Price Discovery in Commodity Futures and Cash Markets with Heterogeneous Agents

Sophie van Huellen*

Abstract

Since 2004, commodity futures markets have seen an unprecedented liquidity inflow linked to noise traders that follow global liquidity cycles rather than market fundamentals. This paper develops a price discovery model for commodity futures markets that incorporates noise trader effects by assuming two forms of limits to arbitrage: transaction costs and noise trader risk. It is shown that under these assumptions, commodity prices are driven by both market fundamentals and noise trader positions. Further, noise trader effects spill over to the cash market if limits to arbitrage due to transaction costs are imperfect but are confined to the futures market otherwise. The model is empirically tested using data from six grain and soft commodity markets.

Keywords: commodity futures, index investment, price discovery, speculation
JEL: D84, G13, Q02, Q11

* School of Oriental and African Studies (SOAS), University of London, Thornhaugh Street, Russell Square, London WC1H 0XG. Email: sv8@soas.ac.uk. Tel.: +44 (0) 20 7898 4543

1. Introduction

Since 2004, commodity futures markets have seen an unprecedented liquidity inflow linked to financial investors that seek exposure to commodities for portfolio diversification. The inflow has been spurred by emerging evidence that commodity futures provide a natural hedge against inflation (Gorton and Rouwenhorst 2006) and in anticipation of high returns due to increasing demand from emerging market economies resulting in a commodity 'super price cycle' (Kaplinsky 2006). The inflow has further been facilitated by deregulation of U.S. American commodity futures markets and the introduction of innovative trading instruments, such as commodity indices. These developments had a lasting effect on the volume and nature of trades in these markets[1] and triggered a debate around potential implications for price discovery. Commodity markets have been reported to progressively behave like traditional asset markets (Domanski and Heath 2007) and prices are conjectured to follow global liquidity cycles rather than market fundamental, most prominently in 2008 when commodity markets served as the main transmission channel of the global financial crisis to emerging and developing markets (Nissanke 2012).

With the intention to explore the effects of novel trading instruments such as commodity indices on price discovery mechanisms in commodity markets, I develop a price discovery model with heterogeneous agents. Following the tradition of De Long, et al. (1990) and the more recent literature that emerged from the above debate, e.g. see Gilbert (2008), Nissanke (2012) and Irwin and Sanders (2012), I distinguish between informed traders, arbitrage traders and uninformed noise traders. The latter trader category is associated with index traders as their position taking is largely synchronised with global liquidity cycles but unrelated to market fundamentals (Alam and Gilbert 2017). The model builds on the price discovery model by Garbade and Silber (1983) and extends the model's initial assumption of limits to arbitrage due to transaction costs by the additional assumption of limits to arbitrage due to noise trader risk. I show that under these assumptions, the relationship between cash and futures markets and the price discover in both markets depends on the elasticity of arbitrage and the weight of uninformed noise traders in the market.

Specifically, four different market regimes are identified: (1) effective arbitrage, (2) transaction cost induced limits to arbitrage and no noise trader risk, (3) effective arbitrage and noise trader risk, and (4) limits to arbitrage and noise trader risk. Price levels and changes are driven by both changes in market fundamentals and changes in noise trader positions under the latter two regimes. Noise trader effects spill over to the cash market if arbitrage between cash and futures markets is effective but is confined to the futures market if arbitrage between the two markets involves transaction costs. If transaction costs hinder arbitrage, cointegration between cash and futures prices breaks down and the market basis follows a random walk process. If further, noise trader risk is present, the market basis reflects price impulses unrelated to market fundamentals. The latter also applies to the

---

[1] The liquidity inflow was visibly curbed with regards to over-the-counter transactions after the Dodd-Frank Act of 2010, but the general interest in commodity derivatives did not fade (BIS 2016).

calendar spread between two consecutively traded contracts if the composition of traders in both contracts differs.

The number of studies that theoretically explore the potential effects of heterogeneous traders including noise traders on price discovery mechanisms in commodity futures markets is small yet growing. Among notable contributions are studies by Basak und Pavlova (2016), Brunetti and Reiffen (2014), and Hamilton and Wu (2014, 2015). These authors build on the conjecture that the presence of noise traders, if acting systematic and not at random, dilutes the market's information content, therefore feeding unwarranted speculative bubbles. This paper contributes to the growing literature by development of a model that considers jointly the presence of noise traders and limits to arbitrage due to transaction costs. This contribution is particularly valuable as it formalises the link between index traders and an excessive market basis that was suggested for several US grain markets over recent years (e.g. see Irwin, et al. 2011, Garcia, et al. 2015, van Huellen 2018a) as well as anomalies in commodities' term structures (e.g. see Heidorn, et al. 2015, van Huellen 2018b). [2]

The empirical validity of the model is tested for six grain and soft commodity markets traded at the Chicago Board of Trade (CBT): soft red winter wheat, number two yellow corn, number one yellow soybeans, cocoa, coffee 'c', and number two cotton. Arbitrage effectiveness is found to be low for corn and wheat markets, where non-convergence linked to bottlenecks in the delivery system occurred over recent years (Garcia, et al. 2015). Trader position data is found to have a significant effect on the price discovery process in the short-run and feedback effects to cash markets are strongest for markets for which the futures market is leading the price discovery process and arbitrage is effective on average. The coincidence of noise trader risk and perfectly inelastic arbitrage, corresponding to regime four, is rare. However, evidence for noise trader risk is found in all six markets with varying elasticities to arbitrage resulting in spill-over effects to cash markets.

The paper is structured in five parts. After this short introduction, the paper continues with the derivation of the price discovery model for commodity markets with heterogeneous agents and limits to arbitrage due to transaction costs. The third part presents data and methodology for testing the model. The forth part reports empirical results and the fifth part concludes on the evidence.

2.   A Heterogeneous Trader Model

Eq. (1) summarises the no-arbitrage condition between cash and futures prices, with F_t being the futures price at time t with maturity date T, S_t being the cash price at t, r_t and w_t being continuously compounded risk free interest rate and storage costs over time $\tau = T - t$ and y_t being the convenience yield; a utility based reward that accrues to the holder of inventories.

---

[2] See Irwin (2013), Fattouh, et al. (2013), Cheng and Xiong (2014) and Boyd, et al. (2018) for a comprehensive summary of the empirical literature on the potential effects of index investors on price dynamics in commodity futures markets.

$$F_{t,T} = S_t e^{(r_t + w_t - y_t)\tau} \tag{1}$$

Taking logs of Eq. (1), with $\tau = 1$ and net-carry costs $c_t = r_t + w_t - y_t$ for simplicity, yields Eq. (2). At maturity $\tau \to 0$ so that $f_{t,t} = s_t$ and market basis $b_t \equiv s_t - f_{t,t} = 0$.

$$f_{t,t+1} = s_t + c_t \tag{2}$$

It is useful to distinguish between two arbitrage mechanisms. *Spatial arbitrage* exploits deviations between cash and futures prices beyond the carry relationship in Eq. (2). *Fundamental arbitrage* exploits deviations of price levels from the underlying fundamental value, but not necessarily relative prices. The former relies on arbitrage traders' willingness to trade the derivative and the physical commodity. The latter relies on the assumption that traders in both futures and cash markets make decisions based on the same information set so that a common (fundamental) factor drives both markets.

The spatial no-arbitrage condition is described by Eq. (2), while Eq. (3) must hold under the fundamental no-arbitrage condition with $(1, -C)$ being the cointegrating vector and φ_t the common factor, in our case fundamental information. The common factor implies cointegration of the two price-series so that f_t ($f_t = f_{t,t+1}$ for simplicity hereafter) and s_t are decomposed into a permanent common information component φ_t and a transitory noise component $\tilde{f}_t$ and $\tilde{s}_t$ respectively; see Granger (1986) and Gonzalo and Granger (1995).[3]

$$\begin{bmatrix} f_t \\ s_t \end{bmatrix} = \begin{bmatrix} C \\ 1 \end{bmatrix} \varphi_t + \begin{bmatrix} \tilde{f}_t \\ \tilde{s}_t \end{bmatrix} \tag{3}$$

In the following, I will extend the model by Garbade and Silber (1983) which already incorporates the assumption of limits to spatial arbitrage in the form of transaction costs or market frictions by the additional assumption of limits to fundamental arbitrage. Garbade and Silber (1983) assume two trader types, informed traders that are active in the cash or in the futures market, and arbitrage traders that are active in both futures and cash markets. Eq. (4) summarises the solutions for the market clearing prices in the cash and futures market at time t with $a = \left(\frac{H}{AN_s}\right) / \left[1 + \left(\frac{H}{AN_s}\right) + \left(\frac{H}{AN_f}\right)\right]$ and $b = \left(\frac{H}{AN_f}\right) / \left[1 + \left(\frac{H}{AN_f}\right) + \left(\frac{H}{AN_s}\right)\right]$. N_s is the number of informed traders in the cash market, N_f is the number of informed traders in the futures market, $H > 0$ is elasticity of spatial arbitrage, $A > 0$ is elasticity of demand in both futures and cash markets and r_t^s and r_t^f are the average

---

[3] Note that cointegration only holds if $c_t \sim I(0)$ in Eq. (2). The reasonability of this assumption is discussed in the following section.

reservation prices of cash and futures traders respectively. See Appendix A1 for a detailed derivation of Eq. (4).

$$s_t = (1-a)r_t^s + ar_t^f - ac_t$$
$$f_t = br_t^s + (1-b)r_t^f + bc_t$$
(4)

In addition to the assumption of limits to spatial arbitrage incorporated in the Garbade and Silber (1983) model via finite elasticity of arbitrage H, I make the additional assumption of limits to fundamental arbitrage via noise trader risk. Let three trader types exist in the futures market. Arbitrageurs that exploit any deviations from Eq. (2), informed traders who base their positions on information about market fundamentals Φ, and uninformed traders who follow an impulse Y that is common among all uninformed traders but unrelated to market fundamentals. Only informed traders are present in the cash market, while both informed and uninformed traders are present in the futures market. This composition reflects the arrival of financial liquidity, channelled predominantly via indices, at commodity futures markets. Index positions are largely synchronised over global liquidity cycles but unrelated to market fundamentals (Alam and Gilbert 2017).

With the division between informed and uninformed traders, the reservation prices of the ith and j^{th} trader in the cash and futures market are $r_{i,t}^s = E_{i,t}[s_t|\Phi_i]$ and $r_{j,t}^f = (1-\beta)\{E_{j,t}[f_t|\Phi_j]\} + \beta\{E_{j,t}[f_t|Y_j]\}$, with $Y_j \in Y$ and $\Phi_j \in \Phi$ and with β being the probability that the jth trader is a noise trader. Hence, if $\sum E_{i,t}[s_t|\Phi_i] = \sum E_{j,t}[f_t|\Phi_i] = \varphi_t$ and $\sum E_{j,t}[f_t|Y_i] = \vartheta_t$, the evolution of the average reservation prices can be expressed as in Eq. (5), with the white noise components $w_t \sim I(0)$ being the sum of idiosyncratic errors.[4]

$$r_t^s = \varphi_t + w_t^s$$
$$r_t^f = (1-\beta)\varphi_t + \beta\vartheta_t + w_t^f$$
(5)

By rewriting $\varphi_t = \varphi_{t-1} + \Delta\varphi_t$ and $\vartheta_t = \vartheta_{t-1} + \Delta\vartheta_t$ and following Garbade and Silber (1983) in assuming that each trader expects the last price to be the full information value so that s_{t-1} and f_{t-1} have to be the average reservation prices immediately after the last clearing, it follows that $\varphi_{t-1} = \vartheta_{t-1} = f_{t-1}$ for the futures market and $\varphi_{t-1} = s_{t-1}$ for the cash market. Therefore, Eq. (5) can be transformed into Eq. (6).

$$r_t^s = s_{t-1} + \Delta\varphi_t + w_t^s$$
$$r_t^f = f_{t-1} + (1-\beta)\Delta\varphi_t + \beta\Delta\vartheta_t + w_t^f$$
(6)

---

[4] Note that Eq. (5) can be interpreted as factor decomposition, with common factor φ_t. Under limits to spatial arbitrage $H = 0$, Eq. (5) is equivalent to Eq. (3) with $(1-\beta) = C$ and $s_t = r_t^s, f_t = r_t^f$ if $\vartheta_t \sim I(0)$.

For simplicity, we assume that net-carry costs follow a white noise process with $\bar{c}$ being mean net-carry costs, see Eq. (7).[5] This assumption will be eased in the following section.

$$c_t = \bar{c} + w_t^c \tag{7}$$

Substituting Eq. (6) and Eq. (7) into Eq. (4) yields an error correction decomposition with the long-run dis-equilibrium error of Eq. (2) $e_t = (f_t - s_t - \bar{c})$ and short-run shocks driven by traders' price impulses; $\Delta\varphi_t$ for informed traders and $\Delta\vartheta_t$ for noise traders. For ease of presentation $\pi_s = a\beta$, $\pi_f = (1-b)\beta$, $w_t^{s'} = w_t^s + a(w_t^f - w_t^s - w_t^c)$, and $w_t^{f'} = w_t^f - b(w_t^f - w_t^s - w_t^c)$.

$$\Delta s_t = ae_{t-1} + (1 - \pi_s)\Delta\varphi_t + \pi_s\Delta\vartheta_t + w_t^{s'}$$
$$\Delta f_t = -be_{t-1} + (1 - \pi_f)\Delta\varphi_t + \pi_f\Delta\vartheta_t + w_t^{f'} \tag{8}$$

If noise traders induce systematic price impulses, so that $E[\Delta\vartheta_t] \neq 0$, the effect of these traders on the price discovery process depends on the elasticity of spatial arbitrage H, the weight of the futures market N_f relative to the cash market N_s as well as the weight of noise traders β in the futures market. Note that these impulses by noise traders only enter prices in the short-run but not the long-run.

Following from Eq. (8), Table 1 distinguishes between four different arbitrage regimes: (1) effective spatial arbitrage and no noise trader risk, (2) limits to spatial arbitrage and no noise trader risk, (3) effective spatial arbitrage and noise trader risk and (4) limits to spatial arbitrage and noise trader risk. If noise trader risk is present, as in (3) and (4), price levels and changes are driven by both changes in market fundamentals and changes in speculative demand. Speculative demand spills over to the cash market if spatial arbitrage is effective as in (3) but is restricted to the futures market if spatial arbitrage in limited as in (4). Further, if spatial arbitrage is limited, as in (2) and (4), the price levels are determined by the past settlement price of the respective markets resulting in the market basis being driven by its past values with a unit root process. If noise trader risk is present, as in (4), the market basis is further driven by noise traders' demand. Like the market basis, the calendar spread follows a unit root process under limits to spatial arbitrage with an addition noise component in (4). Several testable implications can hence be inferred from Table 1. If there are no limits to spatial arbitrage, the basis is driven solely by net-carry costs and futures and cash prices are cointegrated so that price dynamics can be expressed in an error correction decomposition. If limits to spatial arbitrage are present, the basis follows a random walk and the error correction coefficient is insignificant due to the break in the cointegrating relationship between cash and futures markets. If, further, noise trader risk is present, changes in the futures price are linked to speculative demand while the cash price remains unaffected by this demand as long as the elasticity of arbitrage is low.

---

[5] See for instance Low, et al. (2002) who impose a similar assumption.

Table 1. Spatial Arbitrage Regimes with and without Noise Trader Risk

	(1) Effective spatial arbitrage, no noise trader risk	(2) Limits to spatial arbitrage, no noise trader risk	(3) Effective spatial arbitrage, noise trader risk	(4) Limits to spatial arbitrage, noise trader risk
H	$H \to \infty$	$H \to 0$	$H \to \infty$	$H \to 0$
β	$\beta = 0, \alpha = 1$	$\beta = 0, \alpha = 1$	$0 < \beta < 1$	$0 < \beta < 1$
a, b	$\lim_{H \to \infty} a = \omega_f,$ $\lim_{H \to \infty} b = \omega_s$	$\lim_{H \to 0} a = 0,$ $\lim_{H \to 0} b = 0$	$\lim_{H \to \infty} a = \omega_f,$ $\lim_{H \to \infty} b = \omega_s$	$\lim_{H \to 0} a = 0,$ $\lim_{H \to 0} b = 0$
π	$\pi_s = 0, \pi_f = 0$	$\pi_s = 0, \pi_f = 0$	$\pi = \pi_s = \pi_f = \omega_f \beta$	$\pi_s = 0, \pi_f = \beta$
s_t	$s_t = [\omega_s s_{t-1} + \omega_f (f_{t-1} - \bar{c})] + \Delta\varphi_t + w_t^{s'}$	$s_t = s_{t-1} + \Delta\varphi_t + w_t^s$	$s_t = [\omega_s s_{t-1} + \omega_f (f_{t-1} - \bar{c})] + (1 - \pi)\Delta\varphi_t + \pi\Delta\vartheta_t + w_t^{s'}$	$s_t = s_{t-1} + \Delta\varphi_t + w_t^s$
f_t	$f_t = [\omega_s (s_{t-1} + \bar{c}) + \omega_f f_{t-1}] + \Delta\varphi_t + w_t^{f'}$	$f_t = f_{t-1} + \Delta\varphi_t + w_t^f$	$f_t = [\omega_s (s_{t-1} + \bar{c}) + \omega_f f_{t-1}] + (1 - \pi)\Delta\varphi_t + \pi\Delta\vartheta_t + w_t^{f'}$	$f_t = f_{t-1} + (1 - \beta)\Delta\varphi_t + \beta\Delta\vartheta_t + w_t^f$
Δs_t	$\Delta s_t = \omega_f e_{t-1} + \Delta\varphi_t + w_t^{s'}$	$\Delta s_t = \Delta\varphi_t + w_t^s$	$\Delta s_t = \omega_f e_{t-1} + (1 - \pi)\Delta\varphi_t + \pi\Delta\vartheta_t + w_t^{s'}$	$\Delta s_t = \Delta\varphi_t + w_t^s$
Δf_t	$\Delta f_t = -\omega_s e_{t-1} + \Delta\varphi_t + w_t^{f'}$	$\Delta f_t = \Delta\varphi_t + w_t^f$	$\Delta f_t = -\omega_s e_{t-1} + (1 - \pi)\Delta\varphi_t + \pi\Delta\vartheta_t + w_t^{f'}$	$\Delta f_t = (1 - \beta)\Delta\varphi_t + \beta\Delta\vartheta_t + w_t^f$
F_t	$F_t = \omega_f F_{t-1} + \omega_s \bar{c} + w_t^F$	$F_t = F_{t-1} + w_t^F$	$F_t = \omega_f [F_{t-1} + (\beta_1 - \beta_2)(\Delta\vartheta_t - \Delta\varphi_t)] + \omega_s \bar{c} + w_t^F$	$F_t = F_{t-1} + (\beta_1 - \beta_2)(\Delta\vartheta_t - \Delta\varphi_t) + w_t^F$
b_t	$b_t = -\bar{c} + w_t^b$	$b_t = b_{t-1} + w_t^b$	$b_t = -\bar{c} + w_t^b$	$b_t = b_{t-1} + \beta(\Delta\varphi_t - \Delta\vartheta_t) + w_t^b$

Notes: Recall, $\pi_s = a\beta$, $\pi_f = (1 - b)\beta$, $w_t^{s'} = w_t^s + a(w_t^f - w_t^s - w_t^c)$, and $w_t^{f'} = w_t^f - b(w_t^f - w_t^s - w_t^c)$. Weights $\omega_f = N_f/(N_s + N_f)$ and $\omega_s = N_s/(N_s + N_f)$ and hence $\omega_s = 1 - \omega_f$. Past error $e_{t-1} \equiv f_{t-1} - s_{t-1} - \bar{c}$. Market basis $b_t \equiv s_t - f_t$. Calendar spread $F_t \equiv f_{t,t+2} - f_{t,t+1}$ so that $\tau = 1$ for simplicity and β_1 and β_2 being the share of noise traders in the deferred and the near to maturity futures contract respectively.

3. Data and Methodology

3.1 Methodology

Before testing the derived hypotheses, several considerations with regards to the specification of the net-carry costs need to be addressed. Previous studies suggest that net-carry costs could be non-stationary due to a root component in the convenience yield or interest rate or both, e.g. Figuerola-Ferretti and Gonzalo (2010) and Brenner and Kroner (1995). Settling the question empirically is difficult as of the three components constituting net-carry costs, convenience yield is latent, while storage cost data is difficult to obtain. An elegant solution was proposed by Figuerola-Ferretti and Gonzalo (2010) who suggest modelling the convenience yield as a linear function of cash and futures prices. The model rests on the assumption of mean reversion of interest rate and storage costs, so that $r_t = \bar{r} + I(0)$, and $w_t = \bar{w} + I(0)$. With the convenience yield being defined as $y_t = \gamma_1 s_t - \gamma_2 f_t$, net-carry costs are specified as $c_t = \overline{rw} - \gamma_1 s_t + \gamma_2 f_t$. However, the mean reversion of interest rate and storage costs might not be supported empirically.

I suggest an alternative approach that involves modelling storage costs and convenience yield as a function of level of inventory. The relationship between inventory and convenience yield is theoretically and empirically confirmed by Pindyck (2001), Bozic and Fortenbery (2011) and Pirrong (2011). Following this approach, $w_t(I_t) = w_0 + \gamma_2' I_t$ and $y_t(I_t) = y_0 - \gamma_2'' I_t$, so that $c_t = (w_0 + y_0) + \gamma_1 r_t + \gamma_2 I_t$, with convenience yield being inversely related to inventory and storage cost being positively related to inventory, which means c_t is increasing with I_t, with $\gamma_2 = \gamma_2' + \gamma_2''$. Both decompositions of net-carry costs can be incorporated into Eq. (8), so that the long-run disequilibrium error e_t is specified as in Eq. (9).

$$e_t = (1 - \gamma_2)\left[f_t - \frac{(1 - \gamma_1)}{(1 - \gamma_2)} s_t - \frac{\overline{rw}}{(1 - \gamma_2)} \right] \tag{9a}$$

$$e_t = f_t - s_t - \gamma_1 r_t - \gamma_2 I_t + (y_0 - w_0) \tag{9b}$$

The combination of Eq. (9) and Eq. (8) can be generalised into Eq. (10), with $\boldsymbol{w_t} \sim IID(\boldsymbol{0}, \Omega)$, $\boldsymbol{\Gamma} = \boldsymbol{\alpha\gamma'}$ and $\boldsymbol{\gamma}$ being the cointegrating vector with $\boldsymbol{\gamma' y_t} \sim I(0)$, $\boldsymbol{y_t}$ being the vector of long-run variables, that is cash and futures prices and net-carry costs and $\boldsymbol{z_t} \sim I(\boldsymbol{0})$ being a set of additional explanatory variables such as changes in fundamental and speculative information.

$$\Delta \boldsymbol{y_t} = \boldsymbol{\Gamma y_{t-1}} + \sum_{j=0}^{r} \boldsymbol{\delta_j z_{t-j}} + \sum_{i=1}^{m} \boldsymbol{\zeta_i \Delta y_{t-i}} + \boldsymbol{w_t} \tag{10}$$

Eq. (10) is evaluated by estimating the cointegrating rank following Johansen (1988, 1991) to gain insight into the elasticity of spatial arbitrage H and by testing for the

significance of information changes in the short-run approximated by trader position indicators.

In a second step, implications for the market basis are tested by use of a regime switching model (Hamilton 2008). Following Table 1, the four regimes can be collapsed into two regimes for the market basis. Under regimes (1–3) the first difference of the basis follows a white noise process, while under regime (4) the basis follows the difference of fundamental and speculative information weighted by the share of uninformed traders in the market as specified in Eq. (11):

$$\Delta b_t = \begin{cases} w_t^{b_1} & (1\text{--}3) \\ \beta\left(\Delta\varphi_t - \Delta\vartheta_t\right) + w_t^{b_2} & (4) \end{cases} \tag{11}$$

The regime-switching regression is specified in Eq. (12), with ζ_t being a random variable that assumes values $\zeta_t = 1$ or $\zeta_t = 2$ to differentiate between regimes (1–3) and regime (4) in Eq. (11). The probabilistic model of what causes the change from $\zeta_t = 1$ to $\zeta_t = 2$ is based on a two-state Markov chain with constant regime-switching probabilities. The set of explanatory variables, $\mathbf{z_t}$ is defined as in Eq. (10). Lagged values of Δb_t are added to control for autocorrelation in the residuals. Coefficients for lagged values are fixed, while the remaining coefficients can vary across regimes.

$$\Delta b_t = \boldsymbol{\beta}'_{\zeta_t} \mathbf{z_t} + w_t^b \tag{12}$$

However, Eq. (11) and the differentiation of the two regimes only holds if net-carry costs are mean reverting so that $c_t = c_{t-1} = \bar{c}$. Given the above considerations summarized in Eq. (9), two alternatives corresponding to regime (1) and (3) are considered as a potentially third regime in Eq. (11):

$$\Delta b_t = \gamma_1 \Delta s_t - \gamma_2 \Delta f_t + w_t^{b_3} \tag{13a}$$

$$\Delta b_t = -\gamma_1 \Delta r_t - \gamma_2 \Delta I_t + w_t^{b'_3} \tag{13b}$$

$\gamma_1 = \gamma_2 = 1$ in Eq. (13a) is implicitly imposed by construction of the market basis. Hence, accounting for the regime as specified in Eq. (13a) is not useful. Instead I incorporate Δf_t into the regime switching regression Eq. (12) to allow for deviations from 1. The two competing strategies for modelling net-carry costs in Eq. (9) and Eq. (13) are used throughout. However, the second strategy is constraint by data availability and results largely confined to the appendix.

3.2 Data

Commodity prices are obtained from Thomson Reuters Datastream. Following Geman and Sarfo (2012), the futures price is constructed as a weighted average of all simultaneously traded contracts. Each contract is weighted by its share in total open interest.

This way τ is held relatively constant over time which avoids cyclical contraction of carry variables due to maturity cycles.

The approximation of informed and uninformed traders' demand is challenging. We follow several empirical studies in using position data provided by the US Commodity Futures Trading Commission (CFTC), weekly Commitments of Traders Supplemental (CIT) Report; e.g. Irwin and Sanders (2010; 2012), Mayer (2012), Silvennoinen and Thorp (2013), and Singleton (2014). The CFTC CIT report differentiates between long and short positions of commercial traders, index or portfolio insurance traders, non-commercial traders and small non-reporting traders. CFTC position data faces several shortcomings discussed elsewhere in the literature; e.g. Irwin and Sanders (2012). It is important to note here that the division of trader types is based on industry affiliation and not trading strategy. The information content of a trade cannot be observed, and we can only assume that commercial traders on average base their positioning on market fundamentals while index or portfolio insurance traders on average base their positioning on non-fundamental information. Against this conjecture, position change indicators are constructed for the commercial and index trader categories by dividing the respective net long positions (long minus short positions) by total open interest. The 3-month LIBOR rate, obtained from Thomson Reuters Datastream, is used as an approximation for interest rate.

The paper focuses on grains and soft commodities which are storable. However, only for soft commodities data on inventories at exchange registered warehouses is available in weekly frequency. As for the price data, inventory data is obtained via Thomson Reuters Datastream. Weekly trader position data is available from the first week of 2006, while weekly inventory data for cocoa, coffee and cotton is available from 2010 week 32, 2011 week 2 and 2008 week 44 respectively. The samples used in this study end in 2016 week 29.

4. Empirical Results

4.1 Model Estimation

Price variables and interest rate are found non-stationary and I(1), while inventory data is found weakly stationary for cocoa and cotton, but not for coffee; see Appendix Table A1. Having established the order of integration, the cointegrating rank is evaluated by the Johansen (1988) trace test. The test is conducted with correspondence to the cointegrating vector γ in Eq. (10). Commodity futures and cash markets are found to be strongly cointegrated except for wheat where cointegration is weaker; see Table 2. Results for the trace test do not change if considering inventory and interest rate in line with Eq. (9b) whereby interest rate is restricted to be exogenous; see Appendix Table A2.

Table 2. Trace Test s_t, f_t

	Wheat (10)	Corn (8)	Soy (7)	Cocoa (4)	Coffee (1)	Cotton (4)
$H_0: r \leq 0$	18.669	27.161	32.385	24.930	32.153	28.733
	[0.081]	[0.004] **	[0.000] **	[0.009] **	[0.000] **	[0.002] **
$H_0: r \leq 1$	7.3573	6.1318	5.8255	6.5158	4.6194	3.7936
	[0.111]	[0.187]	[0.212]	[0.159]	[0.339]	[0.456]

Notes: Lags in parentheses selected by Akaike information criterion. Constant included in the cointegration relationship. P-values in brackets. * indicating 5% and ** indicating 1% significance level. Data from 2006:W01 to 2016:W29 in weekly frequency.

I replicate the derivation of the convenience yield from the cointegrating vector by Figuerola-Ferretti and Gonzalo (2010). In line with observations made by Figuerola-Ferretti and Gonzalo (2010) for metal markets the cointegrating vectors deviate from the standard $(1, -1)$ with $\hat{\gamma}_1 < -1$ for most markets, which indicates that backwardation prevails with a small unit root process in the net-carry variables; see Table 3. The deviation from unity is statistically significant for corn, soybeans, coffee and cotton. Dolatabadi, et al. (2015) present similar findings and account for the remaining unit root process by allowing for long memory in the equilibrium error. Interestingly, the unit root vanishes after addition of the observed carry variables in line with Eq. (9b); see Appendix Table A3. The findings suggest that the long memory component is in interest rate and inventory rather than convenience yield.

Table 3. Cointegrating Vector $[s_t - \gamma_0 - \gamma_1 f_t]$ Eq. (10) with Eq. (9a)

	$\hat{\gamma}_0$	s.e.	$\hat{\gamma}_1$	s.e.	$H_0: \gamma_1 = 1^\dagger$
Wheat (10)	-0.0378	[0.9962]	-0.9768	[0.1553]	0.0079 (0.9292)
Corn (8)	1.6255	[0.2878]	-1.2561	[0.0470]	11.392 (0.0007)**
Soy (7)	0.3795	[0.0920]	-1.0517	[0.0132]	8.2781 (0.0040)**
Cocoa (4)	0.1012	[0.2597]	-1.0280	[0.0331]	0.4812 (0.4879)
Coffee (1)	0.7191	[0.2086]	-1.1283	[0.0415]	6.6131 (0.0101)*
Cotton (4)	0.6410	[0.1047]	-1.1483	[0.0244]	13.579 (0.0002)**

Notes: Lag length in parenthesis determined by AIC (left column). Standard errors in brackets. †test statistic following a Chi-square distribution with one degree of freedom under the null; p-values in parenthesis. * indicating 5% and ** indicating 1% significance level. Data from 2006:W01 to 2016:W29 in weekly frequency.

I further follow Gonzalo and Granger (1995) in identifying the leading market in the price formation process by testing restrictions on $\alpha'_\perp$. Table 4 reports tests on the $\alpha'_\perp$ derived from $\boldsymbol{\alpha}$ in Eq. (10) to identify which market is leading the price formation process. Results suggest that the futures market is leading in most cases. The evidence is less conclusive for grains than for soft commodities and inconclusive for wheat and corn markets, related to the relatively weak cointegration relationships between futures and cash markets. Tests on $\alpha'_\perp$ imply test for weak exogeneity. For Eq. (10) with the long-run specification Eq. (9b) the lead-lag relationship cannot be clearly identified due to the potential endogeneity of the inventory market. Indeed, tests on $\boldsymbol{\alpha}$ suggest that both cash and futures price are weakly exogenous, except for the cocoa market, where futures prices and inventory are found weakly exogenous; see Appendix Table A4.

Table 4. Test for leading market using $\alpha'_\perp$ for (10) with (9a)

	s_t	f_t	$\alpha'_\perp$	$H_0: \alpha'_\perp = (1,0)$ [†]	$H_0: \alpha'_\perp = (0,1)$ [†]
Wheat (10)	-0.0302	0.0047	(0.1344, 0.8656)	0.8014	0.0348
	[0.0203]	[0.0152]		(0.3707)	(0.8520)
Corn (8)	-0.0126	0.0330	(0.7245, 0.2755)	0.0988	0.9159
	[0.0343]	[0.0296]		(0.7533)	(0.3386)
Soy (7)	-0.1863	0.0070	(0.0362, 0.9638)	9.2955	0.0051
	[0.0845]	[0.0880]		(0.0023)**	(0.9431)
Cocoa (4)	-0.0803	0.0599	(0.4274, 0.5726)	2.1833	0.9770
	[0.0440]	[0.0492]		(0.1395)	(0.3230)
Coffee (1)	-0.0801	0.0054	(0.0629, 0.9371)	4.8895	0.0239
	[0.0331]	[0.0319]		(0.0270)*	(0.8772)
Cotton (4)	-0.1262	0.0074	(0.0552, 0.9448)	4.1316	0.0201
	[0.0575]	[0.0483]		(0.0421)*	(0.8874)

Notes: $\alpha'_\perp = (0,1)$ implies the futures market is leading the price discovery process and $\alpha'_\perp = (1,0)$ implies the cash market is leading the price discovery process. $\alpha'_\perp = (\alpha_2/(-\alpha_1 + \alpha_2), -\alpha_1/(-\alpha_1 + \alpha_2))$. Lag length in parenthesis determined by AIC (left column). Standard errors in brackets. [†] Test statistic following a Chi-squared distribution with two degrees of freedom under the null; p-values in parenthesis. * indicating 5% and ** indicating 1% significance level. Data available from 2006: W01 to 2016: W29 in weekly frequency.

Table 5 reports the coefficients on the information shocks. Estimations are based on single equation error correction models. Position data is found significant with regards to price impulses induced by informed commercial hedgers and uninformed impulses by index traders for most markets. This effect is more pronounced in the futures than the cash market as predicted under imperfect spatial arbitrage. For all but the wheat, corn and coffee market, position changes are significantly related to changes in the cash as well as futures prices. For soybeans and cotton, for which the futures market was found to be leading the price formation process, the effect of trader position changes on the cash market price are strongest. Results in Table 5 are based on the full sample averages and ignore the possibility of shifts between different regimes as suggested in Table 1. In the following section, regime shifts are incorporated in form of a Markov regime switching model.

Table 5a. Trader Position Indicators Eq. (10) with Eq. (9a)

	Cash market s_t		Futures market f_t	
	$\Delta\varphi_t$	$\Delta\vartheta_t$	$\Delta\varphi_t$	$\Delta\vartheta_t$
Wheat (4,4)	0.0068	-0.0668	-0.0250	0.0436
	[0.0392]	[0.0657]	[0.0285]	[0.0477]
Corn (7,7)	0.0078	0.0075	-0.0323**	-0.0118
	[0.0177]	[0.0300]	[0.0158]	[0.0315]
Soy (4,5)	0.0514**	0.0041	-0.1006***	-0.0527
	[0.0254]	[0.0398]	[0.0249]	[0.0328]
Cocoa (4,4)	-0.0298*	0.0268	0.0135	-0.0212
	[0.0161]	[0.0438]	[0.0162]	[0.0480]
Coffee (5,5)	-0.0150	-0.0161	-0.0050	0.0053
	[0.0156]	[0.0284]	[0.0163]	[0.0276]
Cotton (1,1)	0.0185*	0.0422	-0.0238**	-0.0577*
	[0.0106]	[0.0342]	[0.0102]	[0.0305]

Table 5b. Trader Position Indicators Eq. (10) with Eq. (9b)

	Cash market s_t		Futures market f_t	
	$\Delta\varphi_t$	$\Delta\vartheta_t$	$\Delta\varphi_t$	$\Delta\vartheta_t$
Cocoa (2,2)	-0.0530**	0.1432	0.0660**	-0.2716***
	[0.0243]	[0.0799]	[0.0302]	[0.0759]
Coffee (2,3)	0.0130	0.0131	-0.0246	0.0019
	[0.0161]	[0.0283]	[0.0172]	[0.0267]
Cotton (0,1)	0.0303**	-0.0183	-0.0381**	-0.0448
	[0.0143]	[0.0507]	[0.0159]	[0.0421]

Notes: Lag length in parenthesis chosen by general to specific modelling. Heteroscedasticity robust standard errors in brackets. * indicating 1%, ** indicating 5% and *** indicating 1% significance level.

4.2 Test for Regime Shift

Regression results for the regime switching model as specified in Eq. (12) are summarized in Table 6. For most commodities only two statistically distinct regimes can be identified instead of three. Hence, Eq. (12) incorporating Eq. (13a) is estimated with two regimes instead of three.

Table 6. Markov Switching Regime Eq. (12)

	Wheat (4)	Corn (2)	Soy (4)	Cocoa (6)	Coffee (7)	Cotton (1)
			Regime 1			
Δf_t	0.125***	0.116***	-0.016	-0.065***	-0.032	0.163***
	[0.0294]	[0.0137]	[0.0156]	[0.0208]	[0.0249]	[0.0173]
$\Delta\varphi_t$	0.062*	0.021	-0.016	0.005	-0.0002	0.010
	[0.0362]	[0.0154]	[0.0112]	[0.0128]	[0.0176]	[0.0107]
$\Delta\vartheta_t$	0.076	-0.008	-0.001	0.038	0.015	0.026
	[0.0528]	[0.0299]	[0.0156]	[0.0352]	[0.0285]	[0.0225]
			Regime 2			
Δf_t	-0.285**	-0.899***	-0.795***	-1.057***	-0.430***	-0.617***
	[0.1298]	[0.0682]	[0.0380]	[0.0352]	[0.0690]	[0.0593]
$\Delta\varphi_t$	-0.770**	-2.426***	-0.266***	-0.095***	-0.305***	-0.109
	[0.3101]	[0.1772]	[0.0679]	[0.0186]	[0.0737]	[0.0736]
$\Delta\vartheta_t$	-1.079***	-1.800***	-0.313***	-0.149***	-0.401***	-0.122*
	[0.2185]	[0.1162]	[0.0802]	[0.0381]	[0.0872]	[0.0690]
			Transition Probabilities			
P11-C	3.903***	-4.446***	-3.150***	-2.295***	-2.553***	-3.115***
	[0.3678]	[0.4568]	[0.3402]	[0.2603]	[0.3125]	[0.4460]
P21-C	1.314*	0.714	-0.845	1-371***	1.075**	-0.312
	[0.7396]	[0.5530]	[0.5645]	[0.2682]	[0.5160]	[0.6132]
i=1 j=1	0.9802	0.9884	0.9589	0.9085	0.9278	0.9575
i=1 j=2	0.0198	0.0116	0.0411	0.0915	0.0722	0.0425
i=2 j=1	0.7882	0.3287	0.6996	0.2025	0.2545	0.5773
i=2 j=2	0.2118	0.6713	0.3004	0.7975	0.7455	0.4227
Duration 1	50.571	82.250	24.335	10.925	13.850	23.536
Duration 2	1.2688	3.0421	1.4294	4.9387	3.9291	1.7323

Notes: Lag length in parenthesis decided by AIC. c is a constant; * for 10 %, ** for 5%, and *** for 1% significance level. Standard errors in brackets; constant transition probabilities: P(i, k) = P(ζ(t) = k | ζ(t-1) = i); Duration is the constant expected duration for regime 1 and 2.

I find that Regime 2, for which the coefficients for trader positions are significant, is of much shorter duration, spanning one to five weeks only, than the prevailing Regimes 1.

Further, Regime 2 is only significant for wheat, cocoa and coffee. Comparing results in Table 6 with regimes identified in Table 1, I find that regimes (4) and (2) in their strong form of zero elasticity to arbitrage do not exist as evident from the significant coefficient for Δf_t in both regimes identified by the regime switching model. Regime 1 corresponds to a mix of regimes (1) and (2) in Table 1 and is associated with contango. Regime 2 corresponds to a mix of regimes (3) and (4) in Table 1 and is associated with backwardation. Further, coefficients for trader positions are predominantly negative indicating that these are associated with the market moving further into backwardation. These findings support the hypothesis that net-long positions contribute to backwardation through a positive price effect on the futures contract; e.g. see Irwin and Sanders (2012). Our findings are broadly confirmed if using the alternative specification of net-carry costs following Eq. (13b), but coefficients are less robust to lag length selection; see Appendix Table A5.

5.   Conclusion

The paper derives a model of price discovery in commodity markets that accounts for limits to arbitrage in two forms: limits to spatial arbitrage and limits to fundamental arbitrage. Limits to spatial arbitrage arise over transaction costs involved in the process of trading in commodity futures or cash markets. If transaction costs are significant, the elasticity of arbitrage is limited. Limits to fundamental arbitrage arise if systematic noise traders are present in the market which induce price deviations from market fundamentals in the short-run.

It is shown that under these assumptions the relationship between cash and futures markets depends on the elasticity of arbitrage and the weight of uninformed traders in the market. Four different market regimes are identified: (1) effective arbitrage and no noise trader risk, (2) limits to arbitrage and no noise trader risk, (3) effective arbitrage and noise trader risk and (4) limits to arbitrage due to both transaction cost and noise trader risk. It is shown that, firstly, price levels and changes are driven by both changes in market fundamentals and changes in speculative demand under the latter two regimes. Speculative demand spills over to the cash market if spatial arbitrage is effective but is restricted to the futures market if spatial arbitrage is limited. Secondly, if spatial arbitrage is limited the cointegrating relationship between cash and futures prices breaks and the market basis follows a random walk process. If further, noise trader risk is present, the market basis is also driven by speculative demand.

Empirical results largely confirm the model's predictions. Findings suggest that regime (4) with zero elasticity to arbitrage and noise trader risk is rare and short lived. However, a variation of regime (3) with imperfect elasticity to arbitrage and some noise trader risk prevails in most markets, which means that noise trader risk is more prevalent in the futures market than in the cash market.

The model and empirical evidence presented in this paper imply that the price discover process of commodity futures markets is affected by the presence of noise traders. The strength of the effect in both futures and cash markets depends on the relative market weight of noise traders in the futures markets and the elasticity of arbitrage. If the elasticity

of arbitrage between cash and futures markets is high, price impulses by noise traders are likely to spill over to the cash market. If the elasticity of arbitrage is low, noise trader risk is revealed in an excessive market basis.

6.   Appendix

6.1   Garbade and Silber (1983) Synopsis

Let $O_{i,t}$ and $O_{j,t}$ be the positions of the ith and j[th] trader in the cash and futures market respectively, $r_{i,t}$ the ith trader's reservation price in the cash market and $r_{j,t}$ the jth trader's reservation price in the futures market. H is elasticity of spatial arbitrage and A is elasticity of demand with $A > 0, H > 0$.

Eq. (A1a-c) summarise the demand schedule for the ith trader participating in the cash market (1a), for the jth trader participating in the futures market (1b) and for the arbitrage traders active in both the cash and futures market (1c). Note that $(f_t - s_t - c_t)$ is the disequilibrium error from the arbitrage condition Eq. (2).

$$O_{i,t} - A(s_t - r_{i,t}) \tag{A1a}$$

$$O_{j,t} - A(f_t - r_{j,t}) \tag{A1b}$$

$$\pm H(f_t - s_t - c_t) \tag{A1c}$$

The market clearing condition in the cash market is summarised in Eq. (A2) with $i = 1, \dots, N_s$ and N_s number of traders in the cash market.

$$\sum_{i=1}^{N_s} O_{i,t} = \sum_{i=1}^{N_s} \{O_{i,t} - A(s_t - r_{i,t})\} + H(f_t - s_t - c_t) \tag{A2}$$

In correspondence to Eq. (A2), market clearing condition in the futures markets are summarised in Eq. (A3) with $j = 1, \dots, N_f$ and N_f number of traders in the futures market.

$$\sum_{j=1}^{N_f} O_{j,t} = \sum_{j=1}^{N_f} \{O_{j,t} - A(f_t - r_{j,t})\} - H(f_t - s_t - c_t) \tag{A3}$$

Eq. (A2) and Eq. (A3) can be solved for f_t and s_t with the reservation prices being expressed as the average reservation price of N_s and N_f traders in the respective markets so that $r_t^s = N_s^{-1} \sum_{i=1}^{N_s} r_{i,t}$ and $r_t^f = N_f^{-1} \sum_{j=1}^{N_f} r_{j,t}$.

$$s_t = \frac{\left[1 + \left(\frac{H}{AN_f}\right)\right] r_t^s + \left(\frac{H}{AN_s}\right)(r_t^f - c_t)}{1 + \left(\frac{H}{AN_f}\right) + \left(\frac{H}{AN_s}\right)} \tag{A4}$$

$$f_t = \frac{\left[1 + \left(\frac{H}{AN_s}\right)\right] r_t^f + \left(\frac{H}{AN_f}\right)(r_t^s + c_t)}{1 + \left(\frac{H}{AN_f}\right) + \left(\frac{H}{AN_s}\right)}$$

Eq. (A4) can be simplified into Eq. (A5), corresponding the Eq. (4). According to Eq. (A5) commodity futures and cash price are a function of the average reservation prices, net carry costs and the parameters a and b defined as: $a = \left(\frac{H}{AN_s}\right) / \left[1 + \left(\frac{H}{AN_s}\right) + \left(\frac{H}{AN_f}\right)\right]$ and $b = \left(\frac{H}{AN_f}\right) / \left[1 + \left(\frac{H}{AN_s}\right) + \left(\frac{H}{AN_f}\right)\right]$.

$$\begin{aligned} s_t &= (1-a)r_t^s + ar_t^f - ac_t \\ f_t &= br_t^s + (1-b)r_t^f + bc_t \end{aligned} \tag{A5}$$

Several implications under different degrees of spatial arbitrage can be derived from Eq. (A5) as summarised in Table 1. Under limits to spatial arbitrage, the cash and futures price is equal to the respective average reservation price and the market basis is the difference of the two reservation prices. Under effective spatial arbitrage, cash and futures prices are the weighted averages of the reservation price of each market plus the net-carry costs and the market basis is equal to the net-carry costs in line with Eq. (2).

Spatial Arbitrage Regimes

	(1) Limits to spatial arbitrage	(2) Effective spatial arbitrage
H	$H \to 0$	$H \to \infty$
a, b	$\lim_{H \to 0} a = 0, \lim_{H \to 0} b = 0$	$\lim_{H \to \infty} a = \omega_f, \lim_{H \to \infty} b = \omega_s$
s_t	$s_t = r_t^s$	$s_t = \omega_s r_t^s + \omega_f(r_t^f - c_t)$
f_t	$f_t = r_t^f$	$f_t = \omega_f r_t^f + \omega_s(r_t^s + c_t)$
b_t	$b_t = r_t^s - r_t^f$	$b_t = -c_t$

Notes: Market basis $b_t \equiv s_t - f_t$. Weights $\omega_f = N_f/(N_s + N_f)$ and $\omega_s = N_s/(N_s + N_f)$ and hence $\omega_s = 1 - \omega_f$.

6.2   Additional Tables

<div align="center">Table A1. Augmented Dickey Fuller Test to Establish Order of Integration</div>

		x_t		Δx_t	
		c	c & t	c	c & t
Futures f_t	Wheat	-1.931 (0)	-1.776 (0)	-29.28** (0)	-29.30** (0)
	Corn	-1.527 (5)	-1.975 (5)	-12.87** (4)	-12.87** (4)
	Soy	-1.699 (2)	-2.490 (2)	-19.23** (1)	-19.22** (1)
	Cocoa	-2.039 (3)	-2.827 (3)	-19.36** (2)	-19.37** (2)
	Coffee	-1.267 (4)	-2.173 (4)	-16.43** (3)	-16.42** (3)
	Cotton	-2.033 (2)	-2.347 (2)	-19.25** (1)	-19.24** (1)
Cash s_t	Wheat	-2.263 (0)	-2.659 (0)	-30.81** (0)	-30.81** (0)
	Corn	-1.561 (5)	-1.999 (5)	-13.02** (4)	-13.02** (4)
	Soy	-1.582 (5)	-2.237 (5)	-13.99** (4)	-13.99** (4)
	Cocoa	-2.116 (3)	-2.670 (3)	-19.13** (2)	-19.15** (2)
	Coffee	-1.402 (0)	-2.791 (0)	-30.11** (0)	-30.10** (0)
	Cotton	-2.005 (0)	-2.331 (0)	-28.89** (0)	-28.87** (0)
Inventory I_t	Cocoa[*]	-4.634** (8)	-4.585** (8)	-6.168** (2)	-4.471** (5)
	Coffee[†]	-0.895 (4)	-0.329 (4)	-4.028** (3)	-4.589** (3)
	Cotton[‡]	-4.090** (2)	-4.291** (2)	-9.438** (1)	-9.438** (1)
Interest r_t	-	-1.431 (2)	-1.071 (2)	-8.104** (5)	-8.147** (5)

Notes: Augmented Dickey Fuller Test with critical values 5% = -2.87, 1% = -3.44 with constant and no trend included (c) and 5% = -3.42, 1% = -3.97 with constant and trend included (c & t). H_0: Time series has a unit root. * indicating 5% and ** indicating 1% significance level. Akaike Information Criteria (AIC) for choice of lag length (max 10 lags); lag length in parenthesis after the test statistic. Data available from [*]2010:W32, [†]2011:W02; [‡]2008:W44. Remaining data from 2006:W01 to 2016:W29 in weekly frequency.

Table A2. Trace Test s_t, f_t, I_t with r_t treated as exogenous

	Cocoa (2) [*]	Coffee (3) [†]	Cotton (2) [‡]
$H_0: r \leq 0$	38.264	36.454	44.859
	[0.021]*	[0.035]*	[0.003]**
$H_0: r \leq 1$	14.417	9.7462	18.511
	[0.268]	[0.669]	[0.085]
$H_0: r \leq 2$	5.4704	1.2308	4.8546
	[0.244]	[0.905]	[0.310]

Notes: Lags in parentheses selected by Akaike information criterion. Constant included in the cointegration relationship. P-values in brackets. * indicating 5% and ** indicating 1% significance level. Data available from [*]2010:W32, [†]2011:W02; [‡]2008:W44. Remaining data from 2006:W01 to 2016:W29 in weekly frequency.

Table A3. Cointegrating Vector $[s_t - \gamma_0 - \gamma_1 f_t - \gamma_2 I_t - \gamma_3 r_t]$ Eq. (10) with Eq. (9b)

	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$H_0: \gamma_1 = 1$ [†]
Cocoa (2) [*]	0.6385	-1.1013	0.0148	0.0073	1.8827 (0.1700)
	[0.4859]	[0.0597]	[0.0106]	[0.0216]	
Coffee (3) [†]	-0.0317	-1.0031	0.0084	-0.0576	0.0213 (0.8839)
	[0.1006]	[0.0178]	[0.0149]	[0.0194]	
Cotton (2) [‡]	0.5352	-1.1139	0.2150	0.0539	2.5227 (0.1122)
	[0.1985]	[0.0480]	[0.0496]	[0.0280]	

Notes: Lag length in parenthesis determined by AIC (left column). Standard errors in brackets. [†]test statistic following a Chi-square distribution with one degree of freedom under the null; p-values in parenthesis. * indicating 5% and ** indicating 1% significance level. Interest rate r_t restricted to be exogenous in 5b. Data available from [*]2010:W32, [†]2011:W02; [‡]2008:W44 (left column).

Table A4. Test for weak exogeneity using α for (10) with (9b)

	s_t	$H_0: \alpha_1 = 0$	f_t	$H_0: \alpha_2 = 0$	I_t	$H_0: \alpha_3 = 0$
Cocoa (2) [*]	-0.1268	7.4133**	0.0772	1.8565	-0.2139	0.8685
	[0.0410]	(0.0065)	[0.0492]	(0.1730)	[0.1853]	(0.3514)
Coffee (3) [†]	-0.3974	3.3839	-0.1692	0.6783	-0.1778	4.3135*
	[0.1854]	(0.0658)	[0.1809]	(0.4102)	[0.0775]	(0.0378)
Cotton (2) [‡]	-0.0267	0.3696	-0.0222	0.4016	-0.1704	12.134**
	[0.0367]	(0.5432)	[0.0315]	(0.5262)	[0.0341]	(0.0005)

Notes: Lag length in parenthesis determined by AIC (left column). Standard errors in brackets. Tests statistic on restrictions following Chi-squared distribution under the null hypothesis with one degree of freedom; p-values in parenthesis. * indicating 5% and ** indicating 1% significance level. Data available from [*]2010:W32, [†]2011:W02; [‡]2008:W44 (left column) to 2016:W29 in weekly frequency.

Table A5. Markov Switching Regime Eq. (12)

	Cocoa (5)		Coffee (5)		Cotton (1)	
	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
			Regime 1			
Δr_t	-0.120*	0.0631	-0.022	0.0326	-0.003	0.0187
ΔI_t	0.038	0.0513	-0.043	0.0566	-0.005	0.0033
$\Delta \varphi_t$	-0.022	0.0265	0.021	0.0149	0.0004	0.0137
$\Delta \vartheta_t$	0.171**	0.0724	-0.006	0.0258	0.012	0.0313
			Regime 2			
Δr_t	0.312**	0.1498	0.754***	0.1559	-0.713*	0.3948
ΔI_t	0.040	0.1192	0.688***	0.2666	0.051***	0.0138
$\Delta \varphi_t$	-0.178***	0.0446	-0.462***	0.0683	0.486***	0.1342
$\Delta \vartheta_t$	-0.586***	0.1149	-0.606***	0.0857	0.961***	0.1793
			Transition Probabilities			
P11-C	1.212***	0.3656	2.955***	0.4569	3.717***	0.5054
P21-C	1.511	1.4625	0.080	0.7778	0.826	1.2941
i=1 j=1; i=2 j=2	0.7706	0.1807	0.9505	0.4799	0.9763	0.3046
i=1 j=2; i=2 j=1	0.2294	0.8193	0.0495	0.5201	0.0237	0.6954
Duration 1; 2	4.3597	1.2206	20.206	1.9228	42.136	1.4380

Notes: Lag length in parenthesis decided by AIC. c is a constant; * for 10 %, ** for 5%, and *** for 1% significance level. Constant transition probabilities: P(i, k) = P(ζ(t) = k | ζ(t-1) = i); Duration is the constant expected duration for regime 1 and 2.

References

Alam, Md Rafayet, and Scott Gilbert. 2017. "Monetary policy shocks and the dynamics of agricultural commodity prices:evidence from structural and factor-augmented VAR analyses." *Agricultural Economics, 48 (1)* 15-27.

Basak, Suleyman, and Anna Pavlova. 2016. "A Model of Financialization of Commodities." *The Journal of Finance, 71(4)* 1511-1556.

BIS. 2016. *OTC Derivatives Statistics At End-June 2016.* Statistical Release, Basel, Switzerland: Bank for International Settlements, Monetary and Economic Department.

Boyd, Naomi E., Jeffrey H. Harris, and Bingxin Li. 2018. "An update on speculation and financialization in commodity markets." *Journal of Commodity Markets, 10* 91-104.

Bozic, Marin, and T. Randall Fortenbery. 2011. "Pricing Options on Commodity Futures: The Role of Weather and Storage." *Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2011 AAEA & NAREA Joint Annual Meeting, Pittsburgh, Pennsylvania, July 24-26, 2011* .

Brenner, Robin J., and Kenneth F. Kroner. 1995. "Arbitrage, Cointegration, and Testing the Unbiasedness Hypothesis in Financial Markets." *The Journal of Financial and Quantitative Analysis, 30 (1)* 23-42.

Brunetti, Celso, and David Reiffen. 2014. "Commodity Index Trading and Hedging Costs." *Journal of Financial Markets, 21* (Federal Reserve Board) 153-180.

Cheng, Ing-Haw, and Wei Xiong. 2014. "The Financialization of Commodity Markets." *Annual Review of Financial Economics, 6* 419-41.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. 1990. "Noise Trader Risk in Financial Markets." *The Journal of Political Economy, 98 (4)* 703-738.

Dolatabadi, Sepideh, Morten Ørregaard Nielsen, and Ke Xu. 2015. "A Fractionally Cointegrated VAR Analysis of Price Discovery in Commodity Futures Markets." *The Journal of Futures Markets, 35 (4)* 339–356.

Domanski, Dietrich, and Alexandra Heath. 2007. "Financial investors and commodity markets." *BIS Quarterly Review.*

Fattouh, Bassam, Lutz Kilian, and Lavan Mahadeva. 2013. "The Role of Speculation in Oil Markets: What Have We Learned So Far? ." *Energy Journal, 34* 7-33.

Figuerola-Ferretti, Isabel, and Jesus Gonzalo. 2010. "Modelling and measuring price discovery in commodity markets." *Journal of Econometrics, 158* 95–107.

Garbade, Kenneth D., and William 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., S. H. Irwin, and A. Smith. 2015. "Futures Market Failure?" *American Journal of Agricultural Economics, 97 (1)* 1-25.

Geman, Hélyette, and Seth Sarfo. 2012. "Seasonality in Cocoa Spot and Forward Markets: Empirical Evidence." *Journal of Agricultural Extension and Rural Development, Vol.4(8)* 164-180.

Gilbert, Christopher L. 2008. "Commodity Speculation and Commodity Investment." *Università degli Studi di Trento, Discussion Paper, No.20.*

Gonzalo, Jesus, and Clive Granger. 1995. "Estimation of Common Long-Memory Components in Cointegrated Systems." *Journal of Business & Economic Statistics, 13 (1)* 27-35.

Gorton, Gary, and K. Geert Rouwenhorst. 2006. "Facts and Fantasies about Commodity Futures." *Financial Analysts Journal, 62 (2)* 47-68.

Granger, Clive. 1986. "Developments in the Study of Cointegrated Economic Variables." *Oxford Bulletin of Economics and Statistics, 48 (3)* 213-228.

Hamilton, James D. 2008. "Regime Switching Models." In *The New Palgrave Dictionary of Economics, 2nd edition*, by N. Durlauf and L.E. Blume. Palgrave Macmillan.

Hamilton, James D., and Jing Cynthia Wu. 2015. "Effects Of Index-Fund Investing On Commodity Futures Prices." *International Economic Review, 56(1)* 187–205.

Hamilton, James D., and Jing Cynthia Wu. 2014. "Risk Premia In Crude Oil Futures Prices ." *Journal of International Money and Finance, 42* 9–37.

Heidorn, Thomas, Frieder Mokinski, Christoph Rühl, and Christian Schmaltz. 2015. "The impact of fundamental and financial traders on the term structure of oil." *Energy Economics, 48* 276-287.

Irwin, S. H., P. Garcia, D. L. Good, and E. L. Kunda. 2011. "Spread and Non-Convergence in Chicago Board of Trade Corn, Soybeans, and Wheat Futures: Are Index Funds to Blame." *Applied Economics Perspective and Policy, 33 (1)* 116-142.

Irwin, Scott H. 2013. "Commodity Index Investment and Food Prices: Does the "Masters Hypothesis" Explain Recent Price Spikes?" *Agricultural Economics, 44 (supplement)* 29-41.

Irwin, Scott H., and Dwight R. Sanders. 2012. "Testing the Masters Hypothesis in commodity futures markets." *Energy Economics, 34* 256–269.

Johansen, Soren. 1991. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Autoregressive Models." *Econometrica, 59 (6)* 1551-1580.

Johansen, Soren. 1988. "Statistical Analysis of Cointegration Vectors." *Journal of Economic Dynamics and Control, 12 (2-3)* 231-254.

Kaplinsky, Raphael. 2006. "Revisiting the Revisited Terms of Trade: Will China Make a Difference? ." *World Development, 34 (6)* 981-995.

Low, Aaron, Jayaram Muthuswamy, Sudipto Sakar, and Eric Terry. 2002. "Multiperiod hedging with futures contracts." *The Journal of Futures Markets, 22 (12)* 1179-1203.

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

Nissanke, Machiko. 2012. "Commodity Market Linkages in the Global Financial Crisis: Excess Volatility and Development Impacts." *The Journal of Development Studies, 48 (6)* 732-750.

Pindyck, Robert S. 2001. "The Dynamics of Commodity Spot and Futures Markets: A Primer." *The Energy Journal, 22 (3)* 1-29.

Pirrong, Craig. 2011. "Stochastic Fundamental Volatility, Speculation, and Commodity Storage." In *Commodity Price Dynamics: A Structural Approach*, by Craig Pirrong, 109-130. Cambridge, UK: Cambridge University Press.

Silvennoinen, Annastiina , and Susan Thorp. 2013. "Financialization, Crisis and Commodity Correlation Dynamics." *Journal of International Financial Markets, Institutions and Money, 24* 42-65.

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

van Huellen, Sophie. 2018a. "How financial investment distorts food prices: Evidence from US grain markets." *Agricultural Economics, (49) 2* 171-181.

van Huellen, Sophie. 2018b. "Too much of a good thing? Speculative effects on commodity futures curves." *Journal of Financial Markets (forthcoming)*.