

$$Yield_t = \beta_0 + \beta_1 year\_index_t + \beta_2 SUM_t + \epsilon_t \quad (3.3)$$

where $Yield_t$ is national final yield estimates in year t; $year\_index_t$ is the time index in year t; SUM_t is the sum of "excellent" and "good" ratings at the end of the season in year t.

With corn as an example, Figure 3.3 illustrates how to provide recursively out-of-sample yield forecasts with the model. Yields forecasts for 2000 week 23 are obtained with the following steps. First, we run IG National model with time index, the percentage of corn rated in "good" and "excellent" conditions at the end of years, and the national final yield estimates from 1986 to 1999. Second, the sum of ratings in week 23 and the year index for 2000 are entered in the model to get the yield forecasts for week 23, 2000.

State-level yield forecasts follow the same procedure of national level. Instead of using national yield estimates, they use state-level final yield estimates to build Irwin and Good State model (IG State model, hereinafter) and eventually receive weekly yield forecasts for each state. For individual state, we can also compare the predictability of BM model and IG State model to examine if there is a significant trade-off between model complexity and forecast accuracy.

Irwin and Good (2017b) pointed out that the disadvantage of this straightforward approach is that it does not consider the bias in the early weeks' condition ratings within the growing season. Their weekly analysis (Irwin and Good, 2017a) showed that in early weeks the correlations coefficients between the sum of "good" and "excellent" ratings and the yields are lower than that of final weeks. The reason behind this observation is that, on average, the early weeks' ratings for corn and soybean are over-estimated. Early in the growing season, crops usually are in a normal or a better than normal condition. However, for a few years (like drought in 2012), when adverse weather conditions occur, crop yields would deteriorate and become worse than normal. This would make crop ratings of "good" and "excellent" in the final week lower than that in early weeks, which makes the final week's ratings lower than the average ratings in early weeks. To measure the size of bias, we follow definition of bias proposed by Irwin and Hubbs (2018a,c) and specify the bias as:

$$bias_t = final\ week\ rating_t - early\ week\ rating_t, \quad (3.4)$$

where $final\ week\ rating_t$ is the current year's sum of "good" and "excellent" ratings at the end of growing season and $early\ week\ rating_t$ is the sum of "good" and "excellent" ratings of each early week in year t. Because on average final ratings are lower than that of early weeks' ratings, we expect the bias to be negative. To adjust the bias in the early weeks, we need to add the bias to the early weeks' ratings as:

$$adj\_early\_rating_t = early\ week\ rating_t + bias_t. \quad (3.5)$$

For both corn and soybean, we consider weeks before 31 as the early weeks that need bias correction. Therefore, these weeks are week 23 – week 30 for corn and week 25 – week 30 for soybean.

We apply the moving-average procedures to estimate the size of bias. With ten-year moving-average approach, we first calculate the weekly rating difference between the final week and each of the early weeks over the previous ten years. Then, as we have the average bias for each week of the early weeks, for the current year, we can add the bias to the reported ratings to have the adjusted ratings for a week that is in the range of early weeks. For some weeks, we do not have consecutive observations in all each year. In these scenarios, we use all the data we have from the previous ten or five years, but we might not have all the ten or five data points to calculate the average bias. These two approaches are considered as two augmented bias-adjusted Irwin and Good models (IG National with Bias Adjustment model, hereinafter).

3.3.4 Bain and Fortenbery model

Bain and Fortenbery (2017) fixed weight model (BF model, hereinafter) assigned fixed weights to each condition category to transform the ordinal condition ratings to a numerous crop condition index (CCI). Below is the definition of fixed weights CCI:

$$CCIndex = 100\% \cdot Excellent + 75\% \cdot Good + 50\% \cdot Fair$$
$$+25\% \cdot Poor + 0 \cdot Very\ Poor \quad (3.6)$$

The ratings for each condition category are in percentages, therefore fixed weights CCI is bounded between 0 and 1. Bain and Fortenbery built the weekly crop yield forecasting model by having the end of season $CCIndex$ in the framework, and the model is specified as:

$$Yield_i = \alpha_0 + \alpha_1 \cdot Trend_i + \beta_1 \cdot CCIndex_i + e_i \quad (3.7)$$

where $Yield_i$ is the final yields in year *i*, $Trend_i$ is the time index for year *i*, $CCIndex_i$ is the end of season $CCIndex$ value for year *i*. For example, the yield forecasts for week 23, 2000 for corn are estimated with the following steps. First, we transform crop conditions of the end of growing season to the fixed weight $CCIndex$ from 1986 through 1999. Second, we run the model with annual final yield estimates as the response variable and year index and the fixed weight $CCIndex$ as explanatory variables. Third, once we obtain the crop condition ratings for week 23, 2000, we transform them to the fixed weight $CCIndex$ and enter them in the model with updated year index for 2000 to have the yield forecasts.

3.4. Model Comparison and Forecast Evaluation

There are two sets of comparisons conducted by our study. First, we compare all five yield forecasting models with naïve trend yield model to evaluate the value of crop condition index as a yield indicator. Second, we set BM model as a benchmark to compare it with other four simpler yield forecasting models. The comparisons are conducted both at state level and national level, by focusing on the forecast errors for week 29 of mid-July, over the out-of-sample period from 2000 through 2020, and the root mean squared percentage error (RMSPE) of each yield forecasting model over the out-of-sample period throughout the entire growing season.

We use the absolute value of the difference between final yield estimates and the yield forecasts for week 29 to measure the forecast error in week 29. The weekly forecast errors $e_{w,t}^{i}$ for model i are defined as the percentage difference between the USDA final yields and this model's yield forecasts:

$$e_{w,t}^{i} = 100 \cdot \frac{(y_t - \widehat{y_{w,t}^{i}})}{y_t} \tag{3.8}$$

where y_t is the final USDA yield estimates and $\widehat{y_{w,t}^{i}}$ is the predicted yields in year t for week w produced by model i. We use the root mean squared percentage error (RMSPE) to measure each model's predicative accuracy. RMSPE is defined as

$$RMSPE_{w,t}^{i} = \sqrt{\frac{1}{n}\sum\left(\frac{(y_t - \widehat{y_{w,t}^{i}})}{y_t}\right)^2} \tag{3.9}$$

where n is the number of observations for each week over the out-of-sample period. One advantage of RMSPE error is that it transforms the error to the positive percentage value, so it avoids offsetting positive and negative errors, and we only need to consider one direction of the error. The other advantage is that RMSPE makes the errors comparable for corn and soybean.

3.4.1 Naïve trend yield model

One of the key factors that determines crop yields is the technology development over the years. Crops tend to increase their yields year by year, which is known as the "trend yield" (Irwin, Good, and Tannura, 2009). Naïve trend yield model serves as the base model that we use to compare with five yield forecasting models. This is because naïve trend yield model only includes time to account for the variations in yields over time, and yet yield forecasting models include time and crop condition ratings to explain the development of yields. The comparisons provide clear evaluations of whether additional crop condition ratings contain valuable yield information as

naïve model only includes year index to account for the variations in national yields. Naïve trend yield model is specified as below:

$$Yield_t = \beta_0 + \beta_{1,t} year\_index_t + \epsilon_t, \quad (3.10)$$

where $Yield_t$ is the national final yield estimates in year t, $year\_index_t$ is the corresponding year index running from 1 to 35 for the year from 1986 to 2020.

The yield forecasts provided by naïve trend yield model also follow the recursively out-of-sample forecasting approach. For example, when we are in year 2000, we use yields and time indices from 1986 to 1999 to train the model. In 2000, we can make yield predictions using the updated year index of 15 for all weeks during the growing season for corn and soybean.

3.4.2 Comparison at state level

Both BM model and IG State model provide the state-level yield forecasts. To compare which model systematically provide better yield forecasts, we compare: (i) the absolute value of yield forecast errors for mid-growing season, that is approximately week 29 for both corn and soybean over the out-of-sample period; (ii) weekly RMSPE over the out-of-sample period for each week during the growing season. We conduct the comparisons for two representative states given their geographic difference: Illinois and South Dakota.

Figure 3.4 and Figure 3.5 present the percentage difference between forecasted yields provided by BM model and IG State model for week 29 over the out-of-sample period for corn and soybean, respectively. Figure 3.4 (a) and Figure 3.5 (a) show that in Illinois, there is no clear pattern of which model outperforms over time. Figure 3.4 (b) and Figure 3.5 (b) show a similar pattern in South Dakota. For year 2012 when crop productions are largely impacted by droughts, we observe that for Illinois, BM model provided more accurate yield forecasts than IG State model in the mid-growing season, whereas for South Dakota, it shows IG State model is more accurate.

Figure 3.6 and Figure 3.7 present the RMSPE for each week during the growing season over the out-of-sample period for corn and soybean. For Illinois, Figure 3.6 (a) show that BM model has better performance since mid-July till the end of growing season for corn; Figure 3.7 (a) shows IG State model outperforms BM model from mid-July to mid-August for soybean. For South Dakota, Figure 3.6 (b) suggests BM model takes the lead from early-June to early-July, then IG State model provides more accurate yield forecasts from early-July till the end of growing season for corn; Figure 3.7 (b) suggests that BM model has better forecasting performance from early-June to late-August, then Irwin and Good model takes the lead till the end of growing season for soybean.

3.4.3 Comparison at national level

All yield forecasting models been discussed in this study provide national-level yield forecasts for each week during the growing season over the out-of-sample period. First, we focus on the forecast errors for mid-growing season from 2000 – 2020. Figure 3.8 and Figure 3.9 presents the forecast error between national yield forecasts provided by five yield forecasting models and the final USDA yield estimates for corn and soybean, respectively. We also present the yield forecasts provided by naïve tend yield model. It suggests that from 2000 – 2020, all forecasting models provide more accurate yield forecasts than naïve trend yield model, which shows the value of crop condition ratings for yield forecasts. To compare the forecast errors of each forecasting model, we observe that there is no clear pattern to show which model has the superior forecasting performance.

Table 3.1 summarizes the RMSPE of five forecasting models for each week during the growing season. RMSPE of all five models for corn are bounded with a maximum level of 9.1% from IG National model, indicating that for week 23, the yield forecasts provided by IG National

model are within 9.1% of the final average yields. A minimum level of 3.85% from BM model for week 39, indicating the yield forecasts provided by Bain and Fortenbery model are within 3.5% of the final yields estimates. The average of RMSPE for corn is about 5% throughout the growing season. For soybean, the pattern is similar. RMSPE are in the range of (3.8%, 7.8%), and the overall average RMSPE across the whole forecasting path is about 6%.

Figure 3.10 shows for corn, all five forecasting models provide more accurate yield forecasts than naïve trend yield model since week 24, about early-June till the end of growing season. Figure 3.11 shows for soybean, all five forecasting models show the forecasting advantage since early-August. Both plots present the pattern of the yield forecasts provided by BM model and its four competitors: near the end of growing season, all models provide the most accurate yield forecasts; and later in the growing season, there is no forecasting improvements. This pattern indicates by the mid-August, yield forecasting models apply the crop condition ratings reach the limits as human observations cannot fully capture the true underlying information in the fields.

3.4.4 Single-horizon yield forecasts comparison

For each week we conduct a pairwise comparison between BM model and its four competitors. We apply the modified Diebold-Mariano (MDM) test for each week to test if BM model provides more accurate yield forecasts at single horizon throughout the growing season. The MDM test is developed by Harvey, Leybourne, and Newbold (1997) with the advantage that the MDM test works well for small samples; and with the increase in forecasting horizons, the over-sized MDM test results remain stable. The MDM test is applied on two models' out-of-sample forecast errors. For each week, there are 21 observations as the out-of-sample period covers years from 2000 – 2020.

The null hypothesis is that two models have the same predictive accuracy, and it lies upon the loss function between two models' errors. To be more specific, we test if the difference in RMSPE between BM model and other of its competitors is significant. Here we assume the loss function to be quadratic, and when we fail to reject the null hypothesis, we have:

$$d_{w,t} = \left(e_{w,t}^2\right)^2 - \left(e_{w,t}^1\right)^2 \tag{3.11}$$

$$E\left(d_{w,t}\right) = 0 \tag{3.12}$$

where $e_{w,t}^1$ represents the yield errors from BM model, and $e_{w,t}^2$ represents the yield errors from one of its competing models.

For the h-step ahead yield forecasts, the statistic of MDM is defined as:

$$MDM = \left[\frac{n+1-2h+n^{-1}h(h-1)}{n}\right]^{\frac{1}{2}} \cdot \bar{d}_w \cdot \left[V(\bar{d}_w)\right]^{-\frac{1}{2}} \tag{3.13}$$

$$V\left(\bar{d}_w\right) = [n^{-1}(\gamma_0 + 2\sum_{s=1}^{h-1}\gamma_s)] \tag{3.14}$$

where $\bar{d}_w$ is the sample mean of $d_{w,t}$, w is the forecast week and $w = 1,2,3,\ldots,17$ for corn and $w = 1,2,3,\ldots,15$ for soybean, $\gamma_0 = n^{-1}\sum_{t=1}^{n}(d_{w,t} - \bar{d}_w)^2$ as the variance of $d_{w,t}$, $\gamma_s = n^{-1}\sum_{t=s+1}^{n}\left(d_{w,t} - \bar{d}_w\right)\left(d_{w,t-s} - \bar{d}_w\right)$, $s = 1,2,3,\ldots,h-1$, as the sth auto-covariance of $d_{w,t}$. As each week we make yield predictions for a year ahead, we have one-step ahead forecasts where $h = 1$. Therefore, the MDM statistics for each forecast week is:

$$MDM_w = [(n-1)]^{\frac{1}{2}} \cdot \bar{d}_w \cdot \left[n^{-1}\left(\sum_{t=1}^{n}(d_{t,w} - \bar{d}_w)^2\right)\right]^{-\frac{1}{2}} \tag{3.15}$$

The MDM test statistics for corn and soybean are summarized in Table 3.3 and Table 3.4. The null hypothesis is that: each week throughout the out-of-sample forecasting period, the forecasting performance of BM model and one of its competing models have the same predictability. Test statistics show that for corn, out of 68 cases of pair-wise yield forecast comparisons, from week 23 to week 39, all test statistics are insignificant. These results suggest

that we fail to reject the null hypothesis that BM model does not have better forecasting performance than other less-computation demanding models. For soybean, out of 60 cases of pair-wise yield forecast comparisons, covering forecast weeks from week 25 – week 39, there is no significant cases. These results suggest for soybeans, BM model does not outperform its competitors for each week throughout the growing season.

3.4.5 Best model selected by the Model Confidence Set (MCS)

Each week, all five yield forecasting models produce weekly yield forecasts for corn and soybean. In the previous section, we apply the MDM test to conduct a pairwise yield performance test between BM model and one of its competing models. To extend the pairwise comparisons, Model Confidence Set (MCS) test allows model selection for all yield forecasting models (Hansen, Lunde, and Nason, 2011). For a given significance level α, MCS test selects the model with best forecasting accuracy from a set of models.

Colino et al. (2012) show that equal-weighted composites provide more accurate forecasts than individual outlook programs for hog prices. Following their approaches, we build the Equal Weighted Model that produce composite forecasts which are the arithmetic average of the five individual yield forecasts. We include the composite forecasts in the set of yield forecasting models and we apply the MCS test to test whether composite forecasts outperform individual forecasts.

As MCS is built on the iterative procedures where each step, it eliminates the worst performing model from the set of six models ($\mathcal{M}_0$) until the last model survives from the tests in all previous five steps. Each step, to select which model should be eliminated, it is based on the t-statistics proposed by Hansen, Lunde, and Nason (2011):

$$t_{i\cdot} = \frac{\overline{d_{i\cdot}}}{\sqrt{\widehat{var}(\overline{d_{i\cdot}})}}, \text{ for } i, j \in \mathcal{M}_0 \tag{3.16}$$

where $\overline{di.} \equiv m^{-1} \sum_{j \in \mathcal{M}_0} \overline{dij}$, $\overline{dij} = n^{-1} \sum_{t=1}^{n} d_{ij,t}$, $d_{ij,t} = L_{i,t} - L_{j,t}$, $L(\cdot)$ is the squared error function. Corresponding p-values are collected from the bootstrap of the test statistics. The best model selected by MCS has p-value equals to 1. When more than one model has p-value equals to 1, we use the equivalence test: $T_D \equiv \sum_{i \in \mathcal{M}_0} (t_i.)^2$ to test if the last survived model outperforms its competitors.

Our study reports the last model selected by MCS test based on the p-values produced by 2,000 bootstrap replicates for each week. We first show MCS test results for the set of models only consists of five individual yield forecasting models; next we show the MCS test results for the set of models with Equal Weight Model and five individual yield forecasting models. The significance level for MCS test is 10%. We also report the p-values for the equivalence test. When p-values greater than 0.1, it suggests the best selected model fails to have superior predictability than its competitors.

Weekly MCS test results for corn and soybean are reported in Table 3.5 and Table 3.6. In Table 3.5, we report the best model that survives four steps of model selections for each week. The set of models consist of five individual yield forecasting models. For both corn and soybean, early in the growing season, the best model is IG National with Bias Adjustment, and by the end of growing season, BM model and BF model provide the most accurate yield predictions. In Table 3.6, we report the MCS test for the set of models including Equal Weighted model and five individual yield forecasting models. For both crops throughout the growing stages, the best model selected is Equal Weighted model that provide equal-weighted composite forecasts of five individual yield forecasts. For Table 3.5 and Table 3.6, each week, the p-value of each step's model elimination is greater than the significance level of 0.1, and there is more than one model has p-value equals to 1. The equivalence test p-values are also reported in Table 3.5 and Table 3.6. They

fail to reject the null hypothesis of whether these models have equal predictive ability. These findings suggest the selected best models indeed provide more accurate yield forecasts, however, they to significantly outperform their competitors.

3.4.6 Weekly forecasting errors correlation

During the growing season, crop conditions ratings are published each week. In the estimation process, the data under preparation are compared with data reported in previous week and in surrounding counties. This procedure makes the weekly condition ratings correlated in a year. We want to test if such dependence is available in the forecasting error between the USDA final yields and the yield forecasts produced by one of our selected yield forecasting models. We conduct the correlation test for corn and soybean forecasting errors using BM model as we assume that for other models because the out-of-sample yield forecasts are also produced recursively, they should follow a similar pattern. We run multiple OLS models between the first week and the weeks ahead. Each OLS estimation, the first week is treated as the independent variable, and each week ahead is the dependent variable.

Correlation test results for corn and soybean are summarized in Table 3.2. We use Heteroskedasticity and Autocorrelation Corrected (HAC) standard errors to produce t-statistic and p-value. Test results show that the first week are correlated significantly to all weeks in the growing season at the 5% significant level. As all competing yield forecasting models have similar patterns of yield forecast errors over the growing season, we can expect this correlation embedded in these forecasting models as well.

3.4.7 Multi-horizon yield forecasts comparison

One limitation of the MDM test is that it only provides comparison test results for two competing models at each horizon w. It is very common to find that at some horizons the first model

outperforms the second, and at some other horizons, such situation reverses. Sometime, only using the single horizon forecast comparison test like the MDM test is likely to conclude contradictory results. For two competing models that cover multi-horizons, it is necessary to perform an omnibus test on all the forecasting horizons. When we argue which model has better forecasting performance, the omnibus test adds more conclusive evidence. Quaedvlieg (2021) introduced the multi-horizon superior predictive ability (SPA) tests that enable the comparison of forecasts of different models jointly, combining these models' predictability across all horizons. The author proposed two tests, the first one is the uniform SPA test that tests if a model has superior forecasting performance at each individual horizon; the second one is the average SPA test that tests if a model has superior forecasting performance considering the whole forecasting path. For our study, we follow Quaedvlieg (2021) average SPA test as we can see in Figure 3.10 and Figure 3.11 that there are some cross-over points between BM model and its competing models for corn and soybean, suggesting for some weeks, BM model has lower forecast errors while in some other weeks, its competing models achieve lower forecast errors. Therefore, it is more appropriate to apply the multi-horizon average Super Predictive Ability (aSPA) test (2021) pair-wisely between BM model and the other four forecasting models. This test extends the MDM test and compare two models' yield forecasts across the whole growing season.

Each year USDA final yields are denoted as y_t, and the weekly yield forecasts produced by model i is denoted as $\widehat{y_t^i}$. In multi-horizon test framework, $\widehat{y_t^i}$ is a 17-dimension vector, $\widehat{y_t^i} = \left[\widehat{y_{1,t}^i}, \widehat{y_{2,t}^i}, \dots \widehat{y_{h,t}^i}, \dots, \widehat{y_{17,t}^i}\right]$, where h indicate the week that produces the yield forecasts; i represents different choice of forecasting models; t is the year when the fixed-event forecasts happen. We define the loss function as $L_t^i = L\left(y_t, \widehat{y_t^i}\right)$, and it projects the final yield estimates onto a 17-dimension space. The loss function is defined in a quadratic form, that is the square of the

percentage difference between the final yield estimates and each week's yield forecasts provided by model i. Here we use notation "1" to stand for BM model, and "2" for its competing model. Then we define the loss differential for the two competing yield forecasts as $d_t = L_t^2 - L_t^1$. D is the loss differentials matrix and its dimensions are 21×17. $D = [d_1^T, \dots, d_t^T, \dots, d_{21}^T]^T$, where $d_t = [d_t^1, \dots, d_t^h, \dots, d_t^{17}]$. Each entry of the matrix D is denoted as d_t^h, and D is specified as:

$$D = \begin{bmatrix} d_1^1 & \dots & d_1^h & \dots & d_1^{17} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_t^1 & \dots & d_t^h & \dots & d_t^{17} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{21}^1 & \dots & d_{21}^h & \dots & d_{21}^{17} \end{bmatrix}_{21\times 17} \tag{3.17}$$

We use the mean loss differentials, $\mu^{aSPA} = \sum_{h=1}^{17} w_h \mu_h$, to compare two models' overall predictability. μ^{aSPA} can be taken as the weighted sum of each week's average differentials, where w_h is the weights for each forecast week; $\mu_h = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} d_t^h$ is the mean of each week's loss differentials and we use $\bar{d}_h = \frac{1}{21} \sum_{t=1}^{21} \widehat{d_t^h}$ to estimate μ_h. The null hypothesis of the aSPA test is $\mu^{aSPA} \leq 0$, meaning considering all horizons, on average, BM model fails to provide better performance than its competitors. The studentized statistic for aSPA test is:

$$t_{aSPA} = \frac{\sqrt{T} \sum_{h=1}^{17} w_h \cdot \bar{d}_h}{\hat{\varsigma}} \tag{3.18}$$

where $\hat{\varsigma} = \sqrt{w' \widehat{\Omega} w}$; $w = [w_1, \dots, w_h, \dots, w_{17}]^T$is the 17-dimentional weight vector. Ω is the variance-covariance matrix of matrix D. We denote $D = [D_t^1, \dots, D_t^h, \dots, D_t^{17}]$, where $D_t^h = [d_1^h, \dots, d_{21}^h]^T$. The variance-covariance matrix Ω of matrix D is defined as:

$$\Omega = \begin{bmatrix} var(D_t^1) & \dots & cov(D_t^1,\ D_t^{17}) \\ \vdots & \vdots & \vdots \\ cov(D_t^1,\ D_t^{17}) & \dots & var(D_t^{17}) \end{bmatrix}_{17\times 17}, \tag{3.19}$$

where $var(D_t^1) = \frac{1}{T}(D_t^1)^T(D_t^1) = \frac{1}{T}\sum_{t=1}^{21}(d_t^1)^2$; $cov(D_t^1,\ D_t^{17}) = \frac{1}{T}(D_t^1)^T(D_t^{17}) = \frac{1}{T}\sum_{t=1}^{21} d_t^1 \cdot d_t^{17}$. For a given year, from our correlation test results, we found each week's differentials are highly correlated. Instead of directly estimating the full variance-covariance matrix Ω, we use the Newey-West HAC estimator to find its estimator, $\widehat{\Omega}$. The choices of weights are flexible. We follow the examples proposed by Quaedvlieg (2021): first, we select the equal weight where $w^h = \frac{1}{17}$ for each week; second, we use "efficient" weights to minimize ς as the yield forecasts during the growing season are based on accumulated crop growing survey information. We assign small weights to early forecasts where variance is high, and we assign large weights to near end-of-season forecasts where variance is low. Therefore, the inverse-variance weights are defined as $w^h = \frac{1}{\sigma_h^2(\sum_{i=1}^{17}\sigma_i^2)}$ and they satisfy the condition that the sum of weights is equal to 1. To obtain the critical values and p-values, we use the moving block bootstrap (MBB) technique to draw the distribution. We focus on the significance level at 5%, and the significance level is the corresponding percentile of the bootstrap distribution.

From the single-horizon MDM test, throughout the growing season, for both corn and soybean, the test results produce mixed evidence. To test if BM model has better forecasting performance throughout the whole growing season, we perform multi-horizon average SPA test. First, we assign equal weights to each horizon for the loss differentials. The null hypothesis of the average SPA test is that considering all horizons, on average, simple yield forecasting model has better performance than BM model. Test results are summarized in Table 3.7. The multi-horizon average SPA test p-values are all greater than 5%, suggesting BM model fails to have significantly better predictability than its four competitors. Second, we conduct the average SPA test with varying weights for each week of the growing season. Test results are summarized in Table 3.8. The findings with average SPA tests are consistent with what we found with the single horizon

MDM test: BM model fails to systematically outperforms its competing models during the growing season for corn and soybean. A plausible argument for this finding is that BM model only controls for the time and spatial variations in the state-level crop condition ratings, so the transformed weekly CCI do not contain more determinate factors to account for the variations in yields than the other yield forecasting models which apply simple approaches to transform the ordinal condition ratings.

3.5. Conclusions

Crop production forecasts have been an important indicator for price changes in agricultural commodity markets. A small group of studies use the crop conditions data to build the yield forecasting models. Condition ratings are the products of human sensors, and they provide consistent subjective assessment about crops conditions that highly correlated with crops' yields.

This study examined the forecasting accuracy of a batch of yield forecasting models that directly transform the ordinal crop condition ratings to the numeric condition index along with a recently developed model introduced by Begueria and Maneta in 2020 that applies the cumulative link mixed model to transform the condition ratings to the continuous condition index. We conduct the out-of-sample yield forecasts recursively for corn and soybean from 2000 through 2020 for all models. We compared each model's yield forecasts with USDA final yield estimates and we used RMSPE to measure each model's forecasting accuracy. We found all models provide a pattern of the forecasting accuracy: in early weeks of the growing season, RMSPE are relatively higher than that in the final weeks. This pattern is reasonable as we move toward the end of season, crops are about to mature, so their conditions connect to yields more closely. The average RMSPE level for all models throughout the growing season is about 5% for corn and 6% for soybean. Our findings suggest this group of models that use crop conditions data provide accurate yield forecasts.

This study compared the forecasting performance of BM model with its four competing yield forecasting models that have already been widely applied by industry practitioners. One disadvantage of BM model is that it is more complex and computational demanding. Our study evaluates if the BM model provides significantly superior yield forecasts than its competitors so its disadvantage can be compensated. We use the modified Diebold Mariano test for the single-horizon pair-wise forecasts comparisons. Test results suggest for both corn and soybean, BM model fails to outperform its competitors. With Model Confidence Set test, we find among individual yield forecasting models, in early weeks IG with Bias Adjustment is usually the best model, and in final weeks, BM model and BF model are selected as the best models. With composite forecasts in the set of models, Equal Weighted model is selected as the best model. However, all best models fail to significantly outperform their competitors. Furthermore, we conduct the multi-horizon average Superior Predictive Ability test to test whether averaging out the forecasting performance over the growing season, BM model has superior predictability. Test results show that for both corn and soybean, BM model fails to provide significantly more accurate yield forecasts than its competing yield forecasting models.

3.6. Tables and Figures

Table 3.1: The RMSPE of weekly yield forecasting models for corn and soybean at national level over 2000 – 2020

Date	BM Model	IG State Model	IG National Model	IG National with Bias Adjustment Model	BF Model
Panel A: Corn					
June 03	7.9	8.6	9.1	7.9	8.7
June 10	7.7	8.1	8.6	7.5	8.4
June 17	7.2	7.6	7.9	6.7	7.7
June 24	6.4	6.8	7.1	5.9	6.8
July 01	5.9	5.9	6.3	5.3	5.8
July 08	5.2	5.0	5.2	4.6	4.7
July 15	4.6	4.5	4.7	4.5	4.3
July 22	4.4	4.2	4.5	4.4	4.1
July 29	4.4	4.2	4.4	4.4	4.0
August 05	4.4	4.1	4.4	4.4	4.1
August 12	4.3	4.0	4.3	4.3	4.0
August 19	4.2	4.0	4.1	4.1	3.8
August 26	4.3	4.1	4.3	4.3	4.1
September 02	4.3	4.2	4.3	4.3	4.2
September 09	4.2	4.2	4.3	4.3	4.1
September 16	3.9	4.1	4.2	4.2	3.8
September 23	3.8	3.9	4.0	4.0	3.5
Panel B: Soybean					
June 17	6.4	7.2	7.5	6.6	7.7
June 24	6.5	7.1	7.4	6.6	7.8
July 01	6.7	7.1	7.3	6.8	7.7
July 08	6.9	7.1	7.3	6.9	7.6
July 15	7.1	7.0	7.2	6.9	7.6
July 22	6.9	6.9	6.9	6.6	7.4
July 29	6.7	6.7	6.7	6.7	7.2
August 05	6.6	6.6	6.7	6.7	7.0
August 12	5.9	5.9	5.9	5.9	6.2
August 19	4.9	5.0	5.0	5.0	5.2
August 26	4.3	4.5	4.5	4.5	4.6
September 02	4.2	4.4	4.3	4.3	4.3
September 09	4.1	4.2	4.2	4.2	4.2
September 16	3.9	4.1	4.0	4.0	3.9
September 23	3.7	3.8	3.9	3.9	3.8

Notes: For each week, there are 21 observations in the out-of-sample period from 2000 – 2021. The RMSPE measures the average forecast errors over the out-of-sample period, and it is measured in percentage (%). BM model is proposed by Begueria and Maneta (2020), IG State model, IG National model, IG National with Bias Adjustment Model are proposed by Irwin and Good (2017a), and BF model is proposed by Bain and Fortenbery (2017).

Table 3.2: Correlation test results for BM model for corn and soybean over 2000 – 2020

Panel A: Corn, independent variable: forecasting error week 23 (June 03)				Panel B: Soybean, independent variable: error in week 25 (June 17)			
Dependent Week Date	Coefficient	t-Statistic	p-Value	Dependent Week Date	Coefficient	t-Statistic	p-Value
June 10	**0.969**	70.881	1.70E-24				
June 17	**0.905**	48.456	2.25E-21				
June 24	**0.799**	21.316	9.97E-15	June 24	**1.003**	51.069	8.35E-22
July 01	**0.714**	15.04	5.25E-12	July 01	**1.007**	29.283	2.84E-17
July 08	**0.587**	10.564	2.16E-09	July 08	**1.004**	19.573	4.71E-14
July 15	**0.454**	7.311	6.22E-07	July 15	**0.974**	14.435	1.08E-11
July 22	**0.395**	5.632	1.98E-05	July 22	**0.906**	11.455	5.66E-10
July 29	**0.396**	6.015	8.71E-06	July 29	**0.9**	13.453	3.67E-11
August 05	**0.39**	6.714	2.04E-06	August 05	**0.857**	12.864	7.93E-11
August 12	**0.408**	7.444	4.80E-07	August 12	**0.786**	13.558	3.21E-11
August 19	**0.388**	6.287	4.91E-06	August 19	**0.622**	9.832	6.91E-09
August 26	**0.376**	6.251	5.29E-06	August 26	**0.546**	8.775	4.13E-08
September 02	**0.376**	6.899	1.41E-06	September 02	**0.521**	8.719	4.56E-08
September 09	**0.36**	7.581	3.69E-07	September 09	**0.491**	8.093	1.41E-07
September 16	**0.333**	7.259	6.88E-07	September 16	**0.474**	8.681	4.88E-08
September 23	**0.327**	6.913	1.36E-06	September 23	**0.457**	9.497	1.20E-08

Notes: Correlation test results with the OLS regression where the independent variable is the forecast error in the first week; and the dependent variable is each one of the other weeks in the growing season. Regression function is specified as: $e_{1+i} = \alpha_i + \beta_i e_1 + \sigma_i, i = 1, \dots, 16$ for corn and $i = 1, \dots, 14$ for soybean. HAC estimation is used to correct for autocorrelation and heteroskedasticity. Coefficients in bold indicate they are significantly at the level that is no greater than 5%. BM model is proposed by Begueria and Maneta (2020).

Table 3.3: The Modified Diebold Mariano (MDM) test statistics between the BM model and other yield forecasting models for Corn

Date	BM vs IG State	BM vs IG National	BM vs IG National with Bias Adjustment	BM vs BF
June 03	0.306	0.664	-0.455	0.434
	(0.763)	(0.514)	(0.654)	(0.669)
June 10	0.056	0.463	-0.846	0.245
	(0.956)	(0.648)	(0.408)	(0.809)
June 17	0.020	0.319	-1.359	0.119
	(0.984)	(0.753)	(0.189)	(0.907)
June 24	0.083	0.385	-1.370	0.082
	(0.935)	(0.704)	(0.186)	(0.935)
July 01	-0.415	0.002	-1.442	-0.493
	(0.683)	(0.998)	(0.165)	(0.628)
July 08	-0.545	-0.194	-1.470	-0.811
	(0.592)	(0.848)	(0.157)	(0.427)
July 15	-0.521	-0.067	-0.710	-0.856
	(0.608)	(0.947)	(0.486)	(0.402)
July 22	-0.659	-0.055	-0.167	-0.871
	(0.517)	(0.957)	(0.869)	(0.394)
July 29	-0.923	-0.326	-0.326	-1.092
	(0.367)	(0.748)	(0.748)	(0.288)
August 05	-0.998	-0.307	-0.307	-0.661
	(0.330)	(0.762)	(0.762)	(0.516)
August 12	-1.484	-0.360	-0.360	-0.790
	(0.153)	(0.723)	(0.723)	(0.439)
August 19	-1.739	-0.801	-0.801	-0.984
	(0.097)	(0.432)	(0.432)	(0.337)
August 26	-1.626	-0.724	-0.724	-0.381
	(0.120)	(0.477)	(0.477)	(0.707)
September 02	-0.540	0.347	0.347	0.091
	(0.595)	(0.732)	(0.732)	(0.929)
September 09	-0.064	0.459	0.459	-0.002
	(0.949)	(0.651)	(0.651)	(0.998)
September 16	0.308	0.773	0.773	-0.123
	(0.761)	(0.449)	(0.449)	(0.903)
September 23	0.096	0.234	0.234	-0.903
	(0.924)	(0.817)	(0.817)	(0.377)

Notes: This table presents the t-statistics and p-values (in parenthesis) for the MDM test. The null hypothesis is that for each week, each of the four competing forecasting models have the same predictability as the BM model. BM model is proposed by Begueria and Maneta (2020), IG State model, IG National model, IG National with Bias Adjustment Model are proposed by Irwin and Good (2017a), and BF model is proposed by Bain and Fortenbery (2017).

Table 3.4: The Modified Diebold Mariano (MDM) test statistics between the BM model and other yield forecasting models for soybean

Date	BM vs IG State	BM vs IG National	BM vs IG National with Bias Adjustment	BM vs BF
June 17	0.224	0.445	-0.210	0.626
	(0.825)	(0.661)	(0.836)	(0.538)
June 24	0.163	0.411	-0.226	0.643
	(0.872)	(0.685)	(0.824)	(0.527)
July 01	0.019	0.233	-0.186	0.450
	(0.985)	(0.818)	(0.855)	(0.657)
July 08	-0.061	0.120	-0.274	0.393
	(0.952)	(0.906)	(0.787)	(0.699)
July 15	-0.588	0.003	-0.694	0.647
	(0.563)	(0.998)	(0.496)	(0.525)
July 22	-0.544	-0.422	-2.387	0.426
	(0.593)	(0.677)	(0.027)	(0.675)
July 29	-0.391	-0.379	-0.379	0.469
	(0.700)	(0.709)	(0.709)	(0.644)
August 05	-0.357	-0.165	-0.165	0.494
	(0.725)	(0.871)	(0.871)	(0.626)
August 12	-0.253	-0.304	-0.304	0.414
	(0.803)	(0.764)	(0.764)	(0.683)
August 19	0.250	0.220	0.220	0.978
	(0.805)	(0.828)	(0.828)	(0.340)
August 26	0.613	0.708	0.708	1.318
	(0.547)	(0.487)	(0.487)	(0.202)
September 02	1.039	0.813	0.813	0.597
	(0.311)	(0.426)	(0.426)	(0.557)
September 09	0.934	0.933	0.933	0.896
	(0.362)	(0.362)	(0.362)	(0.381)
September 16	1.437	0.959	0.959	0.305
	(0.166)	(0.349)	(0.349)	(0.764)
September 23	1.159	0.911	0.911	0.559
	(0.260)	(0.373)	(0.373)	(0.582)

Notes: *, **, *** is the significant level at 10%, 5%, 1% respectively. This table presents the t-statistics and p-values (in parenthesis) for the MDM test. The null hypothesis is that for each week, each of the four competing forecasting models have the same predictability as the BM model. BM model is proposed by Begueria and Maneta (2020), IG State model, IG National model, IG National with Bias Adjustment Model are proposed by Irwin and Good (2017a), and BF model is proposed by Bain and Fortenbery (2017).

Table 3.5: Weekly best model selected by MCS test for corn and soybean from 2000 – 2020

Panel A: Corn			Panel B: Soybean		
Date	Best Model with MCS Test	MCS p-values	Date	Best Model with MCS Test	MCS p-values
June 03	IG National with Bias Adjustment	0.289			
June 10	IG National with Bias Adjustment	0.380			
June 17	IG National with Bias Adjustment	0.362	June 17	IG National with Bias Adjustment	0.262
June 24	IG National with Bias Adjustment	0.322	June 24	IG National with Bias Adjustment	0.257
July 01	IG National with Bias Adjustment	0.570	July 01	IG National with Bias Adjustment	0.444
July 08	IG National with Bias Adjustment	0.662	July 08	IG National with Bias Adjustment	0.403
July 15	BF	0.744	July 15	IG State	0.218
July 22	BF	0.318	July 22	IG National with Bias Adjustment	0.240
July 29	BF	0.540	July 29	IG Nation	0.196
August 05	IG State	0.595	August 05	IG State	0.257
August 12	IG State	0.275	August 12	IG Nation	0.519
August 19	BF	0.691	August 19	BM	0.637
August 26	IG State	0.156	August 26	BM	0.739
September 02	IG State	0.891	September 02	BM	0.799
September 09	IG State	0.814	September 09	BM	0.856
September 16	BM	0.588	September 16	BM	0.739
September 23	BF	0.727	September 23	BM	0.759

Notes: MCS p-values are all greater than the significance level of 0.1, suggesting the selected best performing model fails to significantly outperform other individual yield forecasting models. The best model selected by MCS test is based on the significance level of 0.1, p-values are produced with 2000 bootstrap replicates for the test statistics.

Table 3.6: Weekly best model selected by MCS test for corn and soybean from 2000 – 2020

Panel A: Corn			Panel B: Soybean		
Date	Best Model with MCS Test	MCS p-values	Date	Best Model with MCS Test	MCS p-values
June 03	Equal Weighted model	0.209			
June 10	Equal Weighted model	0.289			
June 17	Equal Weighted model	0.238	June 17	Equal Weighted model	0.217
June 24	Equal Weighted model	0.248	June 24	Equal Weighted model	0.245
July 01	Equal Weighted model	0.460	July 01	Equal Weighted model	0.400
July 08	Equal Weighted model	0.640	July 08	Equal Weighted model	0.404
July 15	Equal Weighted model	0.766	July 15	Equal Weighted model	0.212
July 22	Equal Weighted model	0.246	July 22	Equal Weighted model	0.251
July 29	Equal Weighted model	0.588	July 29	Equal Weighted model	0.158
August 05	Equal Weighted model	0.617	August 05	Equal Weighted model	0.218
August 12	Equal Weighted model	0.290	August 12	Equal Weighted model	0.421
August 19	Equal Weighted model	0.766	August 19	Equal Weighted model	0.613
August 26	Equal Weighted model	0.128	August 26	Equal Weighted model	0.724
September 02	Equal Weighted model	0.896	September 02	Equal Weighted model	0.774
September 09	Equal Weighted model	0.816	September 09	Equal Weighted model	0.795
September 16	Equal Weighted model	0.671	September 16	Equal Weighted model	0.663
September 23	Equal Weighted model	0.779	September 23	Equal Weighted model	0.732

Notes: Equal Weighted model produce yield forecast composites of five yield forecasting models. MCS p-values are all greater than the significance level of 0.1, suggesting Equal Weighted model fails to significantly outperform individual yield forecasting models. The best model selected by MCS test is based on the significance level of 0.1, p-values are produced with 2000 bootstrap replicates for the test statistics.

Table 3.7: The multi-horizon average superior predictive ability (aSPA) test between BM model and other yield forecasting models for corn and soybean with fixed weights

Crop	BM vs IG State	BM vs IG National	BM vs IG National with Bias Adjustment	BM vs BF
corn	-0.210	0.210	-0.745	-0.344
	(0.566)	(0.408)	(0.740)	(0.607)
soybean	0.0304	0.251	-0.285	0.659
	(0.482)	(0.397)	(0.595)	(0.220)

Notes: This table presents the t-statistics and p-values (in parenthesis) for the multi-horizon aSPA test. The null hypothesis is that considering all horizons, on average, the competing yield forecasting model has better performance than BM model. BM model is proposed by Begueria and Maneta (2020), IG State model, IG National model, IG National with Bias Adjustment Model are proposed by Irwin and Good (2017a), and BF model is proposed by Bain and Fortenbery (2017).

Table 3.8: The multi-horizon average superior predictive ability (aSPA) test between BM model and other yield forecasting models for corn and soybean with varying weights

Crop	BM vs IG State	BM vs IG National	BM vs IG National with Bias Adjustment	BM vs BF
corn	-1.062	-0.094	-0.414	-0.995
	(0.869)	(0.574)	(0.678)	(0.834)
soybean	0.836	1.079	0.531	0.943
	(0.269)	(0.135)	(0.312)	(0.170)

Notes: This table presents the t-statistics and p-values (in parenthesis) for the multi-horizon aSPA test. The null hypothesis is that considering all horizons, on average, the competing yield forecasting model has better performance than BM model. BM model is proposed by Begueria and Maneta (2020), IG State model, IG National model, IG National with Bias Adjustment Model are proposed by Irwin and Good (2017a), and BF model is proposed by Bain and Fortenbery (2017).

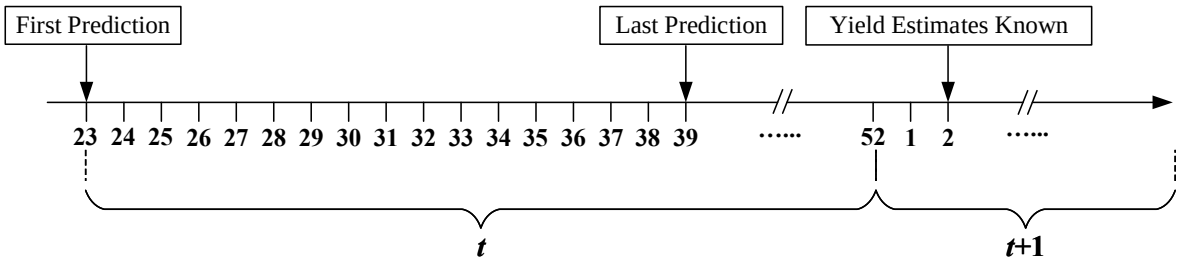


Figure 3.1: Yield forecasting cycle for corn

Notes: we use corn as an example to illustrate the forecasting cycle. For soybean, the first prediction is in week 25 and the last prediction is in week 39.

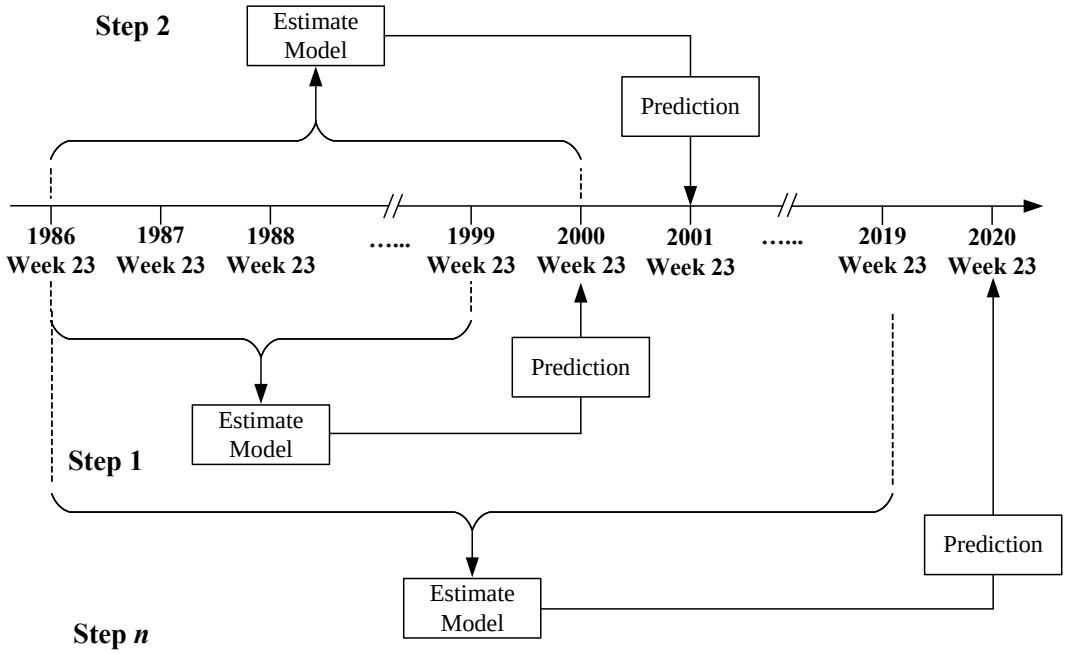


Figure 3.2: Recursive out-of-sample yield forecasts with Begueria and Maneta model (2017)

Notes: we use corn as an example to illustrate the forecasting cycle. For soybean, the first prediction is in week 25 and the last prediction is in week 39.

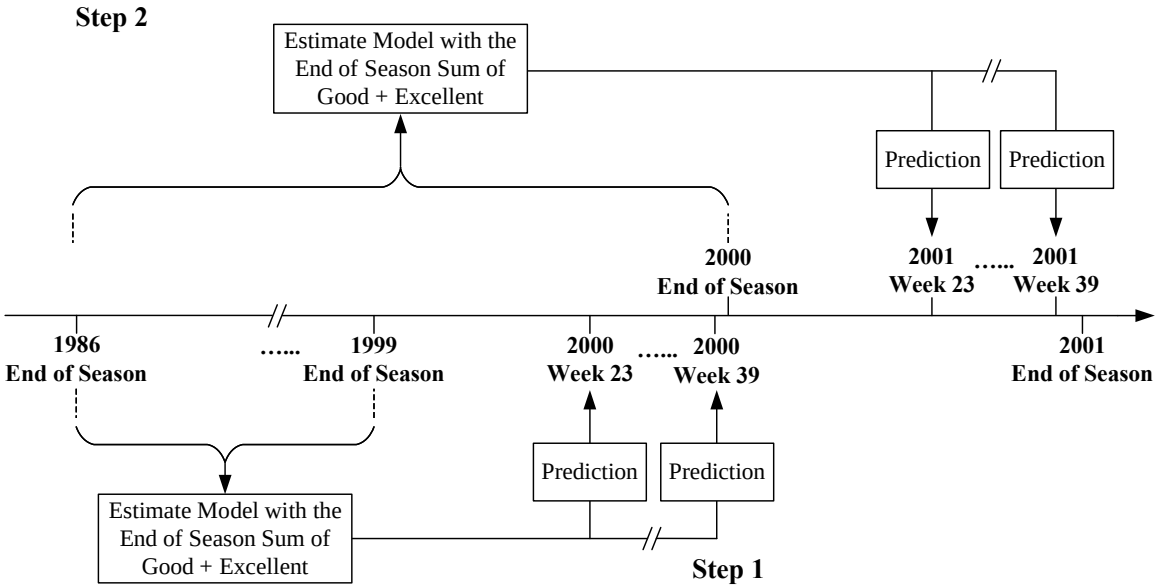
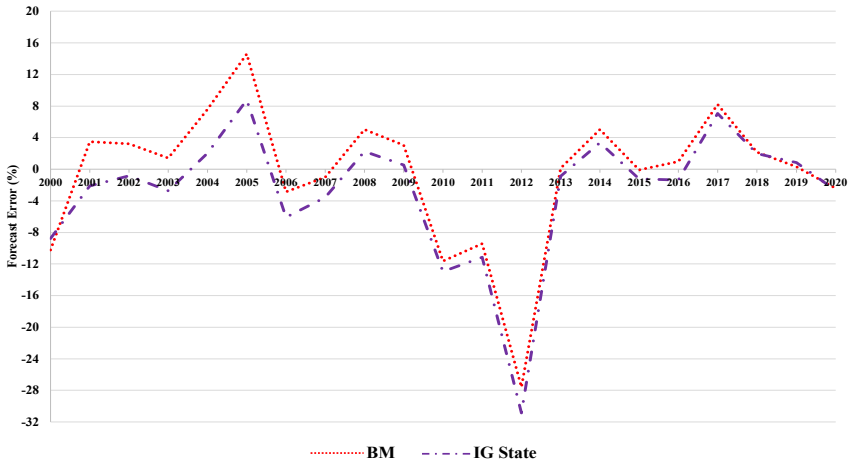
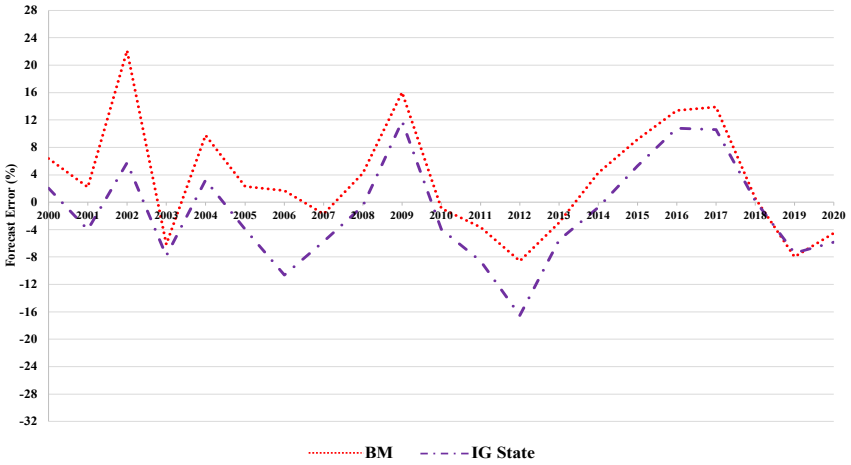


Figure 3.3: Recursive out-of-sample yield forecasts with IG State model and IG National model (2017a)

Notes: we use corn as an example to illustrate the forecasting cycle. For soybean, the first prediction is in week 25 and the last prediction is in week 39.



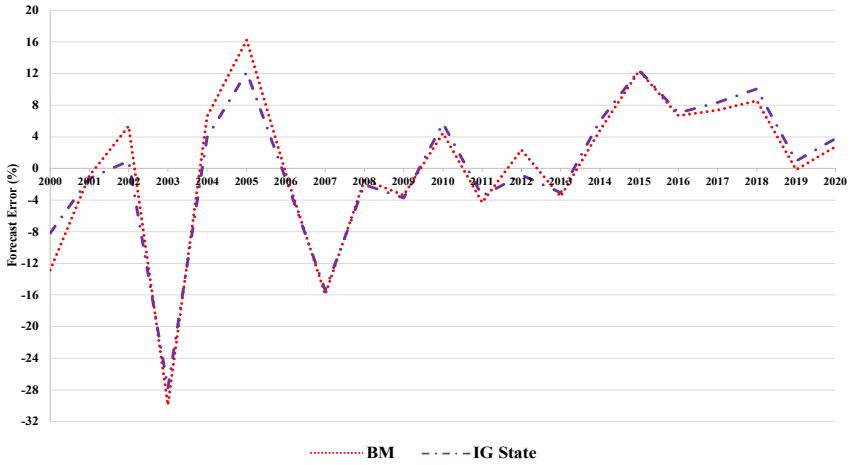
(a) Illinois Corn



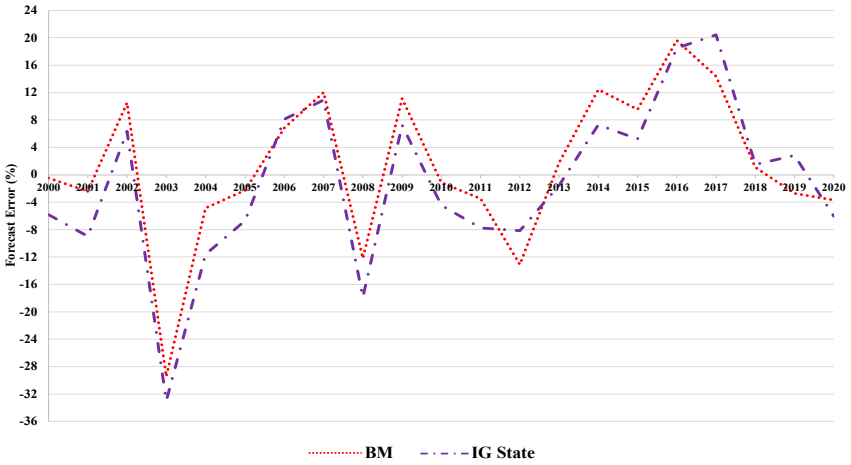
(b) South Dakota Corn

Figure 3.4: The forecast error (%) of BM model and IG State model for week 29 for corn, for Illinois and South Dakota, 2000 – 2020

Notes: BM model is proposed by Begueria and Maneta (2020), IG State model is proposed by Irwin and Good (2017a).



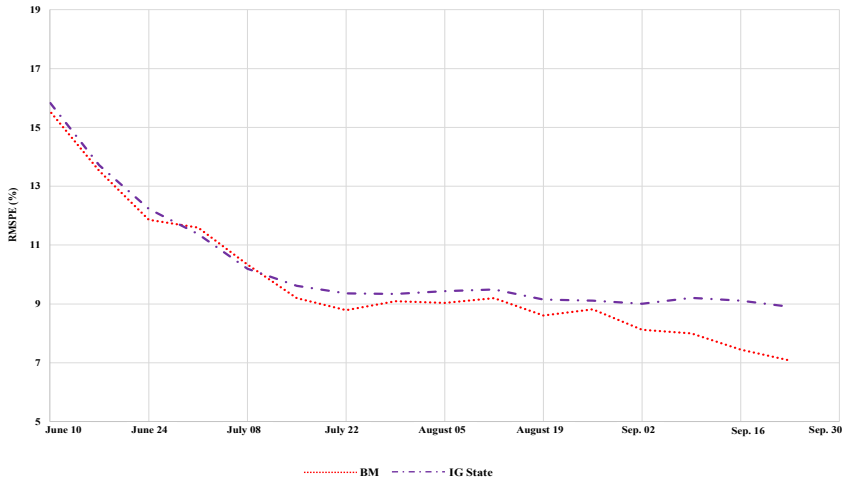
(a) Illinois Soybean



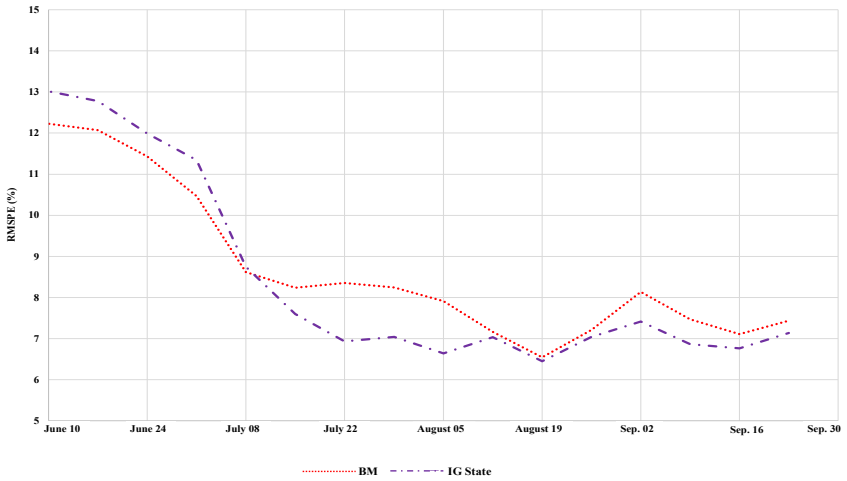
(b) South Dakota Soybean

Figure 3.5: The forecast error (%) of BM model and IG State model for week 29 for soybean, for Illinois and South Dakota, soybean, 2000 – 2020

Notes: BM model is proposed by Begueria and Maneta (2020), IG State model is proposed by Irwin and Good (2017a).



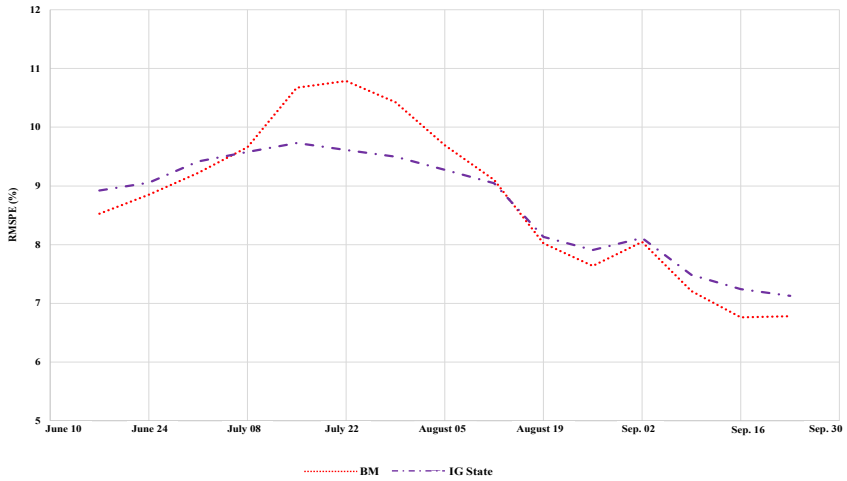
(a) Illinois Corn



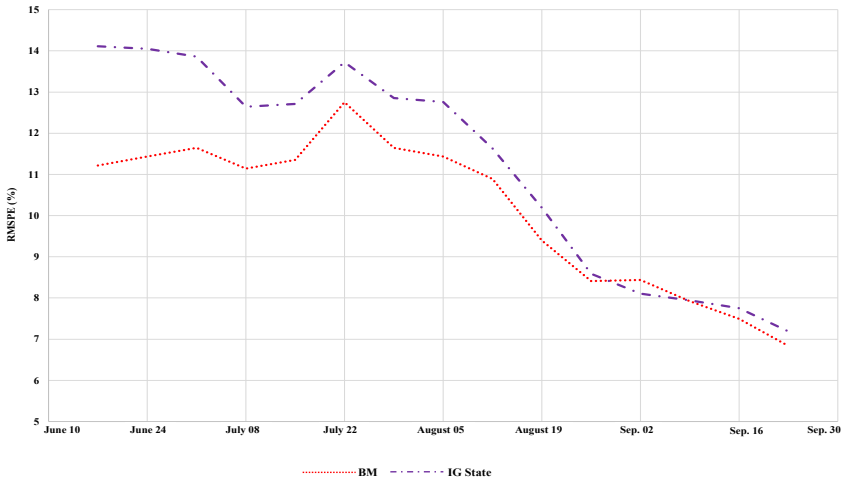
(b) South Dakota Corn

Figure 3.6: Weekly RMSPE of BM model and Irwin and Good model for Illinois and South Dakota for corn, 2000 – 2020

Notes: BM model is proposed by Begueria and Maneta (2020), IG State model is proposed by Irwin and Good (2017a).



(a) Illinois Soybean



(b) South Dakota Soybean

Figure 3.7: Weekly RMSPE of BM model and Irwin and Good model for Illinois and South Dakota for soybean, 2000 – 2020

Notes: BM model is proposed by Begueria and Maneta (2020), IG State model is proposed by Irwin and Good (2017a).

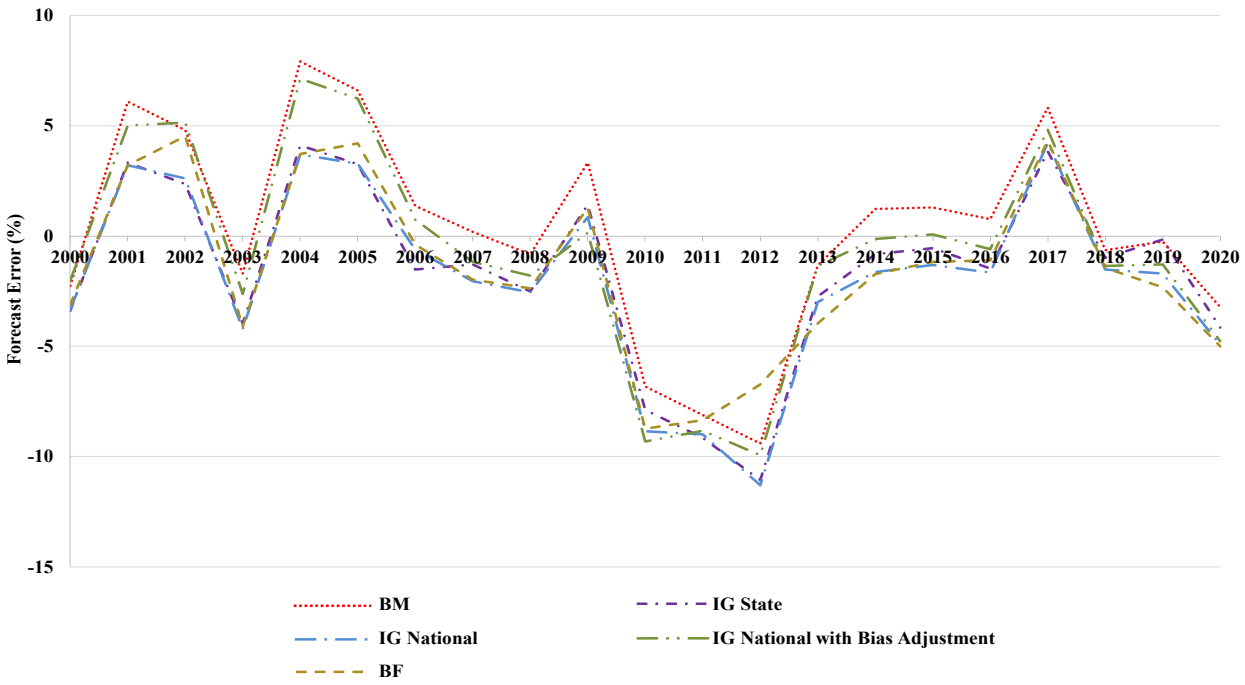


Figure 3.8: The forecast error (%) of five yield forecasting models for week 29 for corn, 2000 – 2020

Notes: BM model is proposed by Begueria and Maneta (2020), IG State model, IG National model, IG National with Bias Adjustment Model are proposed by Irwin and Good (2017a), and BF model is proposed by Bain and Fortenbery (2017).

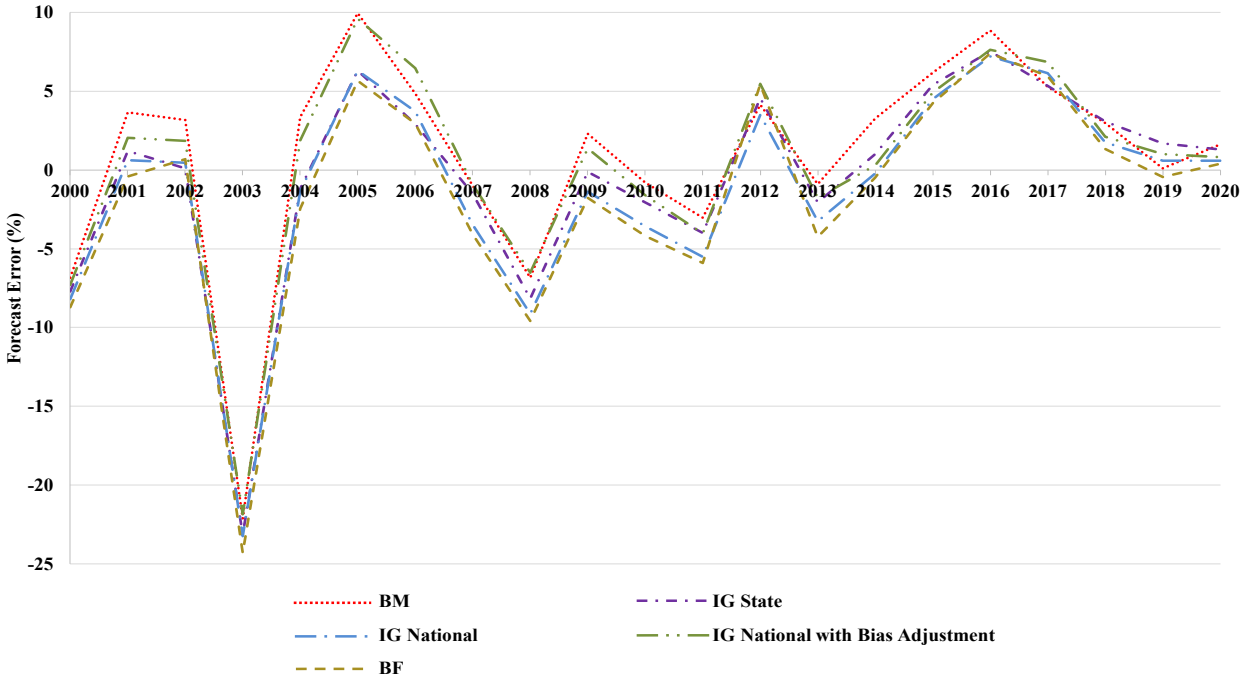


Figure 3.9: The forecast error (%) of five yield forecasting models for week 29 for soybean, 2000 – 2020

Notes: BM model is proposed by Begueria and Maneta (2020), IG State model, IG National model, IG National with Bias Adjustment Model are proposed by Irwin and Good (2017a), and BF model is proposed by Bain and Fortenbery (2017).

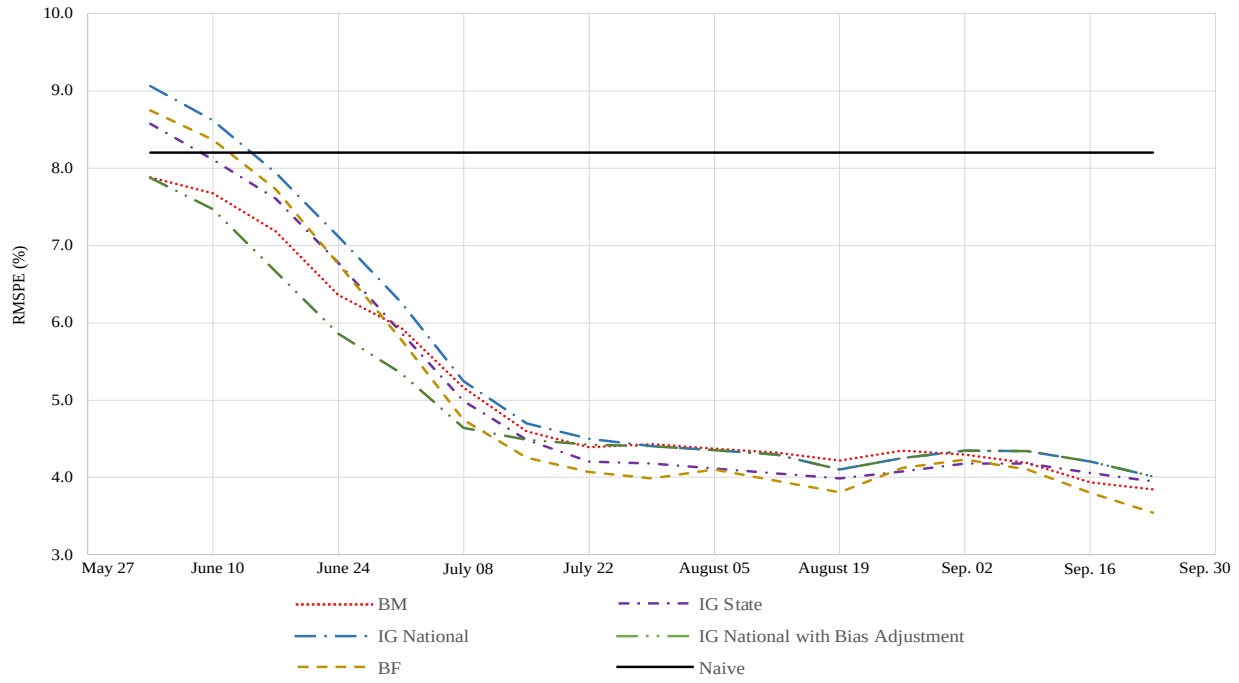


Figure 3.10: RMSPE of five yield forecasting models at national level from 2000 – 2020 for corn

Notes: we also include naïve trend yield model to illustrate the value of crop condition ratings as a yield indicator. BM model is proposed by Begueria and Maneta (2020), IG State model, IG National model, IG National with Bias Adjustment Model are proposed by Irwin and Good (2017a), and BF model is proposed by Bain and Fortenbery (2017).

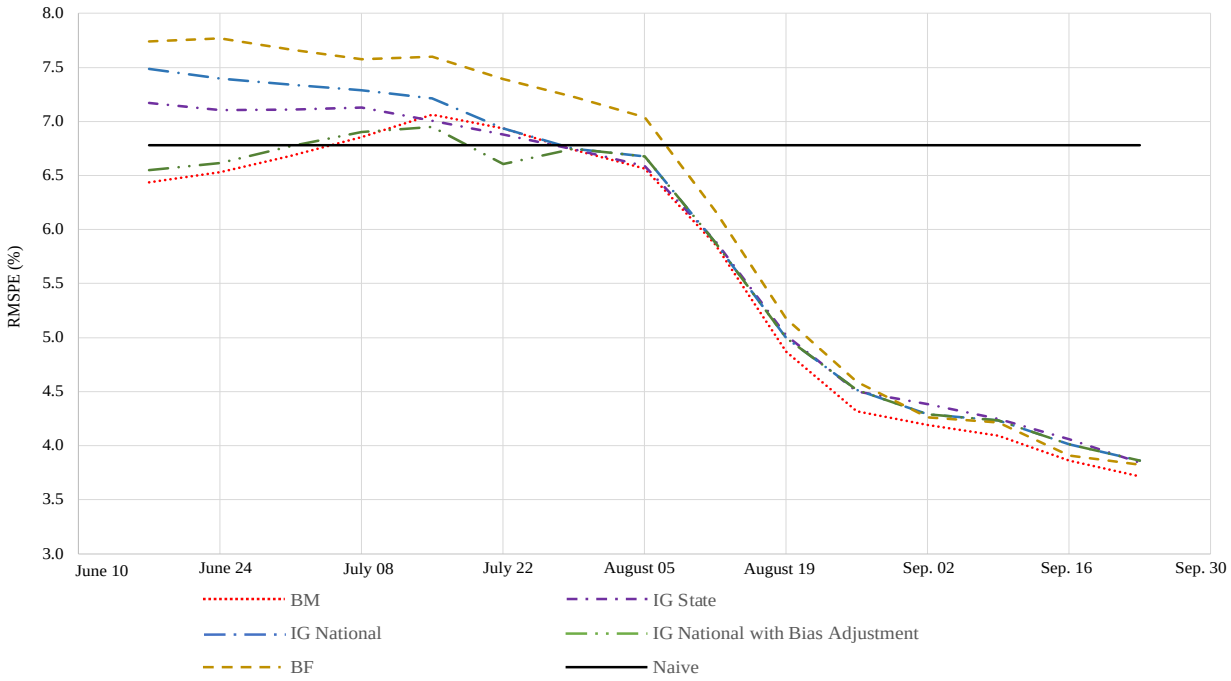


Figure 3.11: RMSPE of five yield forecasting models at national level from 2000 – 2020 for soybean

Notes: we also include naïve trend yield model to illustrate the value of crop condition ratings as a yield indicator. BM model is proposed by Begueria and Maneta (2020), IG State model, IG National model, IG National with Bias Adjustment Model are proposed by Irwin and Good (2017a), and BF model is proposed by Bain and Fortenbery (2017).

CHAPTER 4:

DO EXTREME CIT POSITION CHANGES MOVE PRICES IN GRAIN FUTURES MARKETS?

4.1. Introduction

A global controversy erupted during the 2007-08 spike in commodity prices about the role of new participants in futures markets—financial index investors. A variety of commodity investment instruments typically are lumped together under the heading "financial index investment" (Engelke and Yuen, 2008). Regardless of the form, these investments have the common goal of providing financial investors with long exposure to returns from a basket of commodity futures. The surge in financial index investment led to widespread charges that the investment wave caused irrational and gross mispricing across a wide range of commodities. This has been labeled the "Masters Hypothesis", which according to Sanders and Irwin (2017), has the following tenets: 1) financial index investors were directly responsible for driving commodity futures prices higher; 2) the deviations of futures prices from fundamental value were economically very large; and 3) the impact was pervasive across commodity futures markets. These claims have been used to justify the need for tighter regulations on speculation in commodity futures markets around the world.

Some studies find evidence in support of the Masters Hypothesis in agricultural futures markets (e.g., Mayer, 2012; Gilbert and Pfuderer, 2014; Tadesse et al., 2014). However, a much longer list of studies fails to find a significant price impact of commodity index traders (CITs). Many of these studies use linear Granger causality tests between weekly futures returns and CIT positions reported by the U.S. Commodity Futures Trading Commission (CFTC). Noteworthy examples include Stoll and Whaley (2010), Sanders and Irwin (2011), Aulerich, Irwin, and Garcia (2014), Lehecka (2015), and Hamilton and Wu (2015). Over a wide range of markets, data, and

methods, these studies find, at best, very limited evidence of a direct link between CIT positions and returns in agricultural futures markets.

Despite the weight of the evidence against the Masters Hypothesis, it continues to resonate with a number of market participants, civic organizations, and policymakers. This may reflect the fact that the impact of financial index investment in agricultural futures markets is more complicated and nuanced than can be detected by relatively simple linear Granger causality tests commonly used in prior literature. Instead of the linear causality at the mean, the relationship between index investment and futures prices may be non-linear and/or hidden in the tails of the data. As noted by Lee and Yang (2012), some statistical relationships may fail to present at the mean of the data but can show up in the tails of the distribution.

To date, only two studies in the CIT literature have used statistical tests to detect these more subtle relationships. Palazzi et al. (2020) applied non-linear Granger causality tests to CIT positions and returns in 12 agricultural futures markets, finding that the more sophisticated non-linear causality test also failed to find evidence of a significant relationship. Algieri, Kalkuhl, and Koch (2017) estimate a multinomial logit model to investigate which factors are associated with the propagation of extreme events in agricultural futures markets, and once again, do not find evidence of an impact of CIT positions. However, neither of these studies analyzed the relationship across different quantiles of the distributions. Given that the discussion on the Master Hypothesis mostly centers around episodes with significant upward price movements, there is clearly a need for additional research to investigate whether the linkage between CIT positions and prices differs under various pricing scenarios.

Our study applies a recently-developed cross-quantilogram (CQ) test to examine the impact of CIT positions on returns in four agricultural futures markets. Han et al. (2016) developed the

CQ test to thoroughly analyze the causal relationship between two series in all parts of their distributions, especially the tail quantiles. This test has several advantages, as it: i) captures the lead-lag relationships across all parts of distributions; ii) does not require moment conditions; iii) only requires the time series to be stationary; and iv) includes long lags in the model specification to avoid concerns about degrees-of-freedom. The CQ test has been applied under a variety of contexts, including the spillovers between the U.S. and Chinese agricultural futures markets (Jiang et al., 2016), the spillover of spot gold prices to U.S. stock prices (Baumöhl and Lyócsa, 2017), the quantile dependence and predictability between various energy prices (Scarcioffolo and Etienne, 2021), among others. To the best of our knowledge, the present study is the first to apply the CQ test to analyze the price impact of CIT positions in any type of commodity futures market.

The data for the study consists of weekly CIT positions and returns from January 6, 2004 through December 31, 2019 for Chicago Board of Trade (CBOT) corn, wheat, soybeans, and Kansas City Board of Trade (KCBOT) wheat. We first conduct three types of linear causality tests to provide a baseline for the relationship between CIT positions and prices movements. We fail to reject the null of no causality in most of the cases, across the different tests, measures of position pressure, or the sample period considered. Next, we apply the CQ test of directional predictability in the tails of the distributions of the CIT positions and price movements. Similar to the linear tests, we find very little evidence of a directional relationship in the extremes of the distributions. Our results add to the growing evidence that the Masters Hypothesis is not a useful description of the price impact of CITs in agricultural futures markets.

4.2. Data

4.2.1 Commodity index trader positions

The Supplemental Commitment of Traders (SCOT) report published by the CFTC provides weekly CIT positions for CBOT corn, CBOT wheat, CBOT soybeans, and KCBOT wheat. Every Friday at 3:30 p.m. Eastern time, the CFTC publishes SCOT reports in conjunction with the traditional Commitments of Traders (COT) report. CIT positions in the SCOT report are released as the number of long and short contracts held by index traders as of the previous Tuesday's market settlement. One potential issue with the CIT position data is the internal netting of positions by swap dealers that offer index products to investors. In some markets, short swap positions for certain commodity products tend to offset long swap positions associated with commodity index investments. Fortunately, previous research shows that netting of swap activity is minimal in agricultural markets, and therefore, CIT positions in the SCOT report are generally regarded as accurate measures of aggregate CIT positions (Irwin and Sanders, 2012; Sanders and Irwin, 2013).

The CIT position data are publically available starting from January 2006. Previous studies argue that using post-2006 data may lead to biased results because the buildup of CIT positions in grain futures markets was concentrated in the previous two years (Sanders and Irwin, 2011; Irwin and Sanders, 2011). The CFTC collected additional data for selected grain futures markets over 2004-2005 at the request of the U.S. Senate Permanent Subcommittee on Investigations (USS/PSI, 2009) and the additional data is used for this study. Specifically, weekly CIT positions for the four grain futures markets are available from January 6, 2004 to December 31, 2019, for a total of 853 weekly observations for each market.

We consider two widely used measures that directly reflect the "weight" of index positions in grain futures markets. To start, we compute the net long CIT position for a given market as:

$$CIT\ Net\ Long_t = CITL_t - CITS_t, \tag{4.1}$$

where $CITL_t$ and $CITS_t$ are the numbers of long and short contracts held by CITs at week t, respectively. In general, CITs hold relatively small short positions in grain futures markets, so the difference between long and net long positions is not large. The first measure of CIT pressure is the change in CIT net positions for a given market:

$$\Delta CIT\ Net\ Long_t = (CITL_t - CITS_t) - (CITL_{t-1} - CITS_{t-1}), \tag{4.2}$$

The second measure of pressure is the weekly percentage growth of CIT net long positions, defined as,

$$\%CIT\ Net\ Long_t = \frac{(CITL_t - CITS_t) - (CITL_{t-1} - CITS_{t-1})}{(CITL_{t-1} - CITS_{t-1})}, \tag{4.3}$$

Descriptive statistics for the two measures of index position pressure are presented in Table 4.1. For net long positions, distributions for all four commodities are left-skewed. They each have positive kurtosis, indicating heavy-tailed distributions. The Jarque-Bera (JB) test suggests that none of the series are normally distributed. The two index position measures both have heavy tails. Augmented Dickey-Fuller (ADF) tests results are not surprising, indicating that CIT net long positions are non-stationary, while the change in net long positions and percent growth in positions are stationary.

4.2.2 Futures prices and returns

We collect nearby futures prices and compute weekly returns (percentage change in prices) for each of the four markets. To avoid inconsistency in price series when contract rollover occurs, we always calculate returns using the same nearest-to-expiration contract. Since the CFTC compiles the data for SCOT reports as of Tuesday each week, we use Tuesday's closing price to represent the price observation for a week.

Descriptive statistics for futures prices and returns are also presented in Table 4.1. All nearby futures prices are right-skewed with heavy tails, and non-normally distributed. For return

distributions, corn and soybeans are left-skewed, and the two wheat markets are right-skewed. All returns have heavy tails and the JB test suggests none of them are normally distributed. ADF test results suggest that nearby futures prices are non-stationary while returns are stationary.

4.2.3 Sample break

As noted above, our data covers index trader positions and nearby futures prices from the beginning of 2004 to the end of 2019. Figure 4.1 plots the total notional value of CIT positions summed across the four grain markets. Notional value for a given week is computed by multiplying the CIT position in a market by the corresponding nearby futures price, and after adjusting for contract size, summing across the four markets. We also include the stages of "financialization," recently proposed by Irwin, Sanders, and Yan (2022), that overlap with the sample period for this study. The first is the growth stage of financialization from 2004 to 2011, during which we observe a rapid increase in commodity index investment. Two spikes in the notional value of CIT positions are observed during the growth stage, one in 2007-2008, and the other one in 2010-2011. These peaks are between $35 and $40 billion. The second stage is the post-financialization period from 2012 to 2019, where we observe CIT notional value decreasing steadily to around $15 billion in the last three years of the sample. If price pressure from CITs exists, it would make the most sense for it to be evident in the growth stage. In the statistical analysis that follows, we report results for the full sample and the two sub-samples based on the growth stages of financialization. This accounts for the very different structural dynamics of index investment before and after 2011.

4.3. Linear Tests

Plots of CIT positions and futures prices for the four commodities are shown in Figure 4.2. The plots confirm no contemporaneous increase in futures prices during the large build-up of index traders' positions during 2004 – 2005. Thereafter, if anything, there appears to be a negative

relationship between CIT positions and futures prices. Of course, graphical evidence like this is only suggestive. It is important to test for direct statistical links between CIT positions and prices. We begin with the standard linear Granger causality test that has been used in numerous studies in the literature on CIT positions and movements in agricultural futures prices. While these tests have been conducted numerous times in the past, we include them here to provide a benchmark using the same data for the later CQ tests.

4.3.1 Linear Granger causality tests

In the widely-used linear causality framework (Granger, 1980), a time-series regression is used to determine if one series is useful in forecasting another, or simply, "Granger causing." The specification of the test for returns and CIT pressure in grain futures markets is shown below for a given market:

$$Return_t = \alpha_t + \sum_{i=1}^{m} \gamma_i \, Return_{t-i} + \sum_{j=1}^{n} \beta_j \, \Delta Position_{t-j} + \epsilon_t, \qquad (4.4)$$

where $Return_t$ is the log-difference in nearby weekly futures prices for a given market at time *t*, and $\Delta Position_t$ is the measure of CIT pressure in the same market. All series are stationary (see Table 4.1). The null hypothesis is that all β_j are jointly zero, suggesting that CIT positions do not Granger-cause returns. Alternatively, if CIT pressure indeed drives up futures prices, then β_j will be greater than zero. The optimal lag order based on Akaike Information Criterion (AIC) is one for both returns and the growth of positions (*m=1*, *n=1*) for each of the four grain futures markets.

The results of the linear Granger causality test estimated over the full sample period and two subsample periods are presented in Table 4.2. For the full sample period (2004 – 2009), in only two out of the eight cases the null hypothesis of no Granger-causality is rejected at the 5% significance level. Both cases are in the CBOT wheat market. Note that the direction of the estimated relationship is negative, suggesting that lagged changes in CIT positions negatively

correlate with price changes, just the opposite of that implied by the Masters price pressure hypothesis. In the first subsample (2004 – 2011) or the growth stage of financialization, in all eight cases we fail to reject no Granger-causality from positions to returns at the 5% significance level. In the second subsample from 2012 to 2019, i.e., the post-financialization stage, significant directional predictability from positions to returns is once again only found for the CBOT wheat market. The estimated directional impact is negative.

4.3.2 Augmented Granger causality tests

The second set of tests in the linear Granger causality framework is the augmented test of Toda and Yamamoto (1995). This method estimates a VAR model in levels to detect the dynamic causal relationship between two processes that may be integrated or cointegrated of arbitrary order. When two time series are cointegrated or are not strictly stationary, the traditional Granger causality test may detect a spurious relationship, invalidating the results. To avoid such inconsistency, the Toda and Yamamoto (1995) test for Granger causality in a VAR model that accounts for cointegration and stationarity. The model is specified below for a given market:

$$\begin{bmatrix} Price_t \\ Position_t \end{bmatrix} = \sum_{i=1}^{p+d_{max}} \begin{bmatrix} \gamma_{1,i} & \gamma_{2,i} \\ \gamma_{3,i} & \gamma_{4,i} \end{bmatrix} \begin{bmatrix} Price_{t-i} \\ Position_{t-i} \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + t \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}, \qquad (4.5)$$

where $Price_t$ is the nearby futures price and $Position_t$ is the net long CIT position. We conduct the augmented Granger Causality test in the following steps: i) each series is tested for the order of integration using the ADF test; ii) determine the value d_{max}, which is the maximum order of integration of two series; iii) set up the VAR model and use the AIC to determine the optimal lags p for the system; iv) use the augmented lag $p + d_{max}$ to estimate the VAR system; and v) apply the Wald test to determine if the position coefficients are significantly different from zero.

Augmented GC test results are presented in Table 4.3. Note that only one set of results is presented since this test is based on the level of net long CIT positions instead of the change or percent growth in positions. We focus on the direction from index positions to futures prices as the VAR model avoids invalid estimates when two series have different integration orders. We first estimate $d_{max} = 1$ based on the ADF test. Then to determine the lag orders of the VAR model, we use AIC to find the appropriate lags with a maximum lag order of 20 lags and select two lags for the bivariate VAR model. As shown in Table 4.3, we fail to reject the null of no causality in all cases when utilizing the augmented Granger causality test.

4.3.3 Long-horizon regression tests

Both the standard and augmented Granger causality tests are designed to detect relationships between weekly CIT positions and returns. Such tests may have low power when detecting relationships over longer horizons (e.g., Summers 1986). Index trader positions may flow in "waves" that build up slowly that eventually push up prices, and then fade slowly as the process is reversed (Sanders and Irwin, 2011). In this situation, horizons longer than a week may be needed to fully capture the relationship between CIT position pressure and futures returns. We follow Sanders and Irwin (2014, 2016) and implement the long-horizon framework developed by Valkanov (2003). The model is defined for a given market as:

$$\sum_{i=0}^{k-1} Return_{t+i} = \alpha + \beta \sum_{i=0}^{k-1} \Delta Position_{t+i-1} + \epsilon_t, \quad (4.6)$$

where all variables are the same as before. To obtain the estimated coefficients, we run an OLS regression of the long-horizon dependent variable on the long-horizon independent variable. Once the horizon k is determined, the dependent variable is the sum of futures returns from t to $t + k - 1$, and the independent variable is the sum of growth/change in CIT positions from $t - 1$ to $t + k$. In essence, equation (6) is an OLS regression of a *k*-period moving sum of the dependent variable

at time *t* against a *k*-period moving sum of the independent variable in the previous period, time *t*-1. If the estimated β is positive, this indicates a fads-style model where prices tend to increase slowly over a relatively long period after widespread index fund buying. To be consistent with previous studies, we choose $k = 4$ and $k = 12$ to represent monthly and quarterly time horizons using the weekly data (Singleton, 2014; Hamilton and Wu, 2015; Sanders and Irwin, 2014, 2016).

Valkanov (2003) demonstrates that the OLS slope estimator in equation (4.6) is consistent and converges at a high rate as the sample size increases. However, this specification obviously creates an overlapping horizon problem for testing. Valkanov shows that Newey-West *t*-statistics do not converge to well-defined distributions and suggests using the re-scaled *t*-statistic, $t/\sqrt{T}$, along with simulated critical values for inference. Valkanov also demonstrates that the re-scaled *t*-statistic generally is the most powerful among several alternative long-horizon test statistics. Valkanov (2003) provides the simulated critical values for the re-scaled statistic under various scenarios. For sample size $T = 750$, and two nuisance parameters $c = 0$ and $\delta = 0$, the critical values at the 5% significance level at two tails are $(-0.672, 0.727)$.

We report the estimated OLS slope coefficients and the rescaled t-statistics for the Valkanov test at the monthly and quarterly horizons in Table 4.4. When we compare the test statistics with provided critical values, there is no single case where the rescaled t-statistics are outside the range of the critical values. Once again, estimation results from this linear test suggest no evidence that CIT positions pressure grain futures prices upward.

4.4. Cross-Quantilogram (CQ) Tests

In the previous section, we conducted three linear causality tests to provide a comprehensive baseline for the relationship between CIT positions and futures prices movements. We fail to reject the null of no causality in most cases, for different testing methods, measures of position pressure,

and the sample period considered. These findings are consistent with most prior studies that use similar linear tests (e.g., Stoll and Whaley, 2010; Sanders and Irwin, 2011; Hamilton and Wu, 2015; Lehecka, 2015).

As noted earlier, a concern with these findings is that the relationship between CIT positions and returns may be more subtle and difficult to detect than is possible with linear tests. In particular, linear tests may fail to detect a causal relationship hidden in the tails of the distribution (Lee and Yang, 2012). To address this limitation, we apply the recently developed CQ test to investigate the directional predictability from the change in CIT net long positions to futures returns in the four grain futures markets. We also apply the CQ test to examine the directional impact of futures returns to the change in CIT net long positions.

Linton and Whang (2007) introduced the quantilogram, which measures the directional predictability of a stationary time-series based on different parts of the distribution of a time-series variable. The quantilogram method provides estimates of sample lead-lag correlation of quantiles and a Box-Pierce-type statistic that aggregates the individual correlations across lags. Based on the concept of the quantilogram for a single series, Han et al. (2016) developed the cross-quantilogram (CQ) to measure the directional predictability of a pair of stationary times-series in all parts of the distributions and a Box-Ljung version of a portmanteau test for overall directional predictability. According to Han et al. (2016), the CQ method has several advantages, as it: i) captures the directional lead-lag relationships across all parts of distributions; ii) does not require moment conditions of series; iii) only requires the time series to be stationary; and iv) includes long lags in the model specification to avoid concerns about degrees-of-freedom.

Specifically, for two strictly stationary time-series variables, $x_{1,t}$ and $x_{2,t}$, we define their cumulative distribution as $F_i(\cdot)$, and their density function as $f_i(\cdot)$. Next, we define the quantile

function of each series as $q_i\,(\alpha_i) = inf(v: F_i(v) \geq \alpha_i)\,, \forall \alpha_i \in (0,1)$ for $\alpha \equiv (\alpha_1, \alpha_2)^T$. This quantile function returns the minimum quantile of x_i for the probability at α_i. The CQ for quantile α and lag k is specified as:

$$\rho_\alpha(k) = \frac{E[\psi_{\alpha_1}(x_{1,t} - q_{1,t}(\alpha_1))\psi_{\alpha_2}(x_{2,t-k} - q_{2,t-k}(\alpha_2))}{\sqrt{E[\psi^2_{\alpha_1}(x_{1,t} - q_{1,t}(\alpha_1))]}\sqrt{E[\psi^2_{\alpha_2}(x_{2,t-k} - q_{2,t-k}(\alpha_2))]}}, \tag{4.7}$$

where $\psi_{\alpha_i}(u) \equiv 1(u < 0) - \alpha_i$ is a check function that captures the direction of deviation for a given quantile; $k = 0, \pm 1, \pm 2,$ .... Inside the check function, $\{1\big[[x_{i,t} \leq q_{i,t}(\cdot)]\big]\}$ is an indicator function, also known as the quantile-hit or quantile-exceedance process in the literature, that takes on a value of one when $[x_{i,t} \leq q_{i,t}(\cdot)]$ and zero otherwise. The $\psi_{\alpha_i}(\cdot)$ function transforms the indicator observations into a sorted sequence for a given quantile level. When an observation is smaller or equal to a given quantile, $\psi_{\alpha_i}(\cdot)$ returns $1 - \alpha_i$; whereas when an observation is greater than a given quantile, $\psi_{\alpha_i}(\cdot)$ returns $-\alpha_i$. In essence, the CQ is the cross-correlation of two quantile-hit processes (Han et al., 2016).

Empirically, we have two series of interests—the change in CIT net long positions and returns. We denote these two stationary time series as $\{x_{1,t}, x_{2,t}\}_{t=1}^T$, respectively. First, we estimate the unconditional quantile functions $\widehat{q_i}(\cdot)$ for each series by solving for the following minimization functions:

$$\widehat{q_i}(\alpha_i) = \underset{v_i \in \mathbb{R}}{\operatorname{argmin}} \sum_{t=1}^T \pi_{\alpha_i}(x_{i,t} - v_i), \tag{4.8}$$

where$\pi_{\alpha_i}(u) \equiv u(\alpha_i - 1[u < 0]), i = 1, 2.$ For a set of quantiles of two series $\{\hat{q}_{1,t}(\alpha_1), \hat{q}_{2,t-k}(\alpha_2)\}$, the sample CQ is defined:

$$\hat{\rho}_\alpha(k) = \frac{\sum_{t=k+1}^T \psi_{\alpha_1}(x_{1,t} - \hat{q}_{1,t}(\alpha_1))\psi_{\alpha_2}(x_{2,t-k} - \hat{q}_{2,t-k}(\alpha_2))}{\sqrt{\sum_{t=k+1}^T \psi^2_{\alpha_1}(x_{1,t} - \hat{q}_{1,t}(\alpha_1))}\sqrt{\sum_{t=k+1}^T \psi^2_{\alpha_2}(x_{2,t-k} - \hat{q}_{2,t-k}(\alpha_2))}}, \tag{4.9}$$

where $k = 0, \pm 1, \pm 2, ...$The CQ estimates, $\hat{\rho}_\alpha(k)$, capture the directional predictability between two series at a given quantile set $\{\alpha_1, \alpha_2\}$. Further, $\hat{\rho}_\alpha(k) \in [-1,1]$. For example, when $\hat{\rho}_\alpha(1) = 0$, this indicates that when the change in CIT net long positions at time t is above or below the quantile $\hat{q}_{2,t-1}(\alpha_2)$ there is no correlation with returns at time $t+1$ being above or below the quantile $\hat{q}_{1,t}(\alpha_1)$. When $\hat{\rho}_\alpha(1) > 0$, it suggests there is directional predictability between the change in CIT net long positions at time t and returns at time $t+1$, given the two series hit in the quantiles of α_1 and α_2.

An example for corn over the full sample period is presented in Figure 3.3 to help illustrate how CQ statistics are computed. This plot shows an example for a pair of observations that both hit the quantile with $\alpha_1 = \alpha_2 = 0.1$. On September 27, 2011, the change in corn CIT net long positions is -15,920 contracts and hits in the 0.1 quantile for position changes. One week later on October 4, 2011 we observe a corn return of -10.41% and it hits in the 0.1 quantile for returns as well. The arrow shows that when changes in CIT net long positions are below the 0.1 quantile, it is followed by a return one week later that is also below its 0.1 quantile. This type of comparison is repeated for all observations to compute a CQ statistic for $\alpha_1 = \alpha_2 = 0.1$.

To test for the directional predictability of two series in different quantiles up to k lags, we follow the quantile version of the portmanteau statistical test proposed by Han et al. (2016). To test if there is overall directional predictability from $x_{2,t-k}$ to $x_{1,t}$, for $k \in \{1,2,\dots,p\}$, the null hypothesis is $H_0: \rho_\alpha(1) = \rho_\alpha(2) = \cdots = \rho_\alpha(p) = 0$, against the alternative hypothesis $H_a: \rho_\alpha(k) \neq 0$. The test statistics is:

$$\hat{Q}_a^{(p)} = \frac{T(T+2)\sum_{k=1}^{p} \hat{\rho}_\alpha^2(k)}{T-k}, \tag{4.10}$$

where $\hat{Q}_a^{(p)}$ is the portmanteau test statistic for overall directional predictability. The corresponding critical values for the portmanteau test (Han et al., 2016) are derived from the stationary bootstrap

of Politis and Romano (1994). The stationary bootstrap is a block bootstrap procedure, and the length of each block is randomly determined. The strength of the block bootstrap is that it can reach a high convergence rate using nonparametric estimation to find critical values regardless of the distribution (Han et al., 2016).

The cross-quantilograms (CQ) for the full sample period are presented in Figures 4.4 through 4.7 for CBOT corn, CBOT soybeans, CBOT wheat, and KCBOT wheat, respectfully.[6] We consider four quantiles for both returns and CIT positions: 0.10, 0.25, 0.75, 0.90, resulting in 16 pairs of CQ results for each commodity. These four quantiles represent extreme large decreases, large decreases, large increases, and extreme large increases for the two series. Within each figure, there are 16 subplots that visualize how returns in extreme quantiles respond to the dynamics of lagged extreme changes in CIT net long quantiles. These plots are organized in four panels: (a)–(d), where each panel presents the estimated sample CQ estimates from one of the four quantiles of position changes to all four extreme levels of returns.

Consider Figure 4.4(a) as an example. Here, the four CQ estimates for the lagged changes in CIT net long positions at the extreme low quantile ($\alpha_2 = 0.1$) and returns at the extreme low ($\alpha_1 = 0.10$), low ($\alpha_1 = 0.25$), high ($\alpha_1 = 0.75$), and extreme high ($\alpha_1 = 0.90$) quantiles for corn over the full sample period are presented. The black bar is the estimated sample CQ statistic at lag k, i.e., $\hat{\rho}_{\alpha}(k)$. The null hypothesis is that at lag k there is no predictability from the large negative movements in CIT position changes to large movements in futures returns. The red-dashed lines represent the 95% bootstrapped confidence intervals for no directional predictability with 1,000

[6] To save space, we only discuss the CQ estimates for the full sample period and the change in CIT net long positions in the paper. Results for all other tests, including two subsample periods are presented in the Appendix B. These results do not differ materially from the full sample results presented here.

bootstrapped replicates. We include 13 lags as this is approximately the same quarterly horizon we used in the long-horizon linear tests in the previous section.

Caution is needed when interpreting the sign of the CQ estimates. For CIT position changes in the two low quantiles ($\alpha_2 = 0.1$ or 0.25) and returns in two low quantiles ($\alpha_1 = 0.1$ or 0.25), which corresponds to the top row of plots in Figures 4.4-4.7, a positive CQ estimate suggests that a large drop in CIT net long positions are likely to predict large decreases in futures prices; on the other hand, when the sign is negative, a large drop in CIT net positions are less likely to predict a subsequent large decrease in futures prices. Meanwhile, for CIT position changes in the two low quantiles and returns in two high quantiles ($\alpha_1 = 0.75$ or 0.9), a positive CQ estimate (as plotted in the second row of Figures 4.4-4.7) suggests when a large drop in CIT net positions occurs, the likelihood of predicting a large increase in futures prices is low; whereas when CQ estimate is negative, the likelihood of predicting a large price increase is high.

CQ test statistic is mostly non-significant in panels (a) and (b) of Figures 4.4-4.7. This suggests that over a 13-week horizon, whether CIT position changes are smaller or greater than the 0.1 or 0.25 quantiles cannot predict returns located in either the left (quantiles 0.1 and 0.25) or right tails (quantiles 0.75 and 0.9) of the distribution. For the few cases where the CQ estimates are significant, empirical evidence for different commodity markets is mixed. For example, during the full sample period in the soybean market, we observe that large decreases in CIT net long positions positively predict large decreases in returns. By contrast, in the CBOT wheat market we observe that large CIT net long decreases are followed by large increases in returns.

Panels (c) and (d) in Figures 4.4-4.7 plot the CQ estimates when the lagged CIT positions are in the two high quantiles (0.75 and 0.9). For returns located in the two low quantiles, i.e., $\alpha_1 = 0.1$ or 0.25, a positive CQ estimate suggests that a large increase in CIT net positions is less likely

to predict a large drop in futures prices, whereas a negative estimate suggests that a large increase in CIT net long positions is more likely to be followed by large decreases in futures prices. For returns in two high quantiles ($\alpha_1 = 0.75$, $\alpha_1 = 0.9$), when CQ estimates are positive, it indicates that when CIT net long positions exceed high quantiles, they are likely to predict returns located in high quantiles. Meanwhile, a negative CQ estimate suggests that a large increase in CIT net long positions is less likely to predict returns with a large increase. For most cases when CIT positions experience a substantial increase, there is no significant directional predictability from the change in CIT net long positions to returns. Taken together, Figures 4.4-4.7 suggest that there are no systematic lead-lag relationships from CIT positions to futures prices when both series are in extreme quantiles.

The portmanteau test statistics for directional predictability from changes in CIT net long positions to returns are presented in Table 4.5, covering the full sample period and two subsample periods. As noted earlier, the portmanteau test is an omnibus test that aggregates the CQ test statistics from 1 to 13 lags for each pair of quantiles of the two series. During the full sample period, only one case out of 64 has a significant relationship from positions to returns. For the first and second subsamples, there are six and two cases, respectively, out of 64 that fail to reject the null hypothesis of no directional predictability. In total, there are only 9 cases out of 192 (or 4.7%) with a significant portmanteau statistic, slightly less than the number of significant test statistics one would expect at random for a 5% significance level.

For the cases with a significant portmanteau statistic in Table 4.5, the dominant sign of the underlying CQ estimates over the 1-13 lags is reported in parentheses. Dominance is defined as the sign that appeared more frequently for the 13 estimates. We do this to aid in interpreting these few cases with overall significance. With one exception, the dominant sign is consistent with CIT

position changes having the price pressure impact expected under the Masters Hypothesis. For example, during the growth stage of financialization (panel B), the portmanteau statistic for soybeans is significant for $\alpha_1 = \alpha_2 = 0.1$ and the dominant sign is positive. This implies that large decreases in CIT positions directionally predict large decreases in soybean returns. The one exception is CBOT wheat futures during the post-financialization period (panel C) and $\alpha_1 = 0.1, \alpha_2 = 0.75$. Here, the dominant sign is negative, implying that large decreases in CIT positions tend to directionally predict large increases in wheat returns. It is important to emphasize that the number of significant cases, regardless of the dominant sign, is basically what one would expect based on random chance.

As the final part of the analysis, we examine the direction of predictability of futures returns to CIT positions. There is a documented tendency for large non-commercial speculators in agricultural futures markets to be trend-followers (Sanders, Irwin, and Merrin, 2009). That is, positions of large speculators in agricultural futures markets tend to increase after futures prices increase, and *vice versa*. The available evidence for CITs is not as strong. For instance, Auerlich, Irwin, and Garcia (2014) find a significant but small impact of past returns on daily CIT positions in 12 agricultural markets, but this disappears when the analysis is limited to roll windows. Lehecka (2015) analyzes weekly CIT positions in the same 12 agricultural futures markets and reports that past returns do not significantly impact CIT positions.

The portmanteau test statistics for directional predictability from returns to CIT net position changes are presented in Table 4.6 for the full and two subsample periods.[7] During the full sample period, only two cases out of 64 have a significant relationship from positions to returns. For the first and second subsample periods, there are five and three cases, respectively, out of 64 that fail

[7] Results for the percentage changes in CIT net long positions are included in the Appendix B.

to reject the null hypothesis of no directional predictability. In total, there are only 10 cases out of 192 with a significant portmanteau statistic. Mirroring the results for positions leading returns, this is only 5.2% of the total cases, slightly greater than the number of significant test statistics one would expect at random for a 5% significance level. Furthermore, in the 10 significant cases, only five show evidence that CIT net long positions have a large decrease following a drop in futures prices. The above results provide scant support for the idea that extreme CIT position changes have a trend-following component. This is actually not all that surprising given that financial index investment is motivated by long-term investment objectives rather than short-term trading (e.g., Stoll and Whaley, 2010).

4.5. Conclusions

The price impact of financial index investment in agricultural futures markets continues to be a concern to many market participants, civic organizations, and policymakers. The concern that waves of financial index investment have led to irrational and gross mispricing in agricultural futures markets has been labeled the "Masters Hypothesis." While the bulk of the evidence suggests this hypothesis is not well-founded, it's also possible that the impact of index investment in agricultural futures markets is more complicated and nuanced than can be detected by relatively simple linear causality tests used in many studies. The relationship between index investment and futures prices may be non-linear and/or hidden in the tails of the data. The purpose of this study was to use the cross-quantilogram (CQ) test recently developed by Han et al. (2016) to examine whether predictability exists between the change in commodity index trader (CIT) positions and returns in the tails of the distributions for four agricultural futures markets. In addition to making no assumptions of the distributions of the data, the CQ test is able to determine if there is a causal

relationship between two series in all parts of the distributions of the series, especially the tail quantiles.

The data for the study consists of weekly CIT positions and returns from January 6, 2004 through December 31, 2019 for Chicago Board of Trade (CBOT) corn, CBOT wheat, CBOT soybeans, and Kansas City Board of Trade (KCBOT) wheat. We first conduct three types of linear causality tests to provide a comprehensive baseline for the relationship between CIT positions and agricultural futures prices movements. The null of no causality was not rejected in the majority of the cases across the different tests, measures of position pressure, or the sample period considered. Next, we apply the CQ test to the same data to determine if there is a relationship between the tails of the distributions of index positions and price movements. Consistent with the standard linear causality tests, we find no evidence of a relationship between index positions and returns.

Commodity markets continue to attract investors who seek to diversify their portfolios and hedge against inflation. Given the increasing complexity of global commodity markets, concerns remain on the role that different types of traders play in shaping commodity prices. The present paper adds to the growing evidence that the Masters Hypothesis is not a useful description of the price impact of CITs in agricultural futures markets, even when prices underwent extreme movements. Future studies may wish to examine other types of traders on both the long- and short-term pricing of commodity markets.

4.6. Tables and Figures

Table 4.1: Summary statistics for weekly commodity index traders (CIT) positions and nearby futures prices in four grain futures markets, January 6, 2004 to December 31, 2019

Commodity (units)	obs	Min	Max	Mean	Std. Dev	Skewness	Kurtosis	JB Test	ADF Test
Panel A: CIT Net Long Positions (number of contracts)									
CBOT Corn	835	64646	503937	332391	85529	-0.822	3.561	105.074***	-2.700
CBOT Soybean	835	27101	201251	128727	36529	-0.848	3.804	122.643***	-3.036
CBOT Wheat	835	33696	229565	149459	42852	-0.258	2.564	15.885***	-3.093
KCBOT Wheat	835	12055	66592	37162	12303	-0.242	2.187	31.173***	-3.413**
Panel B: Change in CIT Net Long Positions (number of contracts)									
CBOT Corn	834	-44788	60317	213	9195	0.291	8.569	1089.39***	-12.535***
CBOT Soybean	834	-23250	27251	138	4520	-0.218	9.125	1310.35***	-13.138***
CBOT Wheat	834	-33227	15010	85	3862	-0.660	10.635	2086.52***	-13.451***
KCBOT Wheat	834	-6400	14342	45	1641	0.812	12.361	3136.5***	-14.525***
Panel C: Percent Change in CIT Net Long Positions (%)									
CBOT Corn	834	-14.007	21.958	0.159	3.052	0.516	9.622	1560.83***	-12.807***
CBOT Soybean	834	-20.146	23.204	0.197	3.697	0.342	10.090	1762.9***	-12.903***
CBOT Wheat	834	-20.405	14.166	0.136	2.811	-0.206	9.132	1312.65***	-13.884***
KCBOT Wheat	834	-19.574	26.412	0.165	4.223	0.473	8.231	981.879***	-14.155***
Panel D: Price ($/bushel)									
CBOT Corn	835	1.863	8.313	4.154	1.459	0.856	3.061	101.998***	-1.969
CBOT Soybean	835	5.035	17.683	10.157	2.753	0.237	2.447	18.468***	-2.100
CBOT Wheat	835	2.898	12.230	5.468	1.616	0.822	3.548	104.553***	-2.627
KCBOT Wheat	835	3.170	12.610	5.729	1.764	0.819	3.087	93.69***	-2.409
Panel E: Return (%)									
CBOT Corn	835	-16.493	18.410	-0.151	3.954	-0.002	5.183	165.606***	-13.441***
CBOT Soybean	835	-15.668	11.337	0.064	3.365	-0.233	4.128	51.802***	-14.239***
CBOT Wheat	835	-17.625	16.837	-0.225	4.330	0.204	4.048	43.955***	-14.166***
KCBOT Wheat	835	-16.373	16.215	-0.169	4.131	0.126	3.782	23.448***	-14.393***

Notes: * indicates statistical significance at 5%. Skewness measures the symmetry of a series' distribution; when it is negative (positive), it indicates the distribution is skewed to the left (right). Kurtosis measures the tail shape of the distribution; when it is negative (positive), it indicates a thin (heavy)-tailed distribution. Jarque-Bera (JB) test is a "goodness of fit" test with the null hypothesis that a series follows a normal distribution. The null of the Augmented Dickey-Fuller (ADF) test is that a series has a unit root.

Table 4.2: Granger causality test results for weekly commodity index traders (CIT) positions and nearby futures prices in four grain futures markets, January 6, 2004 to December 31, 2019

Commodity	F-statistic		
	Full Sample	2004-2011	2011-2019
Panel A: dependent variable: returns, independent variable: growth in CIT net long positions			
CBOT corn	1.976	3.471	0.017
	(0.160)	(0.063)	(0.895)
CBOT soybean	0.214	0.942	0.020
	(0.644)	(0.332)	(0.888)
CBOT wheat	5.366*	2.044	3.931*
	(0.021)	(0.154)	(0.048)
KCBOT wheat	0.235	0.453	0.061
	(0.628)	(0.501)	(0.805)
Panel B: dependent variable: returns, independent variable: percentage growth in CIT net long positions			
CBOT corn	1.004	1.977	0.003
	(0.317)	(0.160)	(0.959)
CBOT soybean	0.042	0.106	0.000
	(0.837)	(0.745)	(0.993)
CBOT wheat	6.201*	2.572	4.365*
	(0.013)	(0.110)	(0.037)
KCBOT wheat	0.121	0.116	0.034
	(0.728)	(0.734)	(0.854)

Notes: * indicates statistical significance at 5%. F-test statistics are reported in the table, with the corresponding p-values in the parenthesis below. The full sample period consists of 834 weekly observations. For the growth stage of financialization, there are 416 weekly observations from January 13 to December 27, 2011. The post-financialization period runs from January 3, 2012 to December 31, 2019, resulting in weekly observations. The null hypothesis is that there is no Granger causality from CIT position changes to futures returns. The estimated coefficients associated with the position variable are negative for cases with significant test statistics.

Table 4.3: Augmented Granger causality test results for weekly commodity index traders (CIT) positions and nearby futures prices in four grain futures markets, January 6, 2004 to December 31, 2019

Commodity	Wald statistic		
	Full Sample	2004-2011	2011-2019
Dependent variable: CIT Net Long Positions, independent variable: price			
CBOT corn	1.758	4.590	0.649
	(0.415)	(0.101)	(0.723)
CBOT soybean	0.159	1.954	1.163
	(0.923)	(0.376)	(0.559)
CBOT wheat	2.229	1.992	5.838
	(0.328)	(0.369)	(0.054)
KCBOT wheat	3.356	1.436	3.403
	(0.187)	(0.488)	(0.182)

Notes: * indicates statistical significance at 5%. Wald test statistics are reported in the table, with the corresponding p-values are in the parenthesis below. The full sample period consists of 834 weekly observations. For the growth stage of financialization, there are 416 weekly observations from January 13 to December 27, 2011. The post-financialization period runs from January 3, 2012 to December 31, 2019, resulting in weekly observations. The null hypothesis is that there is no Granger causality from CIT position changes to futures returns.

Table 4.4: Long-horizon regression tests for weekly commodity index traders (CIT) positions and nearby futures prices in four grain futures markets, January 6, 2004 to December 31, 2019

Commodity	Full Sample		2004-2011		2011-2019	
Horizon (k)	Slope	Rescaled t-statistic	Slope	Rescaled t-statistic	Slope	Rescaled t-statistic
Panel A: dependent variable: returns, independent variable: CIT Net Long Change						
CBOT corn						
Monthly (k=4)	0.0000289	0.0825	0.0000337	0.0557	0.0000237	0.0615
Quarterly (k=12)	0.0000498	0.156	0.0000660	0.129	0.0000248	0.0653
CBOT soybean						
Monthly (k=4)	0.000150	0.252	0.000366	0.295	0.0000613	0.107
Quarterly (k=12)	0.000238	0.440	0.000507	0.520	0.0000928	0.170
CBOT wheat						
Monthly (k=4)	0.0000106	0.0110	-0.0000346	-0.0218	0.0000547	0.0473
Quarterly (k=12)	0.0000333	0.0400	-0.0000395	-0.0308	0.000135	0.118
KCBOT wheat						
Monthly (k=4)	0.000476	0.222	0.000622	0.126	0.000426	0.199
Quarterly (k=12)	0.000468	0.204	0.000452	0.0962	0.000452	0.201
Panel B: dependent variable: returns, independent variable: CIT Net Long Pct Change						
CBOT corn						
Monthly (k=4)	0.000282	0.0988	0.000614	0.0673	0.000175	0.0720
Quarterly (k=12)	0.000238	0.178	0.000595	0.139	0.0000486	0.0932
CBOT soybean						
Monthly (k=4)	0.000232	0.248	0.000629	0.257	0.000105	0.109
Quarterly (k=12)	0.000143	0.447	0.000753	0.459	-0.000176	0.168
CBOT wheat						
Monthly (k=4)	0.000487	0.0183	0.000907	-0.0214	0.000354	0.0556
Quarterly (k=12)	0.000485	0.0640	0.000842	-0.0239	0.000303	0.161
KCBOT wheat						
Monthly (k=4)	0.000476	0.229	0.000622	0.129	0.000426	0.200
Quarterly (k=12)	0.000468	0.252	0.000452	0.143	0.000452	0.215

Notes: * indicates statistical significance at 5%. Critical values for the rescaled t-statistics shown in the table (-0.672, 0.727) are available in Valkanov (2003) table 4.4 for case 2, $c = 0$, $\delta = 0$, $T = 750$. The full sample period consists of 834 weekly observations. For the growth stage of financialization, there are 416 weekly observations from January 13 to December 27, 2011. The post-financialization period runs from January 3, 2012 to December 31, 2019, resulting in weekly observations. The null hypothesis is that there is no Granger causality from CIT position changes to futures returns.

Table 4.5: Cross-quantilogram portmanteau test results for weekly commodity index traders (CIT) positions and nearby futures prices in four grain futures markets, positions leading returns, January 6, 2004 to December 31, 2019

Dependent variable: returns, independent variable: CIT Net Long Change									
	Returns Quantile Level					Returns Quantile Level			
CIT Net Long Change Quantile Level	0.1	0.25	0.75	0.9	CIT Net Long Change Quantile Level	0.1	0.25	0.75	0.9
Full sample: 2004 - 2019									
Panel A: CBOT Corn					Panel B: CBOT Soybean				
0.1	14.696	8.382	16.902	13.818	0.1	16.027	21.238	20.012	17.588
0.25	6.558	13.742	8.83	16.982	0.25	16.350	21.836	17.382	10.200
0.75	9.044	10.408	15.002	19.715	0.75	12.115	13.051	12.311	10.016
0.9	14.793	9.388	11.803	7.915	0.9	17.077	15.167	23.623	6.825
Panel C: CBOT Wheat					Panel D: KCBOT Wheat				
0.1	26.306	20.02	22.123	31.059	0.1	12.627	8.782	22.295	14.923
0.25	19.723	14.367	31.918*(-)	26.411	0.25	10.151	9.578	5.167	4.548
0.75	20.299	7.924	19.906	11.742	0.75	15.508	8.598	11.476	5.327
0.9	24.085	8.181	27.142	7.944	0.9	31.224	8.386	13.363	9.604
Growth stage of financialization: 2004 - 2011									
Panel A: CBOT Corn					Panel B: CBOT Soybean				
0.1	33.952	21.98	16.405	21.96	0.1	61.417*(+)	36.780*(+)	13.836	28.311
0.25	11.059	17.225	22.603	18.313	0.25	31.591	27.531	14.437	22.619
0.75	14.251	7.82	17.027	36.734*(+)	0.75	16.737	23.009	9.837	18.608
0.9	19.438	14.157	11.508	14.341	0.9	12.425	17.715	12.694	5.516
Panel C: CBOT Wheat					Panel D: KCBOT Wheat				
0.1	24.058	33.127*(+)	18.146	24.211	0.1	11.996	9.92	23.353	20.377
0.25	35.379	38.026*(+)	41.714*(-)	31.944	0.25	12.105	17.746	8.849	9.176
0.75	12.972	11.383	10.239	12.092	0.75	12.846	14.904	11.813	7.952
0.9	21.654	10.602	25.324	7.375	0.9	13.178	9.455	11.749	10.329
Post-financialization stage: 2012 - 2019									
Panel A: CBOT Corn					Panel B: CBOT Soybean				
0.1	16.443	14.257	17.503	11.078	0.1	18.461	21.019	8.895	9.584
0.25	13.724	16.065	7.057	4.225	0.25	15.887	16.915	16.198	18.809
0.75	12.124	15.192	14.21	16.211	0.75	12.85	6.735	6.562	7.078
0.9	7.149	11.058	15.096	12.283	0.9	9.699	11.842	17.141	10.333
Panel C: CBOT Wheat					Panel D: KCBOT Wheat				
0.1	19.802	34.054*(-)	26.624	29.637	0.1	15.054	15.226	26.481	34.934
0.25	11.049	18.288	30.470*(-)	12.305	0.25	14.334	11.551	21.385	19.391
0.75	10.592	12.088	14.66	15.66	0.75	31.934	7.347	9.162	12.869
0.9	11.031	9.163	12.301	19.458	0.9	21.349	10.567	9.73	19.403

Notes: * indicates statistical significance at 5%. Box-Ljung test statistics for 13 lags are in the table. The sign (+/-) next to the test statistics indicates the dominant sign of the underlying CQ estimates for the Box-Ljung test statistics.

Table 4.6: Cross-quantilogram portmanteau test results for weekly commodity index traders (CIT) positions and nearby futures prices in four grain futures markets, returns leading positions, January 6, 2004 to December 31, 2019

Dependent variable: CIT Net Long Change, independent variable: returns									
	CIT Net Long Change Quantile Level					CIT Net Long Change Quantile Level			
Returns Quantile Level	0.1	0.25	0.75	0.9	Returns Quantile Level	0.1	0.25	0.75	0.9
Full sample: 2004 - 2019									
Panel A: CBOT Corn					Panel B: CBOT Soybean				
0.1	10.965	15.793	17.586	30.847 *(+)	0.1	19.468	16.947	24.280 *(+)	9.061
0.25	18.625	17.175	17.414	22.69	0.25	13.381	18.636	18.341	11.876
0.75	16.102	18.174	7.292	7.666	0.75	17.855	13.487	21.334	12.803
0.9	15.922	17.321	12.7	11.683	0.9	14.016	8.316	8.761	8.494
Panel C: CBOT Wheat					Panel D: KCBOT Wheat				
0.1	28.253	17.624	12.26	9.607	0.1	15.269	12.492	20.846	15.167
0.25	11.002	22.932	17.1	17.531	0.25	15.087	23.241	15.497	14.437
0.75	24.846	13.878	13.715	10.709	0.75	17.181	17.759	15.269	7.543
0.9	31.025	18.937	8.128	20.215	0.9	14.654	14.634	17.523	8.096
Growth stage of financialization: 2004 - 2011									
Panel A: CBOT Corn					Panel B: CBOT Soybean				
0.1	9.129	8.654	13.778	23.934 *(+)	0.1	57.767 *(+)	49.56 *(+)	18.089	4.943
0.25	21.342	9.917	14.718	19.629	0.25	28.917	26.240	23.573	15.417
0.75	11.98	14.241	7.379	8.12	0.75	32.921 *(-)	17.368	15.897	7.270
0.9	16.692	6.971	18.737	13.692	0.9	20.317	7.457	8.127	12.885
Panel C: CBOT Wheat					Panel D: KCBOT Wheat				
0.1	12.513	14.255	27.695	20.689	0.1	13.91	8.621	17.088	15.278
0.25	6.569	19.882	24.256	25.37	0.25	13.87	16.705	12.767	9.107
0.75	37.02 *(-)	25.503	16.24	17.313	0.75	19.317	19.617	13.947	10.424
0.9	21.543	23.11	21.714	30.078	0.9	15.731	17.396	11.868	11.472
Post-financialization stage: 2012 - 2019									
Panel A: CBOT Corn					Panel B: CBOT Soybean				
0.1	11.470	8.762	13.696	11.082	0.1	29.250 *(-)	36.448 *(-)	14.505	6.986
0.25	11.396	9.172	19.546	14.857	0.25	11.194	23.157	18.155	9.767
0.75	13.891	17.085	6.331	6.407	0.75	11.551	11.793	23.743	16.455
0.9	19.059	20.057	9.893	16.304	0.9	13.258	9.55	20.099	20.056
Panel C: CBOT Wheat					Panel D: KCBOT Wheat				
0.1	31.157	23.023	12.623	12.714	0.1	16.593	11.461	24.700	20.069
0.25	13.377	27.779 *(+)	20.501	15.769	0.25	21.15	8.219	17.220	27.112
0.75	12.473	24.576	16.978	8.097	0.75	29.309	16.392	13.124	11.271
0.9	28.042	19.352	15.083	11.494	0.9	17.361	19.085	8.303	11.376

Notes: * indicates statistical significance at 5%. Box-Ljung test statistics for 13 lags are in the table. The sign (+/-) next to the test statistics indicates the dominant sign of the underlying CQ estimates for the Box-Ljung test statistics.

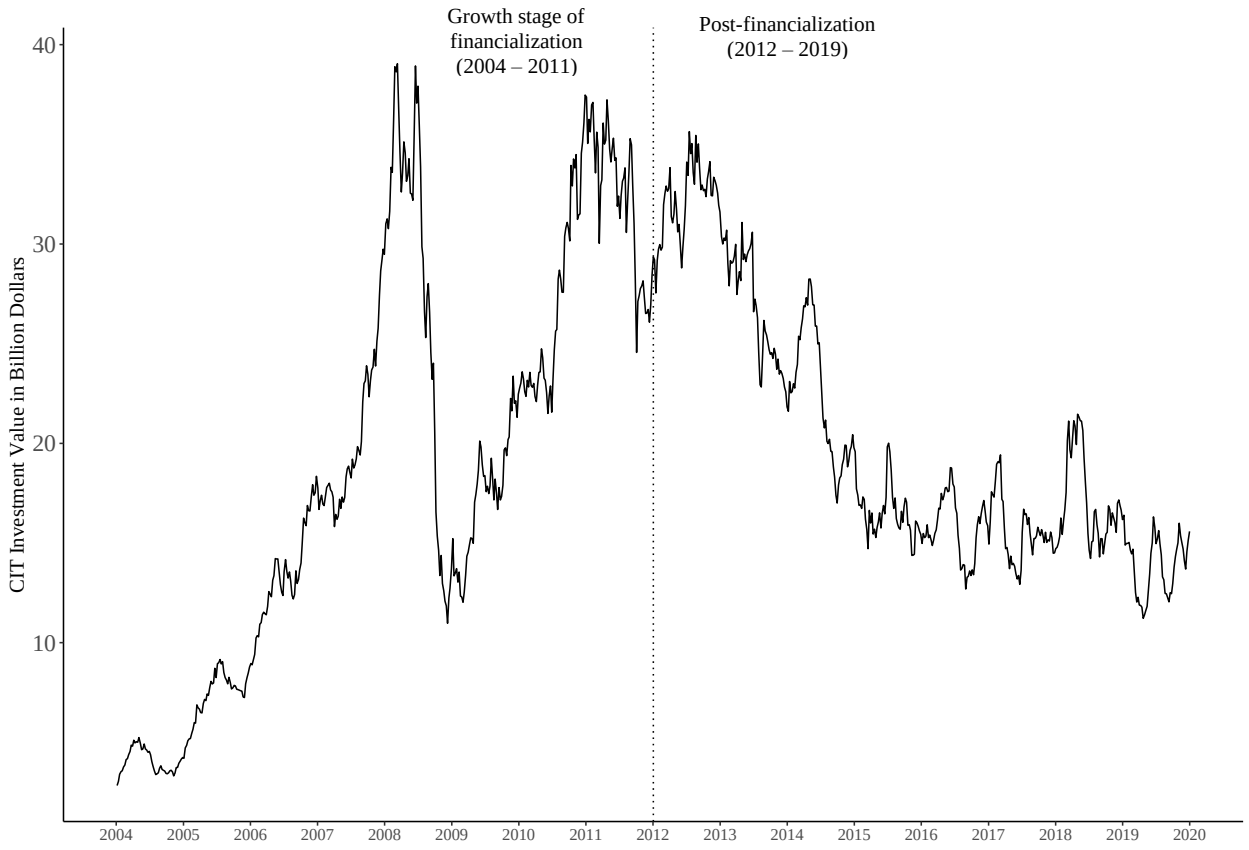


Figure 4.1: Notional value of commodity index investment in four grain futures markets

Notes: The notional value of commodity index investment is calculated using the index positions retrieved from SCOT report and corresponding nearby futures prices during the sample period. The growth stage of financialization and the post-financialization stage is defined following Irwin, Sanders, and Yan (2022).

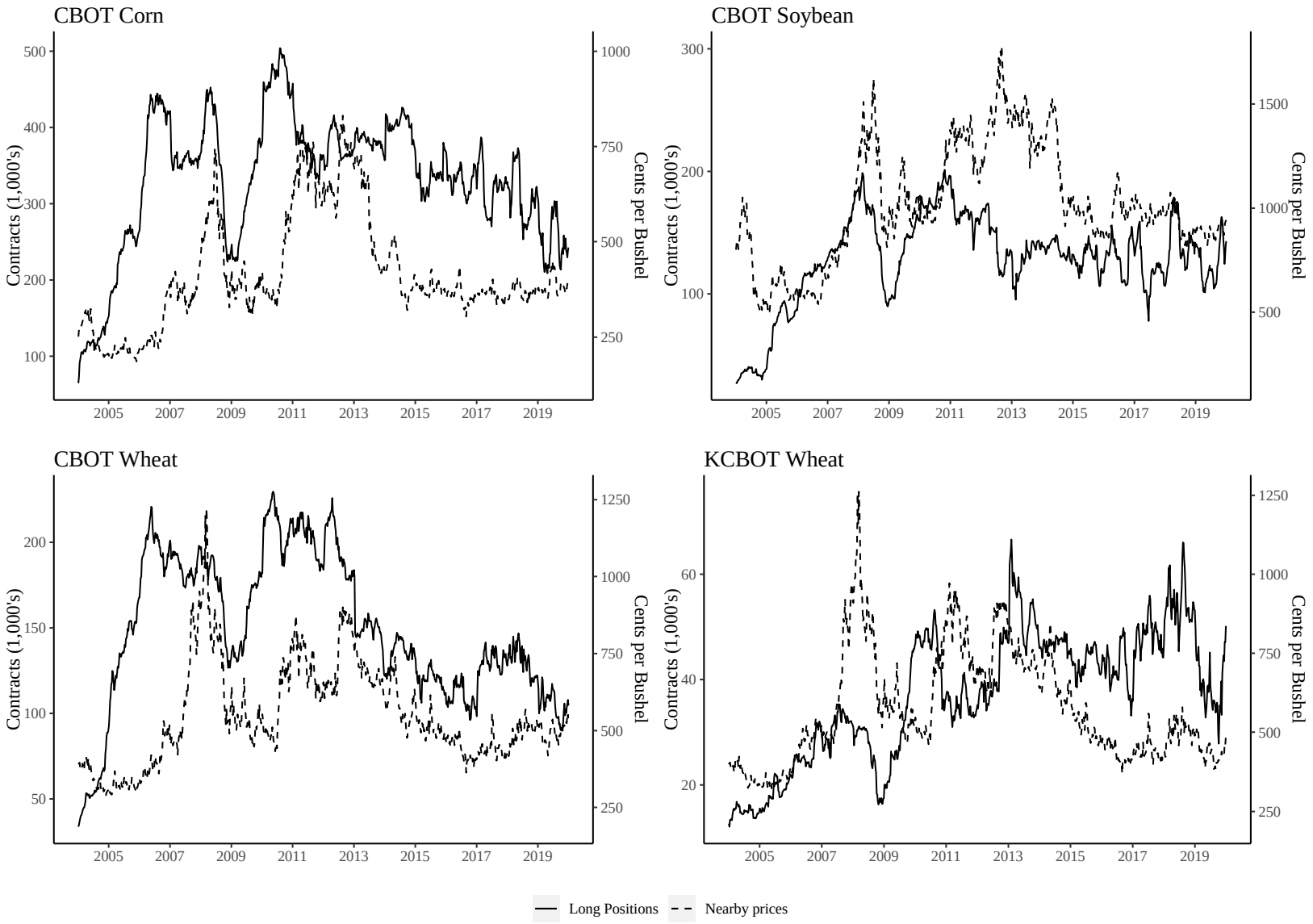


Figure 4.2: Weekly commodity index trader positions and nearby futures prices of CBOT corn, soybean, wheat, and KCBOT wheat, January 2004 to December 2019

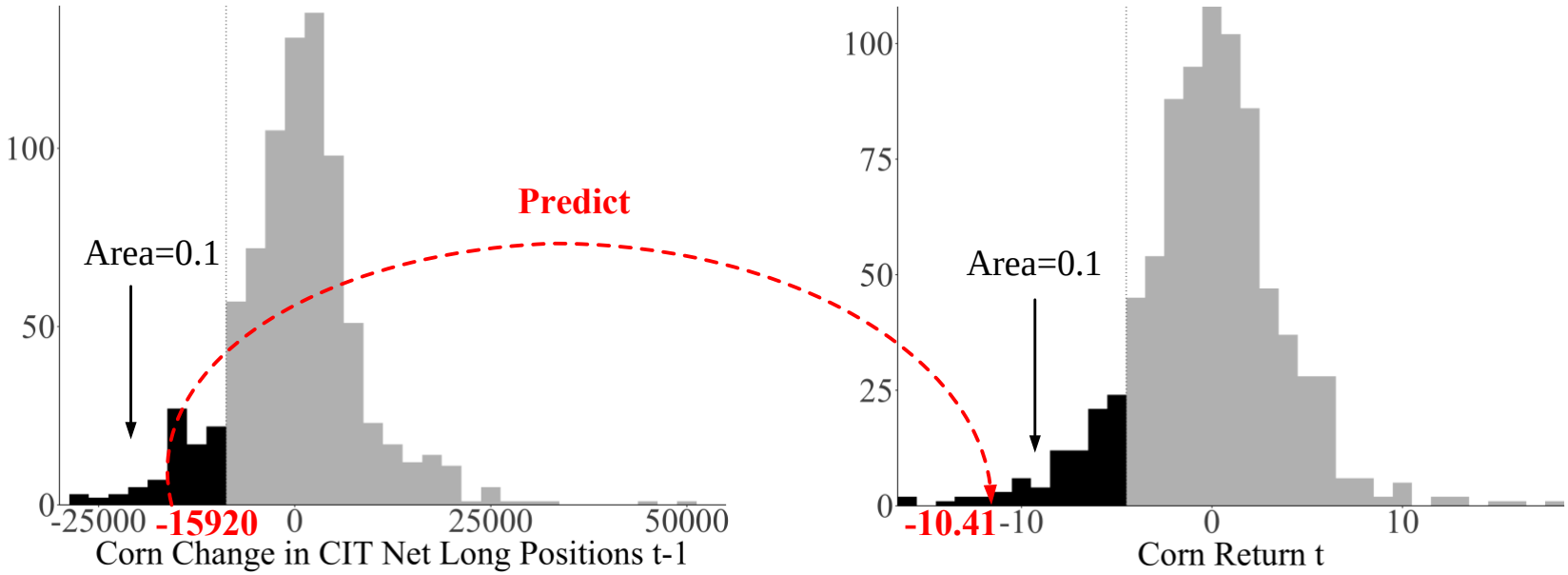
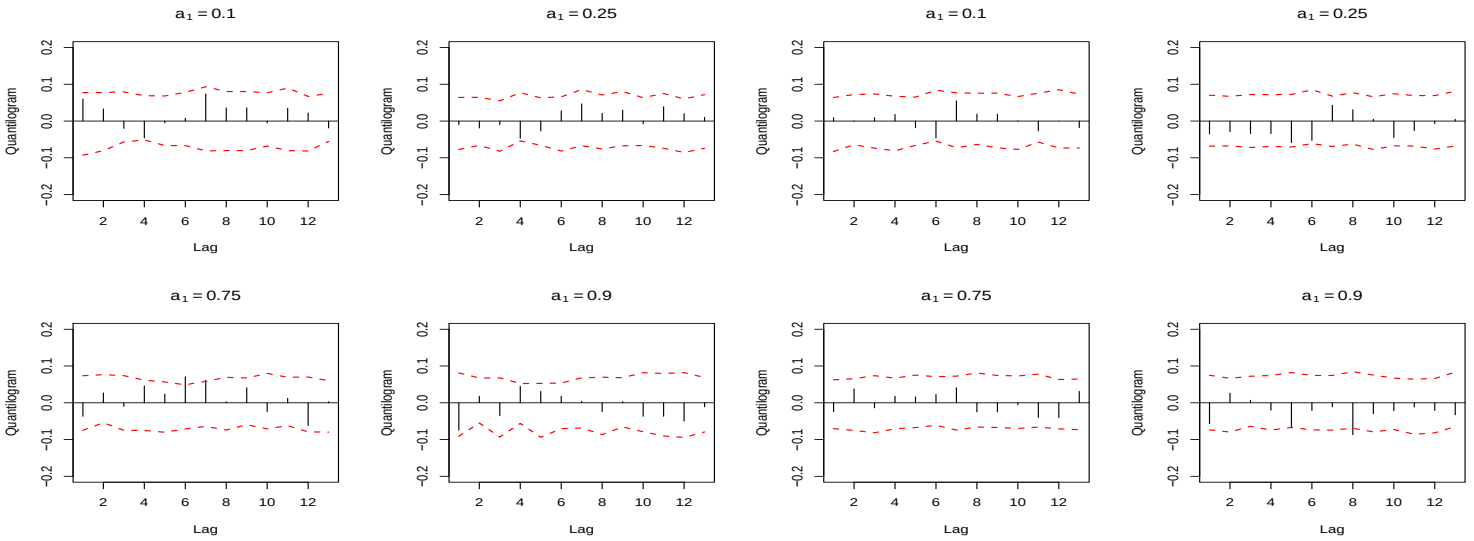
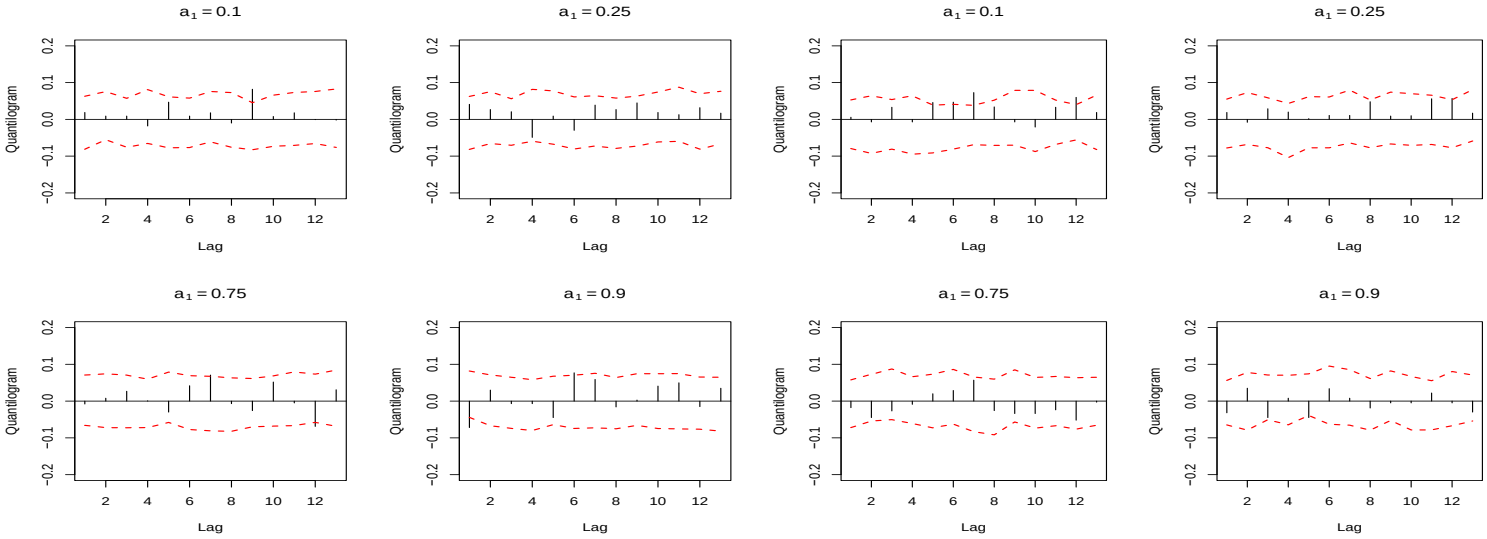


Figure 4.3: Illustration of the lead-lag dependence from CIT net long position changes at $t-1$ to futures returns at t when both are in the low quantile of 0.1, full sample period in the corn market

Notes: On September 27, 2011, the change in corn CIT net long positions is -15,920 contracts and hits in the 0.1 quantile for position changes. One week later on October 4, 2011 we observe a corn return of -10.41% and it hits in the 0.1 quantile for returns as well. The arrow shows that when changes in CIT net long positions are below the 0.1 quantile, it is followed by a return one week later that is also below its 0.1 quantile. This type of comparison is repeated for all observations to compute a CQ statistic for $\alpha_1 = \alpha_2 = 0.1$.

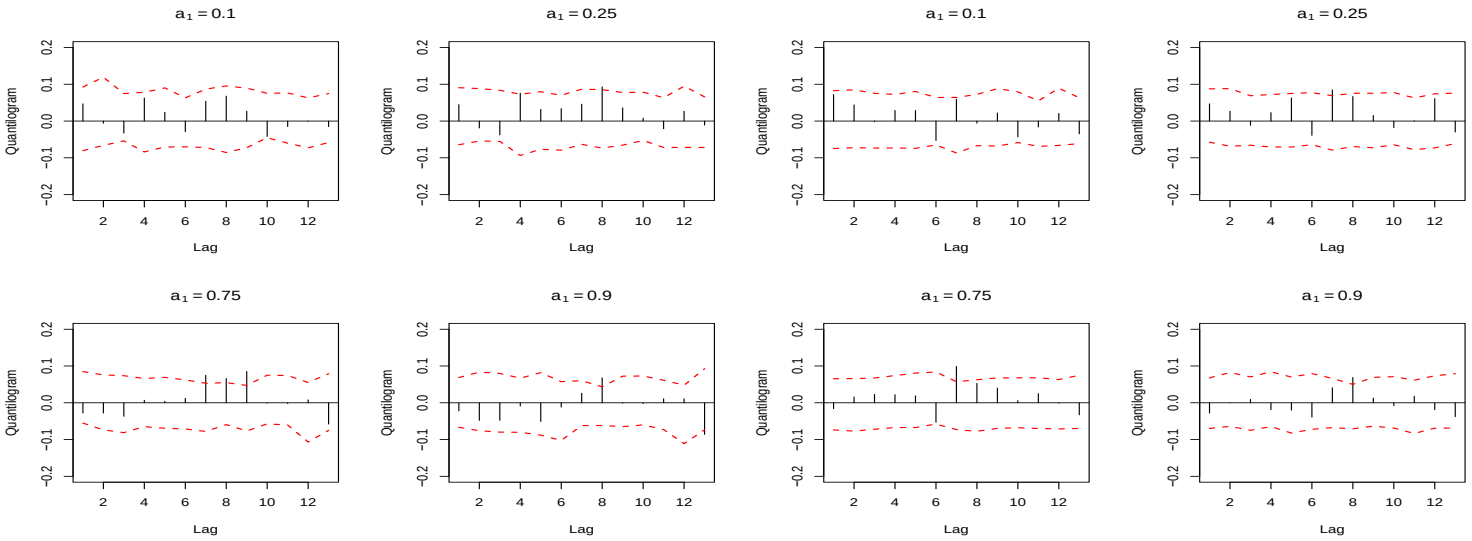


(a) Position change at quantile level 0.1 (b) Position change at quantile level 0.25



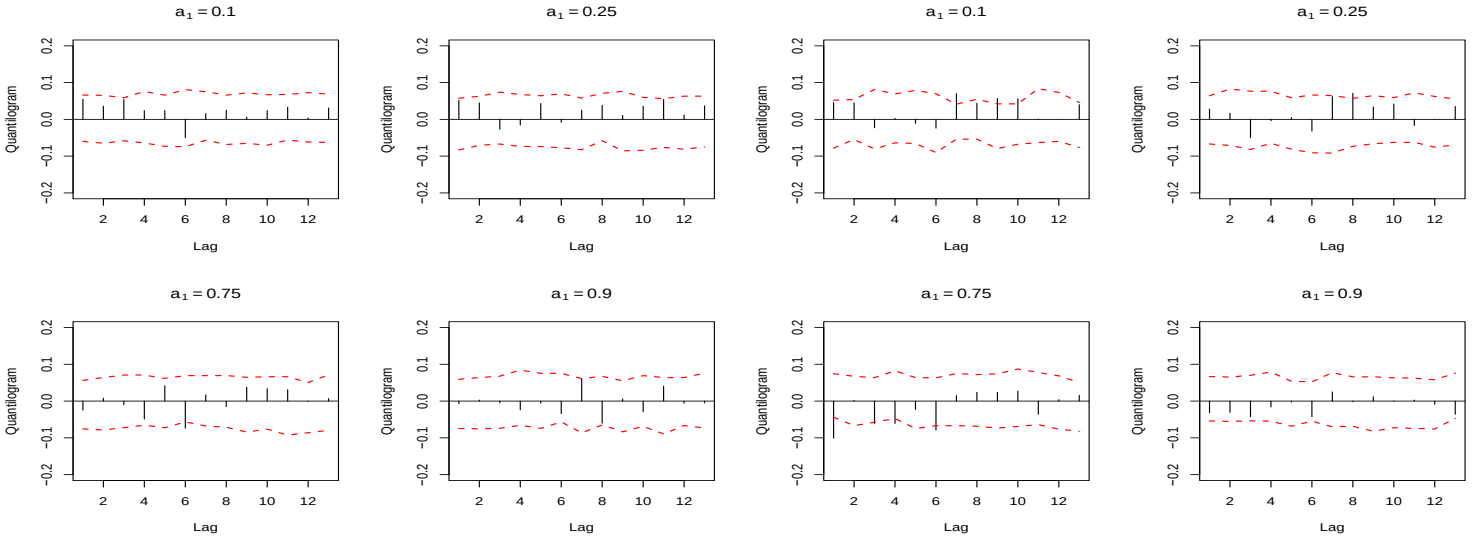
(c) Position change at quantile level 0.75 (d) Position change at quantile level 0.9

Figure 4.4: Cross-quantilogram from changes in CIT net long positions to returns in the CBOT corn futures market, 2004 – 2019



(a) Position change at quantile level 0.1

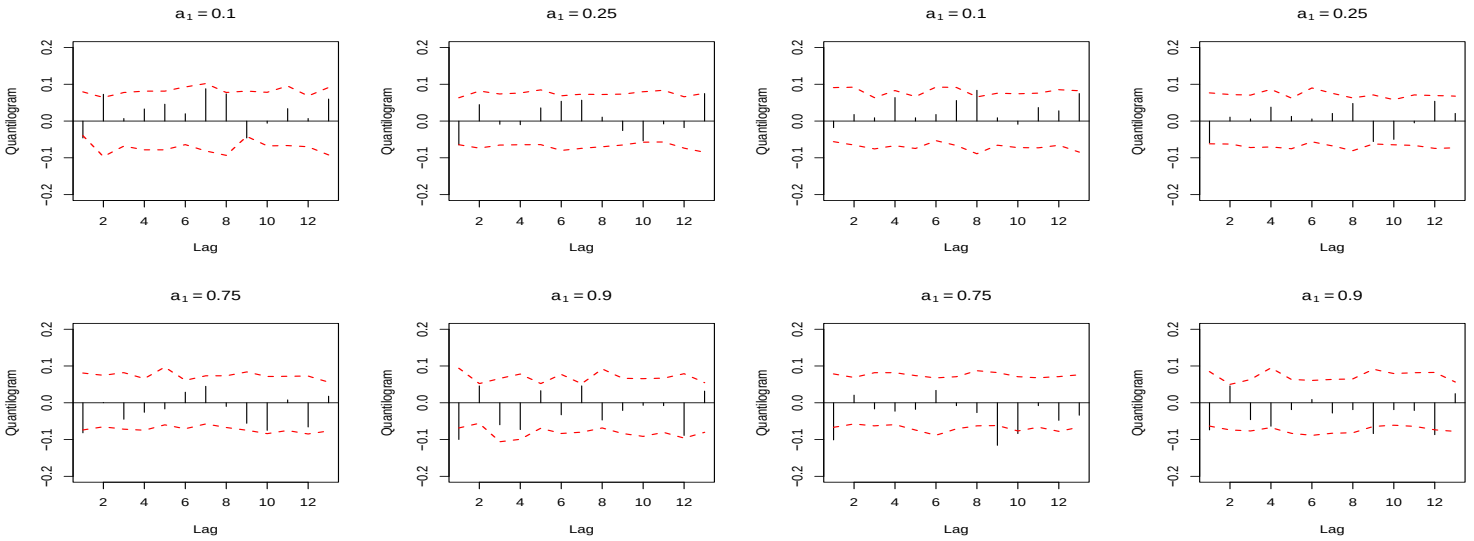
(b) Position change at quantile level 0.25



(c) Position change at quantile level 0.75

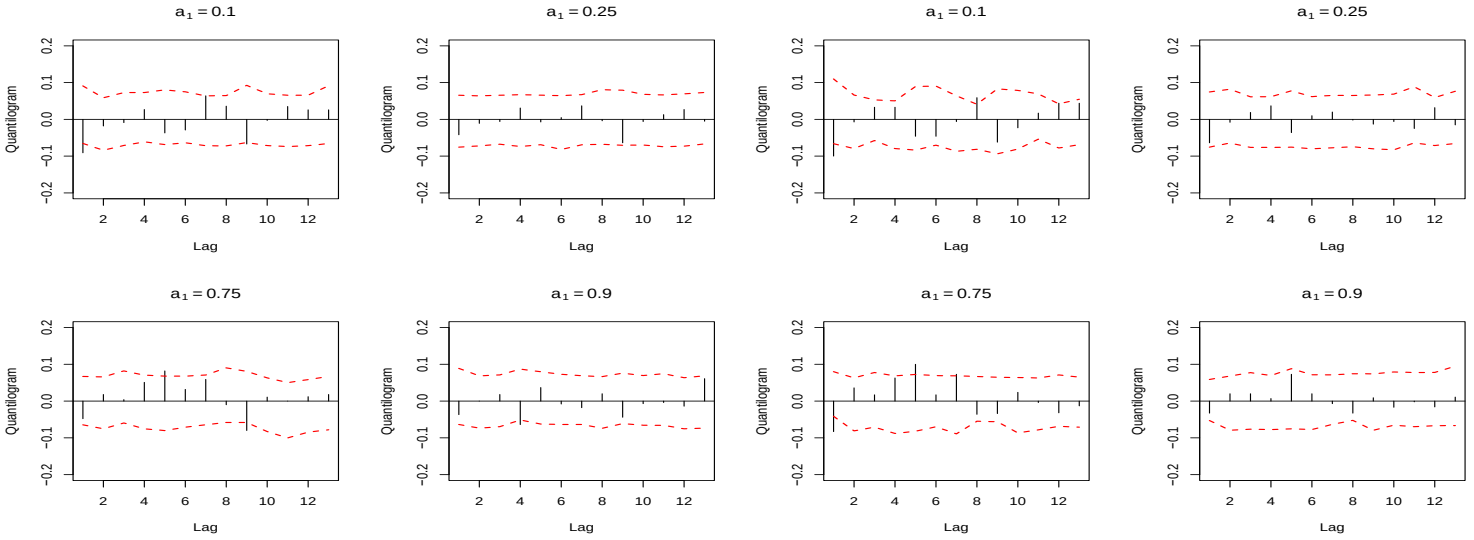
(d) Position change at quantile level 0.9

Figure 4.5: Cross-quantilogram from changes in CIT net long positions to returns in the CBOT soybean futures market, 2004 – 2019



(a) Position change at quantile level 0.1

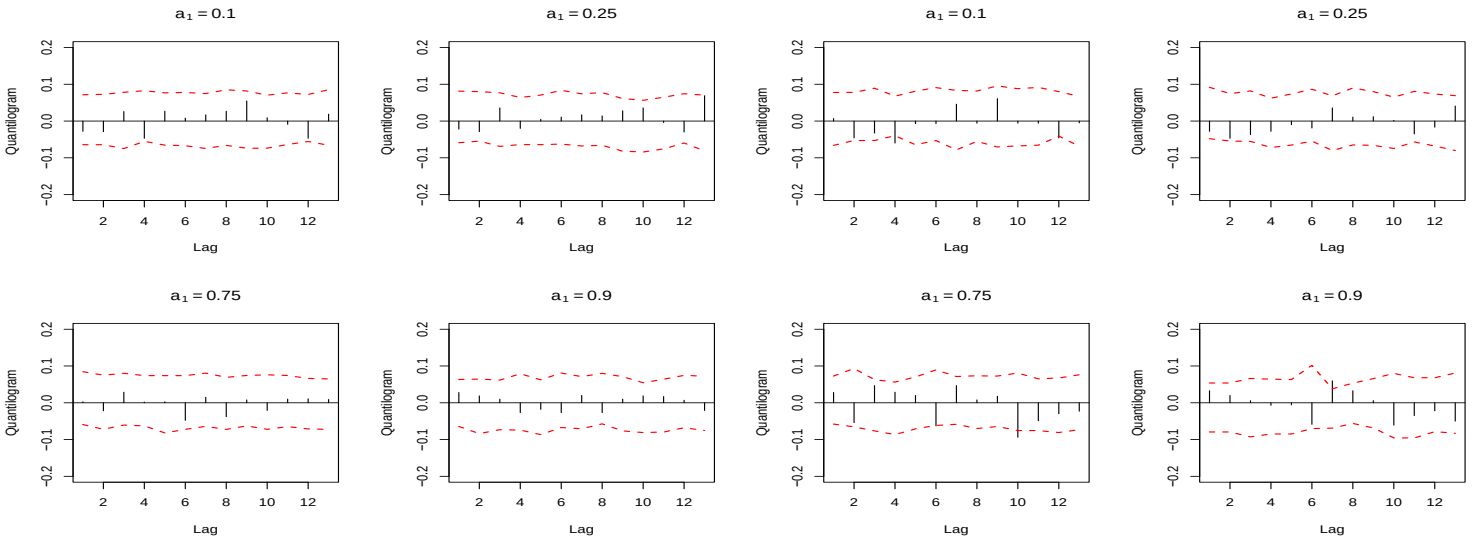
(b) Position change at quantile level 0.25



(c) Position change at quantile level 0.75

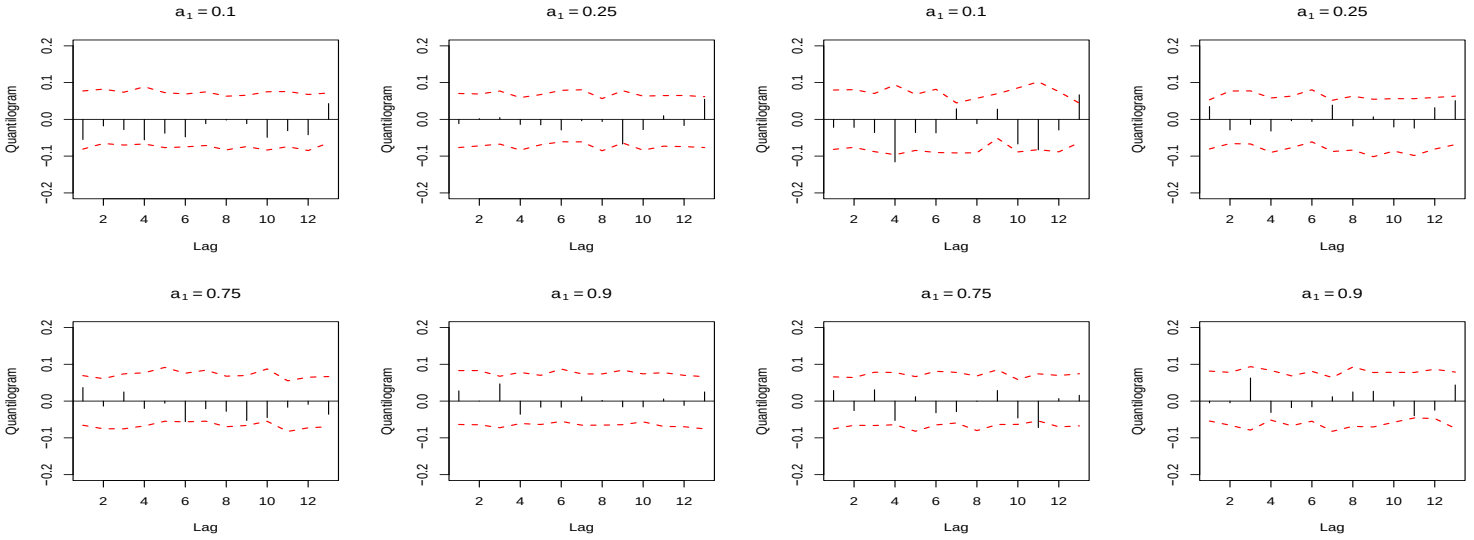
(d) Position change at quantile level 0.9

Figure 4.6: Cross-quantilogram from changes in CIT net long positions to returns in the CBOT wheat futures market, 2004 – 2019



(a) Position change at quantile level 0.1

(b) Position change at quantile level 0.25



(c) Position change at quantile level 0.75

(d) Position change at quantile level 0.9

Figure 4.7: Cross-quantilogram from changes in CIT net long positions to returns in the KCBOT wheat futures market, 2004 – 2019

CHAPTER 5:

CONCLUSIONS

The first essay of the dissertation estimates the pass-through equilibrium and the short-run pass-through dynamics from LCFS credit prices to California wholesale gasoline and diesel prices. Our analysis focus on the period from 2016 through 2020 when LCFS was readopted with more ambitious goals of reducing carbon emissions. During this period, LCFS credit prices have inflated fluctuations that we apply them to estimate the pass-through in wholesale prices. For wholesale gasoline, after pooling three spreads, the long-run estimated coefficient is 1.06 with a Newey-West HAC standard error of 0.2, suggesting in wholesale gasoline markets, fuel providers and fully recoup their LCFS obligations by passing the extra credit costs to their downstream buyers. For wholesale diesel, the pooled estimated coefficient is about 0.5, with a Newey-West HAC standard error of 0.1, showing that suppliers of wholesale diesel can only pass down half of the credit costs to their downstream buyers. For short-run pass-through, in wholesale gasoline markets, obligated parties have complete pass-through after 4 business days, but in wholesale diesel markets, in the first few days, pass-through rates are negative and after 15 days, they can only recoup 60% of the LCFS credit costs. Evidence in wholesale gasoline market for both long-run and short-run shows downstream customers pay for the LCFS credit prices and shows the effectiveness of the program by discouraging the consumption of traditional fuels. The evidence of incomplete pass-through in wholesale diesel for both long-run and short-run shows obligated parties can only pass down a portion of LCFS credit costs to their buyers. The more elastic demand for diesel, the lower credit prices, and the lack of volatility for diesel pool make obligated parties do not have the capacity and incentives to pass down LCFS credit costs. This essay contributes to the literature by being

the first paper that applies the econometric methods to measure the magnitude and speed of the pass-through rates under LCFS in California. We also include a placebo analysis to show that only fuels prices in California respond to this state-level policy. Future work may wish to investigate the market structure of two fuels and estimate the elasticities of demand and supply of wholesale fuels in California.

The second essay of the dissertation examines the forecasting accuracy of a batch of yield forecasting models that directly transform the ordinal crop condition ratings to the numeric condition index and compares the recently proposed BM model with other four simpler models. We conduct the out-of-sample yield forecasts recursively for corn and soybeans from 2000 through 2020 for all models. We compare each model's yield forecasts with USDA final yield estimates and we used RMSPE to measure each model's forecasting accuracy. We find all models provide accurate yield forecasts as throughout the growing season, the average RMSPE is 5% for corn and 6% for soybean. We apply modified Diebold Mariano test and Model Confidence Set test for each week of the growing season to compare the forecasting performance of BM model and its competitors. Test results suggest BM model fail to significantly outperform simpler models. We also apply the average SPA test to compare models' overall performance across the growing season. Test results fail to show BM model provides significantly more accurate yield forecasts. This essay contributes to the literature by using extended crop condition data to build recursively out-of-sample yield forecasts and show if there is a trade-off between model complexity and forecasting accuracy. As all models use the same set of data, we illustrate more complex model does not necessarily provide more accurate forecasts. Future work may wish to improve the yield forecasts by adding more information like crop progress data and to extend the discussion to other agricultural commodities.

The third essay use the recently developed cross-quantilogram (CQ) test to examine whether predictability exists between the change in commodity index trader (CIT) positions and returns in the tails of the distributions for four agricultural futures markets. CQ test has the advantages that it does not need to make any assumptions of the distributions of the data, it captures the lead-lag relationship across all parts of distributions, and it includes long lags without concerns about degrees-of-freedom. We first conduct three types of linear causality tests to provide a comprehensive baseline for the relationship between CIT positions and agricultural futures prices movements. The null of no causality was not rejected in the majority of the cases across the different tests, measures of position pressure, or the sample period considered. Next, we apply the CQ test to the same data to determine if there is a relationship between the tails of the distributions of index positions and price movements. Consistent with the standard linear causality tests, we find no evidence of a relationship between index positions and returns. This essay contributes to the literature by providing evidence against Master Hypothesis using extreme movements in index traders' positions and returns. Future studies may wish to examine other types of traders on both the long- and short-term pricing of commodity markets.

REFERENCES

Algieri, B., M. Kalkuhl, and N. Koch. "A Tale of Two Tails: Explaining Extreme Events in Financialized Agricultural Markets." *Food Policy* 69(2017):256-269.

Armstrong, J.S. *Principles of Forecasting: A Handbook for Researchers and Practitioners.* New York: Springer Science+Business Media, Inc., 2001.

Aulerich, N.M., S.H. Irwin, and P. Garcia. "Bubbles, Food Prices, and Speculation: Evidence from the CFTC's Daily Large Trader Data Files." *The Economics of Food Price Volatility*. J.P. Chavas, D. Hummels, and B.D. Wright, eds. University of Chicago Press, 2014, pp. 211-253.

Bachmeier, L.J., and J.M. Griffin. "New Evidence on Asymmetric Gasoline Price Responses." *The Review of Economics and Statistics* 85,3(2003):772–776.

Bain, R., and T.R. Fortenbery. "Impact of Crop Condition Reports on National and Local Wheat Markets." *Journal of Agricultural and Applied Economics* 49,1(2017):97-119.

Baumöhl, E., and Š. Lyócsa. "Directional Predictability from Stock Market Sector Indices to Gold: A Cross-Quantilogram Analysis." *Finance Research Letters* 23(2017):152-164.

Beguería, S., and M.P. Maneta. "Qualitative Crop Condition Survey Reveals Spatiotemporal Production Patterns and Allows Early Yield Prediction." *Proceedings of the National Academy of Sciences* 117,31(2020):18317-18323.

Borenstein, S., A.C. Cameron, and R. Gilbert. "Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes?" *The Quarterly Journal of Economics* 112,1(1997):305–339.

Burkhardt, J. "Incomplete Regulation in an Imperfectly Competitive Market: The Impact of the Renewable Fuel Standard on US Oil Refineries." Unpublished Manuscript, School of Forestry and Environmental Studies, Yale University, 2016.

Burkholder, D. A Preliminary Assessment of RIN Market Dynamics, RIN Prices, and Their Effects. Washington, DC: U.S. Environmental Protection Agency/Office of Transportation and Air Quality. Internet site: https://www.grassley.senate.gov/download/epa-hq-oar-2015-0111-0062burkholderrin-analysis (Accessed March 5, 2022).

California Air Resource Board, Low Carbon Fuel Standard Regulation (CARB-LCFS). "Final Regulation Order." California, 2015. Internet site: https://ww3.arb.ca.gov/regact/2015/lcfs2015/lcfs2015.htm (Accessed March 5, 2022).

California Air Resource Board (CARB). "Staff Report: Initial Statement of Reasons. Public Hearing to Consider Proposed Amendments to the Low Carbon Fuel Standard Regulation and to the Regulation on Commercialization of Alternative Diesel Fuels." California, 2018. Internet site:https://ww3.arb.ca.gov/regact/2018/lcfs18/isor.pdf?_ga=2.13627127.1522303650.1591469888-1877975938.1591469888 (Accessed March 5, 2022).

Chesnes, M. "Asymmetric Pass-Through in US Gasoline Prices." *Energy Journal* 37,1(2016).

Colino, E.V., S.H. Irwin, P. Garcia, and X. Etienne. "Composite and Outlook Forecast Accuracy." *Journal of Agricultural and Resource Economics* (2012): 228-246.

Engelke, L., and J.C. Yuen. "Types of Commodity Investments." *The Handbook of Commodity Investing*. F.J. Fabozzi, R. Fuss, and D. Kaiser, eds. Hoboken, NJ: John Wiley and Sons, 2008, pp. 549-569.

Fackler, P.L., and B. Norwood. "Forecasting Crop Yields and Condition Indices." Paper presented at NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management, 1999.

Funk, C., and M.E. Budde. "Phenologically-Tuned MODIS NDVI-Based Production Anomaly Estimates for Zimbabwe." *Remote Sensing of Environment*, 113,1(2009):115-125.

Ganapati, S., J.S. Shapiro, and R. Walker. "The Incidence of Carbon Taxes in US Manufacturing: Lessons from Energy Cost Pass-Through." Working paper No. w22281, National Bureau of Economic Research, 2016.

Gilbert, C.L., and S. Pfuderer. "The Role of Index Trading in Price Formation in the Grains and Oilseeds Markets." *Journal of Agricultural Economics* 65,2(2014):303-322.

Granger, C.W.J. "Testing for Causality: A Personal Viewpoint." *Journal of Economic Dynamics and Control* 2(1980):329-352.

Hamilton, J.D., and J.C. Wu. "Effects of Index-Fund Investing on Commodity Futures Prices." *International Economic Review* 56,1(2015):187-205.

Han, H., O. Linton, T. Oka, and Y.J. Whang. "The Cross-Quantilogram: Measuring Quantile Dependence and Testing Directional Predictability between Time Series." *Journal of Econometrics* 193,1(2016):251-270.

Hannan, E.J., R.D. Terrell, and N.E. Tuckwell. "The Seasonal Adjustment of Economic Time Series." *International Economic Review* 11,1(1970):24–52.

Hansen, P.R., A. Lunde, and J.M. Nason. "The Model Confidence Set." *Econometrica* 79.2(2011): 453-497.

Harvey, D., S. Leybourne, and P. Newbold. "Testing the Equality of Prediction Mean Squared Errors." *International Journal of Forecasting* 13,2(1997):281-291.

Hoogenboom, G. "Contribution of Agrometeorology to the Simulation of Crop Production and Its Applications." *Agricultural and forest meteorology* 103,1-2(2000):137-157.

Huang, H., M. Khanna, H. Onal, and X. Chen. "Stacking Low Carbon Policies on the Renewable Fuels Standard: Economic and Greenhouse Gas Implications." *Energy Policy* 56(2013):5–15.

Irwin, S.H., and D.L. Good. "When Should We Start Paying Attention to Crop Condition Ratings for Corn and Soybeans?" *farmdoc daily* (7):96, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, May 24, 2017a.

Irwin, S.H., and D.L. Good. "How Should We Use Within-Season Crop Condition Ratings for Corn and Soybeans?" *farmdoc daily* (7):101, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, June 1, 2017b.

Irwin, S.H., D.L. Good, and M. Tannura. "2009 Final Corn and Soybean Yield Forecasts." Marketing and Outlook Briefs, MOBR 09-06, Urbana-Champaign: Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, 2009.

Irwin, S.H., and T. Hubbs. "What to Make of High Early Season Crop Condition Ratings for Corn?" *farmdoc daily* (8):108, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, June 13, 2018a.

Irwin, S.H., and T. Hubbs. "What to Make of High Early Season Crop Condition Ratings for Soybeans?" *farmdoc daily* (8):109, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, June 14, 2018b.

Irwin, S.H., and T. Hubbs. "Does the Bias in Early Season Crop Condition Ratings for Corn and Soybeans Vary with the Magnitude of the Ratings?" *farmdoc daily* (8):113, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, June 20, 2018c.

Irwin, S.H., and T. Hubbs. "Measuring the Accuracy of Forecasting Corn and Soybean Yield with Good and Excellent Crop Condition Ratings" *farmdoc daily* (8):118, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, June 27, 2018d.

Irwin, S.H., and D.R. Sanders. “Testing the Masters Hypothesis in Commodity Futures Markets.” *Energy Economics* 34,1(2012):256-269.

Irwin, S.H., and D.R. Sanders. “Index Funds, Financialization, and Commodity Futures Markets.” *Applied Economic Perspectives and Policy* 33,1(2011):1-31.

Irwin, S.H., D.R. Sanders, and L. Yan. “The Order Flow Cost of Index Rolling in Commodity Futures Markets.” Working Paper. Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, 2022.

Jiang, H., J.J. Su., N. Todorova, and E. Roca. “Spillovers and Directional Predictability with a Cross-Quantilogram Analysis: The case of US and Chinese Agricultural Futures.” *Journal of Futures Markets* 36,12(2016):1231-1255.

Kauffman, R.K., and S. Snell. “A Biophysical Model of Corn Yield: Integrating Climatic and Social Determinants.” *American Journal of Agricultural Economics* 79(1997):178-190.

Knittel, C.R., B.S. Meiselman, and J.H. Stock. “The Pass–Through of RIN Prices to Wholesale and Retail Fuels under the Renewable Fuel Standard.” *Journal of the Association of Environmental and Resource Economists* 4,4(2017):1081–1119.

Kruse, J., and D. Smith. “Yield Estimation Throughout the Growing Season.” Center for Agricultural and Rural Development, Iowa State University, 1994.

Lade, G.E., and J. Bushnell. “RIN Pass-Through to Retail E85 Prices Under the Renewable Fuel Standard.” Working paper, Dept. of Econ., Iowa State University, 2016.

Lee, T.H., and W. Yang. “Money–Income Granger-Causality in Quantiles.” *Advances in Econometrics* 30(2012):385-409.

Lehecka, G.V. “Do Hedging and Speculative Pressures Drive Commodity Prices, or the Other Way Round?” *Empirical Economics* 49,2(2015):575-603.

Lehecka, G. V. “The Value of USDA Crop Progress and Condition Information: Reactions of Corn and Soybean Futures Markets.” *Journal of Agricultural and Resource Economics* 39,1(2014):88-105.

Lewis, M.S. “Asymmetric Price Adjustment and Consumer Search: An Examination of the Retail Gasoline Market.” *Journal of Economics and Management Strategy* 20,2(2011):409–449.

Linton,O., and Y.J. Whang. “The Quantilogram: With an Application to Evaluating Directional Predictability.” *Journal of Econometrics* 141,1(2007):250-282.

Marshall, A. *Principles of Economics: Unabridged Eighth Edition.* New York: Cosimo, Inc, 2009.

Mayer, J. “The Growing Financialisation of Commodity Markets: Divergences between Index Investors and Money Managers.” *Journal of Development Studies* 48,6(2012):751-767.

Palazzi, R.B., A.C.F. Pinto, M.C. Klotzle, and E.M. De Oliveira. “Can We Still Blame Index Funds for the Price Movements in the Agricultural Commodities Market?” *International Review of Economics and Finance* 65(2020):84-93.

Politis, D.N., and J.P. Romano. “The Stationary Bootstrap.” *Journal of the American Statistical Association* 89,428(1994):1303-1313.

Pouliot, S., A. Smith, and J. H. Stock. “RIN Pass-Through at Gasoline Terminals.” Working paper, Dept. of Econ., Iowa State University, 2017.

Quaedvlieg, R. “Multi-Horizon Forecast Comparison.” *Journal of Business & Economic Statistics* 39,1(2021):40-53.

Sakamoto, T., A.A. Gitelson, and T.J. Arkebauer. “MODIS-Based Corn Grain Yield Estimation Model Incorporating Crop Phenology Information.” *Remote Sensing of Environment* 131(2013):215-231.

Sanders, D.R., and S.H. Irwin. “Bubbles, Froth and Facts: Another Look at the Masters Hypothesis in Commodity Futures Markets.” *Journal of Agricultural Economics* 68(2017):345-365.

Sanders, D.R., and S.H. Irwin. “The ‘Necessity’ of New Position Limits in Agricultural Futures Markets: The Verdict from Daily Firm-level Position Data.” *Applied Economic Perspectives and Policy* 38(2016):292-317.

Sanders, D.R., and S.H. Irwin. “Energy Futures Prices and Commodity Index Investment: New Evidence from Firm-Level Position Data.” *Energy Economics* 46(2014):S57–S68.

Sanders, D.R., and S.H. Irwin. “Measuring Index Investment in Commodity Futures Markets.” *Energy Journal* 34(2013):105-127.

Sanders, D.R., and S.H. Irwin. “New Evidence on the Impact of Index Funds in US Grain Futures Markets.” *Canadian Journal of Agricultural Economics* 59,4(2011):519-532.

Sanders, D.R., S.H. Irwin, and R.P. Merrin. “Smart Money: The Forecasting Ability of CFTC Large Traders in Agricultural Futures Markets.” *Journal of Agricultural and Resource Economics* 34(2009): 276-296.

Scarcioffolo, A.R., and X. Etienne. “Testing Directional Predictability between Energy Prices: A Quantile-Based Analysis.” *Resources Policy* 74(2021): 102258.

Shaw, L.H. “The Effect of Weather on Agricultural Output: A Look at Methodology.” *Journal of Farm Economics* 46(1964):218-230.

Singleton, K.J. “Investor Flows and the 2008 Boom/Bust in Oil Prices.” *Management Science* 60,2(2014):300-318.

Stock, J.H., and M.W. Watson. “A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems.” *Econometrica: Journal of the Econometric Society* (1993):783-820.

Stoll, H.R., and R.E. Whaley. "Commodity Index Investing and Commodity Futures Prices." *Journal of Applied Finance* 20,1(2010):7-47.

Stolper, S. "Who Bears the Burden of Energy Taxes?" *Proceedings Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, 109(2016):1–9.

Summers, L.H. "Does the Stock Market Rationally Reflect Fundamental Values?" *Journal of Finance* 41,3(1986):591-601.

Tadesse, G., B. Algieri, M. Kalkuhl, and J.V. Braun. "Drivers and Triggers of International Food Price Spikes and Volatility." *Food Policy* 47(2014):117-128.

Toda, H.Y., and T. Yamamoto. "Statistical Inference in Vector Autoregressions with Possibly Integrated Processes." *Journal of Econometrics* 66,1-2(1995):225-250.

U.S. Department of Agriculture, National Agricultural Statistics Service (USDA-NASS). *Surveys: Crop Progress and Conditions*. Retrieved December 13, 2021. Internet site: https://www.nass.usda.gov/Surveys/Guide_to_NASS_Surveys/Crop_Progress_and_Condition/index.php

United States Senate, Permanent Subcommittee on Investigations (USS/PSI). "Excessive Speculation in the Wheat Market. Majority and Minority Staff Report." Washington, D.C.: US Government Printing Office, 2009. Internet site: https://www.hsgac.senate.gov/imo/media/doc/REPORTExcessiveSpecullationintheWheatMarketwoexhibitschartsJune2409.pdf?attempt=2 (Accessed February 7, 2022).

Valkanov, R. "Long-Horizon Regressions: Theoretical Results and Applications." *Journal of Financial Economics* 68,2(2003):201-232.

Walker, G.K. "Model for Operational Forecasting of Western Canada Wheat Yield." *Agricultural and Forest Meteorology* 44(1989):339-351.

Weyl, E.G., and M. Fabinger. “Pass-Through as an Economic Tool: Principles of Incidence under Imperfect Competition.” *Journal of Political Economy* 121.3(2013): 528-583.

Whistance, J., W. Thompson, and S. Meyer. “Interactions between California's Low Carbon Fuel Standard and the National Renewable Fuel Standard.” *Energy Policy* 101(2017):447–455.

Yeh, S., J. Witcover, G.E. Lade, and D. Sperling. “A Review of Low Carbon Fuel Policies: Principles, Program Status and Future Directions.” *Energy Policy* 97(2016):220–234.

APPENDIX A:

SUPPLEMENTAL MATERIALS FOR CHAPTER 2

This appendix presents results supplementing the essay entitled, "The Pass-Through of the Low Carbon Fuel Standard to Wholesale Fuels Prices in California". We present pass-through impacts estimated from the distributed lag model developed by Pouliot, Smith and Stock (2017).

The specification of the distributed lag model captures the long-run and the dynamics of pass-through impacts, as in petroleum markets, tax costs usually cannot be reflected instantaneously in fuel prices, rather it might take a few days to pass down the costs to their downstream buyers. The distributed lag analyses for gasoline and diesel are specified as:

$$Spread_{i,t}^{gas} = \alpha_i^{gas} + \beta_i^{gas} LCFS_{t-d}^{gas} + \sum_{n=0}^{d-1} \beta_{i,n}^{gas} \Delta LCFS_{t-n}^{gas} + \Gamma w_{i,t}^{gas} + u_{i,t}^{gas} \quad \text{(A.1)}$$

and,

$$Spread_{j,t}^{diesel} = \alpha_j^{diesel} + \beta_j^{diesel} LCFS_{t-d}^{diesel} + \sum_{n=0}^{d-1} \beta_{j,n}^{diesel} \Delta LCFS_{t-n}^{diesel} + \Gamma w_{j,t}^{diesel} + u_{j,t}^{diesel} \quad \text{(A.2)}$$

The long-run pass-through of the LCFS credit is measured as β_i^{gas} and β_j^{diesel}; the coefficients $\beta_{i,n}^{gas}$ and $\beta_{j,n}^{diesel}$ estimate the cumulative pass-through effects after n days, which capture the short-run dynamics of how spreads respond to an unexpected shock in LCFS credit prices; we also include the season components as $w_{i,t}^{gas}$ and $w_{j,t}^{diesel}$. We use 15 lags, $d = 15$, in the distributed lag model.

Estimated results are summarized in Table A1 over the full sample period. Estimated results are the pooled regression for two groups: wholesale gasoline and wholesale diesel. For wholesale gasoline, the long-run pass-through rate is about 1.22 and it is not significantly different from 1, indicating complete pass-through. The short-run pass-through rates are less precise, and

the dynamics over first 14 days show a very unstable pattern. For example, after 6 days, the rate is about 0.04, and after 8 days, the rate increases to 1.39. For wholesale diesel, the long-run pass-through is 0.46 and it is significantly smaller than 1, showing obligated parties can only pass down 46% of LCFS credit costs. The short-run pass-through rates show that in the first few days, they have full pass-through, however these estimates are less precise as the standard errors are almost as 10 times as the standard error for long-run pass-through rate. After 15 days, the pass-through rates drop back to 0.22 and it is significantly smaller than 1. For wholesale diesel, both long-run and short-run pass-through estimates show that obligated parties cannot fully pass LCFS credit costs to their buyers.

References

Pouliot, S., A. Smith, and J. H. Stock. "RIN Pass-Through at Gasoline Terminals." Working paper, Dept. of Econ., Iowa State University, 2017.

Table A1: Long-run and short-run dynamics of LCFS pass-through

	Pooled Wholesale Spreads			
	RBOB		ULSE	
	coefficient	(SE)	coefficient	(SE)
Long-run	1.221	(0.203)	0.464 ***	(0.088)
Short-run lag				
0	0.359	(0.923)	1.528	(0.581)
1	-0.535	(1.051)	1.461	(0.589)
2	-0.595	(1.096)	1.313	(0.685)
3	0.923	(1.166)	1.249	(0.692)
4	1.072	(1.066)	1.468	(0.764)
5	0.787	(0.983)	1.466	(0.802)
6	0.041	(1.108)	1.207	(0.732)
7	0.114	(1.222)	0.963	(0.709)
8	1.386	(1.250)	0.879	(0.706)
9	1.235	(1.039)	0.955	(0.621)
10	0.693	(0.988)	1.001	(0.681)
11	1.361	(1.179)	0.842	(0.679)
12	0.425	(1.123)	0.772	(0.539)
13	0.877	(1.176)	0.256 **	(0.355)
14	1.712	(1.190)	0.22 **	(0.353)

Notes:
(1) We use the full sample period to apply the distributed lag analysis, from January 4, 2016 to March 12, 2020.
(2) Newey-West standard error with 30 lags is in the parenthesis next to the estimated coefficients.
(3) ***, **, * indicates whether the estimated coefficients are significantly different from 1 at 1%, 5%, 10% level respectively.

APPENDIX B:

SUPPLEMENTAL MATERIALS FOR CHAPTER 4

This appendix presents results supplementing the essay entitled, "Do Extreme CIT Position Changes Move Prices in Grain Futures Markets?" We provide the Box-Ljung version of portmanteau tests from the percentage growth in commodity index traders (CIT) net long positions to returns, and the other way around, for four agricultural futures markets over the full sample periods, the growth stage of financialization, and the post-financialization period. We plot the CQ estimates for the level growth of CIT net long positions over two subsample periods, and the percentage growth of CIT net long positions over the full sample period and two subsample periods.

Table B1 reports the portmanteau test statistics for directional predictability from the percentage change in CIT net long positions to returns. For the full sample and two subsample periods, there are only 10 cases out of 192 with a significant portmanteau statistic. Only four of the 10 significant cases suggest that a large drop in CIT net long positions predicts the subsequent large drop in futures returns.

Table B2 reports the portmanteau test statistics for directional predictability from returns to the percentage change in CIT net long positions. For the full sample and two subsample periods, there are only 10 cases out of 192 with a significant portmanteau statistic. These significant cases fail to show a systematic pattern that index traders are trend-followers—only 5 cases out of 10 show a large drop in returns predicts the subsequent large drop in percentage change in CIT net long positions.

We have shown the CQ estimates for the full sample period in the paper. In this section, we first present the CQ estimates from the level growth in CIT net long positions to returns for the two subsample periods. Figures B1-B4 present the CQ estimates from the level growth in CIT net long positions to returns for the growth stage of financialization. In general, most CQ estimates are insignificant. For the few significant cases, they do not show a systematic lead-lag relationship from positions to prices. Figures B5-B8 present the CQ estimates from the level growth in CIT net long positions to returns for the post-financialization stage. Only in a very small portion of cases the CQ estimates are significant, yet they fail to provide evidence for the Master Hypothesis.

For the CQ estimates from the percentage change in CIT net long positions to returns, Figures B9-B12 present the CQ estimates for the full sample period, Figures B13-B16 present the CQ estimates for the growth stage of financialization, and Figures B17-B20 present the CQ estimates for the post-financialization stage. Similar to the findings with the change in CIT net long positions, we fail to find evidence supporting the directional predictability from the percentage change in CIT net long positions to returns.

Table B1: Cross-quantilogram portmanteau test results for impacts of percentage changes in CIT net long positions to futures returns in four grain futures markets for full sample, the growth stage of financialization, and the post-financialization stage

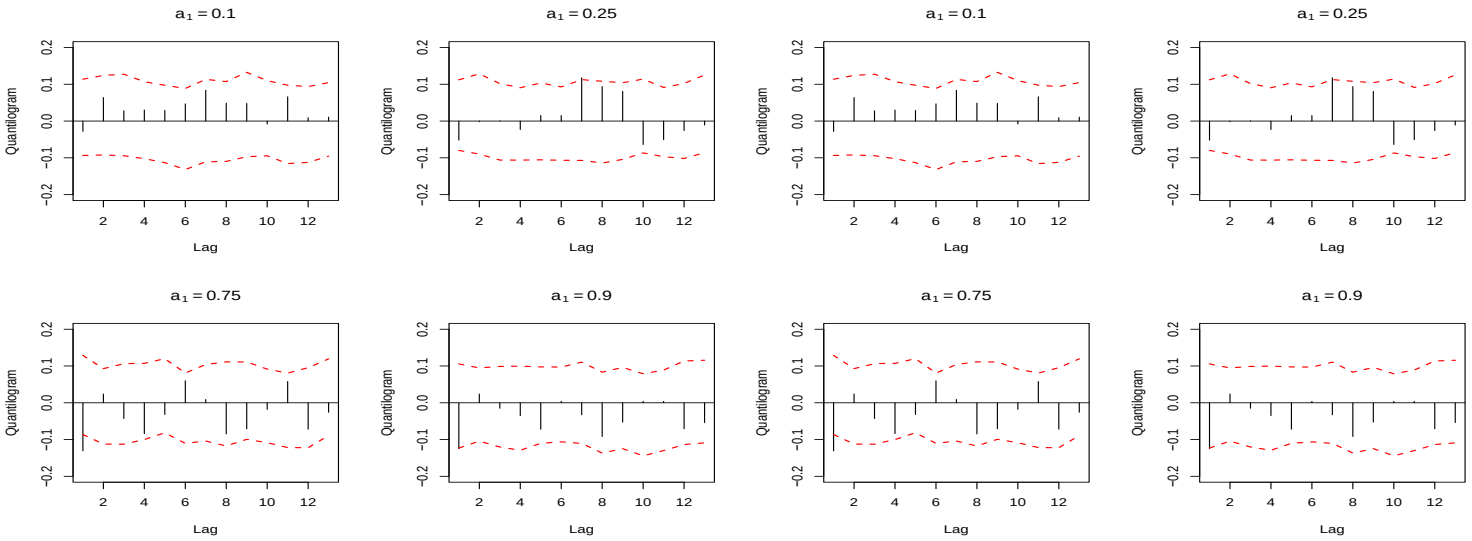
Dependent variable: returns, independent variable: CIT Net Long % Change									
	Returns Quantile Level					Returns Quantile Level			
CIT Net Long % Change Quantile Level	0.1	0.25	0.75	0.9	CIT Net Long % Change Quantile Level	0.1	0.25	0.75	0.9
Full sample: 2004 - 2019									
Panel A: CBOT Corn					Panel B: CBOT Soybean				
0.1	13.449	13.826	25.351	16.679	0.1	17.781	27.351	17.276	16.189
0.25	4.788	14.99	9.342	15.367	0.25	16.252	24.919	16.868	9.707
0.75	7.828	8.342	16.256	12.052	0.75	9.017	9.851	14.836	7.184
0.9	9.541	9.198	9.042	15.512	0.9	6.466	9.726	25.169 *(-)	8.33
Panel C: CBOT Wheat					Panel D: KCBOT Wheat				
0.1	24.993	15.293	30.04 *(-)	27.164	0.1	9.100	6.357	15.009	20.261
0.25	21.599	8.025	20.952	15.213	0.25	11.800	11.192	9.117	6.000
0.75	22.95	8.332	18.689	15.004	0.75	24.987	7.593	8.607	7.656
0.9	22.762	8.352	20.317	6.549	0.9	26.697	9.761	6.963	9.431
Growth stage of financialization: 2004 - 2011									
Panel A: CBOT Corn					Panel B: CBOT Soybean				
0.1	27.24	26.772	17.915	25.053	0.1	37.219	39.794 *(+)	21.991	33.291
0.25	13.064	22.496	21.414	18.345	0.25	35.885 *(+)	35.938 *(+)	16.2	25.637
0.75	11.545	5.89	12.812	18.839	0.75	14.57	16.913	12.44	11.763
0.9	9.604	10.972	18.259	18.382	0.9	10.744	11.159	16.022	5.249
Panel C: CBOT Wheat					Panel D: KCBOT Wheat				
0.1	40.535	34.081 *(+)	14.822	25.71	0.1	6.725	11.343	13.566	11.105
0.25	29.595	28.706	27.808	28.598	0.25	14.098	24.417	13.705	11.691
0.75	22.344	7.783	25.153	21.008	0.75	14.284	11.137	8.342	7.308
0.9	22.896	10.741	21.812	12.502	0.9	14.928	10.895	8.637	13.588
Post-financialization stage: 2012 - 2019									
Panel A: CBOT Corn					Panel B: CBOT Soybean				
0.1	14.799	14.29	13.274	9.493	0.1	22.444	13.883	17.286	12.096
0.25	14.069	20.406	4.873	8.686	0.25	13.541	13.345	15.209	16.196
0.75	12.023	18.083	12.415	16.775	0.75	13.527	6.328	9.501	6.501
0.9	8.898	14.188	13.131	10.791	0.9	12.939	7.403	12.456	7.435
Panel C: CBOT Wheat					Panel D: KCBOT Wheat				
0.1	17.224	28.496 *(-)	26.036	21.386	0.1	14.174	15.111	21.053	28.241
0.25	13.182	20.747	29.407 *(-)	7.853	0.25	10.708	11.255	20.907	18.533
0.75	9.107	15.148	14.77	15.304	0.75	31.541 *(-)	7.661	9.327	19.751
0.9	14.095	9.699	17.082	15.094	0.9	19.445	10.073	15.602	25.202

Notes: * indicates statistical significance at 5%. Box-Ljung test statistics for 13 lags are in the table. The sign (+/-) next to the test statistics indicates the dominant sign of the underlying CQ estimates for the Box-Ljung test statistics.

Table B2: Cross-quantilogram portmanteau test results for impacts of futures returns to the percentage change in CIT net long positions in four grain futures markets for full sample, the growth stage of financialization, and the post-financialization stage

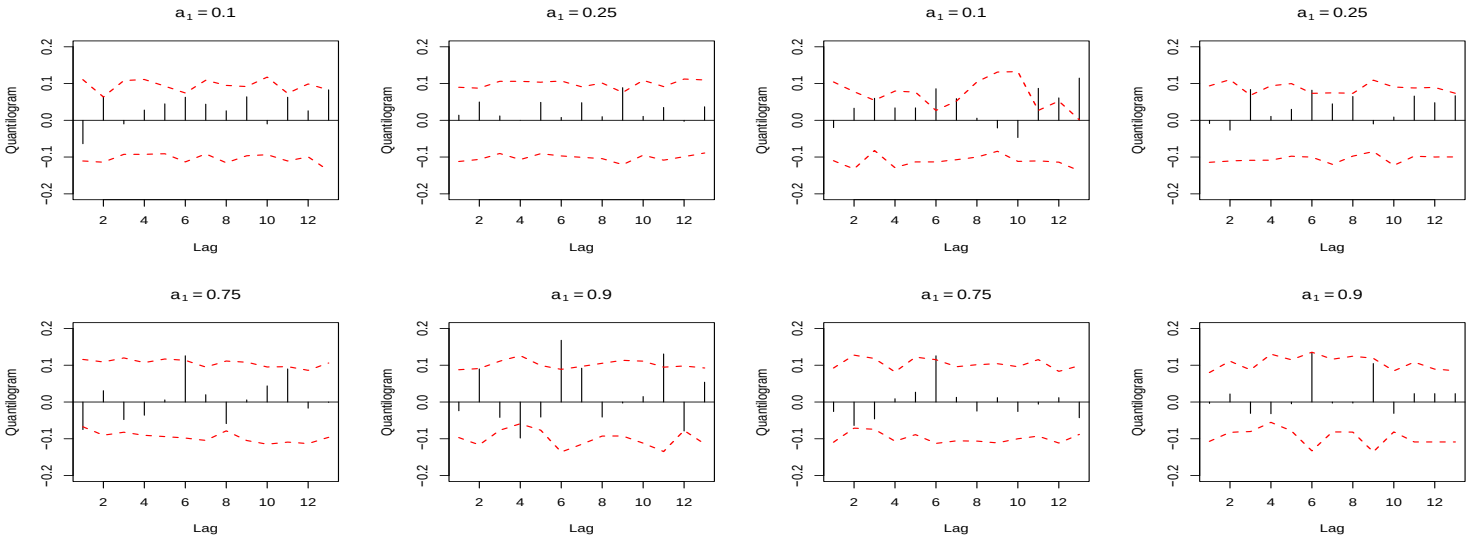
Dependent variable: CIT Net Long % Change, independent variable: returns									
	CIT Net Long % Change Quantile Level					CIT Net Long % Change Quantile Level			
Returns Quantile Level	0.1	0.25	0.75	0.9	Returns Quantile Level	0.1	0.25	0.75	0.9
Full sample: 2004 - 2019									
Panel A: CBOT Corn					Panel B: CBOT Soybean				
0.1	8.586	18.635	12.928	29.433 *(+)	0.1	17.959	23.64	12.465	4.945
0.25	19.736	20.748	12.198	24.686	0.25	15.048	24.334	13.77	6.53
0.75	20.023	26.130	8.209	5.605	0.75	19.324	16.573	23.638	16.09
0.9	16.200	17.824	12.995	12.112	0.9	10.748	16.04	6.694	7.336
Panel C: CBOT Wheat					Panel D: KCBOT Wheat				
0.1	32.49 *(+)	16.822	10.545	6.815	0.1	16.079	14.808	19.759	14.548
0.25	15.547	18.91	16.239	8.515	0.25	22.741	18.672	13.55	20.206
0.75	16.791	17.47	15.36	15.618	0.75	14.674	19.28	16.077	8.801
0.9	24.223	20.397	11.66	15.53	0.9	18.434	17.267	17.279	5.529
Growth stage of financialization: 2004 - 2011									
Panel A: CBOT Corn					Panel B: CBOT Soybean				
0.1	8.681	8.952	13.565	24.105	0.1	71.098 *(+)	51.831 *(+)	8.008	4.931
0.25	23.258	9.901	7.358	13.99	0.25	47.296 *(+)	28.169	23.505	21.537
0.75	10.38	17.637	12.771	11.235	0.75	27.484	20.052	18.206	12.917
0.9	13.941	9.726	20.594	14.573	0.9	21.698	11.887	6.368	12.58
Panel C: CBOT Wheat					Panel D: KCBOT Wheat				
0.1	24.331	14.927	17.995	11.269	0.1	12.084	12.226	18.486	19.558
0.25	8.175	19.379	29.811 *(+)	10.892	0.25	17.198	20.972	15.398	18.075
0.75	31.305 *(-)	21.467	15.438	19.219	0.75	8.423	16.776	14.458	8.342
0.9	24.392	20.056	14.852	28.442 *(-)	0.9	14.359	21.182	16.791	7.192
Post-financialization stage: 2012 - 2019									
Panel A: CBOT Corn					Panel B: CBOT Soybean				
0.1	11.762	9.516	15.824	13.135	0.1	28.896 *(-)	37.041 *(-)	15.441	12.786
0.25	13.616	10.172	22.837	13.826	0.25	12.359	19.769	19.782	9.605
0.75	17.795	18.747	9.355	9.509	0.75	13.111	9.846	23.758	19.025
0.9	9.624	19.697	14.972	20.386	0.9	12.439	8.928	22.797	18.437
Panel C: CBOT Wheat					Panel D: KCBOT Wheat				
0.1	27.132	18.719	14.85	11.449	0.1	21.826	14.236	24.764	25.4
0.25	16.68	22.269	18.255	13.514	0.25	25.678	11.677	15.478	24.189
0.75	13.53	19.36	17.689	5.766	0.75	26.76	19.337	10.033	12.271
0.9	19.71	10.446	13.516	10.553	0.9	16.829	17.71	6.55	11.611

Notes: * indicates statistical significance at 5%. Box-Ljung test statistics for 13 lags are in the table. The sign (+/-) next to the test statistics indicates the dominant sign of the underlying CQ estimates for the Box-Ljung test statistics.



(a) Position change at quantile level 0.1

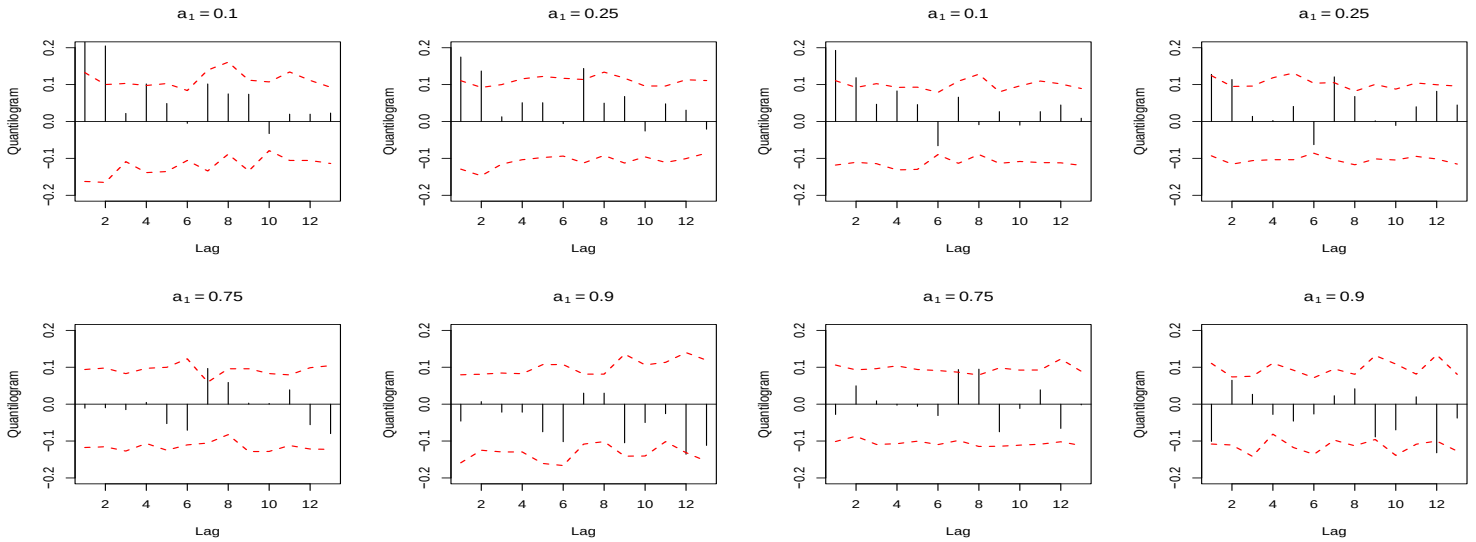
(b) Position change at quantile level 0.25



(c) Position change at quantile level 0.75

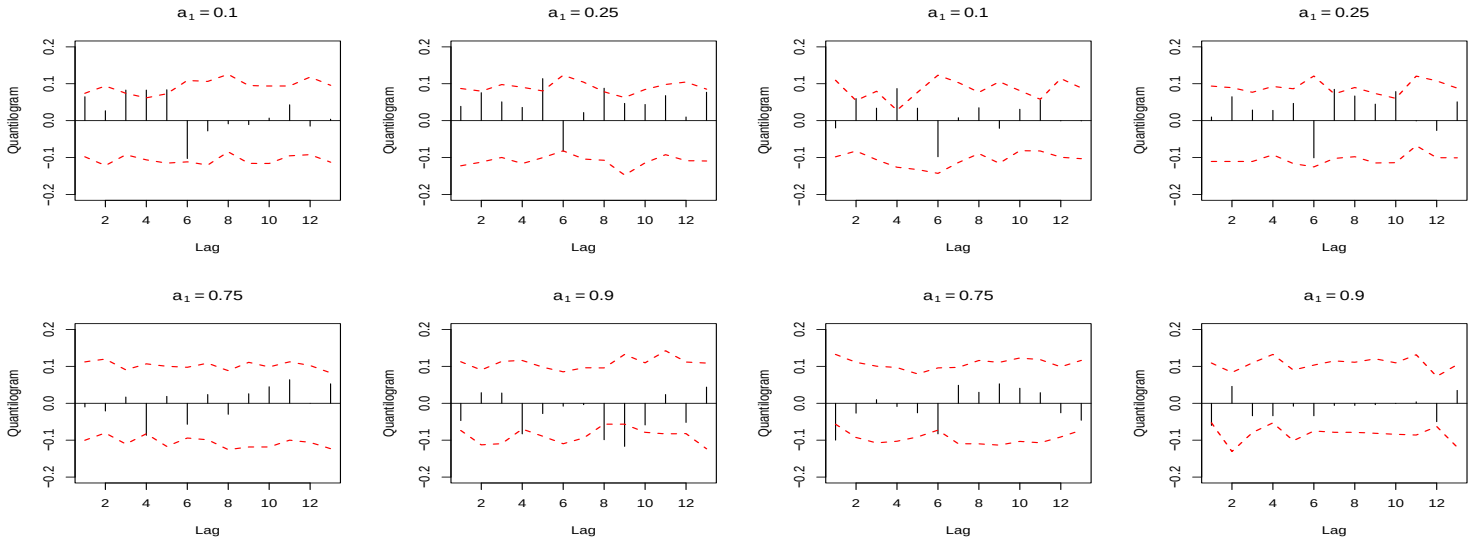
(d) Position change at quantile level 0.9

Figure B1: Cross-quantilogram from changes in CIT net long positions to returns in the CBOT corn futures market, 2004 – 2011



(a) Position change at quantile level 0.1

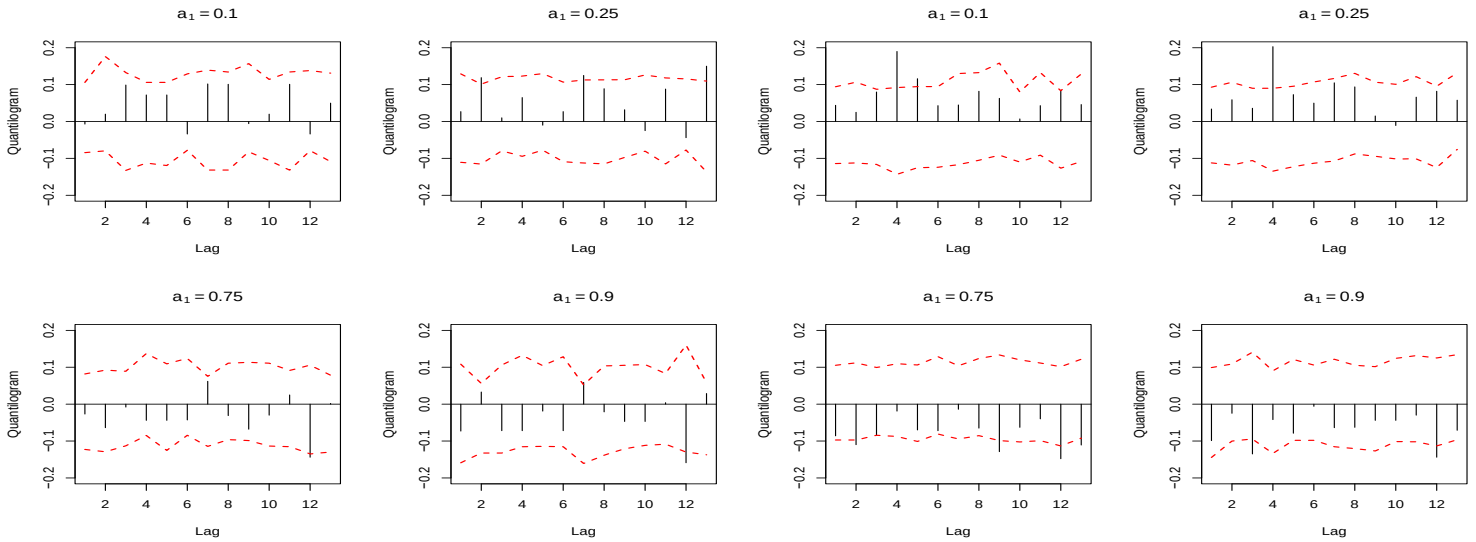
(b) Position change at quantile level 0.25



(c) Position change at quantile level 0.75

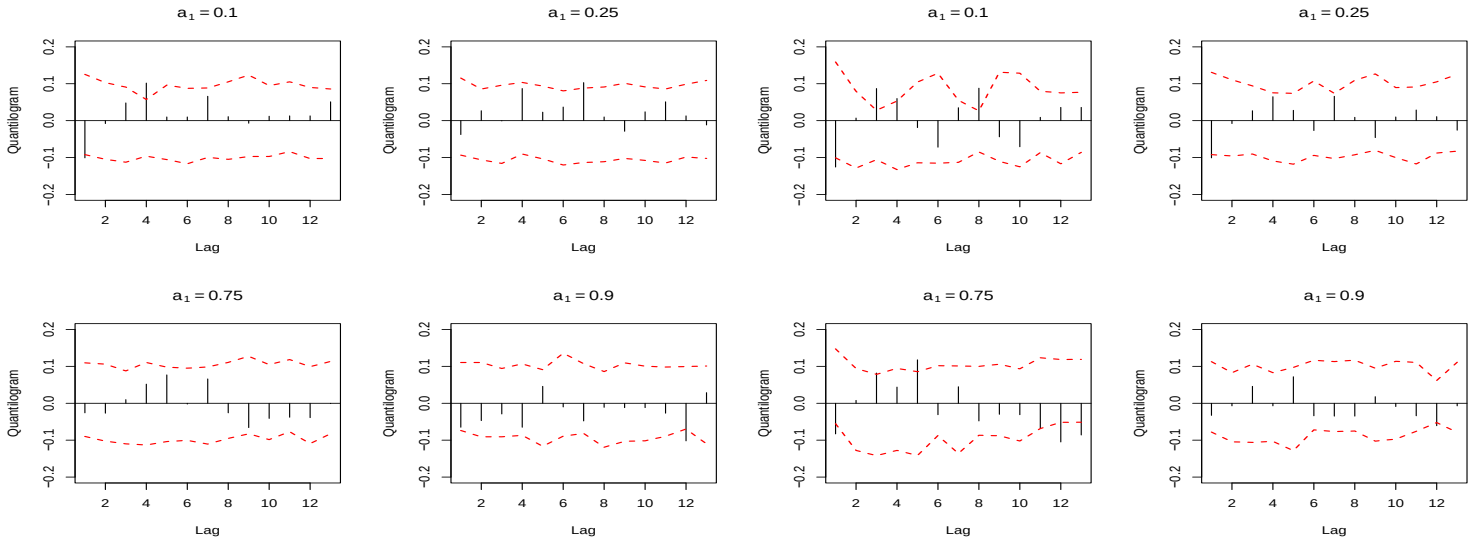
(d) Position change at quantile level 0.9

Figure B2: Cross-quantilogram from changes in CIT net long positions to returns in the CBOT soybean futures market, 2004 – 2011



(a) Position change at quantile level 0.1

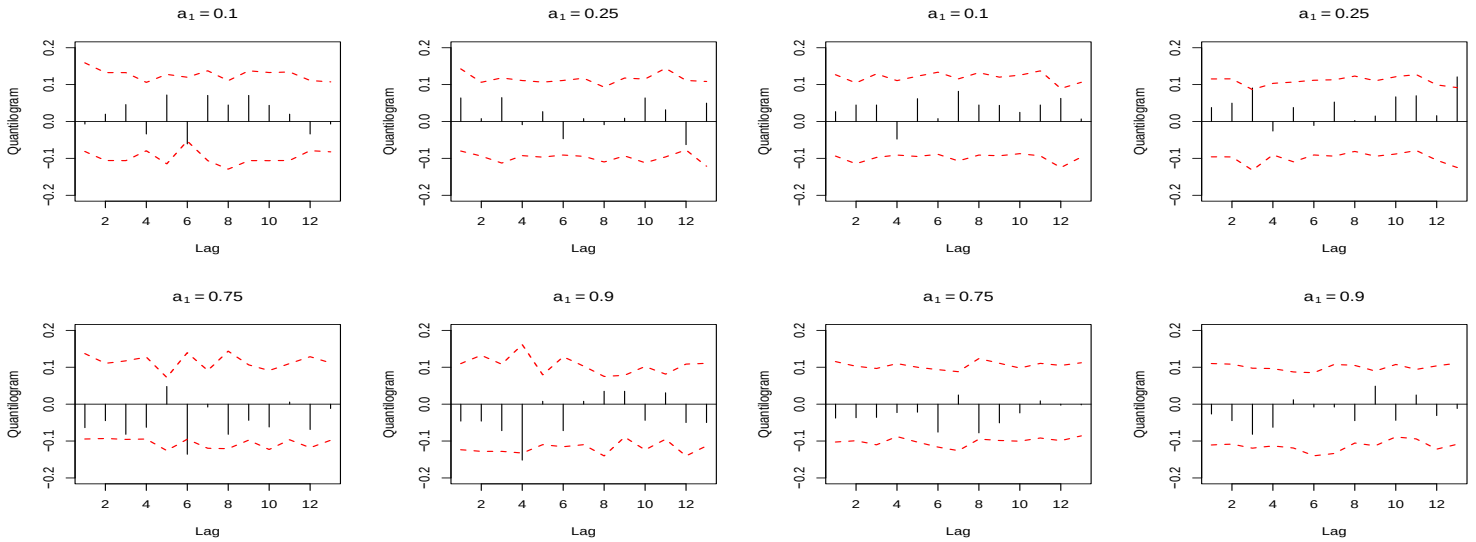
(b) Position change at quantile level 0.25



(c) Position change at quantile level 0.75

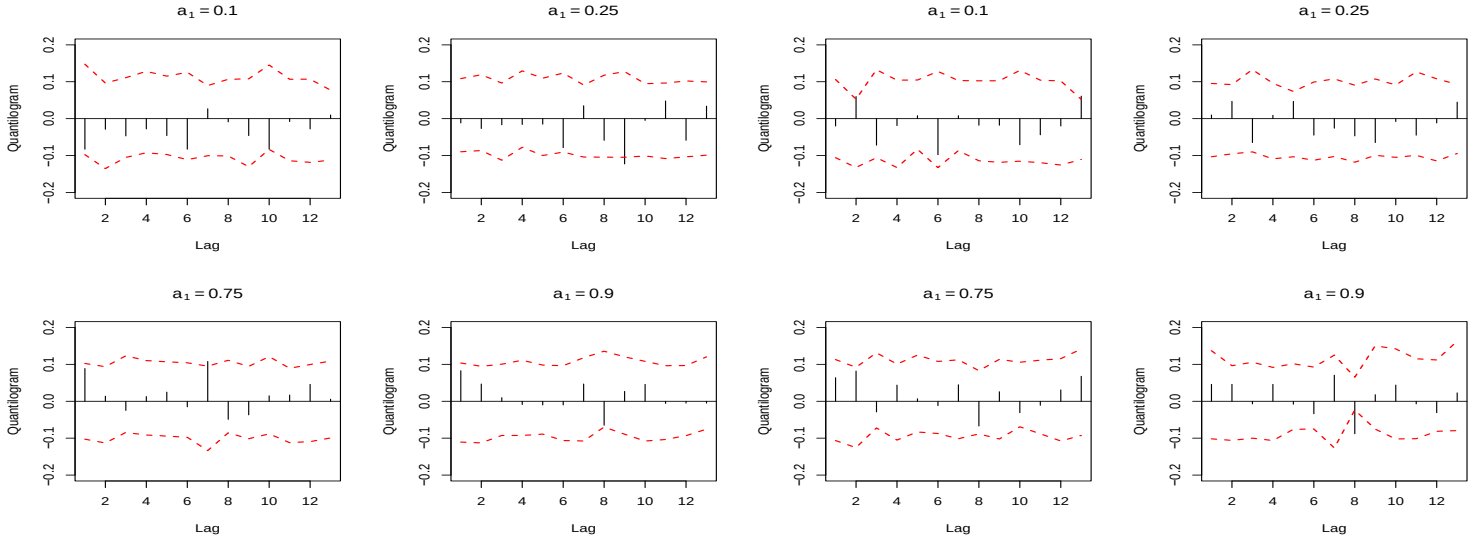
(d) Position change at quantile level 0.9

Figure B3: Cross-quantilogram from changes in CIT net long positions to returns in the CBOT wheat futures market, 2004 – 2011



(a) Position change at quantile level 0.1

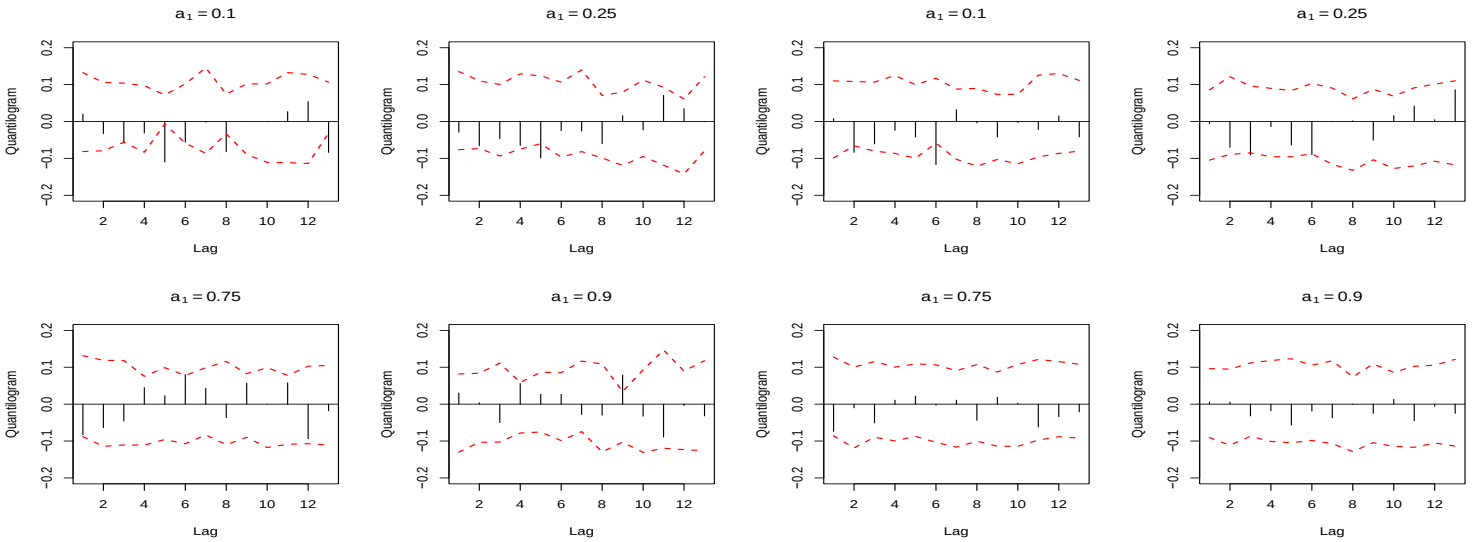
(b) Position change at quantile level 0.25



(c) Position change at quantile level 0.75

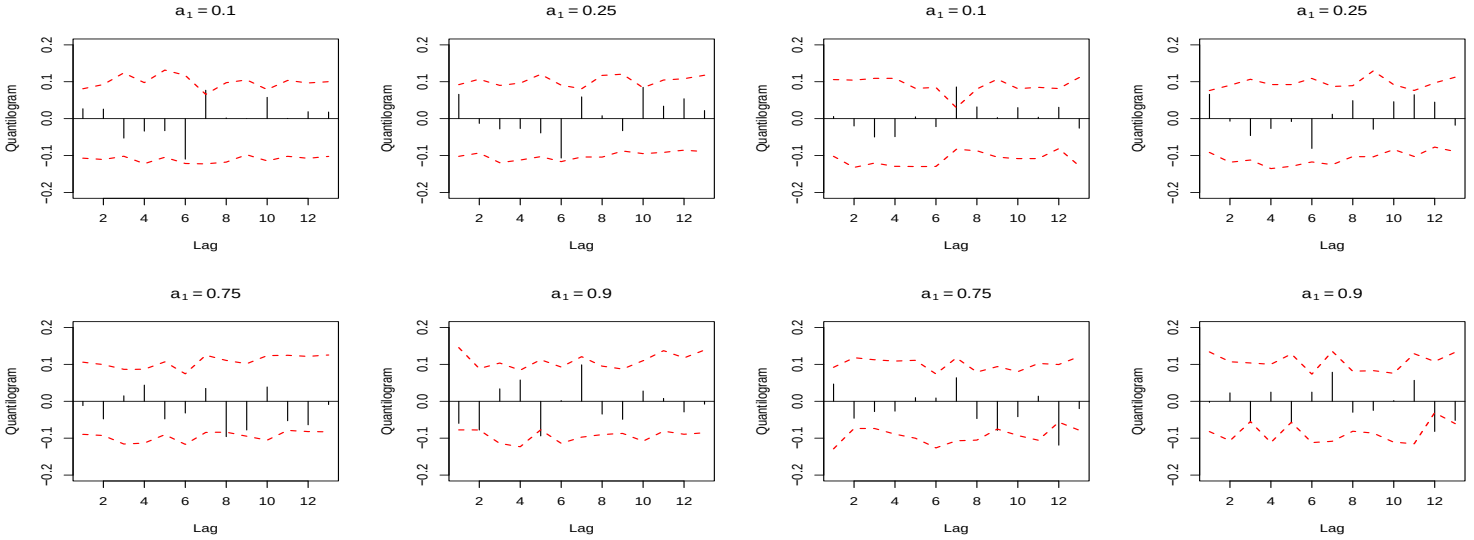
(d) Position change at quantile level 0.9

Figure B4: Cross-quantilogram from changes in CIT net long positions to returns in the KCBOT wheat futures market, 2004 – 2011



(a) Position change at quantile level 0.1

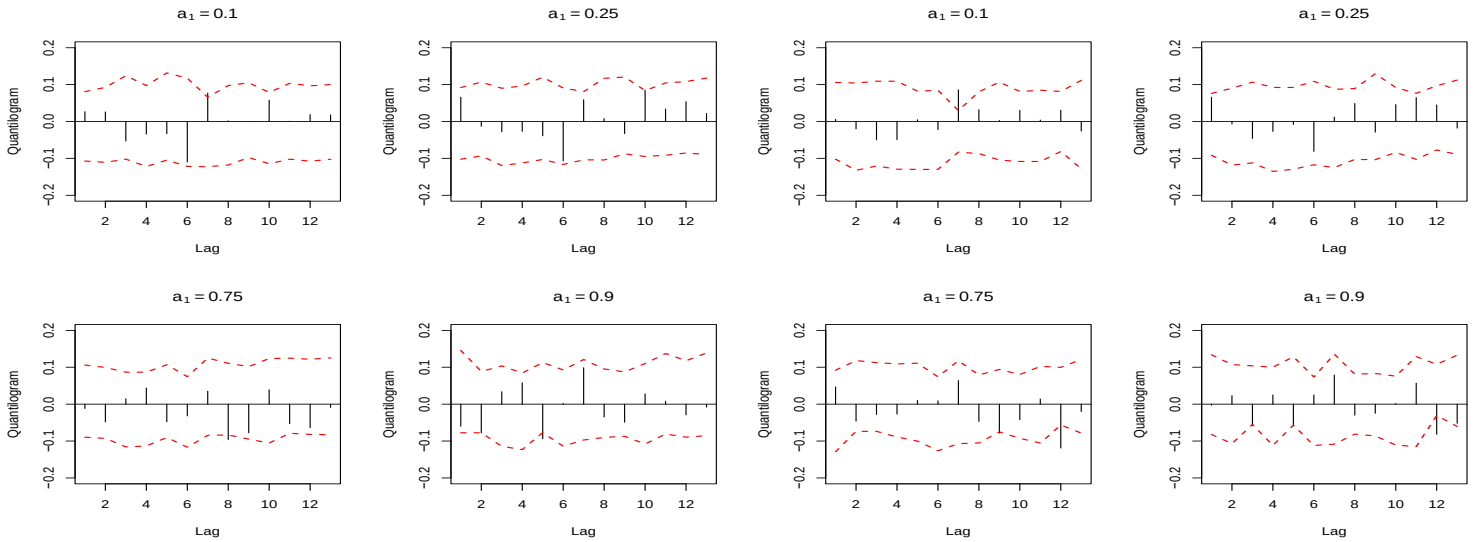
(b) Position change at quantile level 0.25



(c) Position change at quantile level 0.75

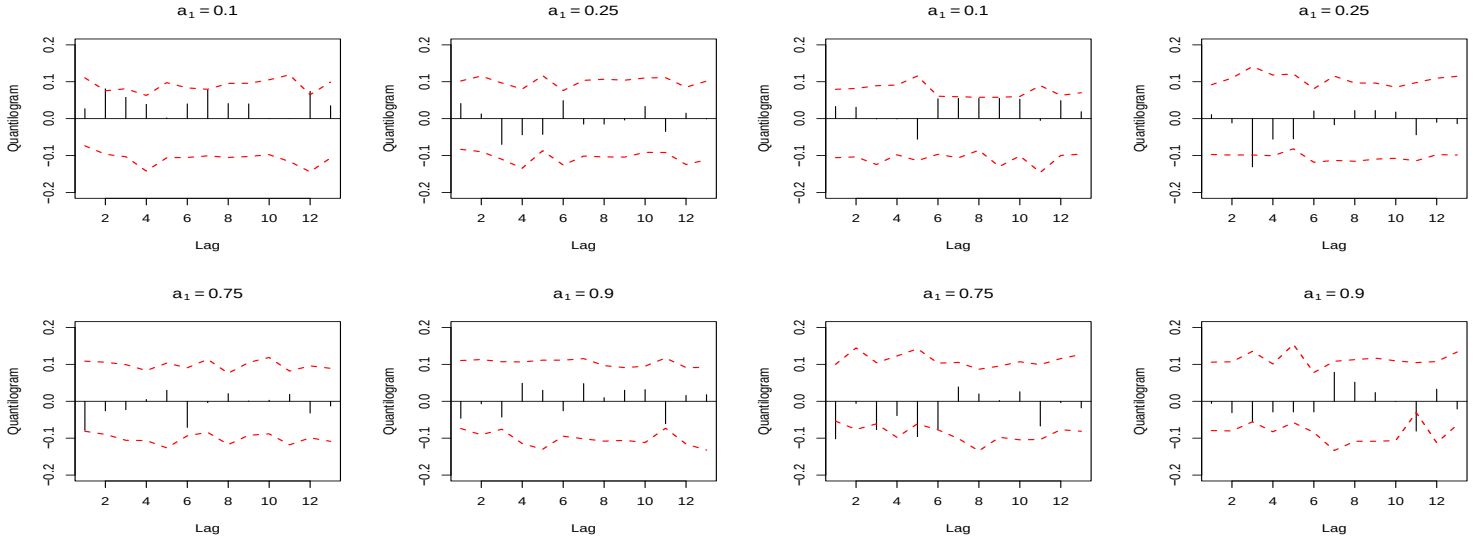
(d) Position change at quantile level 0.9

Figure B5: Cross-quantilogram from changes in CIT net long positions to returns in the CBOT corn futures market, 2012 – 2019



(a) Position change at quantile level 0.1

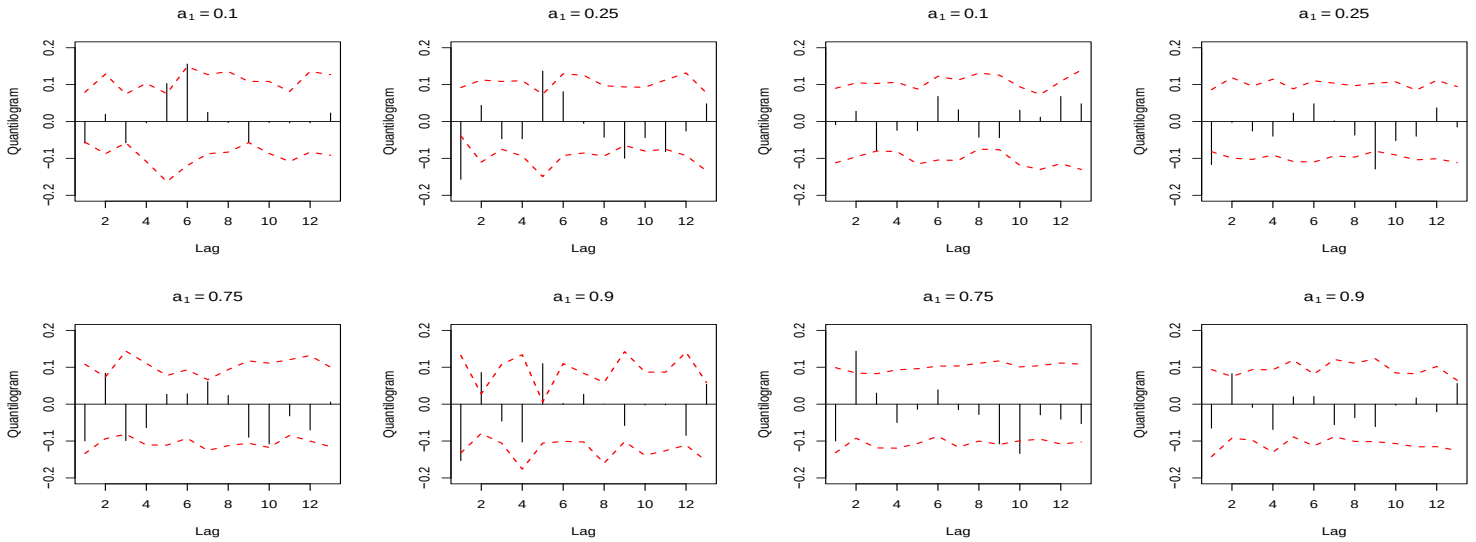
(b) Position change at quantile level 0.25



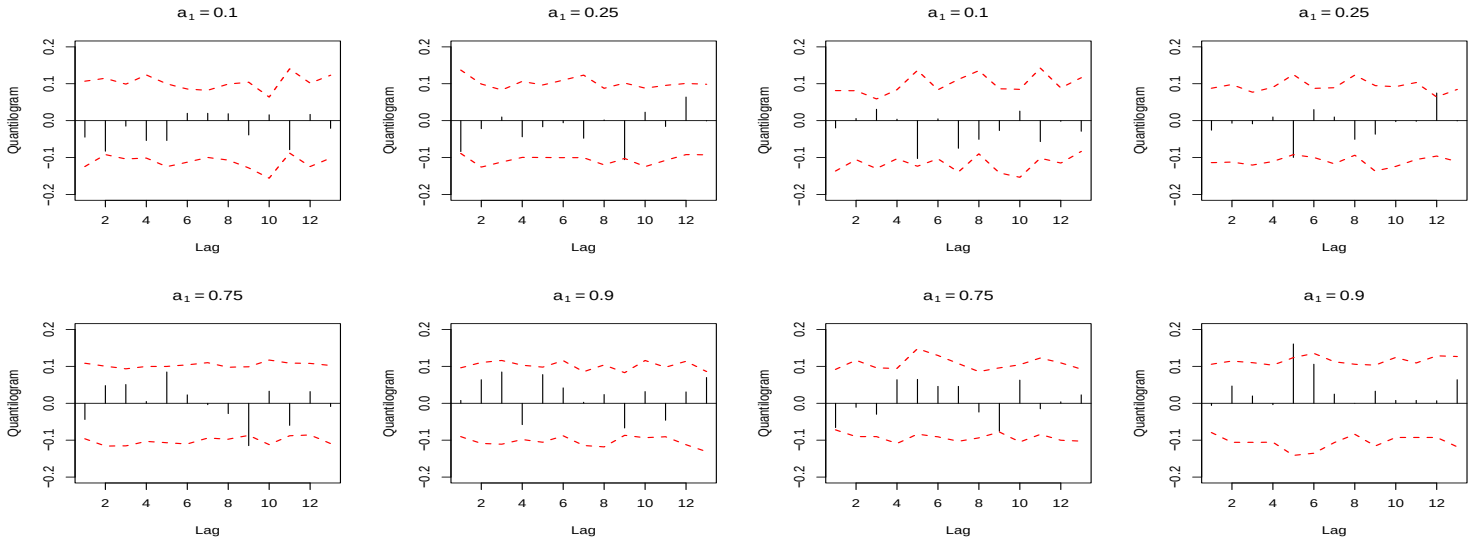
© Position change at quantile level 0.75

(d) Position change at quantile level 0.9

Figure B6: Cross-quantilogram from changes in CIT net long positions to returns in the CBOT soybean futures market, 2012 – 2019

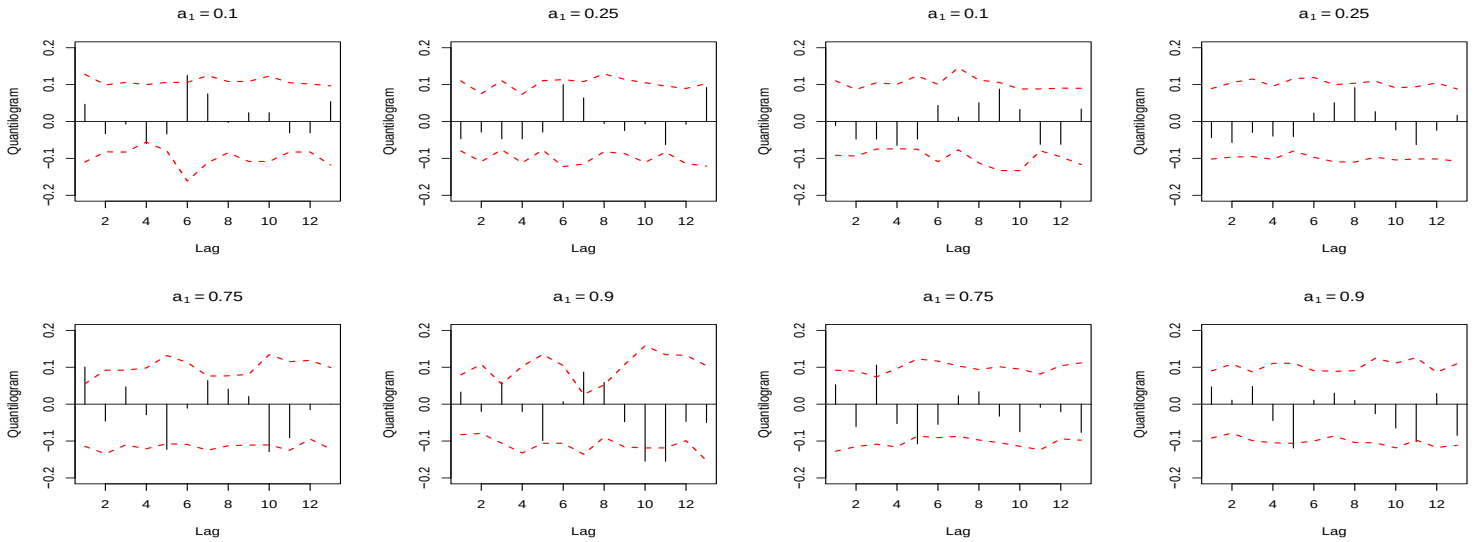


(a) Position change at quantile level 0.1 (b) Position change at quantile level 0.25



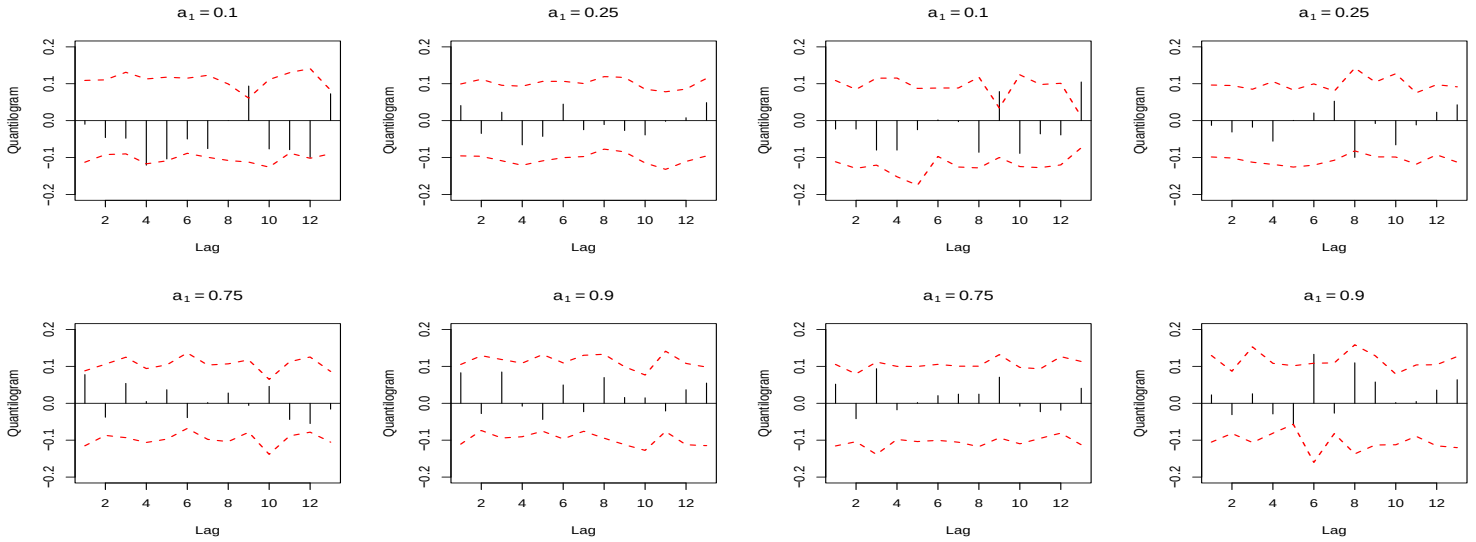
(c) Position change at quantile level 0.75 (d) Position change at quantile level 0.9

Figure B7: Cross-quantilogram from changes in CIT net long positions to returns in the CBOT wheat futures market, 2012 – 2019



(a) Position change at quantile level 0.1

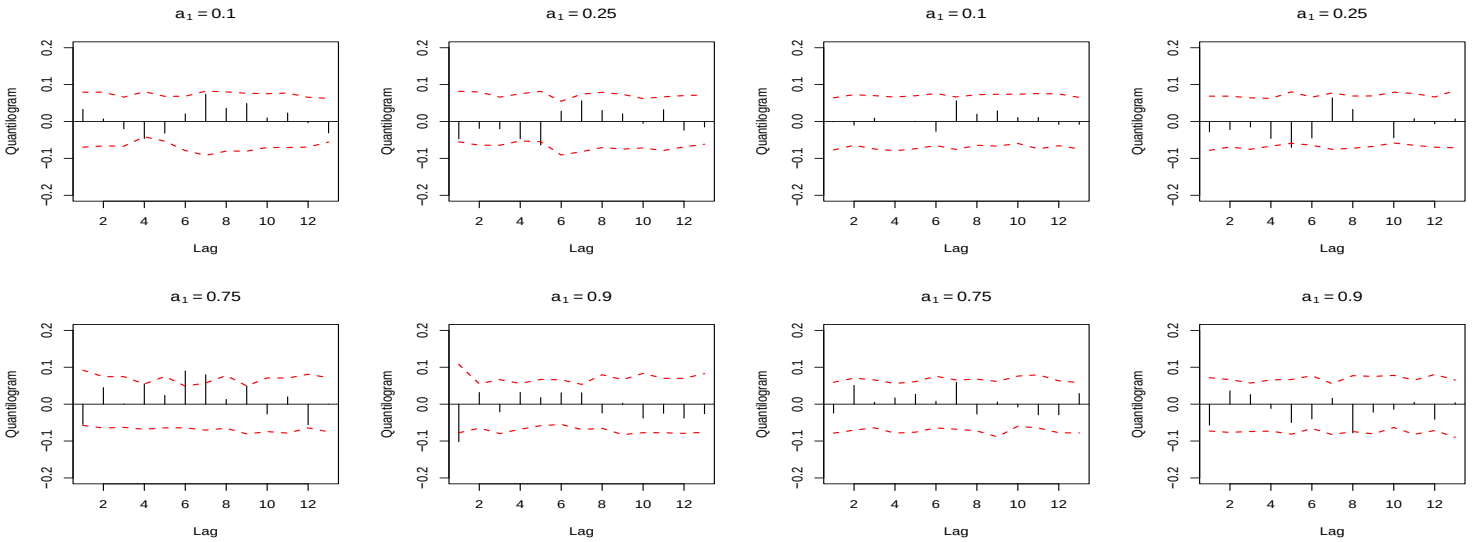
(b) Position change at quantile level 0.25



(c) Position change at quantile level 0.75

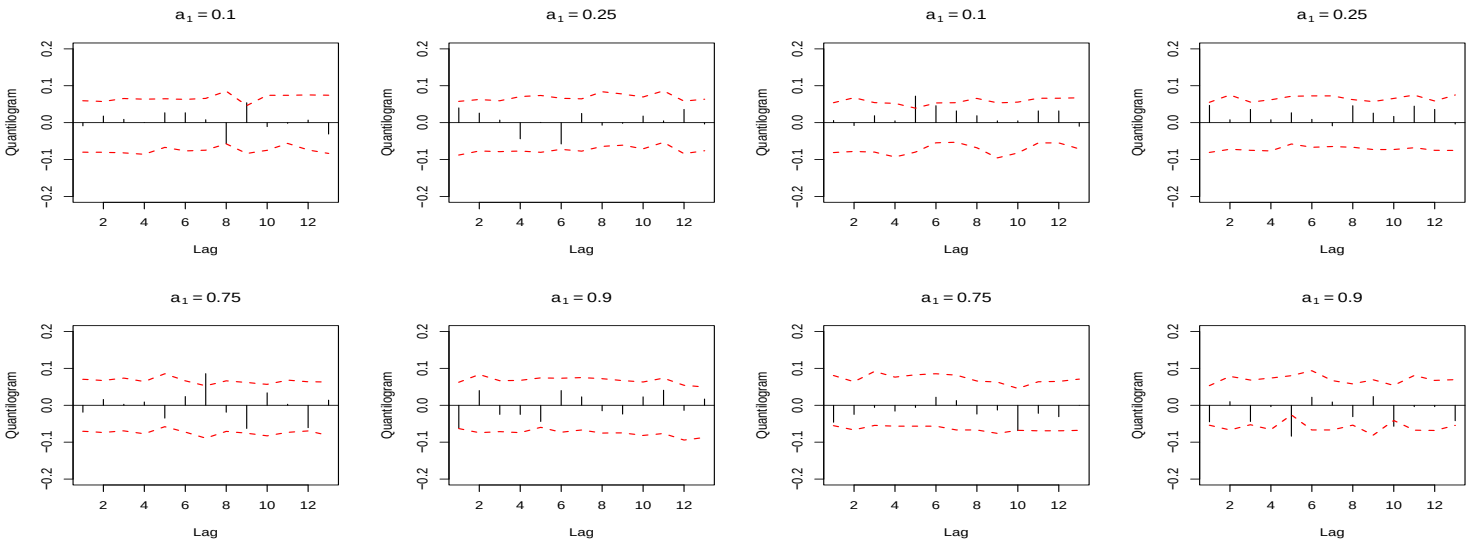
(d) Position change at quantile level 0.9

Figure B8: Cross-quantilogram from changes in CIT net long positions to returns in the KCBOT wheat futures market, 2012 – 2019



(a) Position change at quantile level 0.1

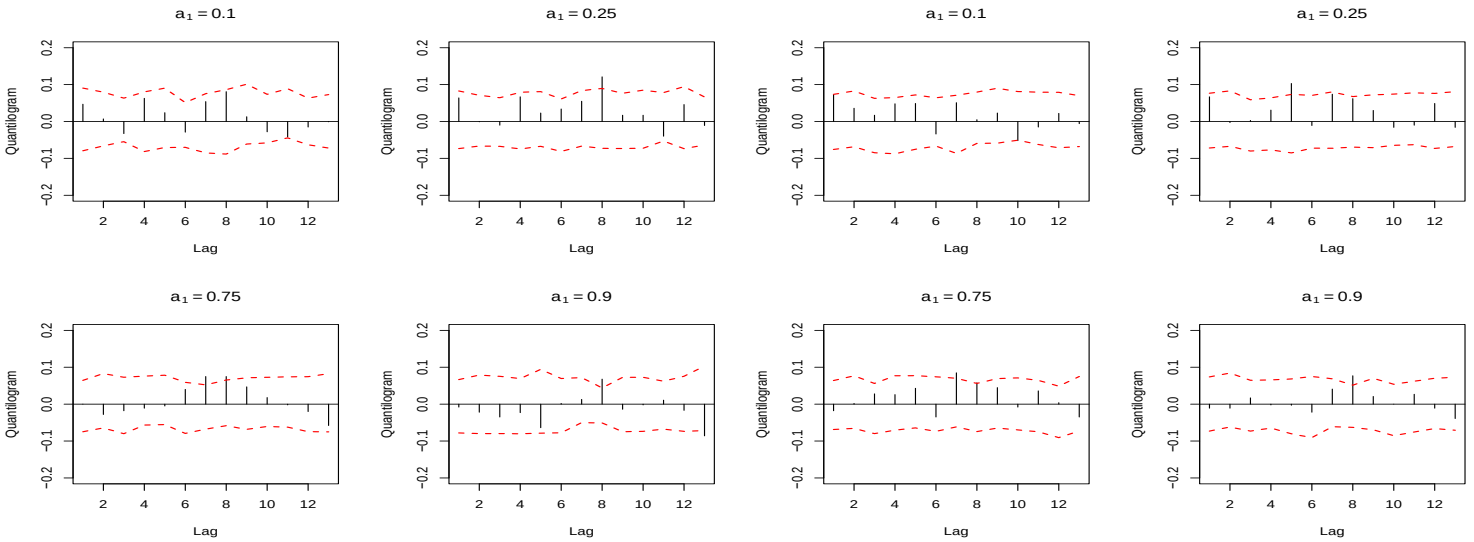
(b) Position change at quantile level 0.25



(c) Position change at quantile level 0.75

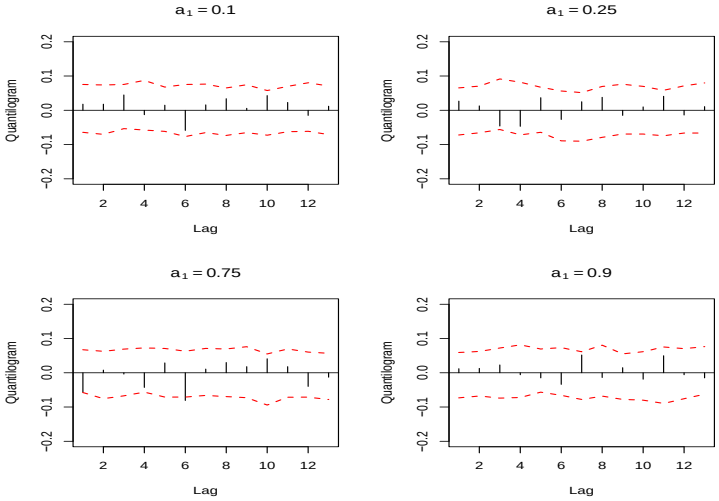
(d) Position change at quantile level 0.9

Figure B9: Cross-quantilogram from percentage changes in CIT net long positions to returns in the CBOT corn futures market, 2004 – 2019

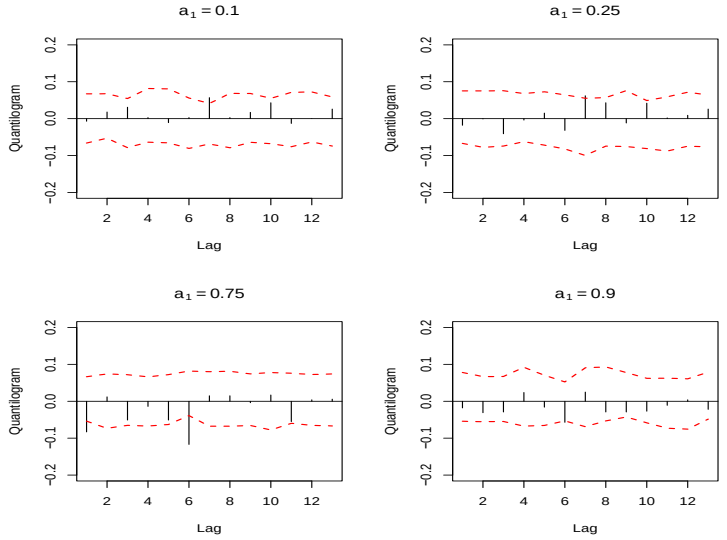


(a) Position change at quantile level 0.1

(b) Position change at quantile level 0.25

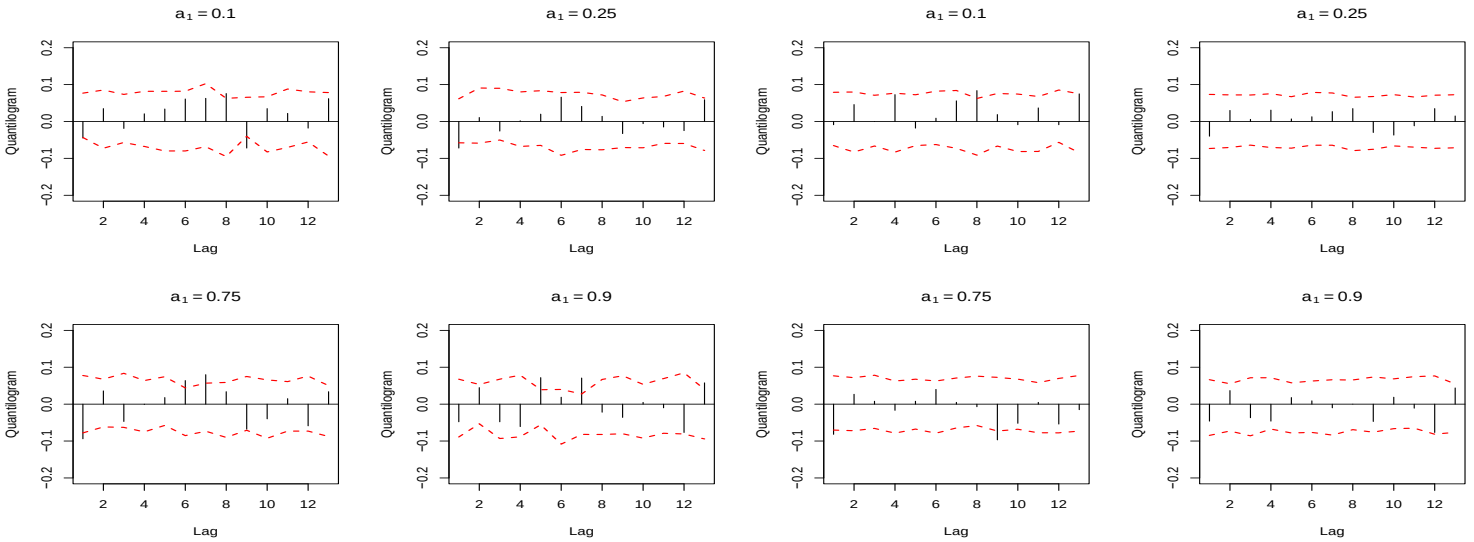


(c) Position change at quantile level 0.75

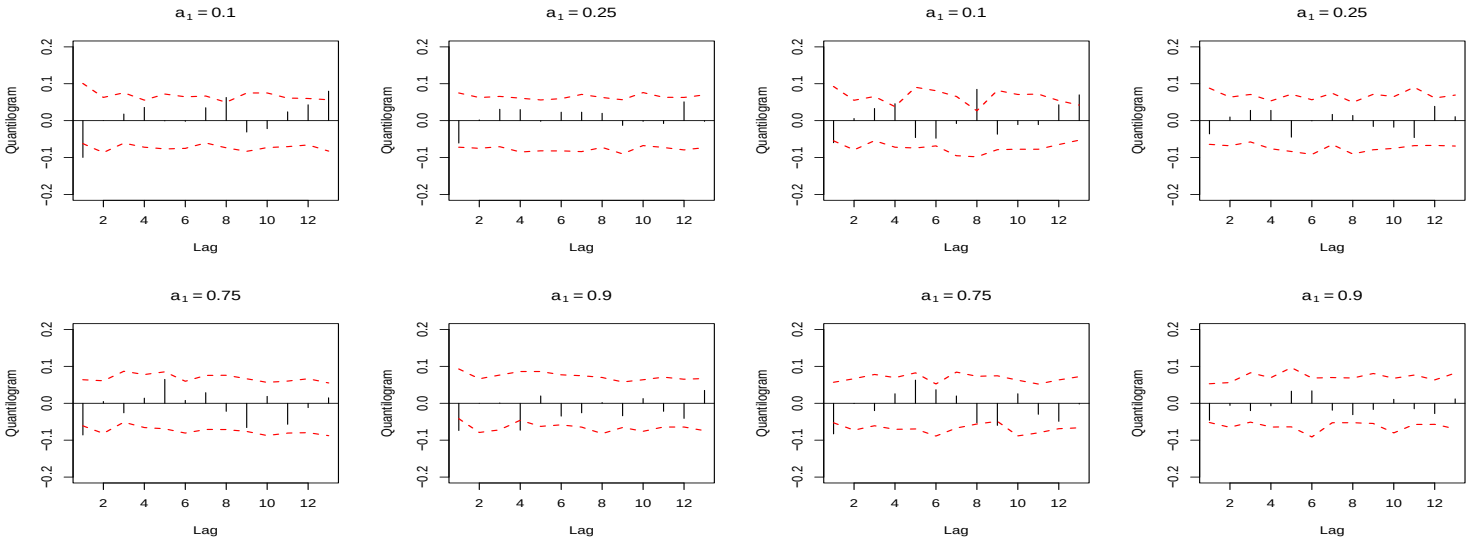


(d) Position change at quantile level 0.9

Figure B10: Cross-quantilogram from percentage changes in CIT net long positions to returns in the CBOT soybean futures market, 2004 – 2019

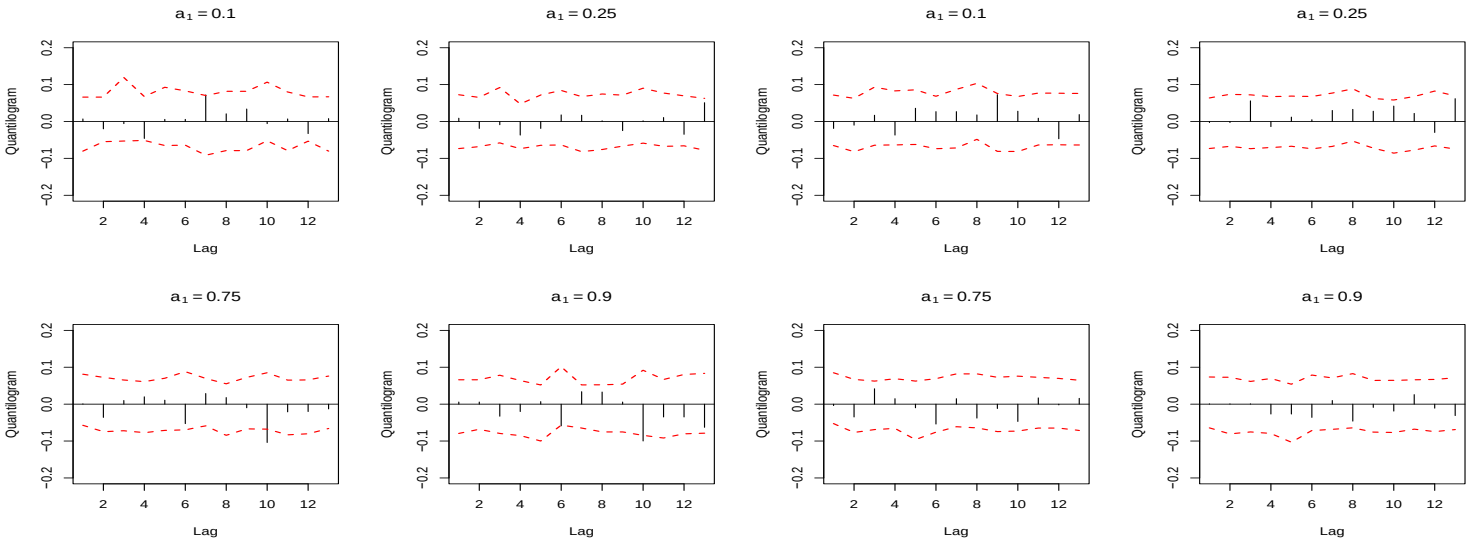


(a) Position change at quantile level 0.1 (b) Position change at quantile level 0.25



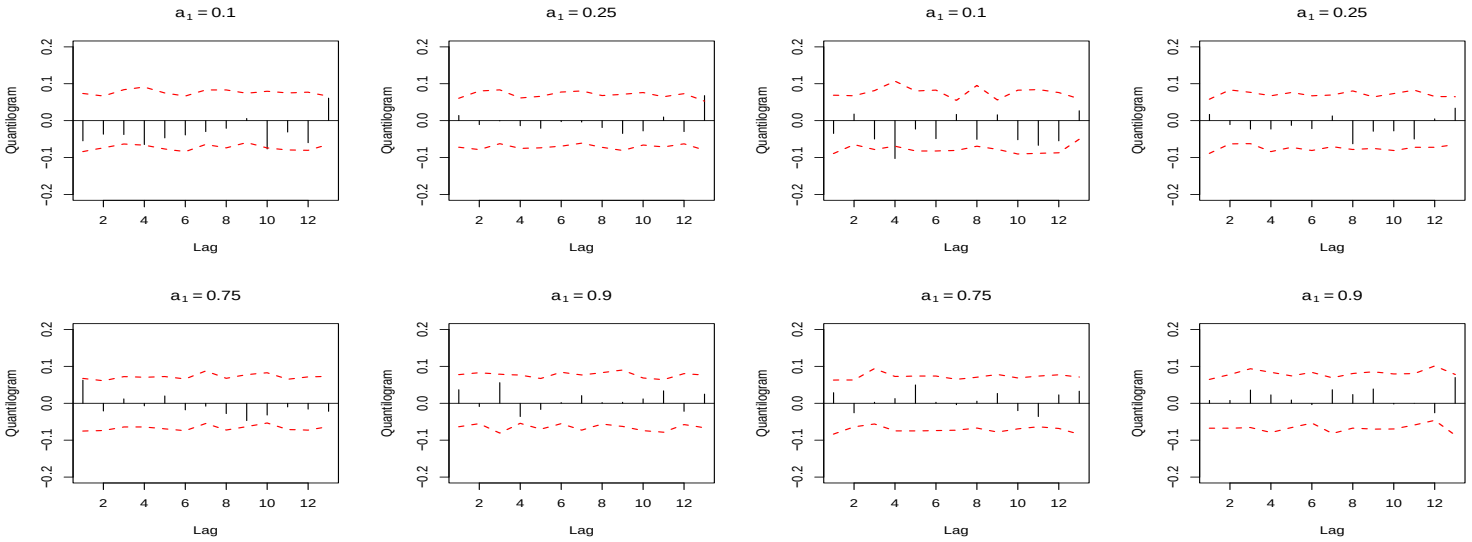
(c) Position change at quantile level 0.75 (d) Position change at quantile level 0.9

Figure B11: Cross-quantilogram from percentage changes in CIT net long positions to returns in the CBOT wheat futures market, 2004 – 2019



(a) Position change at quantile level 0.1

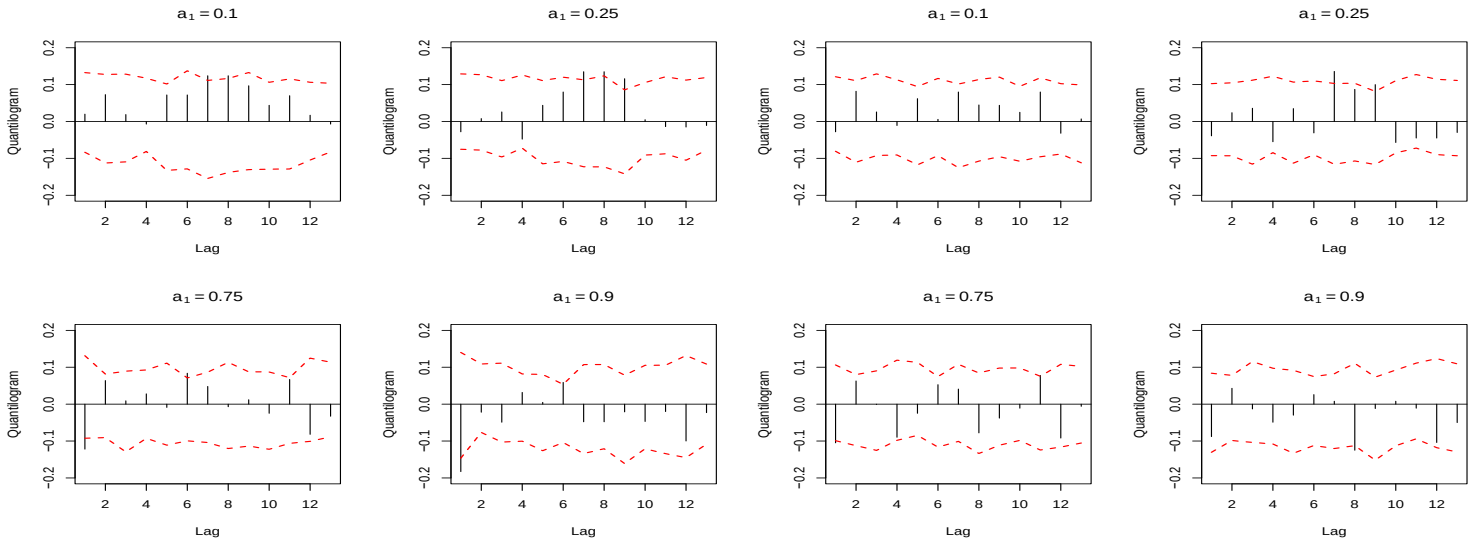
(b) Position change at quantile level 0.25



(c) Position change at quantile level 0.75

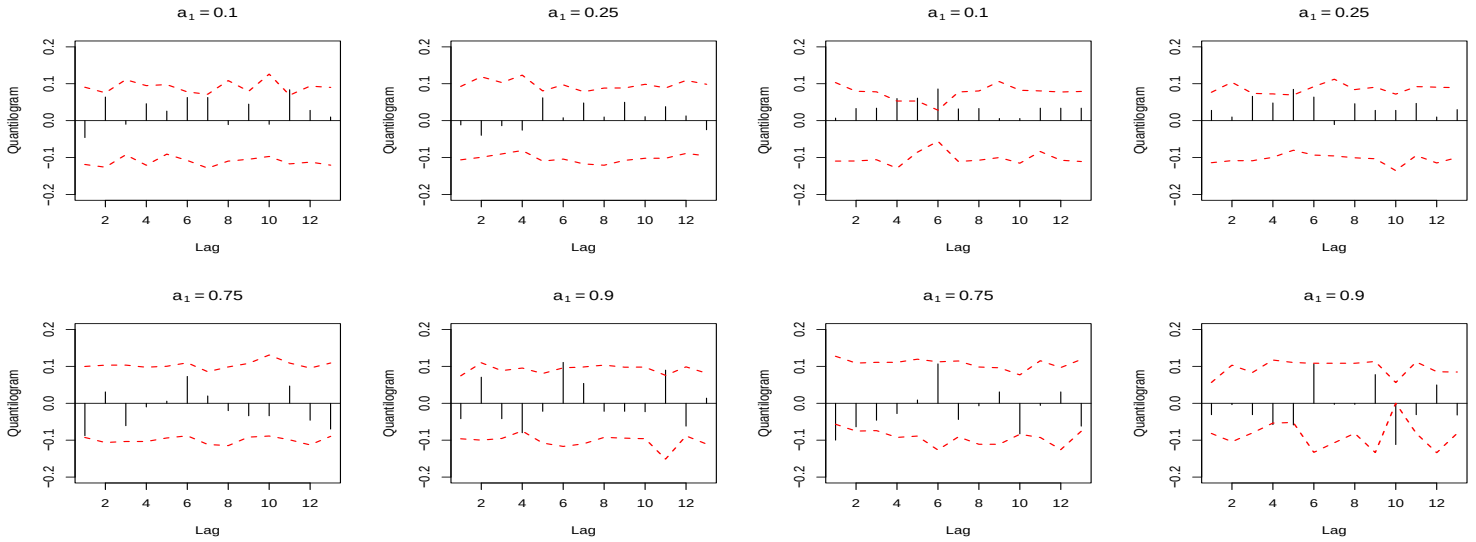
(d) Position change at quantile level 0.9

Figure B12: Cross-quantilogram from percentage changes in CIT net long positions to returns in the KCBOT wheat futures market, 2004 – 2019



(a) Position change at quantile level 0.1

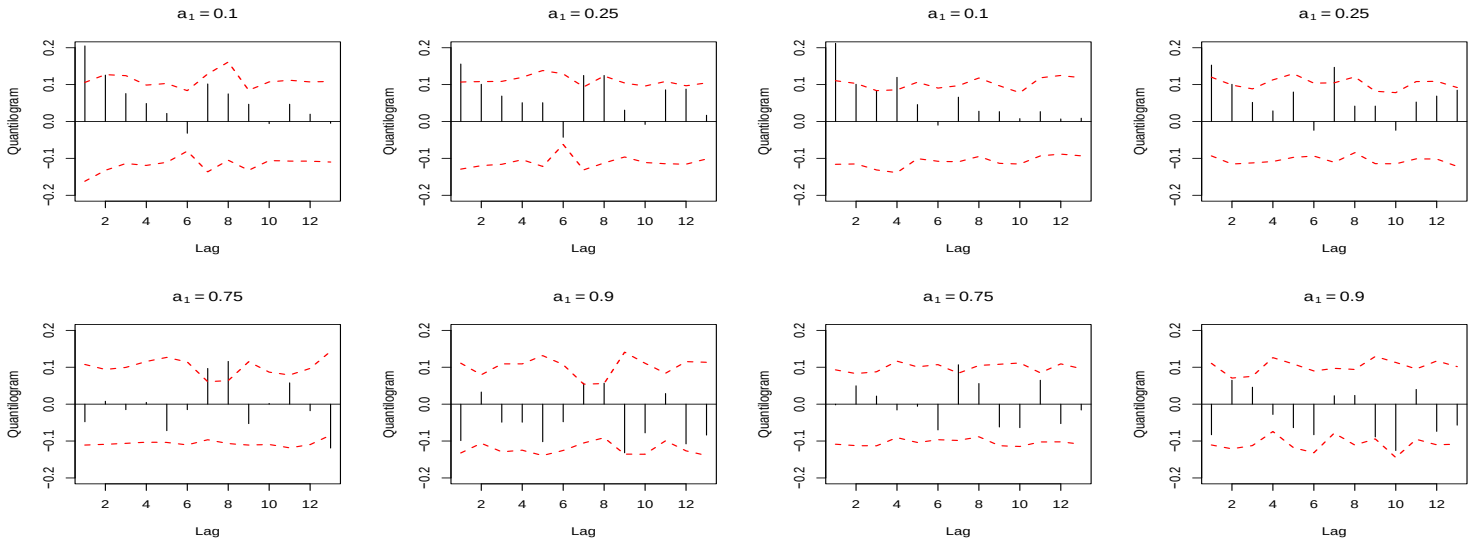
(b) Position change at quantile level 0.25



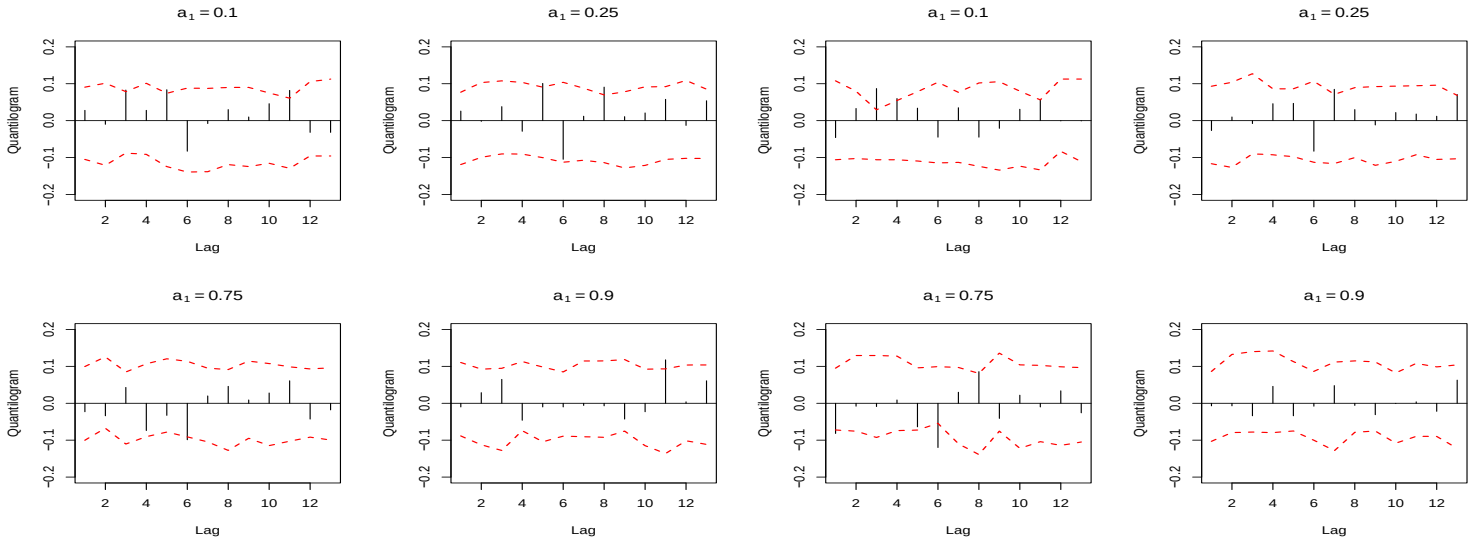
(c) Position change at quantile level 0.75

(d) Position change at quantile level 0.9

Figure B13: Cross-quantilogram from percentage changes in CIT net long positions to returns in the CBOT corn futures market, 2004 – 2011

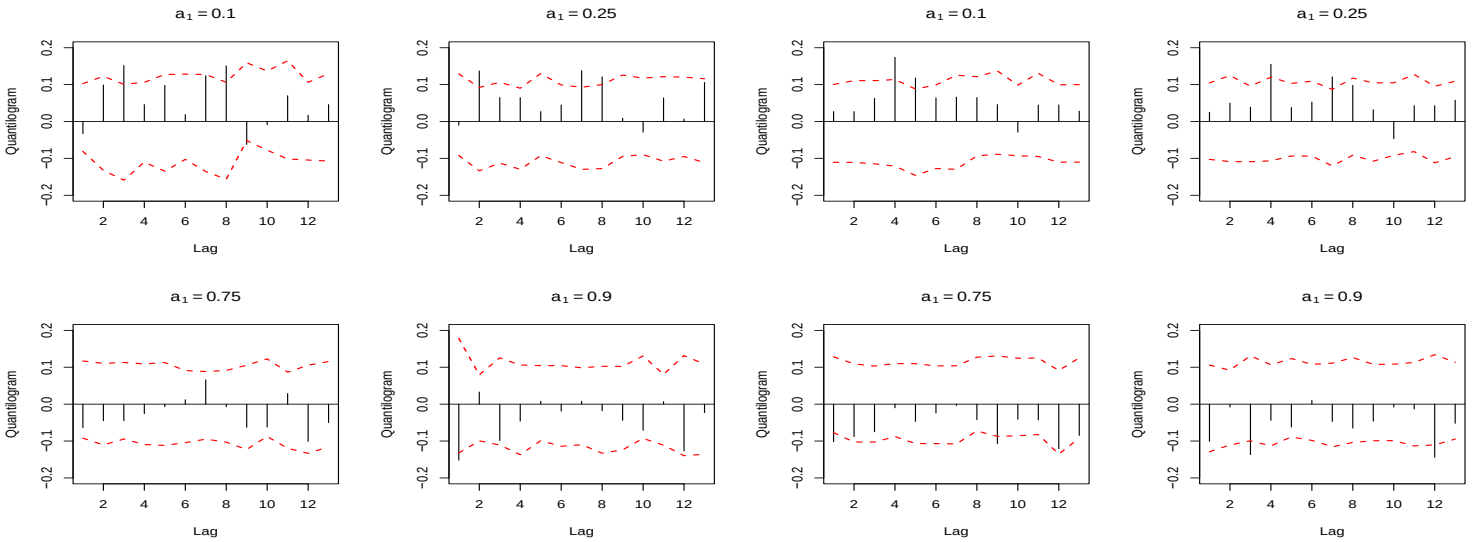


(a) Position change at quantile level 0.1 (b) Position change at quantile level 0.25

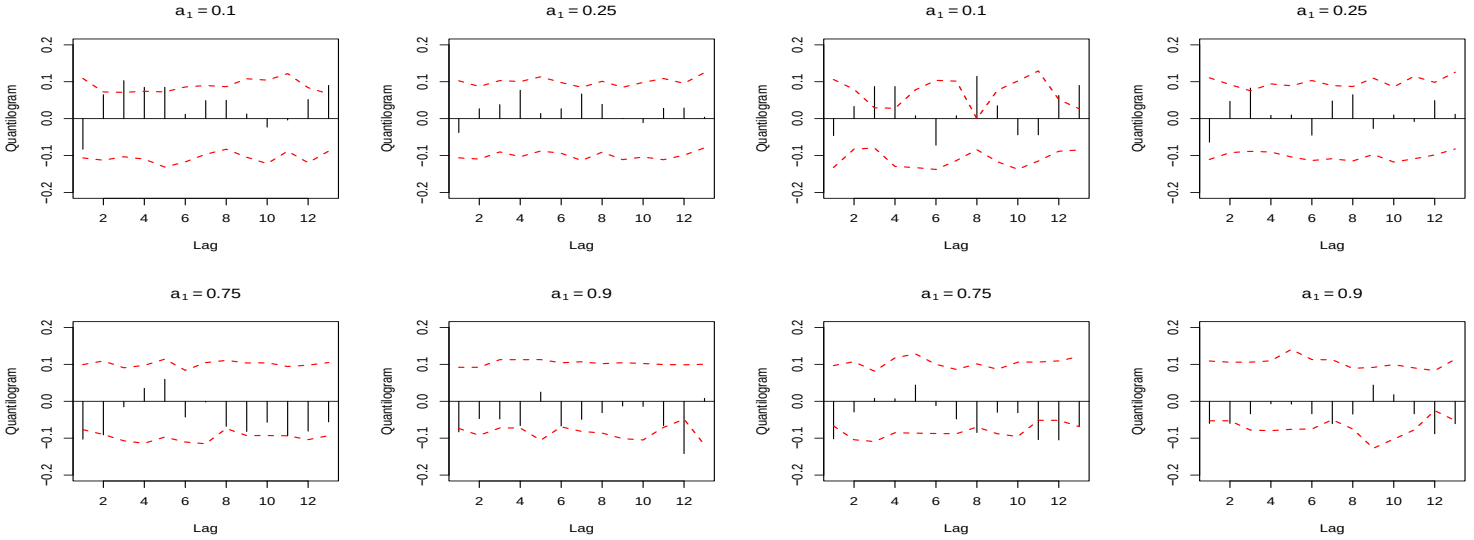


(c) Position change at quantile level 0.75 (d) Position change at quantile level 0.9

Figure B14: Cross-quantilogram from percentage changes in CIT net long positions to returns in the CBOT soybean futures market, 2004 – 2011

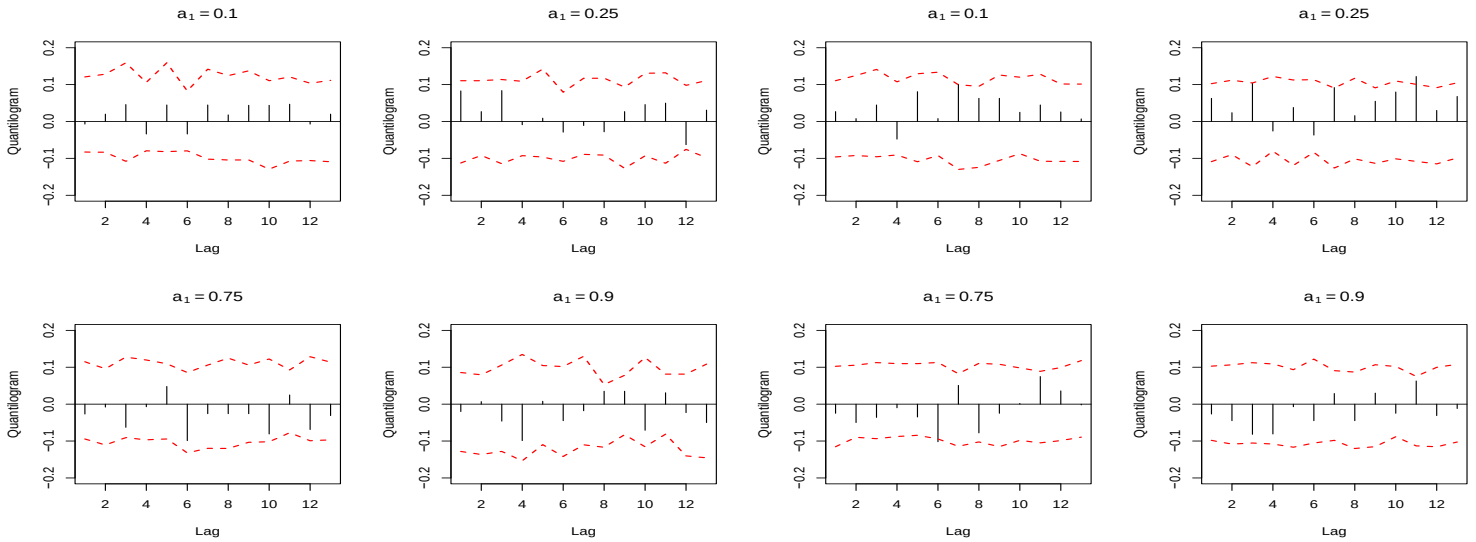


(a) Position change at quantile level 0.1 (b) Position change at quantile level 0.25



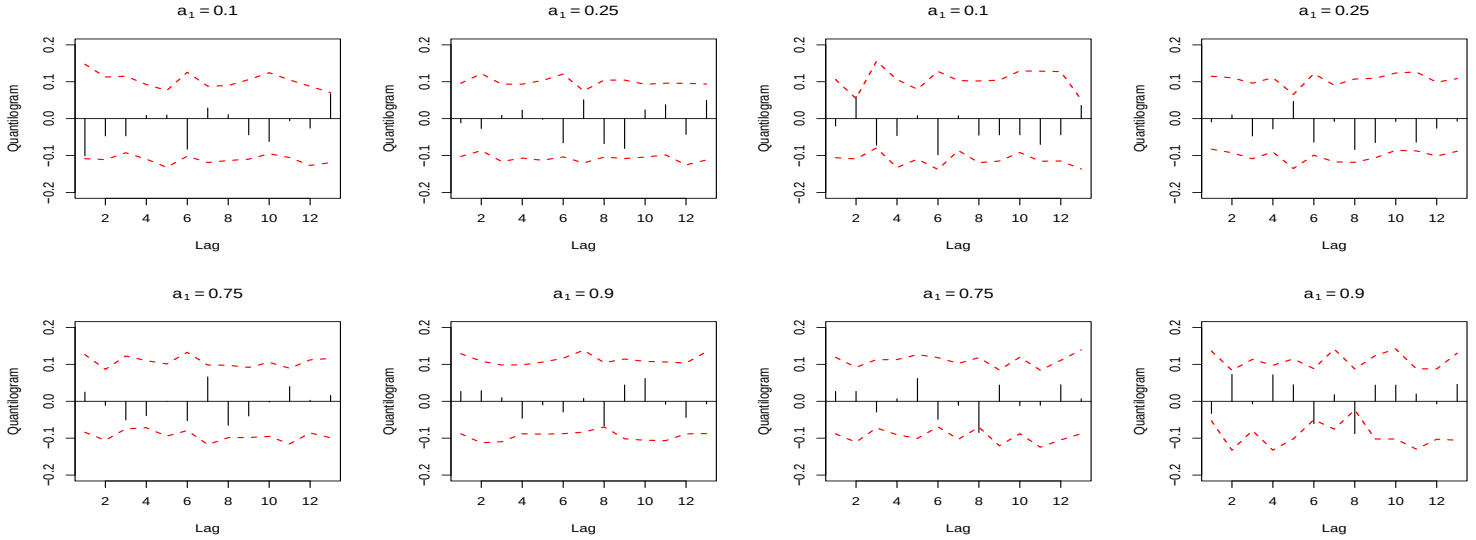
(c) Position change at quantile level 0.75 (d) Position change at quantile level 0.9

Figure B15: Cross-quantilogram from percentage changes in CIT net long positions to returns in the CBOT wheat futures market, 2004 – 2011



(a) Position change at quantile level 0.1

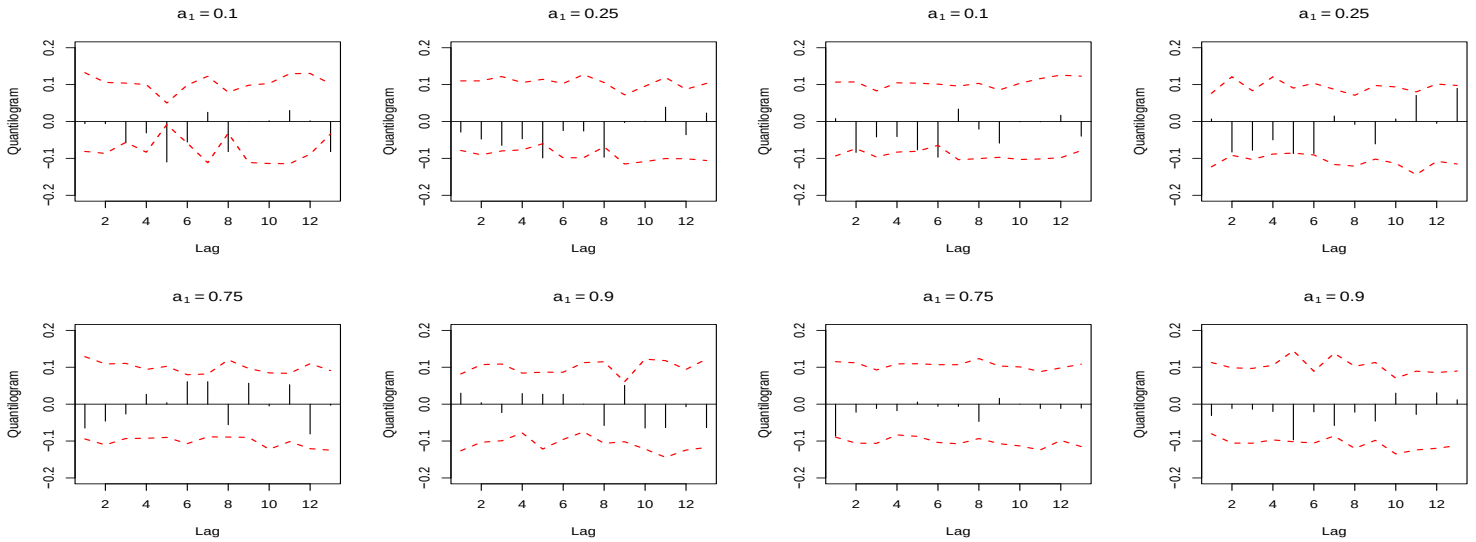
(b) Position change at quantile level 0.25



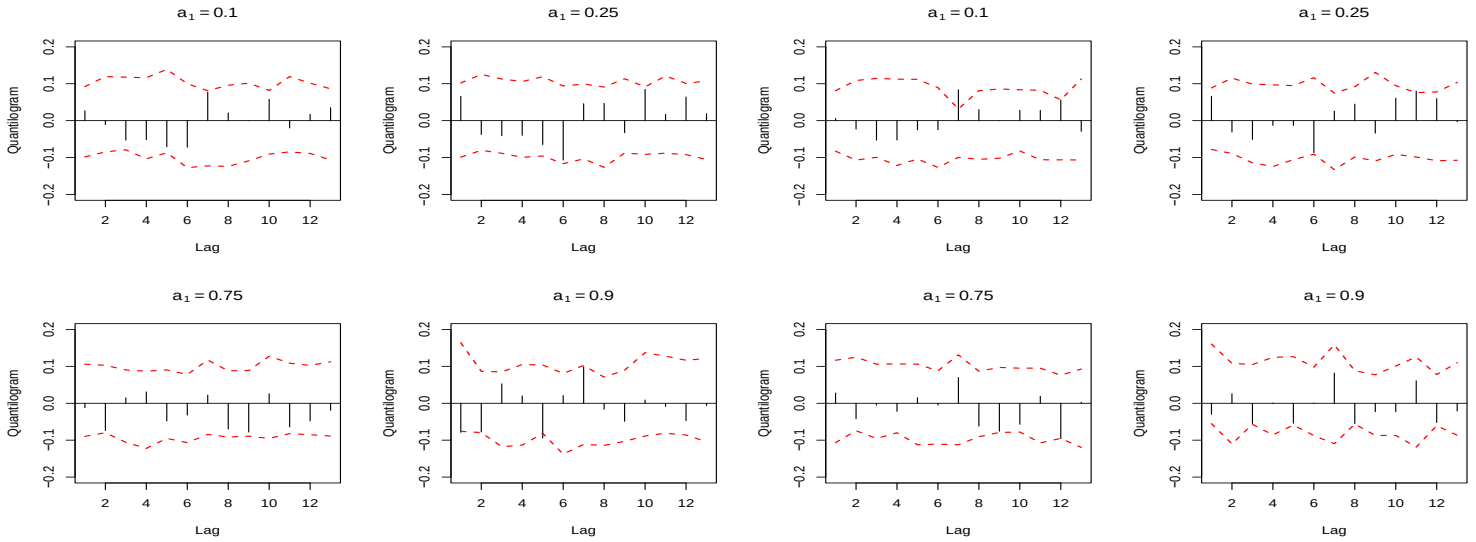
(c) Position change at quantile level 0.75

(d) Position change at quantile level 0.9

Figure B16: Cross-quantilogram from percentage changes in CIT net long positions to returns in the KCBOT wheat futures market, 2004 – 2011

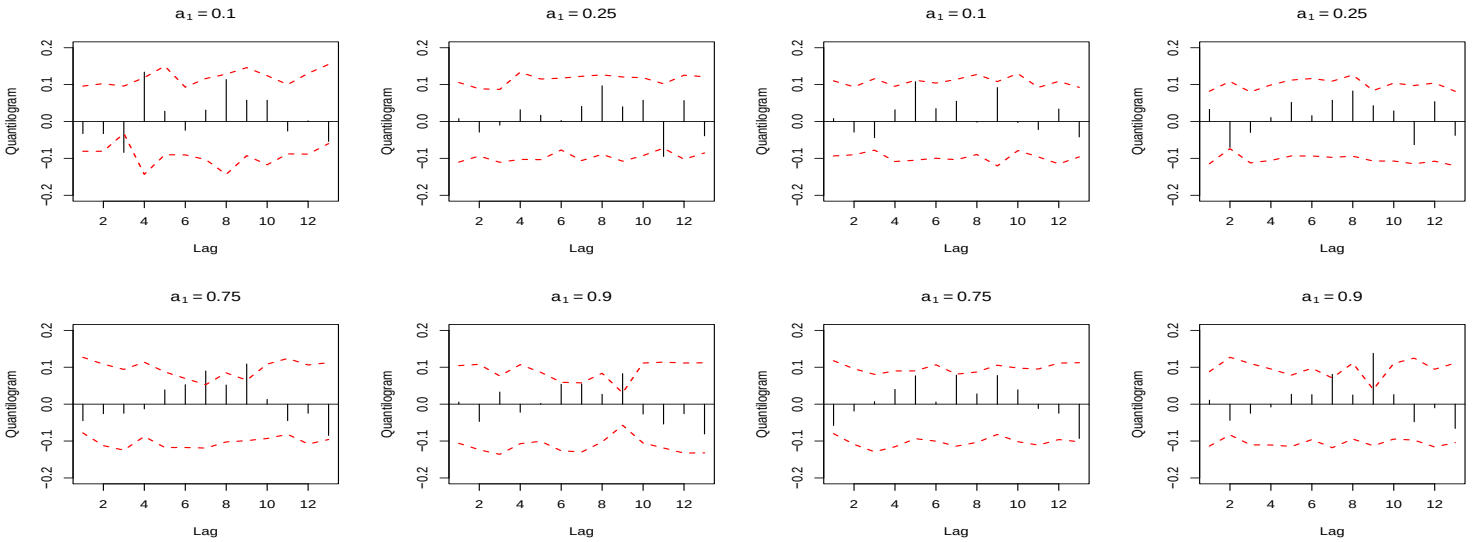


(a) Position change at quantile level 0.1 (b) Position change at quantile level 0.25



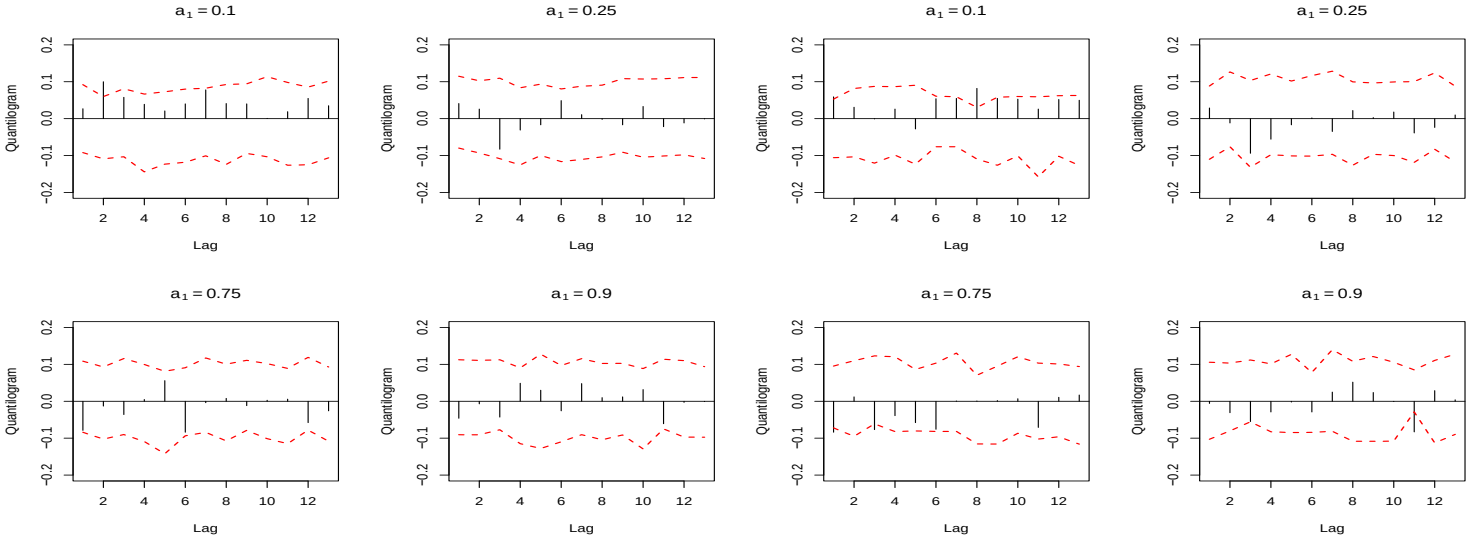
(c) Position change at quantile level 0.75 (d) Position change at quantile level 0.9

Figure B17: Cross-quantilogram from percentage changes in CIT net long positions to returns in the CBOT corn futures market, 2012 – 2019



(a) Position change at quantile level 0.1

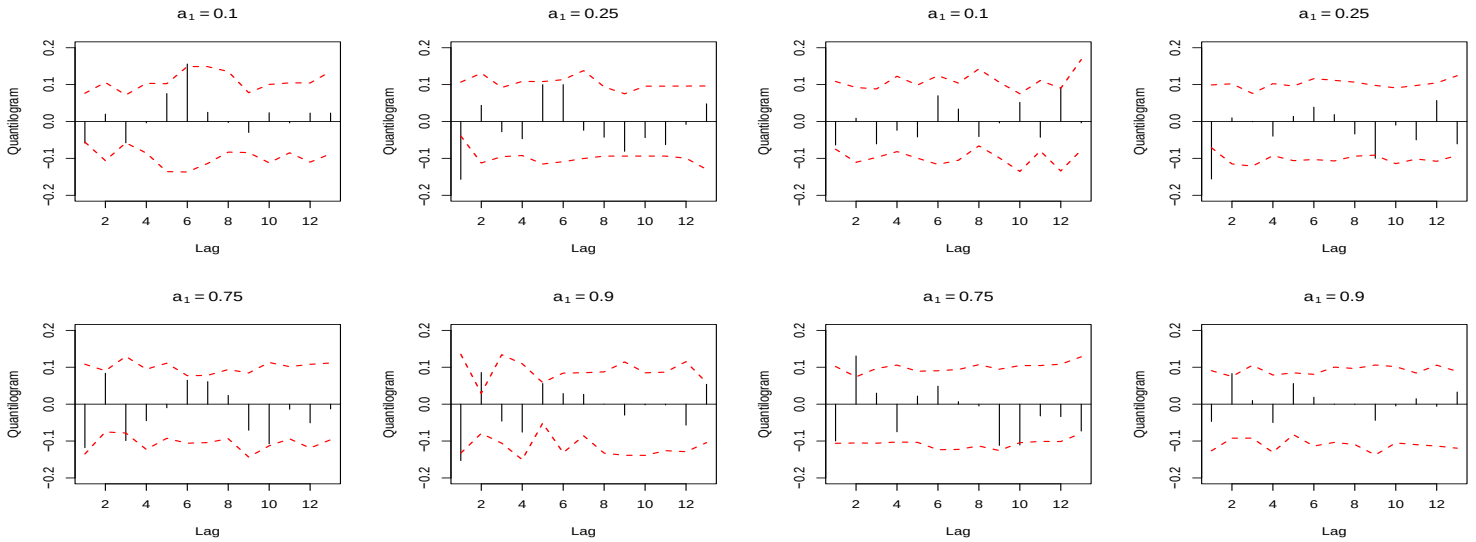
(b) Position change at quantile level 0.25



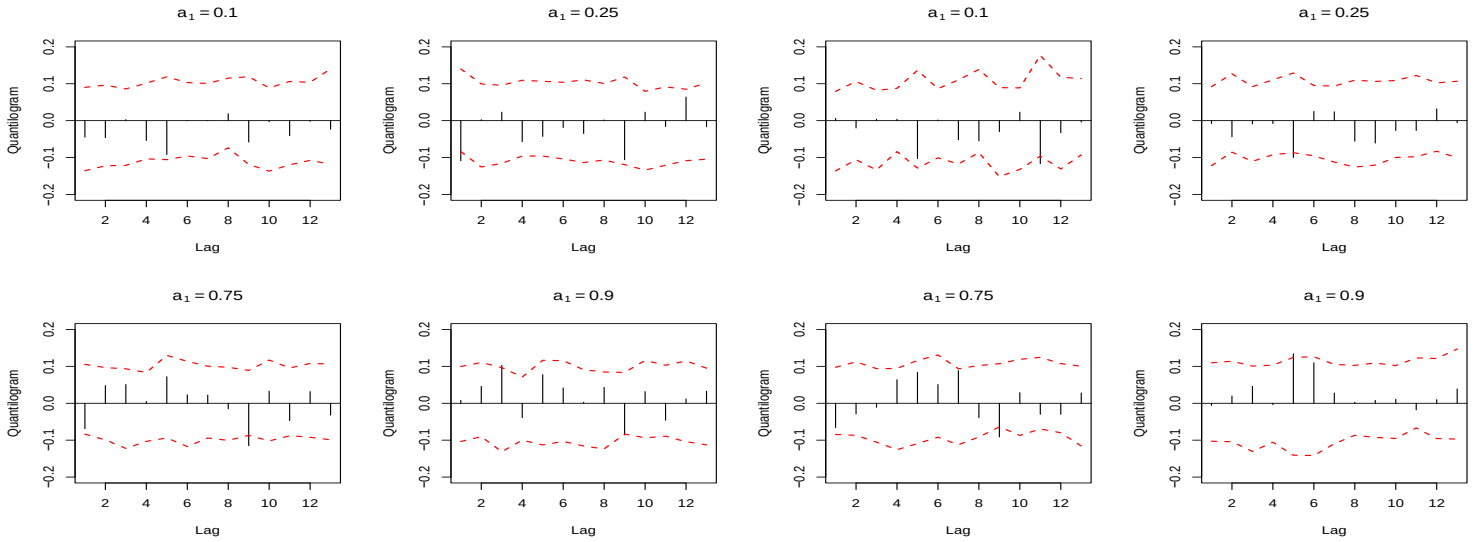
(c) Position change at quantile level 0.75

(d) Position change at quantile level 0.9

Figure B18: Cross-quantilogram from percentage changes in CIT net long positions to returns in the CBOT soybean futures market, 2012 – 2019

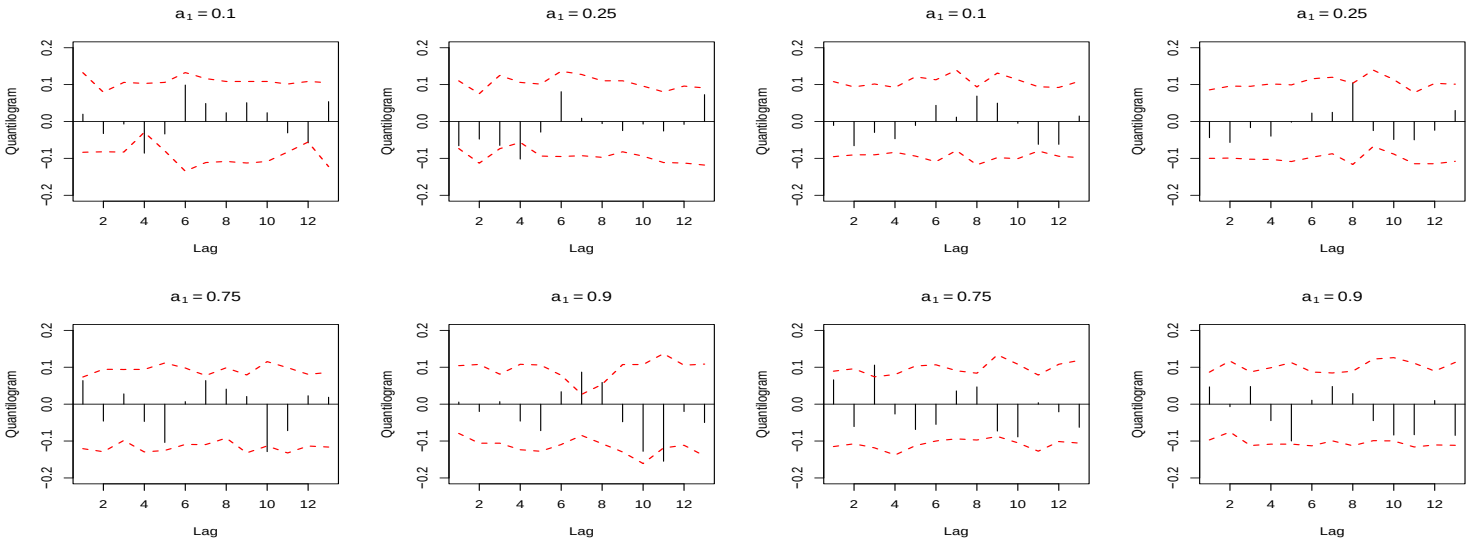


(a) Position change at quantile level 0.1 (b) Position change at quantile level 0.25



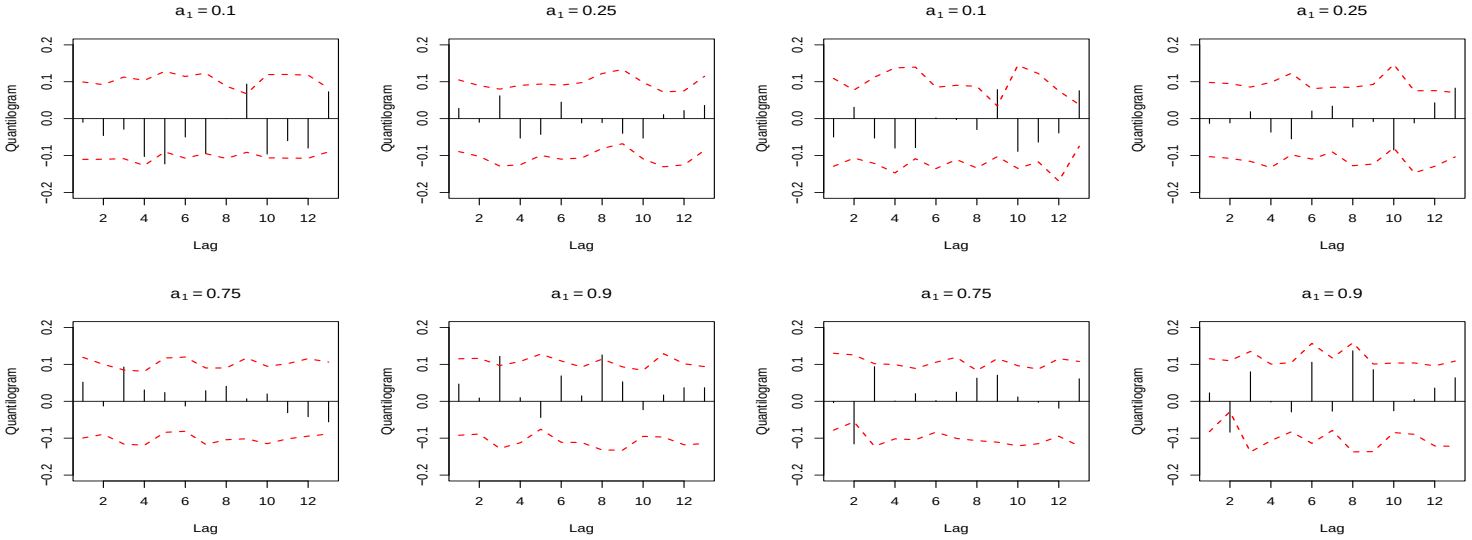
(c) Position change at quantile level 0.75 (d) Position change at quantile level 0.9

Figure B19: Cross-quantilogram from percentage changes in CIT net long positions to returns in the CBOT wheat futures market, 2012 – 2019



(a) Position change at quantile level 0.1

(b) Position change at quantile level 0.25



(c) Position change at quantile level 0.75

(d) Position change at quantile level 0.9

Figure B20: Cross-quantilogram from percentage changes in CIT net long positions to returns in the KCBOT wheat futures market, 2012 – 2019