A Dynamic Analysis of the Distribution of Commodity Futures and Spot Prices

by

Jian Li

Jean-Paul Chavas*

Abstract: This paper studies the evolving nature of commodity prices, with a focus on the dynamics of the joint distribution of futures price and spot price. It proposes a flexible two-step method to evaluate the joint movement of futures price and spot price: first, estimate a quantile vector autoregression (QVAR) model of the marginal distribution of each price; second, estimate a copula of their joint distribution. The approach is applied to the US soybean market over the last four decades. We study the evolving nature of the marginal and joint price distributions, including dynamic own and cross-price effects. We also investigate the presence of nonlinear dynamics and the effects of maturity of the futures contract. The application to the US soybean market illustrates the usefulness of the approach. Some of our key results include: a/ we find evidence of local dynamic instability in the upper-tail of the price distributions; b/ we estimate nonlinear cointegration relationships between futures price and spot price and find that they vary with market conditions; c/ we report quantile-specific impulse response functions documenting the importance of the futures market in soybean price discovery; d/ we evaluate the patterns exhibited by the basis and examine the convergence properties of the futures and spot price.

Keywords: futures price, spot price, dynamics, quantile, soybean

JEL: C31, G13, Q11

---

* Jian Li is Associate Professor in the College of Economics and Management, Huazhong Agricultural University. Email: hzaulj@126.com. Jean-Paul Chavas is Anderson-Bascom Professor in the Department of Agricultural and Applied Economics, University of Wisconsin. Email: jchavas@wisc.edu. Jian Li acknowledges support from the National Natural Science Foundation of China (Grant No. 71803058) and the Fundamental Research Funds from the Central Universities of China (Grant No. 2662017QD023).

A Dynamic Analysis of the Distribution of Commodity Futures and Spot Prices

1. Introduction

There has been much interest in studying the dynamics of markets and their implications for evolving price volatility (Working, 1949; Telsa, 1958; Routledge, 2002; Garcia and Leuthold, 2004; Chavas et al., 2014). The development of futures market has focused attention on the joint determination of futures and spot prices in commodity markets. This paper presents a refined analysis of the dynamic evolution of the distribution of commodity futures and spot prices. Many intriguing questions arise. What are the shapes of the marginal distributions for futures and spot prices? How do they evolve over time? Do they exhibit strong contemporaneous codependence? How do they relate to each other in the short, intermediate and long run? Do these relationships vary with market conditions? Are dynamic adjustments stable? Do the futures price and the spot price converge under different market conditions? While many of these questions are not new (e.g., Garbade and Silber, 1983; Fama and French, 1987; Wang and Ke, 2005; Garcia et al., 2014)[1], our paper proposes a refined approach that provides new and useful insights into the dynamics of the distribution of futures price and spot price.

Our paper proposes a two-step approach to model the evolving distribution of commodity prices, with an application to the US soybean futures and spot markets. In a first step, we specify and estimate a quantile vector autoregression (QVAR) model that provides a flexible representation of the marginal distributions of futures price and spot price and their temporal evolution. In a second step, we estimate a copula that links the marginal distributions with the joint distribution. Putting these two steps together, our QVAR-Copula approach provides all the information needed to evaluate the dynamics of prices. This includes useful implication on a series of issues: dynamic stability, response to shocks, cointegration and convergence properties.

Recent literature has witnessed expanding applications of quantile regression in a range of issues including modeling distributions (Koenker, 2005; Lee et al., 2011; Ramsey, 2020; Chavas, 2020; Huang et al., 2020). Compared with traditional mean regressions, quantile regression estimates the whole price distribution, thus going beyond the simple estimation of central tendency or spreads (e.g., mean or variance) and offering a flexible tool to assess price volatility and its determinants. Applied to dynamic systems, Koenker and Xiao (2006) proposed the quantile autoregression (QAR) model as a flexible representation of an evolving distribution. In this paper, we expand on the simple QAR model in three ways: 1/ we apply it in a multivariate context to capture dynamic cross price effects using a QVAR model; 2/ we allow for nonlinear dynamics; and 3/ building on the QVAR estimation of marginal distributions, we use a copula to explore linkages between marginal and joint distributions.[2]

Our QVAR-Copula approach generates a refined and flexible representation of the dynamics of the futures and spot price distributions. It is less restrictive than standard time series models (e.g., conventional Vector Autoregression (VAR) models or Generalized Autoregressive Conditional Heteroscedastic (GARCH) models; see Hamilton (1994) or

Enders (2010)) in the sense that our approach allows for arbitrary shapes of distribution functions and arbitrary evolution of any moments (including not only mean and variance, but also skewness and kurtosis). Indeed, neither the QVAR nor the copula imposes a priori restriction on the shape of marginal and joint distributions. All those arguments unveil the advantages and contributions of our QVAR-Copula approach in modeling commodity price distributions.

The usefulness of our approach is illustrated in an application to the US soybean market. Based on weekly price data over the period of 1980-2019, our QVAR-Copula approach uncovers several new and interesting results about the price dynamics in the soybean market. First, our analysis evaluates dynamic stability. We find that the futures market contributes to improving stability under nearby contract maturity. We also uncover evidence of local dynamic instability in the upper tail of the price distributions (importantly, this finding could not be obtained from traditional VAR or GARCH models). Second, we investigate the nature of cointegration reflecting the long-term relationship between futures price and spot price. Our analysis shows the existence of nonlinear cointegration and documents how the long-term relationship varies with market conditions. Third, we report quantile-specific impulse response functions showing the important role of the futures market in soybean price discovery. Finally, we evaluate the basis and its dynamic properties. In a way consistent with previous research (e.g., Hoffamn and Aulerich, 2013; Garcia et al., 2014), we find evidence of non-convergence of the futures and spot price and discuss its implications.

The rest of the paper is organized as follows. Section 2 provides a conceptual analysis of the determination of commodity prices in futures and spot markets. Section 3 presents the QVAR-Copula econometric model, and Section 4 describes the data. Applied to the US soybean market, Section 5 reports the econometric estimates of marginal and joint distributions. Economic implications are further discussed in Section 6. Section 7 concludes.

2. Conceptual Approach

Consider a commodity priced in two markets: a futures market and a spot market. At time t, denote the futures price for a futures contract with delivery at time $(t+m)$ by pf_{mt} and the spot price by ps_t. Let K be the set of agents participating in these two markets. At time t, the k-th market participant trades the quantity Qf_{kmt} on the futures market and the quantity Qs_{kt} on the spot market, $k \in K$. We let $(Qf_{kmt}, Qs_{kt},)$ be net quantities defined to be positive for quantity supplied and negative for quantity demanded. In this context, market equilibrium conditions at time t are given by $\sum_{k \in K} Qf_{kmt} = 0$ and $\sum_{k \in K} Qs_{kt} = 0$.

Letting $\boldsymbol{p}_t = (pf_{mt}, ps_t)$. Consider the case where the k-th agent chooses (Qf_{kmt}, Qs_{kt}) using the following decision rules: $Qf_{kmt} = Qf_k\,(\boldsymbol{p}_t, \boldsymbol{p}_{t-1}, \dots, \boldsymbol{p}_{t-n}, \boldsymbol{x}_t, \boldsymbol{e}_t)$ and $Qs_{kt} = Qs_k(\boldsymbol{p}_t, \boldsymbol{p}_{t-1}, \dots, \boldsymbol{p}_{t-n}, \boldsymbol{x}_t, \boldsymbol{e}_t), k \in K$, where n is the number of lags reflecting dynamic adjustments, $\boldsymbol{x}_t$ is a vector of factors affecting the behavior of market participants[3] and $\boldsymbol{e}_t$ is a vector of random variables representing uncertainty. We assume that the functions $(Qf_k(\cdot), Qs_k(\cdot))$ are differentiable and that $\begin{bmatrix} \frac{\partial[\sum_{k \in K} Qf_k(\boldsymbol{p}_t,\cdot)]}{\partial pf_{mt}} & \frac{\partial[\sum_{k \in K} Qf_k(\boldsymbol{p}_t,\cdot)]}{\partial p_{st}} \\ \frac{\partial[\sum_{k \in K} Qs_k(\boldsymbol{p}_t,\cdot)]}{\partial pf_{mt}} & \frac{\partial[\sum_{k \in K} Qs_k(\boldsymbol{p}_t,\cdot)]}{\partial p_{st}} \end{bmatrix}$ is a positive definite matrix, i.e. that aggregate supplies respond positively to rising prices. The market clearing conditions are

$$\sum_{k \in K} Qf_k\,(\boldsymbol{p}_t, \boldsymbol{p}_{t-1}, \dots, p_{t-n}, \boldsymbol{x}_t, \boldsymbol{e}_t) = 0 \tag{1a}$$

and

$$\sum_{k \in K} Qs_k\,(\boldsymbol{p}_t, \boldsymbol{p}_{t-1}, \dots, \boldsymbol{p}_{t-n}, \boldsymbol{x}_t, \boldsymbol{e}_t) = 0. \tag{1b}$$

The solution of equations (1a)-(1b) for $\boldsymbol{p}_t$ are the market-clearing prices denoted by

$$\boldsymbol{p}_t = \begin{bmatrix} pf_{mt} \\ ps_t \end{bmatrix} = \begin{bmatrix} g_f(\boldsymbol{p}_{t-1}, \dots, \boldsymbol{p}_{t-n}, \boldsymbol{x}_t, \boldsymbol{e}_t) \\ g_s(\boldsymbol{p}_{t-1}, \dots, \boldsymbol{p}_{t-n}, \boldsymbol{x}_t, \boldsymbol{e}_t) \end{bmatrix}. \tag{2}$$

Equation (2) provides a representation of the joint price dynamics in both the spot market and the futures market. Either equations (1) or equation (2) are valid models of price determination. The econometric analysis presented in this paper will rely on equation (2) for a simple reason: our empirical analysis is based on weekly price data, but weekly data are not available on the quantities that appear in equation (1).

Since equation (2) is the solution of equations (1a)-(1b), price dynamics reflects the behavior of market participants.[4] In the presence of dynamics and uncertainty, marketing decisions depend on the information available to the market participants. For the k-th agent, the decision rules $Qf_{kmt} = Qf_k\,(\boldsymbol{p}_t, \boldsymbol{p}_{t-1}, \dots, \boldsymbol{p}_{t-n}, \boldsymbol{x}_t, \boldsymbol{e}_t)$ and $Qs_{kt} = Qs_k(\boldsymbol{p}_t, \boldsymbol{p}_{t-1}, \dots, \boldsymbol{p}_{t-n}, \boldsymbol{x}_t, \boldsymbol{e}_t)$ depend on the information available at time t to the k-th agent, $k \in K$.

Much interest has focused attention on the role of speculation in market activities (e.g., Kaldor, 1939; Du et al., 2011; Sockin and Xiong, 2015). In general, speculation involves market participants acting on arbitrage opportunities. An economic analysis of speculation is presented in Appendix A. Three special cases are relevant in our analysis. In the first case, the k-th agent is involved in storage activities in the spot market. In this context, Qs_{kt} is the quantity purchased at time t, stored and sold back on the spot market at time $(t+1)$. Using equation (A1) in Appendix A with $Q_{kt} = Qs_{kt}$, $p_{at} = ps_t$, $\tau = 1$ and $p_{b,t+1} = ps_{t+1}$, the optimal choice of $Qs_{kt} \neq 0$ satisfies

$$r\,E_{kt}(ps_{t+1}) - ps_t = C'_{kt} + R'_{kt}. \tag{3a}$$

where $r \in (0,1)$ is a discount factor, E_{kt} is the expectation operator based on the information available to the k-th agent at time t, C'_{kt} is the marginal cost of storage and R'_{kt} is the marginal risk premium. The solution of (3a) for $Qs_{kt} \neq 0$ gives the k-th agent's net supply function $Qs_k(\boldsymbol{p}_t, \boldsymbol{p}_{t-1}, \dots, \boldsymbol{p}_{t-n}, \boldsymbol{x}_t, \boldsymbol{e}_t)$ mentioned above.

Equation (3a) shows that storage activities generate an intertemporal price arbitrage in the spot market (e.g., Working, 1949; Fama and French, 1987; Routledge et al., 2002). In general, the exercise of arbitrage limits the possibilities of large changes in spot price. Indeed, if a large price increase was anticipated, then the k-th agent would have incentives to buy the commodity on the spot market at time t (thus putting upward pressure on ps_t) to sell it back at time $(t+1)$ (thus putting downward pressure on price ps_{t+1}). But two qualifications apply. First, the price increase must be anticipated, meaning that the incentive to store would not hold for unanticipated price change, thus stressing the importance of information. Second, In cases where $[C'_{kt} + R'_{kt}]$ is large, then there would be no incentive to store when $[r\,E_{kt}(ps_{t+1}) - ps_t]$ is small or negative. This would reflect situations where storage activities would have no implications for intertemporal arbitrage for the spot price ps. Finally, from equation (3a), the properties of the marginal cost of storage C'_{kt} and of the marginal risk premium R'_{kt} play a role (Szymanowska et al., 2014). As shown in Appendix A (in equation (A2)), under risk aversion, the marginal risk premium R'_{kt} is positive when $Qs_{kt} > 0$ but negative when $Qs_{kt} < 0$. Thus, R'_{kt} depends on the trade position of the k-th agent, implying that it cannot be treated as a constant. More generally, when $[C'_{kt} + R'_{kt}]$ is not constant (e.g., the case of increasing marginal cost of storage or of time-varying marginal risk premium), then price arbitrage condition given in (3a) would vary depending on market conditions, implying that a time-varying relationship between ps_t and $E_{kt}(ps_{t+1})$. We will empirically investigate how price dynamics varies under different market conditions below.

In the second case, the k-th agent is a speculator involved in intertemporal arbitrage activities in the futures market. Let Qf_{kmt} be a futures contract purchased at time t and sold on the futures market at time $(t+1)$. Using equation (A1) in Appendix A with $\tau = 1$, $Q_{kt} = Qf_{kmt}$, $p_{at} = pf_{mt}$ and $p_{b,t+1} = pf_{m,t+1}$, the choice of $Qf_{kmt} \neq 0$ satisfies

$$r\, E_{kt}\big(pf_{m,t+1}\big) - pf_t = C'_{kt} + R'_{kt}, \tag{3b}$$

where C'_{kt} is the marginal cost of Qf_{kmt} and R'_{kt} is the associated marginal risk premium for the k-th agent. The solution of (3b) for $Qf_{kmt} \neq 0$ gives the k-th agent's net supply function $Qf_k(\boldsymbol{p}_t, \boldsymbol{p}_{t-1}, \dots, \boldsymbol{p}_{t-n}, m, \boldsymbol{x}_t, \boldsymbol{e}_t)$ mentioned above. Equation (3b) states an intertemporal arbitrage condition for the futures price pf: the expected discounted change in price $[r\, E_{kt}\big(pf_{m,t+1}\big) - pf_{mt}]$ is equal to the marginal cost plus the marginal risk premium: $C'_{kt} + R'_{kt}$. It indicates that speculation generates an intertemporal price arbitrage in the futures market, which limits the possibilities of large changes in the futures price. Again, the implications of such arbitrage opportunities for the dynamics of pf depend on the magnitude of the marginal cost C'_{kt} and of the marginal risk premium R'_{kt}.

In the third case, the k-th agent is involved in arbitrage activities between the futures market and the spot market. Let Q_{kt} be the quantity purchased on the futures market at price pf_{mt} and settled by selling it on the spot market at time $(t+m)$ at price ps_{t+m}. Using equation (A1) in Appendix A with $\tau = m$, $p_{at} = pf_{mt}$, and $p_{b,t+1} = ps_{t+m}$, the optimal choice of $Q_{kt} \neq 0$ satisfies

$$r^m E_{kt}(ps_{t+m}) - pf_{mt} = C'_{kt} + R'_{kt}. \tag{3c}$$

where C'_{kt} is the marginal cost of Q_{kt} and R'_{kt} is the associated marginal risk premium for the k-th agent. Equation (3c) states an arbitrage condition between the futures price pf and the spot price ps: the expected price difference $[r^m E_{kt}(ps_{t+m}) - pf_{mt}]$ equals the marginal cost plus the marginal risk premium $C'_{kt} + R'_{kt}$. When $m \to 0$, equation (3c) yields

$$E_{kt}(ps_t) - \lim_{m\to 0} pf_{mt} = C'_{kt} + R'_{kt}. \tag{3c'}$$

When $C'_{kt} + R'_{kt} = 0$, equation (3c') would imply that $\lim_{m\to 0} pf_{mt} = E_{kt}(ps_t)$, i.e. that the futures price and spot price would converge at the maturity of the futures contract. This suggests the existence of cointegration relationship between the futures and spot markets (e.g., Wang and Ke, 2005; Hernandez and Torero, 2010). While previous analyses have typically focused on linear cointegration, we will study nonlinear quantile cointegration relationships and investigate how such relationships vary with market conditions.

But equation (3c) also shows that such convergence properties would not hold when $C'_{kt} + R'_{kt} \neq 0$. This indicates two scenarios where non-convergence would occur: 1/ when transaction cost (e.g., delivery cost) is high and $C'_{kt} > 0$; and 2/ when uncertainty about the spot price remains around delivery time and $R'_{kt} \neq 0$. Evidence of non-convergence between

ps_t and $\lim_{m \to 0} pf_{mt}$ has been explored in the literature (Hoffman and Aulerich, 2013; Garcia et al., 2014). These issues will be investigated in our empirical analysis presented below.

Finally, in situations where $m = 1$ and $C'_{kt} + R'_{kt} = 0$ in equation (3c), the following price arbitrage condition would hold: $r\, E_{kt}(ps_{t+1}) = pf_{1t}$. Substituting this result into equation (3a) yields

$$pf_{1t} - ps_t = C'_{kt} + R'_{kt}, \tag{4}$$

stating that the basis[5] (defined as $(pf_{1t} - ps_t)$) is equal to $C'_{kt} + R'_{kt}$, where C'_{kt} is the marginal cost of storage and R'_{kt} is the associated marginal risk premium. These results are well known in the literature (e.g., Working, 1949; Telser, 1958; Fama and French, 1987). Equation (4) gives the standard result that the basis $(pf_{1t} - ps_t)$ can be interpreted as the "price of storage": it is the market signal guiding storage activities. When $R'_{kt} = 0$ and $C'_{kt} > 0$, equation (4) reduces to $pf_{mt} - ps_t = C'_{kt}$, indicating that the basis would be positive in the presence of storage activities. When $(pf_{1t} - ps_t)$ is observed to be negative, the presence of this "inverse-carrying charge" has generated some controversy in the literature. Working (1949) has argued this reflects a "convenience yield" that provides incentives to keep positive inventory even under any declining prices. But allowing for risk aversion, equation (4) provides an alternative interpretation. As stated in equation (A2) in Appendix A, under risk aversion, R'_{kt} can be either positive or negative depending on the market position of the k-th agent. It follows that $C'_{kt} + R'_{kt}$ can be negative under some scenarios (e.g., when elevators buy and store grains at harvest time), in which case the basis $(pf_{1t} - ps_t)$ would also be negative. This discussion indicates arbitrage opportunities across markets would imply that the basis exhibits time-varying patterns. These issues are explored empirically next.

3. Econometric method

This section discusses the econometric method used to model the price distributions of futures and spot prices. Our analysis of futures price pf and the spot price ps is based on equation (2) discussed above, stating that $\boldsymbol{p}_t = \boldsymbol{g}(\boldsymbol{p}_{t-1}, \dots, \boldsymbol{p}_{t-n}, \boldsymbol{x}_t, \boldsymbol{e}_t)$, where $\boldsymbol{p}_t = (pf_{mt}, ps_t)$. Assume that the function $\boldsymbol{g}(\cdot)$ is differentiable, that $\boldsymbol{e}_t$ is a serially independent random vector with distribution function $D(\boldsymbol{v}) = Prob(\boldsymbol{e_t} \leq \boldsymbol{v})$, and that $D(\boldsymbol{v})$ is absolutely continuous. In this context, equation (2) is a general autoregressive process of order n representing the stochastic dynamics of $\boldsymbol{p}_t = (pf_{mt}, ps_t)$. It allows for nonlinear dynamics along with own-price and cross-price effects.

In general, from equation (2), the joint distribution function of (pf_{mt}, ps_t) is

$$F(\boldsymbol{p} \,|\boldsymbol{P}_{t-1}, \boldsymbol{x}_t) = \; Prob[\boldsymbol{g}(\boldsymbol{p}_{t-1}, \dots, \boldsymbol{p}_{t-n}, \boldsymbol{x}_t, \boldsymbol{e}_t) \leq \boldsymbol{p}]. \tag{5}$$

where $\boldsymbol{P}_{t-1} = (\boldsymbol{p}_{t-1}, \dots \boldsymbol{p}_{t-n})$**.** Equation (5) provides all the information relevant to the analysis of (pf_{mt}, ps_t) and its dynamics. This section discusses an empirically tractable way to estimate equation (5). While $F(\boldsymbol{p} \,|\boldsymbol{P}_{t-1}, \boldsymbol{x}_t)$ in (5) is the joint distribution function of (pf_{mt}, ps_t), we are interested in exploring the associated marginal distribution functions

$$F_f\big(p_f \big|\boldsymbol{P}_{t-1}, \boldsymbol{x}_t\big) = \; Prob[\boldsymbol{g}_f(\boldsymbol{p}_{t-1}, \dots, \boldsymbol{p}_{t-n}, \boldsymbol{x}_t, \boldsymbol{e}_t) \leq p_f]. \tag{6a}$$

$$F_s(p_s|\boldsymbol{P}_{t-1}, \boldsymbol{x}_t) = \; Prob[\boldsymbol{g}_s(\boldsymbol{p}_{t-1}, \dots, \boldsymbol{p}_{t-n}, \boldsymbol{x}_t, \boldsymbol{e}_t) \leq p_s]. \tag{6b}$$

From Sklar's theorem (Sklar, 1959; Patton, 2006), consider the following relationship between the marginal distributions (F_f, F_c) and the joint distribution F.

$$F(\boldsymbol{p} \mid \boldsymbol{P}_{t-1}, \boldsymbol{x}_t) = \; C[F_f\big(p_f \big| \boldsymbol{P}_{t-1}, \boldsymbol{x}_t\big), \; F_s(p_s | \, \boldsymbol{P}_{t-1}, \boldsymbol{x}_t) | \, \boldsymbol{x}_t], \tag{7}$$

where $C[F_f, F_s | \cdot]$ is a copula corresponding to the mapping $(F_f, F_s) \rightarrow [0, 1]$ (Nelson, 2006). Our analysis will rely on equation (7) and proceed in two steps: first, estimate the marginal distributions F_s in (6a) and F_f in (6b) using a QVAR model; second, evaluate the copula $C[F_f, F_s | \cdot]$ and use (7) to recover the joint distribution F. This two-step QVAR-Copula approach will be used below in studying the dynamics of (pf_{mt}, ps_t).

Note the generality of this approach. First, the estimation of marginal distributions allows for own-price and cross-price effects and nonlinear dynamics. And it allows for the effects of other variables (as captured by the variables $\boldsymbol{x}_t$). Second, the copula allows for arbitrary contemporaneous codependence between pf_{mt} and ps_t (see Nelson, 2006) and the codependence to vary with $\boldsymbol{x}_t$.

3.1 Estimating marginal distributions

Our first step involves estimating the marginal distributions for pf_{mt} and ps_t. Consider the associated quantile functions defined as the inverse of the corresponding marginal distributions. The quantile function for pf_{mt} evaluated at $q_f \in [0, 1]$ is $p_{fq}(q_f | \boldsymbol{P}_{t-1}, \boldsymbol{x}_t) \equiv inf_{p_f} \{p_f: F_f(p_f | \boldsymbol{P}_{t-1}, \boldsymbol{x}_t) \geq q_f\}$. And the quantile function for ps_t evaluated at $q_s \in [0, 1]$ is $p_{sq}(q_s | \boldsymbol{P}_{t-1}, \boldsymbol{x}_t) \equiv inf_{p_s} \{p_s: F_s(p_s | \boldsymbol{P}_{t-1}, \boldsymbol{x}_t) \geq q_s\}$. Next, consider the following specification for these quantile functions

$$p_{fq}(q_f | \boldsymbol{Y}_{t-1}, \boldsymbol{x}_t) = \boldsymbol{\beta}_{0,f}(q_f) + \boldsymbol{\beta}_{1,f}(q_f)\, \boldsymbol{P}_{t-1} + \boldsymbol{\beta}_{2,f}(q_f)\, h_f(\boldsymbol{P}_{t-1}, \boldsymbol{x}_t) \tag{8a}$$

$$p_{sq}(q_s | \boldsymbol{P}_{t-1}, \boldsymbol{x}_t) = \boldsymbol{\beta}_{0,s}(q_s) + \boldsymbol{\beta}_{1,s}(q_s)\, \boldsymbol{P}_{t-1} + \boldsymbol{\beta}_{2,s}(q_s)\, h_s(\boldsymbol{P}_{t-1}, \boldsymbol{x}_t) \tag{8b}$$

where the $\boldsymbol{\beta}(q)$'s are parameters to be estimated and $(h_s(\cdot), h_f(\cdot))$ are known functions. Equations (8a)-(8b) constitute a quantile vector autoregression (QVAR) model providing a flexible representation of the marginal distributions of (pf_{mt}, ps_t) and the underlying price dynamics. Indeed, when $\boldsymbol{\beta}_2 = 0$ and $\boldsymbol{\beta}_1(q)$ does not vary with q, then (8a)-(8b) reduce to a standard vector autoregression (VAR) model commonly used in time series analysis (e.g., Hamilton, 1994; Enders, 2010). Note that allowing $\boldsymbol{\beta}_0(q)$ to vary with q does not impose a priori restrictions on the shape of the distribution function (e.g., it allows for any skewness or kurtosis). In addition, allowing for $\boldsymbol{\beta}_1(q)$ to vary with q permits dynamics to affect the variance or skewness of prices. Finally, allowing $\boldsymbol{\beta}_2(q)$ to be non-zero allows for nonlinear dynamics when the h functions in (8a)-(8b) are nonlinear in $\boldsymbol{P}_{t-1}$ or include interaction effects between $\boldsymbol{P}_{t-1}$ and $\boldsymbol{x}_t$.

Following Koenker (2005) and Koenker and Xiao (2006), equations (8a)-(8b) can be estimated by quantile regression. Consider a sample of T observations denoted as $(p_{ft}, p_{st}, \boldsymbol{P}_{t-1}, \boldsymbol{x}_t)$, $t \in \{1, \dots, T\}$. The parameters $\boldsymbol{\beta}$'s in (8a)-(8b) can be estimated as follows

$$\boldsymbol{\beta}_i^e(q_i) \in argmin_\beta \{\textstyle\sum_{t=1}^T \rho_{q_i}\left(p_{it} - p_{iq}(q_i|\boldsymbol{P}_{t-1}, \boldsymbol{x}_t)\right)\}, i \in \{f, s\}, \tag{9}$$

where $\rho_{q_i}(w) = w\,[q_i - I(w < 0)]$, $I(\cdot)$ is an indicator function and $p_{iq}(q_i\,|\boldsymbol{P}_{t-1}, \boldsymbol{x}_t)$ is parametrized as in (8a)-(8b). As shown in Koenker and Xiao (2004, 2006), the quantile estimator given in (9) provides consistent estimates of the parameters under general conditions. Yet, asymptotic properties of the estimator become non-standard in the presence of unit roots. On that basis, we rely on bootstrapping in conducting hypothesis testing.

Based on the QVAR estimates discussed above, we are interested in exploring the analysis of price dynamics. Note that equation $\boldsymbol{p_t} = \boldsymbol{g}(\boldsymbol{p_{t-1}}, \dots, \boldsymbol{p_{t-n}}, \boldsymbol{x_t}, \boldsymbol{e_t})$ in (2) can be alternatively written as the first-order difference equation

$$\boldsymbol{P_t} \equiv \begin{bmatrix} \boldsymbol{p_t} \\ \boldsymbol{p_{t-1}} \\ \vdots \\ \boldsymbol{p_{t-n+1}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{g}(\boldsymbol{p_{t-1}}, \dots, \boldsymbol{p_{t-n}}, \boldsymbol{x_t}, \boldsymbol{e_t}) \\ \boldsymbol{p_{t-1}} \\ \vdots \\ \boldsymbol{p_{t-n+1}} \end{bmatrix} = \boldsymbol{G}(\boldsymbol{P_{t-1}}, \boldsymbol{x_t}, \boldsymbol{e_t}). \tag{10}$$

Under differentiability, consider the matrix $\boldsymbol{DG}(\boldsymbol{P_{t-1}}, \boldsymbol{x_t}, \boldsymbol{e_t}) \equiv \boldsymbol{\partial G}(\boldsymbol{P_{t-1}}, \boldsymbol{x_t}, \boldsymbol{e_t})/\boldsymbol{\partial P_{t-1}}$ and its eigenvalues $(\lambda_1, \lambda_2, \dots)$ evaluated at point $(\boldsymbol{P_{t-1}}, \boldsymbol{x_t}, \boldsymbol{e_t})$. The modulus of the dominant eigenvalue $|\lambda_1|$ provides useful information about dynamics of $\boldsymbol{p_t}$ in the neighborhood of the $(\boldsymbol{P_{t-1}}, \boldsymbol{x_t}, \boldsymbol{e_t})$. Under nonlinear dynamics, the eigenvalues (or roots) vary with the valuation point. The dominant root then provides information about local dynamics: prices are locally stable (unstable) in the neighborhood of the evaluation point when the modulus of dominant root $|\lambda_1|$ is less than (greater than) 1 (Sayama, 2015, p. 94-95). In this case, $|\lambda_1|$ measures the speed of price adjustments: $\log(|\lambda_1|)$ can be interpreted as the local rate of divergence in prices along a forward path. In a stochastic world, these evaluations can vary depending on stochastic shocks. In our QVAR model, this translates into the question: does the dominant root vary across quantiles? If it does not, then price dynamics does not depend on stochastic shocks (e.g., as found in standard linear VAR models). If it does, then price dynamics would vary with stochastic shocks, stressing the need to distinguish between "positive" versus "negative" shocks. This discussion indicates the general flexibility of our approach. It extends previous research exploring regime-switching models or smooth transitions dynamics (e.g., the TAR model of Caner and Hansen (2001) or ESTAR model of

Kapetanios et al. (2003)). As further discussed below, our approach allows the investigation of how dynamic stability can vary across quantiles and across market situations.

Our QVAR estimate can also be used to evaluate the long-run relationship between pf and ps. To see that, consider a first-order Taylor series approximation of equations (8a)-(8b) with respect to lagged prices evaluated at point $\boldsymbol{z} = (\boldsymbol{Y}_{t-1}, \boldsymbol{x}_t)$

$$\boldsymbol{p}_t(\boldsymbol{q}) \approx \boldsymbol{A}(\boldsymbol{q}, \boldsymbol{z}) + \sum_{j=1}^{n} \boldsymbol{B}_j(\boldsymbol{q}, \boldsymbol{z})\, \boldsymbol{p}_{t-j} \tag{11a}$$

where $\boldsymbol{q} = \left(q_s, q_f\right) \in [0, 1]^2$. Equation (11a) can be equivalently written in first differences

$$\Delta\boldsymbol{p}_t(\boldsymbol{q}) \approx \boldsymbol{A}(\boldsymbol{q}, \boldsymbol{z}) + \boldsymbol{\Pi}(\boldsymbol{q}, \boldsymbol{z})\, \boldsymbol{p}_{t-1} + \sum_{j=1}^{n-1} \boldsymbol{\Gamma}_j(\boldsymbol{q}, \boldsymbol{z})\, \Delta\boldsymbol{p}_{t-j}, \tag{11b}$$

where $\Delta\boldsymbol{p}_t = \boldsymbol{p}_t - \boldsymbol{p}_{t-1}$, $\Delta\boldsymbol{p}_{t-j} = \boldsymbol{p}_{t-j} - \boldsymbol{p}_{t-j-1}$, $\boldsymbol{\Pi}(\boldsymbol{q}, \boldsymbol{z}) = \sum_{j=1}^{n} \boldsymbol{B}_j(\boldsymbol{q}, \boldsymbol{z}) - \boldsymbol{I}_n$ and $\boldsymbol{\Gamma}_j(\boldsymbol{q}, \boldsymbol{z}) = -\sum_{k=j+1}^{n} \boldsymbol{B}_k(\boldsymbol{q}, \boldsymbol{z})\,, j = 1, \dots, m-1$. When the specification in (8a)-(8b) is linear in lag prices, then equations (11a)-(11b) are globally valid. If in addition, the lagged coefficient $\boldsymbol{B}_j$ do not vary with $\boldsymbol{q}$, then the matrix $\boldsymbol{\Pi}$ is a constant. This is the scenario evaluated by Engle and Granger (1987) and Johansen (1995) in their investigation of cointegration: cointegration then arises when the matrix $\boldsymbol{\Pi}$ has rank 1, and the long run relationship between ps and pf is given by the right eigenvector associated with the largest root of $\boldsymbol{\Pi}$ (Engle and Granger, 1987; Johansen, 1995). Our analysis extends this approach in two ways: the matrix $\boldsymbol{\Pi}(\boldsymbol{q}, \boldsymbol{z})$ can vary with the quantiles $\boldsymbol{q}$; and it can vary with the evaluation point $\boldsymbol{z}$. These are scenarios of nonlinear cointegration (e.g., Pizzi, 2010; Tjøstheim, 2020). The situation where $\boldsymbol{\Pi}(\boldsymbol{q}, \boldsymbol{z})$ varies with the quantiles $\boldsymbol{q}$ means that the nature of price adjustments toward their long run equilibrium depends on the nature of the shocks. And having $\boldsymbol{\Pi}(\boldsymbol{q}, \boldsymbol{z})$ varying with the evaluation point $\boldsymbol{z}$ means that changing market conditions (represented by $\boldsymbol{z}$) affects how the markets adjust in the long term. In either case, the right eigenvector associated with the largest root of $\boldsymbol{\Pi}(\boldsymbol{q},\boldsymbol{z})$ reflects how the longer run dynamics of (pf, ps) can change with $(\boldsymbol{q},\boldsymbol{z})$. These issues will be investigated in our empirical analysis.

3.2 Estimating the joint price distribution

Once the marginal distributions (F_f, F_s) in (6a)-(6b) (or their associated quantile functions in (8a)-(8b)) have been estimated, we can obtain the joint distribution using the copula $F = C\left(q_f, q_s \middle| \boldsymbol{x}_t\right)$ in equation (7). This can be done using parametric as well as nonparametric methods. See Nelsen (2006) or Joe (2014) for an overview of alternative approaches. In our case, the first step quantile estimation gives consistent estimates of $\boldsymbol{\beta}_i^e(q_i)$ in (9) and corresponding quantiles $p_{iq}^e(q_i\, |\boldsymbol{P}_{t-1}, \boldsymbol{x}_t)$ in (8), $i \in \{f, s\}$. Following Chavas (2020), solving $p_{iq}^e(q_i\, |\boldsymbol{P}_{t-1}, \boldsymbol{x}_t) = p_{it}$ for q_i gives consistent estimates of $q_{it}^e, i \in \{f, s\}, t \in \{1, \dots, T\}$. In turn, in a second step, we use the q_{it}^e's to estimate the copula $C\left(q_f, q_s \middle| \boldsymbol{x}_t\right)$ in (7). This is done by estimating the conditional distribution function $D\left(q_f \middle| q_s, \boldsymbol{x}_t\right) = Prob\left[q_{it}^e \leq q_f \middle| q_{is}^e = q_s, \boldsymbol{x}_t\right]$. Under differentiability, this conditional distribution satisfies $D\left(q_f \middle| q_s, \boldsymbol{x}_t\right) = \partial C\left(q_f, q_s \middle| \boldsymbol{x}_t\right)/\partial q_s$ (Joe, 2015, p. 30), or equivalently $C\left(q_f, q_s \middle| \boldsymbol{x}_t\right) = \int_0^{q_s} D\left(q_f \middle| q, \boldsymbol{x}_t\right) dq$, making it clear that the conditional distribution $D\left(q_f \middle| q_s, \boldsymbol{x}_t\right)$ provides all the information related to the contemporaneous codependence between p_f and p_s. Below, going beyond Chavas (2020), we permit the copula to change with $\boldsymbol{x}_t$. Letting $\boldsymbol{x}_t = ma$, we allow the variable "time to maturity" ma to affect the copula and consider the following specification[6]

$$D(q_f | q_s, ma) = \gamma_0(q_f) + \gamma_1(q_f)\, q_s + \gamma_2(q_f)\, q_s^{\mathrm{s}} + \gamma_3(q_f)\, ma, \tag{12}$$

where we include both a linear term q_s and a nonlinear term defined as $q_s^s = \left(q_s - 0.5\right)^2$. Using (q_f^e, q_s^e) and applying quantile regression to (12) yield consistent estimates of the parameters $\boldsymbol{\gamma}^e$ and of the associated distribution function $D^e(q_f|q_s, ma)$ in (12). Noting that the distribution of the second-step estimates depends on the statistical properties in the first step (Murphy and Topel, 2002), we employ the empirical bootstrap for hypothesis testing.

Finally, from equations (7) and (12) along with $C\left(q_f, q_s\middle|\, ma\right) = \int_0^{q_s} D\left(q_f\middle|\, q, ma\right) dq$, the codependence between futures and spot prices at any given quantile set $(q_{f,} q_s)$ can be measured by:

$$R\left(q_f, q_s\middle|\, ma\right) \equiv C\left(q_f, q_s\middle|\, ma\right) - q_f \cdot q_s. \tag{13}$$

Conditional on ma and evaluated at point $\left(q_f, q_s\right)$, $R\left(q_f, q_s\middle|\, ma\right) = 0$ when the futures price and spot price are contemporaneously independent. Alternatively, $R\left(q_f, q_s\middle|\, ma\right) > 0\ (< 0)$ implies positive (negative) codependence of pf_{mt} and ps_t. The estimates (q_f^e, q_s^e) are used to obtain $C^e\left(q_f, q_s\middle|\, ma\right) \equiv \int_0^{q_s} D^e\left(q_f\middle|\, q, ma\right) dq$ and $R^e\left(q_f, q_s\middle|\, ma\right) \equiv C^e\left(q_f, q_s\middle|\, ma\right) - q_f \cdot q_s$ as consistent estimates of $C\left(q_f, q_s\middle|\, ma\right)$ and $R\left(q_f, q_s\middle|\, ma\right)$, respectively. Using bootstrapping, empirical testing for whether $R\left(q_f, q_s\middle|\, ma\right) = 0$ is a test of the null hypothesis of independence evaluated at (q_f, q_s, ma). Equation (13) can also shed light on the nature of dependence in the tails of the marginal distributions. Using copula, the investigation of tail dependence has been examined in previous research (e.g., Joe, 2014). The importance of tail dependence and its role in understanding cross-market pricing patterns have been explored by Patton (2006) and Reboredo (2012) using a parametric approach. We want to stress two advantages of our semi-parametric approach: 1/ it is flexible in the sense that it allows for arbitrary codependence; and 2/ it does not require imposing a priori restrictions on the shape of the copula. The usefulness of our QVAR-Copula approach is illustrated next in an application to futures price and spot price in the US soybean market.

4. Data

Our analysis relies on weekly futures and spot prices in the US soybean markets over the last four decades of 1980-2019. The weekly spot prices were collected from the Historic Grain Reports (HGR) of the Iowa Department of Agricultural and Land Stewardship (IDALS). As stated by Fama and French (1987), good spot-price data are not available for many commodities. The HGR reports representative soybean spot price data consistently recorded in Iowa over the last few decades. For our research purposes (e.g., studying the futures-spot basis), it is common to rely on representative regional spot prices as national averaging tends to weaken the underlying futures-spot linkages (e.g., Garcia et al., 2014). The weekly Chicago Board of Trade (CBOT) futures closing prices over the same period were collected from the *Wind* financial database. To match the basis data reported from IDALS, we choose the continuous futures price that is conducted using the front-month rolling method, which rolls nearby futures contracts on expiration dates in contract months.[7] These data are used to evaluate the underlying futures-spot price relationships in the soybean markets.

Table 1 reports the summary statistics of price and other variables used in our analysis. While original prices are expressed in US dollar per bushel, our empirical study is based on the logarithm of prices. Table 1 shows that the futures price has a mean value of 1.994, slightly higher than that of the spot price of 1.938. Other moments (standard deviation, skewness, kurtosis) show similar patterns: moments are generally higher for futures price than spot price.

Figure 1 reports the trajectories of soybean futures and spot price over the last four decades. The prices exhibit large volatility throughout the sample period. Two important factors are worth noting: 1/ the rise of soybean production in Brazil starting in the 1990s, and 2/ the significant impacts of US biofuel policy starting in 2005. Since the late 1990s, Brazil has emerged as a major soybean producer and exporter, resulting in significant changes in global soybean supply and prices (Cattlelan and Agnol, 2018). Also, dramatic price changes have occurred in response to US biofuel policy after 2005, which contributed to price swings during the period of 2007-2014 (Du et al., 2011). As shown in Figure 1, the co-movement of soybean futures and spot prices generally holds. However, recent research has pointed out the growing difference between contemporaneous futures and spot prices (the basis), reflecting convergence problems in agricultural futures markets (Hoffamn and Aulerich, 2013; Garcia et al., 2015). The evolving relationship between futures and spot prices are examined next.

5. Econometric estimation

We employ the QVAR-Copula approach to investigate the dynamic linkages between US soybean futures and spot prices. Expanding on previous research (e.g., Garbade and Silber, 1983; Hernandez and Torero, 2010), our QVAR model considers both linear and nonlinear dynamic effects. We introduce nonlinearity by including the square of the difference between lagged price and its median (i.e., $p^s_{i,t-j} = (p_{i,t-j} - median(p_{i,t-j}))^2, i \in (f, s)$), which reflects how extreme prices affect the price distributions.

Other key variables include the "time to maturity" variable ma measuring the number of weeks to maturity in the futures contract. As shown in Table 1, in our sample, ma takes the values from zero to nine, with a mean value of 3.499. We also include the interaction of ma and lagged futures price (Pf_1) to capture the interaction between past price and futures contract characteristics.[8] Quarterly dummies (Q_{1t}, Q_{2t}, Q_{3t}) are also introduced to reflect seasonality. To capture the structural changes caused by technological and institutional change, Brazil production expansion and biofuel policy (Du et al., 2011; Cattelan et al., 2018), we define three time trends: an overall time trend tt, a linear trend $tt1$ starting in 1998 and a linear trend $tt2$ starting in 2005.[9] The number of price lags *n* that enter equations (8a) and (8b) were chosen using the Bayesian Information Criterion (BIC). In this study, BIC suggested that the QVAR(2) model provides the best fit to the data.

5.1 Estimates of marginal distributions

Based on the specified QVAR(2) model, we estimated the marginal price distributions. Tables 2 and 3 report the QVAR estimates for the futures price pf and spot price ps, respectively, under selected quantiles: q=(0.1, 0.3, 0.5, 0.7, 0.9). According to the pseudo-R^2 proposed by Koenker and Machado (1999), the pseudo-R^2 values in our model lies within the range of 0.892-0.935, indicating high goodness-of-fit. In these tables, lagged prices are denoted using subscript (e.g., $ps_j = ps_{t-j}, j \in \{1,2\}$). First, we find that the lag own price effects are often statistically significant in both the futures and spot price equations, documenting the importance of dynamic adjustments in the price distributions. Table 2 shows strong positive short-run price response of pf to pf_1, especially in the upper quantiles: $\beta_{pf_1} > 1$ in the higher quantiles (q=0.7, 0.9) while $\beta_{pf_1} < 1$ in the lower quantiles (q=0.1, 0.3), indicating a different short term response to downside shocks (compared to upside shocks). Table 3 also shows strong positive response of ps to ps_1, but with stronger short-run price response to ps_1 in the lower quantiles than in the upper quantiles. These results indicate the presence of heterogeneity in the dynamic response to shocks: the responses to shocks vary both across markets and with the nature of the shocks (i.e., downside shocks versus upside shocks). Tables 2-3 report differences in dynamic cross-price effects: the effects of ps_1 on pf are positive in all quantiles and often statistically significant; and they are typically stronger than the effects of pf_1 on ps. And the effects of ps_2 on pf and ps are all negative, thus dampening the positive ps_1 effects. The implications of these estimates for longer term price adjustments are further explored below.

Second, the nonlinear effects of (pf_1^s, pf_2^s) on pf and of (ps_1^s, ps_2^s) on ps are statistically significant in the low and high quantiles, i.e., q=(0.1, 0.3, 0.9). This is an important finding of this study, uncovering that nonlinear dynamics tend to occur in the tails of the distribution. Third, of special interest is the "time to maturity" variable ma and its interaction with lagged futures price $(ma \cdot pf_1)$. Tables 2-3 show that the effects of these variables are statistically significant for many quantiles, indicating that "time to maturity" affects the dynamics of the distribution of both pf and ps. Fourth, consistent with previous research (e.g., Du et al., 2011), we find evidence of seasonality and intertemporal structural changes as the seasonal dummies and time trends are statistically significant at least in some quantiles. For instance, we find $tt1$ to be significant in the upper tail and $tt2$ to be significant in both tails, reflecting how Brazilian production expansion and biofuel policy have affected the soybean price distributions.

The estimates from a VAR model are also reported in Tables 2-3. Recall that the standard VAR model is obtained a special case of the QVAR model when all non-intercept parameters are the same across quantiles. We formally tested this hypothesis. We found strong statistical evidence against this null hypothesis for both pf and ps, with p-values less than 0.01. Thus, there is strong evidence against the VAR model: non-intercept parameters in the QVAR model differ across quantiles. This illustrated in Tables 2-3 when the parameter estimates differ between the lower quantiles and the upper quantiles. An example is given by the nonlinear effects of pf_i^s on pf in Table 2: the VAR model finds no statistically significant effect on pf, but the QVAR reports statistically significant effects in the tails of the price distribution (at $q = 0.1$ and $q = 0.9$). The failure of VAR model to capture such effects indicates a need to go beyond mean regression analysis. Applied to all quantiles, our QVAR approach further models the whole distribution, allowing for a flexible representation of the marginal price distributions (including mean, variance, skewness and kurtosis) and their evolution over time. Appendix B presents additional results on the marginal distributions.

5.2 Estimates of joint distribution

Having estimated the marginal distributions, we proceed with evaluating the joint distribution. We start with evaluating the conditional distribution $D(q_f|q_s, ma)$ given in equation (12). The quantile estimates of $D(q_f|q_s, ma)$ are reported in Table 4 for selected quantiles. Relying on bootstrapping, we conducted formal tests on the parameters $\boldsymbol{\gamma}(q_f)$ in (12), documenting that the linear term q_s, the nonlinear term q_s^s as well as ma all have statistically significant effects on the conditional distribution of q_f. In Table 4, the coefficient of q_s varies between 0.871 and 0.897, indicating positive contemporaneous codependence between pf and ps. The statistical significance of q_s and q_s^s presents strong evidence of departure from independence. Finally, Table 4 reports that the coefficient of ma is positive and statistically significant, showing that ma affects the codependence (pf, ps). Additional results on the codependence between p_f and p_s are presented in Appendix C.

6. Economic implications

This section shows how our approach provides useful insights into the relationships between the futures price and spot price. They include studying three sets of issues: dynamic stability, cointegration and convergence properties.

6.1 Dynamics and (in)stability

The price dynamics and stability have important implications for determining the effects of stochastic shocks on economic variables, and have been studied intensively in literature (e.g., Wang and Tomek, 2007). Recent work has studied the quantile-based unit root tests, though most of which apply in the univariate and linear contexts (Koenker and Xiao, 2004). Our QVAR approach extends previous research by evaluating dynamics and (in)stability issues in a multivariate and nonlinear context. We rely on the roots associated with equation (10), with a special focus on the dominant root $\lambda_1(\cdot)$. As discussed above, $\ln(|\lambda_1(\cdot)|)$ measures the rate of local divergence along a forward path. In our case, $\lambda_1(\cdot)$ can vary across quantiles and across market conditions.

Table 5 reports the modulus of the dominant root across quantiles and under different maturities of soybean futures. First, when holding ma constant, we find that the dominant roots λ_1 tends to increase with q_f. Table 5 shows dynamic stability (with $|\lambda_1| < 1$) in the lower quantiles of q_f, and dynamic instability (with $|\lambda_1| > 1$) in the upper quantiles of q_f, which are consistent in all ma scenarios. This result has three implications. First, dynamic stability varies with the quantiles q_f and with ma. Importantly, this finding requires nonlinear dynamics: it could not be obtained under a standard VAR model or any model exhibiting linear dynamics. Second, from Table 5, positive shocks in the futures markets (corresponding to the upper quantiles of q_f) are associated with increased instability. This is one of our important results: futures markets can contribute to increased price instability, although only in the upper tail of the futures price distribution. This result found in our QVAR model could not be obtained in any model restricted to mean-variance dynamics. Third, Table 5 shows that dominant root $|\lambda_1|$ increases from being below 1 to being above 1 as ma moves from "low" to "high". This is another important result: the futures market can help stabilize the soybean markets under nearby futures contract maturity; but this stabilizing effect does not apply to futures contracts with distant contract maturity. Such findings likely reflect difficulties in obtaining reliable information about the distant future. The destabilizing effects of long-term futures contracts may reflect limitations of futures markets: such effects are factors that may help explain why long-term futures contracts do not exist.

In the context of our nonlinear model, Table 5 documents the local properties of dynamic market stability. With $|\lambda_1| = 1$ being the threshold between stability and instability, we used bootstrapping to test two null hypotheses: local stability $H_0: |\lambda_1| = 1\ vs.\ H_1: |\lambda_1| <$

1 ; and local instability $H_0: |\lambda_1| = 1\ vs.\ H_1: |\lambda_1| > 1$. The test results reported in Table 5 indicate that a unit root, $|\lambda_1| = 1$, is not rejected around the median of the price distributions, a result commonly obtained in the time series literature (e.g., Enders, 2010). But it shows empirical support for dynamic stability $(|\lambda_1| < 1)$ in the lower quantiles of qf and when ma is low; and it finds statistical support for dynamic instability $(|\lambda_1| > 1)$ when $q_f = 0.9$. Finally, Table 5 indicates that rejecting the null hypothesis of instability becomes less likely under any quantiles (q_f, q_s) when ma is high. These findings make it clear that market stability is a local property that varies with market conditions.

Our quantile-based price dynamics results extend previous findings that typically rely on unit root tests applied to traditional VAR models, models exhibiting structural changes (e.g., Wang and Tomek, 2007) or models exhibiting mean-level nonlinear dynamics (e.g., Beckmann and Czudaj, 2014). It presents evidence that price dynamics vary with the nature of the shocks, stressing the need to distinguishing between upside shocks (e.g., due to tightened supply) versus downside shocks (e.g., due to reduced demand). Our finding that dynamic instability arise under a futures prices surge (when $q_f = 0.9$) can be interpreted as evidence of "local bubbles" located in the upper tail of futures price distribution. Such a finding is consistent with previous evidence on commodity price bubbles (e.g., Etienne et al., 2014). Yet, our quantile-based analysis goes beyond previous research by showing that such "local bubbles" depend on the nature of shocks and market conditions.

6.2 Cointegration and price discovery

Previous literature has explored the evaluation of long run cointegration relationships between futures and spot prices (e.g., Wang and Ke, 2005; Hernandez and Torero, 2010), typically relying on linear cointegration. In this study, given the evidence of local unit root just discussed, we explore the existence of long-run relationships based on nonlinear quantile cointegration (Xiao, 2009). Relying on our nonlinear model and equations (11a)- (11b), we evaluate the nature of local cointegration that can vary across quantiles (q_f, q_s) and depends on the evaluation point $\boldsymbol{z}$ (i.e., ma in our case). The quantile-based cointegration tests involve the rank of the matrix $\boldsymbol{\Pi}(\boldsymbol{q}, \boldsymbol{z})$ in (11b). We investigated whether $\boldsymbol{\Pi}(\boldsymbol{q}, \boldsymbol{z})$ has a reduced rank by testing (using bootstrapping) the null hypothesis that $\det(\boldsymbol{\Pi}(\boldsymbol{q}, \boldsymbol{z})) = 0$. We found strong statistical evidence that $\boldsymbol{\Pi}(\boldsymbol{q}, \boldsymbol{z})$ has reduced rank across quantiles and across ma, reflecting that pf and ps exhibit long-term cointegration relationships. The cointegration is represented by the right Eigenvector (v_1, v_2) of $\boldsymbol{\Pi}(\boldsymbol{q}, \boldsymbol{z})$ associated with its largest Eigenvalue. In our case, as discussed in section 3, the cointegration is "local" as it depends on the evaluation point $(\boldsymbol{q}, \boldsymbol{z})$. Table 6 reports the normalized cointegration vector (v_1 / v_2), showing how the nature of long-term cointegration between pf and ps varies across selected scenarios.

Under the three ma scenarios considered, Table 6 shows that (v_1 / v_2) tends to large and close to one in quantiles around the median, but (v_1 / v_2) becomes smaller in both the upper tails and the lower tails of the distributions. This indicates that futures-spot long term price relations (i.e., contango or backwardation) can vary in response to the types of shocks. This is consistent with the analysis of cointegration across quantiles presented in Xiao (2009) and Lee et al. (2011). But we go beyond their work by exploring how cointegration also varies with other variables (ma in our case). Table 6 shows that (v_1 / v_2) tends to be larger as ma increases. It provides additional evidence that cointegration relationships between pf and ps are nonlinear and complex: they vary depending on the nature of shocks (downside shocks versus upside shocks) and maturity characteristics of futures contracts.

Next, the implications of our analysis for price dynamics and price discovery are explored using a Quantile Impulse Response Function (QIRF). This is done by conducting forward simulations of our estimated model over a period of 150 weeks (approximately three years). Our simulations involve three scenarios: S1/ a base case; S2/ a one-time 30 percent positive price shock to the futures price pf; and S3/ a one-time 30 percent positive price shock to the spot price ps. The simulations were conducted as follows: first, at the initial period $t = t_0$, draw a random number q_s from a uniform distribution in the interval (0, 1), and compute the simulated spot prices ps evaluated at q_s using our QVAR estimates; second, draw a random number q_r from a uniform distribution in the interval (0, 1) and use the estimated conditional distribution function $D^e(q_f|q_s,\cdot)$ to obtain $q_f = D^e(q_r|q_s,\cdot)$; third, compute the simulated futures price pf evaluated at q_f using our QVAR estimates; fourth, replicate the random draws and calculations for 500 times and then move to the next period $t = t + 1$ for 150 weeks. The simulated prices are then used to evaluate the QIRF comparing scenarios S2 and S3 with the base case S1.

Figure 2 reports the QIRF generated from the forward simulations. Figure 2(a) shows that a positive shock from the spot market tends to put upward pressure on the futures price, causing sizeable but transitory increases in pf. In contrast, Figure 2(b) reports a different impulse response to a positive shock in the futures price: while the shock also triggers an increase in the spot price, its impacts are found to be significantly larger and more persistent. It takes much longer for the soybean market to absorb a shock from the futures market than from the spot market. While these findings are qualitatively consistent using impulse responses under different quantiles ($q \in (0.1, 0.5, 0.9)$), our analysis documents quantitively different responses to shocks. Overall, our refined approach provides strong evidence that the futures market plays a leading role in price discovery in the US soybean market.

6.3 Basis and convergence

We further rely on our forward simulations to evaluate the patterns exhibited by the basis, $pf - ps$. Our analysis proceeds with repeating the simulation exercises introduced above under alternative scenarios for ma.[10] Figure 3 reports the forward-path patterns of the basis under different quantiles and ma scenarios. Under the three ma scenarios considered, the results illustrate that the basis varies around zero in lower quantile (q=0.1), and gradually increase around the median (q=0.5) and in upper quantile (q=0.9). This is an important result. First, the case around the median (q=0.5) points to a normal situation where the expiring futures price is higher than the spot price, where $pf_{1t} - ps_t = C'_{kt} + R'_{kt} > 0$ as reflected in equation (4). The presence of a positive basis is consistent with market practice and is well-documented in the theory of storage (e.g., Working, 1949; Telser, 1958; Fama and French, 1987; Routledge et al., 2002). Second, the high-quantile case (q=0.9) shows that the basis can become relatively large, leading to "non-convergence" between futures price and spot price. This issue is discussed in more detail below.

Third, the low-quantile case unveils the possibility of having backwardation or "inverse-carrying charge", corresponding to the situation where $pf_{1t} - ps_t = C'_{kt} + R'_{kt} < 0$ (see equation (4) as discussed in section 2). Previous literature has relied on a "convenience yield" interpretation of this inverse-carrying charge as an explanation for why positive inventory would be held under declining prices (e.g., Working, 1949). In section 2, we provide an alternative interpretation from the perspective of risk: under risk aversion, R'_{kt} can be negative and possibly lead to $pf_{1t} - ps_t < 0$ (from equation (4)). When comparing different ma scenarios, Figure 3 shows that a negative basis can develop when ma is high. This is a case where the soybean futures-spot basis behavior is affected by the futures contract maturity properties. These results indicate that the behavior of the basis is situation-specific and exhibit time-varying patterns depending on market conditions.

Finally, how does the basis change as nearly futures contracts approach maturity? To answer this question, we simulate the case where ma starts high and decreases over time toward zero. Figure 4 reports the simulated convergence patterns across quantiles and over time, exemplified for selected quantiles q$\in$(0.1, 0.5, 0.9) and selected years (1990, 2000 and 2010). As nearby contracts get close to maturity, the basis exhibits similar changes across quantiles but varies significantly over time. Compared with the 1990s and 2000s, the basis became larger in the 2010s, implying that its convergence property deteriorated. This is consistent with the findings of recent research. For instance, Garcia et al. (2014) develop a dynamic commodity storage model to explain how the institutional structure of the delivery system caused convergence failures and trigged surging "price wedge" (the difference between carrying storage price and maximum storage rate). As discussed in section 2, convergence properties depend on $C'_{kt} + R'_{kt}$, indicating that reducing transaction cost and improving information can contribute to reaching convergence. In this context, the Chicago Board of Trade has made several modifications to futures contracts that may have contributed to recent improvements in the functioning of the soybean futures market and its convergence performance (Hoffamn and Aulerich, 2013).

7. Conclusion

This paper presents a dynamic analysis of the distribution of commodity futures and spot prices, with an application to the US soybean market over the period of 1980-2019. We propose a two-step QVAR-Copula method to estimate the marginal and joint price distributions. In the first step, we specify and estimate a quantile vector autoregression (QVAR) model representing the marginal distributions of futures and spot prices. We find strong evidence of price dynamics, including the existence of nonlinear dynamics. We examine how the marginal distributions evolve over time and in response to changing market conditions, reflected by time-varying mean, variance, skewness and kurtosis. In the second step, we estimate a conditional distribution to recover the copula linking the marginal distribution to the joint price distribution. We show evidence of strong positive contemporaneous codependence between futures and spot prices at all evaluation points. We also evaluate how exogenous variables (including futures contract maturity) affect the dynamics and codependence of prices.

The application of our approach to the US soybean market provides an illustration of its flexibility and usefulness. Our analysis evaluates dynamic stability and finds evidence of local dynamic instability in the upper tail of the price distributions. We investigate the nature of cointegration capturing the long-term relationship between futures price and spot price; we find evidence of nonlinear cointegration as the long-term relationship varies with market conditions. We also report quantile-specific impulse response functions and find that shocks from futures market trigger larger and more persistent impacts on prices. This documents the importance of the futures market in soybean price discovery. Finally, we evaluate the basis and discuss its dynamic properties and the (non)convergence of the futures and spot price under different market conditions.

While our QVAR-Copula approach is broadly applicable to the analysis of price dynamics, it could be extended in several directions. First, our application was limited to the US soybean market. It would be useful to apply it to study the price behavior of other commodities and other markets (e.g., financial markets). Second, as our approach allows for a flexible assessment of price risk exposure, it could be used to evaluate risk management, including the economics of hedging, insurance and derivative markets. Third, our investigation has found nonlinearities in short term price adjustments and in long term cointegration. More research is needed to explore the implications of these nonlinearities for market dynamics. Fourth, our econometric analysis has uncovered evidence of local instability in the upper tail of the price distribution. At this point, the global implications of this local instability remain unclear. These seem to be good topics for future research.

References

Beckmann, J. and R. Czudaj. 2014. Non-linearities in the relationship of agricultural futures prices. *European Review of Agricultural Economics*, 41(1), 1-23.

Caner, M., and B.E. Hansen, 2001. Threshold autoregression with a unit root. *Econometrica*, 69(6), 1555-1596.

Cattelan, A.J. and A. Dall'Agnol. 2018. The rapid soybean growth in Brazil. *Oilseeds & Fats Crop Journal*, 25, 1-12.

Chavas, J.P., D. Hummels and B.D. Wright (Eds.). 2014. *The economics of food price volatility*. University of Chicago Press.

Chavas, J.P. 2020. The dynamics and volatility of prices in multiple markets: a quantile approach, *Empirical Economics*, 11, 1-22.

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

Fama, E. F. and K. R. French. 1987. Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage. *Journal of Business,* 60(1), 55 – 73.

Enders, W. 2010. *Applied econometric time-series*. 3rd edtion. Wiley, New York.

Engle, R. and C. Granger. 1987. Co-integration and error correction: Representation, estimation and testing. *Econometrica,* 55(2), 251-276.

Etienne, X. L., S.H. Irwin and P. Garcia. 2015. Price explosiveness, speculation, and grain futures prices. *American Journal of Agricultural Economics*, 97(1), 65-87.

Finkelshtain, I. and J.A. Chalfant. 1991. Marketed surplus under risk: Do peasants agree with Sandmo? *American Journal of Agricultural Economics*, 73(3), 557-567.

Garbade, K. D. and W.L. Silber. 1983. Price movements and price discovery in futures and cash markets. *The Review of Economics and Statistics*, 65(2), 289-297.

Garcia, P. and R.M. Leuthold. 2004. A selected review of agricultural commodity futures and options markets. *European Review of Agricultural Economics*, 31(3), 235-272.

Garcia, P., S.H. Irwin and A. Smith. 2015. Futures market failure? *American Journal of*

Agricultural Economics, 97(1), 40-64.

Kapetanios, G., Y. Shin and A. Snell. 2003. Testing for a unit root in the nonlinear STAR framework. *Journal of Econometrics*, 112(2), 359-379.

Hamilton, J. D. 1994. *Time Series Analysis*. Princeton University Press, Princeton.

Hernandez, M., and M. Torero. 2010. Examining the dynamic relationship between spot and futures prices of agricultural commodities (No. 988). International Food Policy Research Institute (IFPRI).

Huang, J., T. Serra and P. Garcia. 2020. Are futures prices good price forecasts? Underestimation of price reversion in the soybean complex. *European Review of Agricultural Economics*, 47(1), 178-199.

Hoffman, L.A. and N. Aulerich. 2013. Recent convergence performance of futures and spot prices for corn, soybeans, and wheat. FDS-13L-01, Economic Research Service, U.S. Department of Agriculture.

Joe, H. 2014 *Dependence modeling with copulas*. Chapman and Hall, CRC Press, New York.

Johansen, S. 1995. *Likelihood-based inference in cointegrated vector autoregressive models.* Oxford University Press on Demand.

Kaldor, N.1938. Speculation and Economic Stability. *Review of Economic Studies*, 7, 1-27.

Koenker R. 2005. *Quantile regression*. Cambridge University Press, Cambridge.

Koenker, R. and J.A. Machado. 1999. Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94, 1296-1310.

Koenker, R. and Z. Xiao. 2004. Unit root quantile autoregression inference. *Journal of the American Statistical Association*, 99, 775-787.

Koenker, R. and Z. Xiao. 2006. Quantile autoregression. *Journal of the American Statistical Association*, 101, 980-990.

Lee, C. C. and J.H. Zeng. 2011. Revisiting the relationship between spot and futures oil prices: evidence from quantile cointegrating regression. *Energy economics*, 33(5), 924-935.

Murphy, K.M. and R.H. Topel. 2002. Estimation and Inference in Two-Step Econometric

Models. *Journal of Business & Economic Statistics*, 20(1), 88-97.

Nelsen, R.B. 2006. *An Introduction to Copulas*, New York: Springer.

Patton, A.J. 2006. Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2), 527-556.

Pizzi C. 2010. Nonlinear cointegration in financial time series. In: Corazza M., Pizzi C. (eds) *Mathematical and Statistical Methods for Actuarial Sciences and Finance*. Springer, Milano.

Ramsey, A.F. 2020. Probability distributions of crop yields: A bayesian spatial quantile regression approach. *American Journal of Agricultural Economics*, 102(1), 220-239.

Reboredo, J.C. 2012. Modelling oil price and exchange rate co-movements. *Journal of Policy Modeling*, 34(3), 419-440.

Routledge, B.R., D.J. Seppi, and C.S. Spatt. 2002. Equilibrium forward curves for commodities. *Journal of Finance,* 55(3), 1297-1328.

Sandmo, A. 1971. On the theory of the competitive firm under price uncertainty. *American Economic Review*, 61(1), 65-73.

Sayama, H. 2015. Introduction to the Modeling and Analysis of Complex Systems. Open SUNY Textbooks. https://knightscholar.geneseo.edu/oer-ost/14.

Sklar, A. 1959. *Fonctions de répartition à n dimensions et leurs marges*, Publ. Inst. Statist. Univ. Paris, 8, 229-231.

Sockin, M. and W. Xiong. 2015. Informational frictions and commodity markets. *The Journal of Finance*, 70(5), 2063-2098.

Szymanowska, M., F. De Roon, T. Nijman and R. Van Den Goorbergh. 2014. An anatomy of commodity futures risk premia. *The Journal of Finance*, 69(1), 453-482.

Telser, L. 1958. Futures trading and the storage of cotton and wheat. *Journal of Political Economy*, 66(3), 233-255.

Tjøstheim, D. 2020. Some notes on nonlinear cointegration: A partial review with some novel perspectives, *Econometric Reviews*, 39(7), 655-673.

Wang, D. and W.G. Tomek. 2007. Commodity prices and unit root tests. *American Journal of Agricultural Economics*, 89(4), 873-889.

Wang, H. H. and B. Ke. 2005. Efficiency tests of agricultural commodity futures markets in China. *Australian Journal of Agricultural and Resource Economics*, 49(2), 125-141.

Working, H. 1949. The theory of price of storage, *American Economic Review*, 39(6), 1254-1262.

Xiao, Z. 2009. Quantile cointegrating regression. *Journal of Econometrics*, 150(2), 248-260.

Table 1. Summary statistics.

Variables	Mean	St. Dev.	Max	Min
pf	1.994	0.325	2.863	1.418
pf_1	1.994	0.325	2.863	1.418
pf_2	1.994	0.325	2.863	1.418
pf_1^s	0.113	0.162	0.910	0.000
pf_2^s	0.113	0.162	0.910	0.000
ps	1.938	0.332	2.862	1.330
ps_1	1.938	0.332	2.862	1.330
ps_2	1.937	0.332	2.862	1.330
ps_1^s	0.116	0.168	0.993	0.000
ps_2^s	0.116	0.168	0.993	0.000
ma	3.499	2.496	9.000	0.000
$ma \cdot pf_1$	6.975	5.165	23.911	0.000
tt	2.000	1.154	3.998	0.004
tt_1	6.047	7.217	21.981	0.000
tt_2	2.809	4.492	14.981	0.000

Note: The data involve weekly observations over the period 1980-2019.

Table 2. QVAR and VAR estimates: futures price equation pf

Variable	VAR	QVAR: pf equation				
		q = 0.1	q = 0.3	q = 0.5	q = 0.7	q = 0.9
Intercept	0.033***	0.066***	0.049***	0.017	0.003	-0.002
	(0.009)	(0.019)	(0.012)	(0.010)	(0.008)	(0.013)
ps_1	0.246***	0.336**	0.226***	0.139*	0.093	0.060
	(0.053)	(0.151)	(0.083)	(0.082)	(0.080)	(0.098)
pf_1	0.909***	0.757***	0.907***	1.018***	1.088***	1.184***
	(0.055)	(0.161)	(0.080)	(0.077)	(0.077)	(0.099)
ps_2	-0.248***	-0.250	-0.162**	-0.155*	-0.139*	-0.120
	(0.053)	(0.160)	(0.082)	(0.086)	(0.076)	(0.092)
pf_2	0.078	0.112	0.000	-0.009	-0.036	-0.107
	(0.055)	(0.175)	(0.083)	(0.080)	(0.072)	(0.098)
pf_1^s	0.017	0.204***	0.110***	-0.004	-0.071	-0.181**
	(0.034)	(0.075)	(0.041)	(0.047)	(0.049)	(0.071)
pf_2^s	-0.018	-0.201***	-0.111***	0.004	0.063	0.158**
	(0.034)	(0.071)	(0.042)	(0.049)	(0.049)	(0.072)
ma	-0.004***	-0.006*	-0.005***	-0.003*	-0.004**	-0.005**
	(0.001)	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)
$ma \cdot pf_1$	0.002***	0.003*	0.003***	0.001	0.002**	0.002**
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Q_{1t}	0.000	0.002	0.001	0.000	0.000	-0.004**
	(0.002)	(0.003)	(0.001)	(0.001)	(0.002)	(0.002)
Q_{2t}	0.000	-0.001	-0.002	-0.001	0.002	0.001
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Q_{3t}	-0.007***	-0.019***	-0.008***	-0.007***	0.000	0.006***
	(0.002)	(0.004)	(0.002)	(0.002)	(0.003)	(0.002)
tt	0.001	0.001	0.000	0.001	0.003**	0.000
	(0.002)	(0.003)	(0.002)	(0.001)	(0.002)	(0.002)
tt_1	0.001*	0.000	0.001	0.001*	0.002***	0.002***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)
tt_2	-0.001	0.002*	0.000	-0.001**	-0.002***	-0.005***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Goodness of fit	Adjusted R^2	Pseudo R^2				
	0.993	0.892	0.913	0.927	0.933	0.934

Note: Lagged prices are denoted using subscript (e.g., $ps_j = ps_{t-j}, j \in \{1,2\}$). Standard errors (presented in parentheses) are obtained using bootstrapping, with statistical significance represented by stars: * = p<0.1; ** = p<0.05; *** = p<0.01. For the "goodness of fit", we report the adjusted R^2 for OLS estimates and the Pseudo-R^2 proposed by Koenker and Machado (1999) for quantile estimates.

Table 3. QVAR and VAR estimates: spot price equation ps

Variable	VAR	QVAR: ps equation				
		q = 0.1	q = 0.3	q = 0.5	q = 0.7	q = 0.9
Intercept	0.013	0.053***	0.024**	0.012	-0.008	-0.009
	(0.009)	(-0.017)	(0.011)	(0.012)	(0.010)	(0.011)
ps_1	1.148***	1.316***	1.221***	1.144***	1.093***	0.968***
	(0.055)	(0.130)	(0.093)	(0.072)	(0.077)	(0.114)
pf_1	0.009	-0.187	-0.108	0.014	0.095	0.217*
	(0.055)	(0.126)	(0.092)	(0.068)	(0.080)	(0.117)
ps_2	-0.275***	-0.302**	-0.292***	-0.256***	-0.259***	-0.132
	(0.055)	(0.138)	(0.094)	(0.073)	(0.078)	(0.115)
pf_2	0.109**	0.133	0.159*	0.091	0.079	-0.035
	(0.055)	(0.137)	(0.094)	(0.072)	(0.080)	(0.118)
ps_1^s	0.067**	0.218***	0.127***	0.036	-0.016	-0.118**
	(0.033)	(0.080)	(0.045)	(0.061)	(0.042)	(0.055)
ps_2^s	-0.065**	-0.212***	-0.123***	-0.030	0.015	0.099*
	(0.033)	(0.081)	(0.045)	(0.062)	(0.041)	(0.056)
ma	-0.004***	-0.006**	-0.005***	-0.004**	-0.004***	-0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
$ma \cdot pf_1$	0.002**	0.003**	0.002***	0.002*	0.002**	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Q_{1t}	-0.001	-0.002	0.002	0.000	-0.002	-0.004**
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Q_{2t}	0.000	-0.002	0.000	0.000	0.001	0.000
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Q_{3t}	-0.008***	-0.023***	-0.011***	-0.009***	-0.003	0.004
	(0.002)	(0.004)	(0.002)	(0.002)	(0.002)	(0.003)
tt	0.000	0.000	0.001	0.001	0.000	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
tt_1	0.000	-0.001	0.000	0.000	0.001*	0.002***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
tt_2	-0.001	0.002**	0.001	-0.001	-0.002***	-0.005***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Goodness of fit	Adjusted R^2	Pseudo R^2				
	0.993	0.894	0.913	0.926	0.932	0.935

Note: Lagged prices are denoted using subscript (e.g., $ps_j = ps_{t-j}, j \in \{1,2\}$). Standard errors (presented in parentheses) are obtained using bootstrapping, with statistical significance represented by stars: * = p<0.1; ** = p<0.05; *** = p<0.01. For the "goodness of fit", we report the adjusted R^2 for OLS estimates and the Pseudo-R^2 proposed by Koenker and Machado (1999) for quantile estimates.

Table 4. Quantile Estimation of the conditional distribution $D(q_f \mid q_s, ma)$, selected quantiles

Variable	estimates				
	$q_f = 0.1$	$q_f = 0.3$	$q_f = 0.5$	$q_f = 0.7$	$q_f = 0.9$
Cons	-0.109***	-0.022**	0.039***	0.102***	0.241***
	(0.010)	(0.010)	(0.010)	(0.011)	(0.019)
q_s	0.871***	0.891***	0.896***	0.897***	0.857***
	(0.009)	(0.007)	(0.006)	(0.007)	(0.010)
q_s^s	0.576***	0.208***	0.017	-0.161***	-0.638***
	(0.064)	(0.044)	(0.043)	(0.044)	(0.078)
ma	0.002*	0.003***	0.003***	0.004***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Note: q_s^s is defined as $q_s^s = (q_s - 0.5)^2$. Standard errors (in parentheses) are obtained using empirical bootstrapping applied over both steps of the two-step estimation approach, with statistical significance represented by stars: * = p<0.1; ** = p<0.05; *** = p<0.01.

Table 5. Modulus of the dominant roots λ_1 under alternative scenarios

ma_low

q_s \ q_f	q=0.1	q=0.3	q=0.5	q=0.7	q=0.9
q=0.1	0.963	0.970	1.042	1.082*	1.128*
	(0.034)	(0.037)	(0.055)	(0.066)	(0.078)
q=0.3	0.973**	0.976**	1.004	1.030	1.058
	(0.017)	(0.012)	(0.021)	(0.034)	(0.049)
q=0.5	0.976*	0.979**	0.994	1.006	1.012
	(0.016)	(0.013)	(0.012)	(0.016)	(0.027)
q=0.7	0.975*	0.980**	0.993	1.000	0.992*
	(0.017)	(0.011)	(0.013)	(0.009)	(0.018)
q=0.9	0.974**	0.979**	0.995	1.003	0.997*
	(0.016)	(0.010)	(0.013)	(0.011)	(0.015)

ma_medium

q_s \ q_f	q=0.1	q=0.3	q=0.5	q=0.7	q=0.9
q=0.1	0.971	0.974	1.044	1.083*	1.128*
	(0.030)	(0.033)	(0.052)	(0.064)	(0.077)
q=0.3	0.978*	0.982**	1.009	1.033	1.060
	(0.015)	(0.010)	(0.019)	(0.032)	(0.047)
q=0.5	0.979	0.982*	0.997	1.008	1.012
	(0.015)	(0.012)	(0.011)	(0.015)	(0.026)
q=0.7	0.979*	0.983**	0.996	1.003	0.994*
	(0.016)	(0.011)	(0.012)	(0.008)	(0.018)
q=0.9	0.979*	0.985**	1.000	1.008	1.002
	(0.016)	(0.010)	(0.013)	(0.011)	(0.012)

ma_high

q_s \ q_f	q=0.1	q=0.3	q=0.5	q=0.7	q=0.9
q=0.1	0.980	0.984	1.047	1.085*	1.128*
	(0.028)	(0.030)	(0.050)	(0.062)	(0.076)
q=0.3	0.986	0.990	1.015	1.039	1.064
	(0.013)	(0.009)	(0.017)	(0.030)	(0.044)
q=0.5	0.984	0.987	1.001	1.010	1.014
	(0.016)	(0.013)	(0.010)	(0.015)	(0.025)
q=0.7	0.983	0.988	1.000	1.007	0.998
	(0.017)	(0.011)	(0.012)	(0.009)	(0.018)
q=0.9	0.987	0.993	1.008	1.015	1.009
	(0.017)	(0.012)	(0.014)	(0.014)	(0.012)

Note: The results are presented under three ma scenarios: low, medium and high, corresponding to ma being set at the 0.1, 0.5 and 0.9 sample quantiles, respectively. We use bootstrapping to test two null hypotheses: local stability $H_0: |\lambda_1| = 1 \; vs. \; H_1: |\lambda_1| < 1$; and local instability $H_0: |\lambda_1| = 1 \; vs. \; H_1: |\lambda_1| > 1$. Bootstrapped standard errors are presented in parentheses, with statistical significance represented by stars: * = p<0.1; ** = p<0.05; *** = p<0.01.

Table 6. Normalized cointegration vectors (v_1/v_2) under alternative scenarios

ma_low					
q_s \ q_f	q=0.1	q=0.3	q=0.5	q=0.7	q=0.9
q=0.1	0.590	0.591	0.267	0.561	0.563
q=0.3	0.777	0.915	1.437	1.132	0.907
q=0.5	0.841	0.941	1.102	1.038	0.938
q=0.7	0.848	0.900	0.954	0.938	0.898
q=0.9	0.773	0.810	0.841	0.835	0.806
ma_medium					
q_s \ q_f	q=0.1	q=0.3	q=0.5	q=0.7	q=0.9
q=0.1	0.645	0.664	0.342	0.594	0.564
q=0.3	0.814	0.944	1.255	1.021	0.833
q=0.5	0.862	0.948	1.048	0.983	0.886
q=0.7	0.853	0.895	0.926	0.907	0.865
q=0.9	0.770	0.796	0.806	0.798	0.770
ma_high					
q_s \ q_f	q=0.1	q=0.3	q=0.5	q=0.7	q=0.9
q=0.1	0.732	0.789	0.468	0.615	0.553
q=0.3	0.858	0.968	1.073	0.902	0.752
q=0.5	0.885	0.951	0.982	0.917	0.825
q=0.7	0.857	0.886	0.890	0.867	0.824
q=0.9	0.763	0.775	0.764	0.755	0.726

Note: The cointegration vector (v_1, v_2) is the right Eigenvector of the matrix $\boldsymbol{\Pi}(\boldsymbol{q}, \boldsymbol{z})$ associated with its largest Eigenvalue, where $\boldsymbol{\Pi}(\boldsymbol{q}, \boldsymbol{z})$ is given in equation (11b).

Figure 1. Trajectories of soybean spot prices log(ps) and futures prices log(pf), 1980-2019

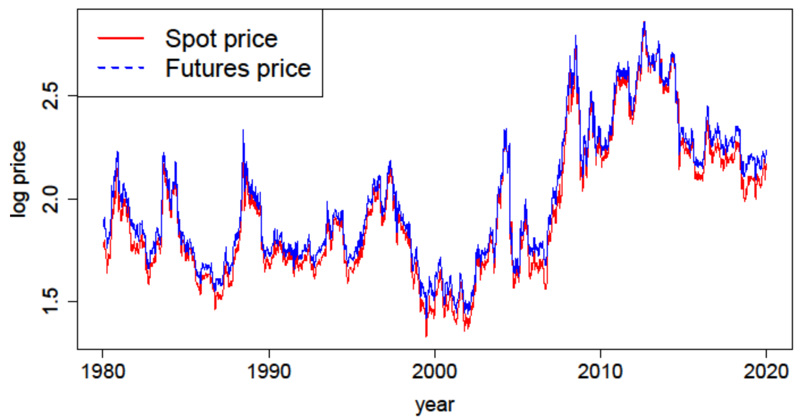
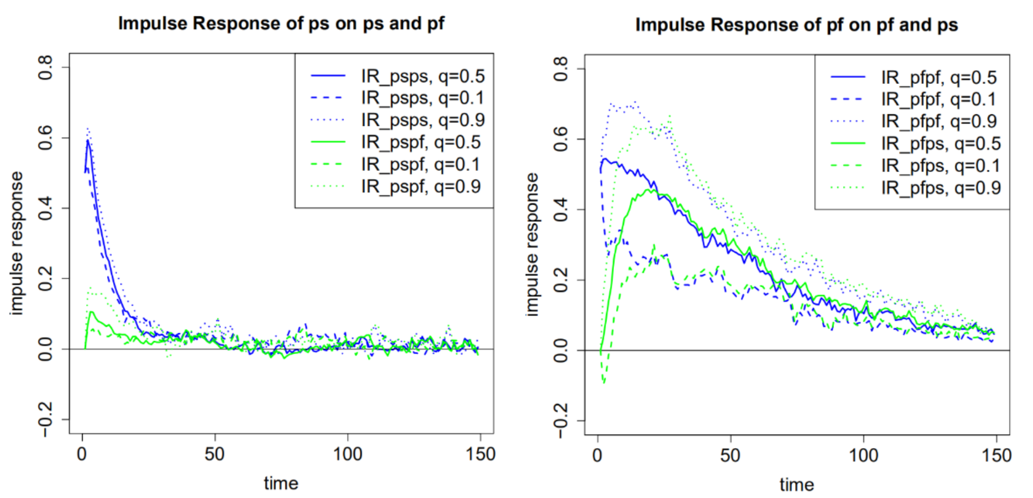
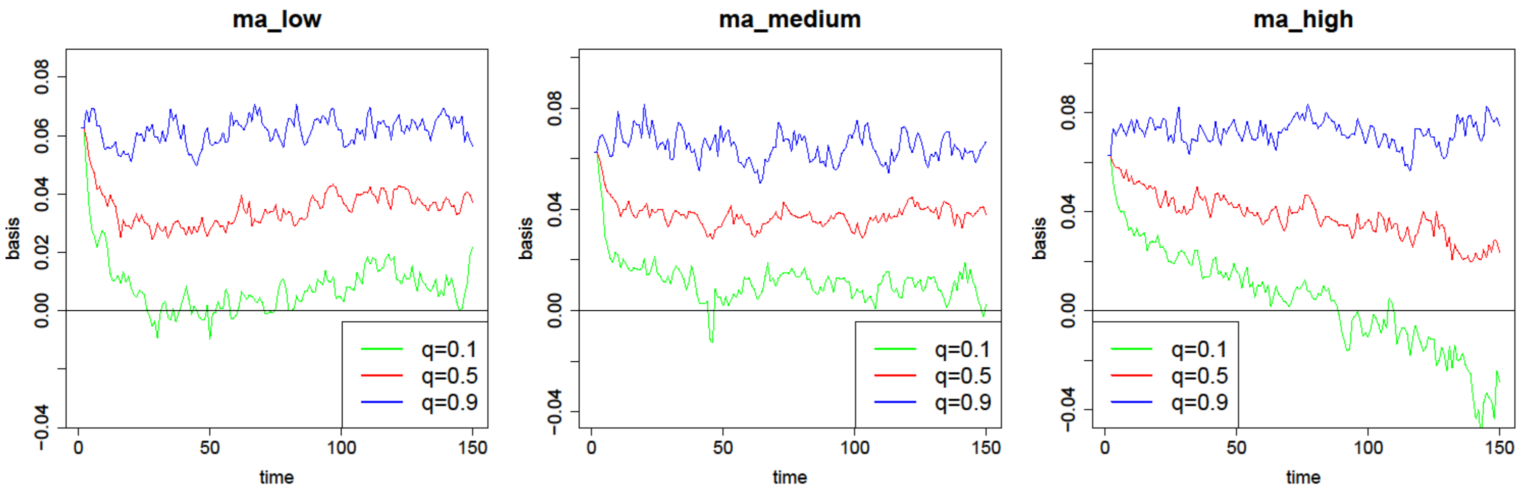


Figure 2. Quantile impulse responses



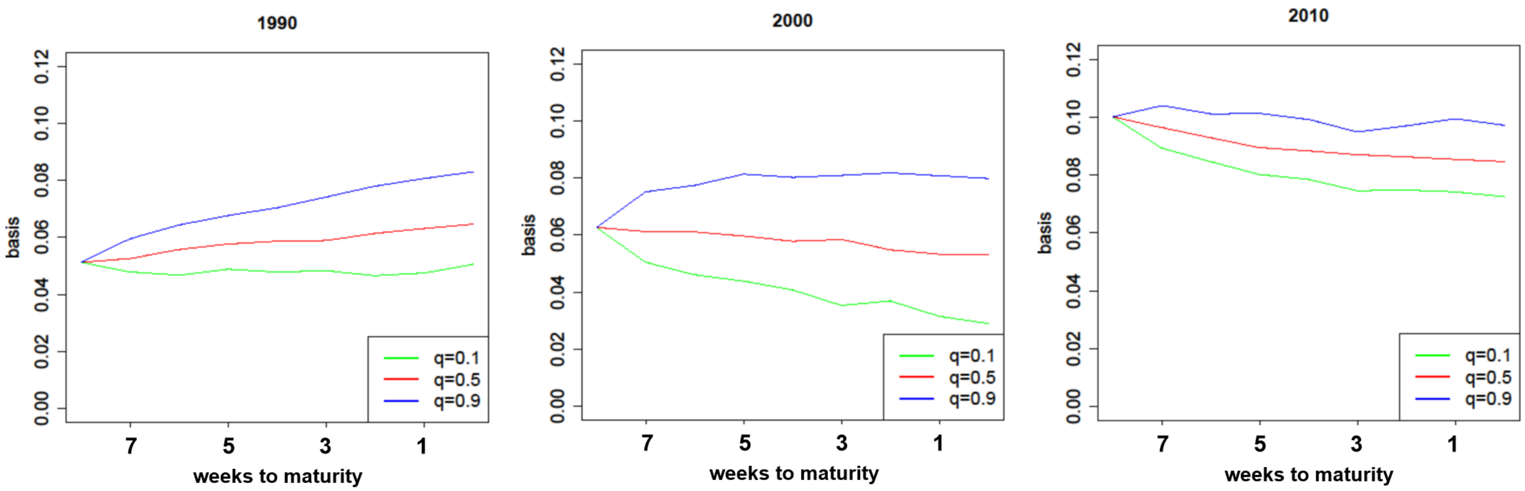
Note: In our notation, "IR_pspf, q=0.5" denotes the impulse response of a one-time 30 percent change in ps on the forward path of pf, evaluated at 0.5 quantile.

Figure 3. Forward-path patterns of basis under alternative scenarios



Note: The three ma scenarios (low, medium and high) correspond to ma being set at the 0.1, 0.5 and 0.9 sample quantiles, respectively.

Figure 4. Forward-path patterns of non-convergence under alternative scenarios



Note: These simulated paths correspond to scenarios where ma is set to decrease over time toward zero.

Appendix A

Consider the case where the k-th agent intends to buy a quantity Q_{kt} at price p_{at} at time t and to sell it at price $p_{b,t+\tau}$ at time $(t+\tau), \tau > 0$. In situations where $Q_{kt} < 0$, this would correspond to the k-th agent selling $|Q_{kt}|$ at price p_{at} at time t and buying it at price $p_{b,t+\tau}$ at time $(t+\tau)$. Alternative interpretations of Q_{kt} and of prices $(p_{at}, p_{b,t+\tau})$ are discussed in the text. Denote by $C_k(|Q_{kt}|)$ the transaction cost of Q_{kt}. The associated present value of profit is $\pi_{kt} = (r^\tau\, p_{b,t+k} - p_{at})\, Q_{kt} - C_k(|Q_{kt}|)$, where $r \in (0,1)$ is a discount factor. The price $p_{b,t+\tau}$ being uncertain at time t, the k-th agent formulates expectation about price $p_{b,t+\tau}$ based on available information. Assume that the k-th agent maximizes the expected utility of profit π_{kt}: $E_{kt}U_k\big[(r^\tau\, p_{b,t+\tau} - p_{at}\big]\, Q_{kt} - C_k(|Q_{kt}|)]$, where E_{kt} is the expectation operator based on the information available to the k-th agent at time t and $U_k(\pi_{kt})$ is a utility function representing the risk preferences of the k-th agent. We assume that $\frac{\partial U_k(\pi_{kt})}{\partial \pi_{kt}} > 0$ and $\frac{\partial^2 U_k(\pi_{kt})}{\partial \pi_{kt}^2} \leq 0$, where $\frac{\partial^2 U_k(\pi_{kt})}{\partial \pi_{kt}^2} < 0$ under risk aversion. When $Q_{kt} \neq 0$ and under differentiability, the optimal decision is given by the first-order condition

$$E_{kt}[U'_{kt} \cdot (r^\tau\, p_{b,t+\tau} - p_{at} - C'_k)],$$

or

$$r^\tau\, E_{kt}\big(p_{b,t+\tau}\big) - p_{at} = C'_{kt} + R'_{kt}, \qquad \text{(A1)}$$

where $U'_{kt} = \frac{\partial U_k(\pi_{kt})}{\partial \pi_{kt}}$, $C'_{kt} = \frac{\partial C_k(|Q_{kt}|)}{\partial |Q_{kt}|}$ is the marginal cost of Q_{kt} and $R'_{kt} = -r^\tau\, Cov(U'_{kt}, p_{b,t+\tau})/E_{kt}(U'_{kt})$ is the marginal risk premium. Equation (A1) states an intertemporal arbitrage condition between the price p_{at} and $p_{b,t+\tau}$: the expected discounted price difference $[r^k E_{kt}(p_{b,t+\tau}) - p_{at}]$ is equal to the marginal cost plus the marginal risk premium: $C'_{kt} + R'_{kt}$. As discussed in the text, the interpretation of equation (A1) varies depending on what Q_{kt} is and on what the prices $(p_{at}, p_{b,t+\tau})$ represent.

As just noted, we interpret the term $R'_{kt} = -r^\tau Cov(U'_{kt}, p_{b,t+\tau})/E_{kt}(U'_{kt})$ as a marginal risk premium. Indeed, under risk neutrality, the utility function $U_k(\pi_{kt})$ is linear, implying that $Cov(U'_{kt}, p_{b,t+\tau}) = 0$ (as U'_{kt} is a constant) and $R'_{kt} = 0$. This raises the question: What is the sign of R'_{kt} under risk aversion? In this case, $U_k(\pi_{kt})$ is strictly concave in π_{kt} and $\frac{\partial^2 U_k(\pi_{kt})}{\partial \pi_{kt}^2} < 0$. Allowing the quantity Q_{kt} to be positive for a sale at time $(t+\tau)$ but negative for a purchase at time $(t+\tau)$, we have the following result.

<u>Lemma 1</u>: Under risk and risk aversion, the marginal risk premium satisfies

$$R'_{kt} \begin{bmatrix} > \\ = \\ < \end{bmatrix} 0 \text{ when } Q_{kt} \begin{bmatrix} > \\ = \\ < \end{bmatrix} 0. \tag{A2}$$

<u>Proof</u>: By definition, given $U'_{kt} > 0$, R'_{kt} is of the sign of $[-Cov(U'_{kt}, pb_{t+\tau})]$. When $p_{b,t+\tau}$ has a non-degenerate distribution, we have

$$sign\{-Cov(U'_{kt}, p_{b,t+\tau})\} = sign\left\{-\frac{\partial U'_{kt}}{\partial p_{b,t+\tau}}\right\} = sign\{-\frac{\partial^2 U_k(\pi_{kt})}{\partial \pi_{kt}^2}\, r^\tau\, Q_{kt}\},$$

which gives (A2).

Q.E.D.

Lemma 1 states that, under risk and risk aversion, the marginal risk premium R'_k is positive when $Q_{kt} > 0$, i.e. when the k-th agent sells the quantity Q_{kt} at price $p_{b,t+\tau}$. Alternatively, it states that the marginal risk premium R'_k is negative when $Q_{kt} < 0$, i.e. when the k-th agent buys the quantity $|Q_{kt}|$ at price $p_{b,t+\tau}$. The case where $R'_k > 0$

when $Q_{kt} > 0$ is a well-known result in the economic literature (e.g., Sandmo, 1971): risk aversion implies a positive marginal risk premium which provides a disincentive to sell at time $(t + \tau)$ in the presence of price risk. As noted by Finkelshtain and Chalfant (1991), Lemma 1 also shows an opposite result when $Q_{kt} < 0$: risk aversion then implies a negative marginal risk premium which provides an extra incentive to buy at time $(t + \tau)$ in the presence of price risk. This result is important for two reasons. First, it makes it clear that, under risk and risk aversion, market position matters: risk affects buyers and sellers differently. Second, by showing that R'_k can change sign, equation (A2) implies that the marginal risk premium cannot be treated as a constant.

Appendix B

Going beyond the estimates reported in Tables 2-3, we re-estimate the QVAR model reported for all quantiles, generating estimates of the quantile functions and their inverse: the distribution functions (Huang et al., 2020; Ramsey, 2020). The estimated distribution functions of pf and ps are shown in Figure B1 for selected years: 1990, 2000 and 2010 (evaluated at the first week of each year). Figure B1 documents several interesting results. First, the estimated distributions are smooth, indicating the quantile estimates based on our specification and data are flexible in modeling the underlying price distributions. Second, Figure B1 shows that the price levels and shapes of the price distribution change over time. For instance, compared with the 1990s and 2000s, the shapes of price distributions in the 2010s exhibit longer tails for both futures and spot price, which is consistent with increasing market volatility associated with the world food crisis over the period of 2007-2014 (Chavas et al., 2014). These findings illustrate the flexibility and usefulness of our QVAR approach in capturing the changing nature of the price distributions.

After estimating the marginal price distributions of pf and ps, we proceed to evaluate its evolution over time. Using equations (8a)-(8b), we define the relative quantiles $pi_q(q_i| \cdot)/pi_q(0.5| \cdot), q_i \in (0, 1),\ i \in \{f, s\}$ as measures of the relative range of prices obtained under evolving market conditions. The results are presented in Figure B2 for selected quantiles. Figure B2 shows how the relative quantiles for futures and spot prices have evolved over time, reporting a widening and asymmetric range during the 2008 world food crisis. Again, going beyond standard regression analysis, this illustrates how our QVAR approach can capture complex price responses to market shocks.

Next, based on the estimated distributions, we calculate the moments (mean, standard deviation, skewness and kurtosis) of estimated price distributions over the sample period. The results are reported in Figure B3. The trajectories of mean prices are very similar to actual prices, reflecting that our QVAR model provides a good fit to the data. The standard deviations capture the evolving price fluctuations in soybean futures and spot markets (e.g., booms and busts during the 2008 world food crisis). Interestingly, we observe deviations from log-normal distributions: Figure B3 shows the existence of non-zero skewness and excessive kurtosis (i.e., when skewness is non-zero and kurtosis is greater than three), which is consistent with findings in recent work (Huang et al., 2020).

Figure B1. Selected futures and spot price distributions in the year of 1990, 2000 and 2010 (evaluated at the first week of each year)

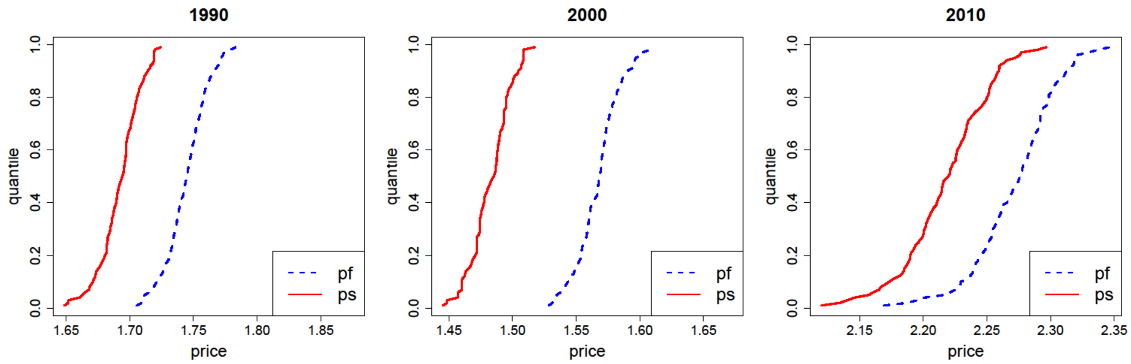


Figure B2. Evolution of relative quantiles (relative to the median) of simulated distributions

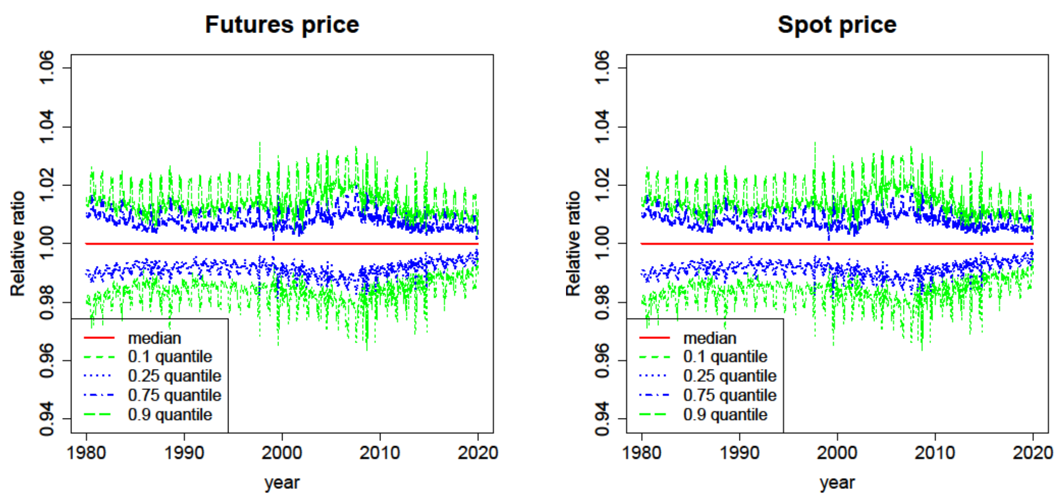
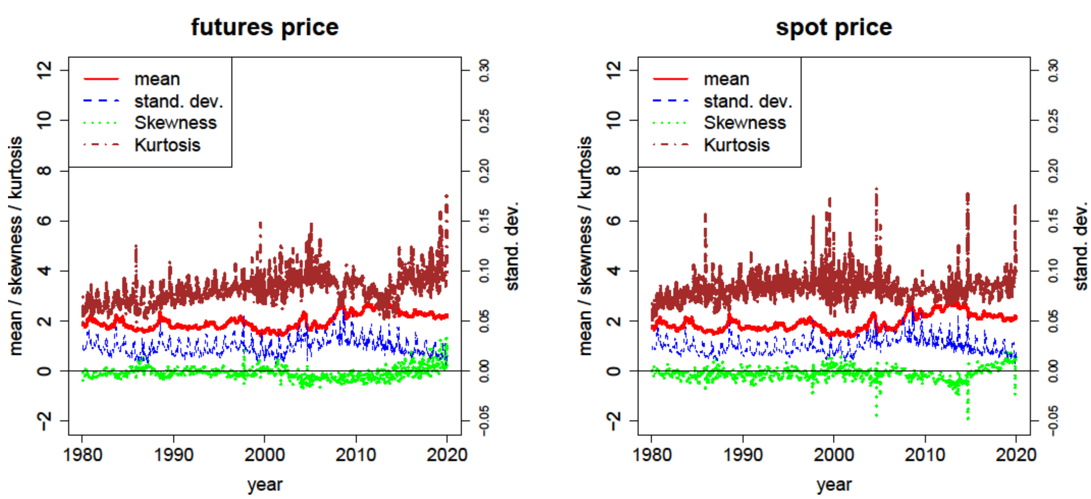


Figure B3. Moments of estimated futures and spot price distributions over time



Appendix C

Using the estimate $D^e(q_f|q_s, ma)$, we further estimate the copula $C^e\big(q_f, q_s|\, ma\big) \equiv \int_0^{q_s} D^e\big(q_f\big|\, q, ma\big)\, dq$ and the associated codependence measure $R^e\big(q_f, q_s|\, ma\big) \equiv C^e\big(q_f, q_s|\, ma\big) - q_f \cdot\, q_s$ given in equation (13). As discussed in the text, $R\big(q_f, q_s|\, ma\big) \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} 0$ under $\begin{Bmatrix} \text{positive dependence} \\ \text{independence} \\ \text{negative dependence} \end{Bmatrix}$ between pf *and* ps evaluated at point (q_f, q_s, ma). Table C1 reports the estimates of codependence $R\big(q_f, q_s|\, ma\big)$ for selected quantiles (q_f, q_s), evaluated at the sample median for ma. It shows that the codependence between ps and pf is positive for all quantiles, reflecting close contemporaneous relationships between the futures price and the spot price. Interestingly, the codependence remains positive in the tails; but it is stronger in the upper tail than in the lower tail: $R^e(0.9{,}0.9|\,\cdot) = 0.079 > 0.067 = R^e(0.1{,}0.1|\,\cdot)$. Our estimates of tail codependence are consistent with previous research using copula to document the role of tail dependence in understanding cross-market pricing patterns, i.e., the parametric evidence presented in Patton (2006) and Reboredo (2012). As noted, our semi-parametric approach has two attractive features: 1/ it is flexible and allows for arbitrary co-dependence; and 2/ it does not require imposing a priori restrictions on the shape of the copula. The results presented in Table C1 is a nice illustration of the usefulness of our approach.

Table C1. Contemporaneous codependence $R(q_f, q_s | ma)$ between pf and ps, selected quantiles

q_s \ q_f	q=0.1	q=0.2	q=0.3	q=0.4	q=0.5	q=0.6	q=0.7	q=0.8	q=0.9
q=0.1	0.067	0.075	0.069	0.060	0.050	0.041	0.031	0.021	0.011
q=0.2	0.072	0.129	0.133	0.118	0.100	0.081	0.062	0.042	0.022
q=0.3	0.064	0.132	0.175	0.171	0.148	0.121	0.093	0.063	0.033
q=0.4	0.054	0.117	0.172	0.203	0.190	0.159	0.123	0.084	0.044
q=0.5	0.044	0.098	0.148	0.192	0.212	0.191	0.151	0.105	0.055
q=0.6	0.035	0.078	0.120	0.159	0.191	0.203	0.175	0.124	0.066
q=0.7	0.024	0.059	0.091	0.121	0.148	0.172	0.179	0.139	0.077
q=0.8	0.014	0.038	0.061	0.082	0.101	0.118	0.136	0.137	0.086
q=0.9	0.004	0.018	0.031	0.042	0.052	0.061	0.071	0.080	0.079

Note: The codependence is evaluated at the sample median for ma. The results are very similar evaluated at the lower and higher quantiles for ma.

Footnotes

[1] Garcia and Leuthold (2004) provides a good review of these related issues in the commodity futures market.

[2] As far as the authors are aware of, the closest-related research to ours include Huang et al., (2020) and Beckmann and Czudaj (2014). The former provides valuable insights on commodity price forecasting relying on quantile autoregression (QAR) in a univariate and linear manner, and the latter sheds light on the nonlinear dynamics in commodity prices using smooth-switching models. Our study extends their analysis in at least three ways as discussed above.

[3] To simplify the notation, we include the futures contract maturity ma among the explanatory variables $\boldsymbol{x}_t$.

[4] Many market participants may specialize in specific activities. For example, producers may specialize in producing for the spot market (with $Qs_{kt} > 0$) while consumers specialize in consuming in the spot market (with $Qs_{kt} < 0$). Agents carrying stocks may specialize in intertemporal arbitrage activities in the spot market, trying to buy when the spot price is low (with $Qs_{kt} < 0$) and sell when the spot price is high (with $Qs_{kt} > 0$). Similarly, some speculators may specialize in intertemporal arbitrage activities in the futures market, trying to buy when the futures price is low (with $Qf_{kmt} < 0$) and sell when the futures price is high (with $Qf_{kmt} > 0$). Other market participants can get involved in both the spot market and the futures market. They include hedgers that take opposite positions in the spot market and the futures markets to reduce their exposure to price risk. And some speculators may speculate on the price difference $(pf - ps)$, trying to modify their market positions to benefit from anticipated changes in $(pf - ps)$.

[5] Throughout the paper, we define the basis as $(pf_{1t} - ps_t)$. The reader should keep this in mind (as the basis is sometimes defined in the literature as $(ps_t - pf_{1t})$, i.e. of opposite sign).

[6] In our empirical analysis, we also considered alternative specifications to (12), including past prices $\boldsymbol{P_{t-1}}$ and their interactions with ma. We found that these additional variables did not have statistically significant effects. The corresponding estimates are available from the authors upon request.

[7] Following conventions (e.g., Beckmann and Czudaj, 2014), we performed robustness checks by trying other rolling method to construct the continuous futures prices (e.g., roll nearby contracts at the end of the month prior to contract expiration). The main findings are found to remain unchanged.

[8] We also explored other interaction terms but found them to be not statistically significant.

[9] We performed sensitivity tests and found our qualitative results to be robust to minor changes in the definitions of the time trends.

[10] Specifically, we consider three ma scenarios: low, medium and high, corresponding to ma being set at the 0.1, 0.5 and 0.9 sample quantiles, respectively. First, we hold ma constant as "low" and use 500 random replications to simulate the forward path of pf and ps for 150 periods. Second, we obtain the simulated distribution of the basis $pf - ps$ from the 500 replications, and evaluate the scenarios corresponding to 0.1, 0.5 and 0.9 quantile of the basis distribution. Third, we allow ma to change to "medium" and "high" and repeat the simulation process. The results are reported in Figure 3.