University of Warsaw | Faculty of Economic Sciences

Working Papers
No. 19/2026 (513)

PRICE DISCOVERY IN SEGMENTED MARKETS: EVIDENCE FROM THE NAIROBI COFFEE EXCHANGE

Gilbert Mbara

Warsaw 2026

ISSN 2957-0506

UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES
WORKING PAPERS

Price Discovery in Segmented Markets: Evidence from the Nairobi Coffee Exchange

Gilbert Mbara[1*]

[1] *University of Warsaw, Faculty of Economic Sciences*
[*] *Corresponding author: g.mbara@uw.edu.pl*

Abstract: We investigate the dynamics of price discovery between physical spot and financial futures markets for green coffee, a commodity uniquely characterized by the geographical and informational separation of its producers from global financial centers. Using dis-aggregated weekly prices from the Nairobi Coffee Exchange (NCE) auction and corresponding International Coffee Exchange (ICE) futures, we estimate a vector error correction model to quantify the transmission of information. Our findings reveal that price discovery predominantly occurs in the physical spot market, which accounts for approximately 83% and 88% of the information and component shares, respectively. These results challenge the conventional view that futures markets are the primary locus of price formation for homogeneous commodities and highlight the critical role of local supply-demand fundamentals in a major producing region. The analysis further uncovers heterogeneity in price discovery contributions across different coffee grades, providing novel insights into the micro-structure of agricultural commodity markets.

Keywords: price discovery, information shares, cointegration, coffee futures, Nairobi Coffee Exchange, commodity markets

JEL codes: C5, C32, C58, C32, E32, Q11

1 Introduction

The determination of primary commodity prices in globally integrated yet geographically fragmented markets presents a complex puzzle for price leadership. For many agricultural commodities, prices are discovered through the interaction of two distinct, yet linked, market structures: the physical spot market, where the actual commodity changes hands, and the forward or futures market, where contracts for future delivery are traded. While futures markets offer near-continuous, transparent price formation, spot markets often operate in discrete, periodic auctions that reveal localized supply and demand conditions. This institutional separation raises a fundamental question: where does price discovery truly occur, and how does information flow between the physical and financial layers of the market?

The global coffee market provides a fertile ground for examining these dynamics. Unlike industrial commodities such as oil or corn, which are largely produced and consumed within regions with well developed financial systems, the production of high-quality "green coffee" is concentrated in developing nations, geographically and institutionally distant from the major financial centers where coffee futures are traded. This separation implies that producers active in the physical spot auctions may have limited direct engagement with futures markets, while speculators and consumers in New York and London trade futures contracts based on global supply-demand expectations. This structural disconnect creates a natural laboratory for testing theories of price leadership and information transmission across physically and financially distinct trading venues.

This study performs an empirical analysis of the price discovery process between the Nairobi Coffee Exchange (NCE) – a physical auction market serving East African producers – and the International Coffee Exchange (ICE) futures contracts for both Arabica and Robusta coffees. We employ a vector error correction model (VECM), building on the theoretical insights of Figuerola-Ferretti and Gonzalo (2010), who demonstrate that commodity spot and futures prices admit an economically meaningful cointegrating relationship driven by the convenience yield. Within this framework, we compute two complementary measures of price discovery: the information share of Hasbrouck (1995), which attributes the variance of innovations to a common efficient price across markets, and the component share of Gonzalo and Granger (1995), which decomposes the permanent component itself into a linear combination of market prices (Baillie, Booth, Tse and Zabotina, 2002).

Our contribution to the literature is threefold. First, we extend the limited empirical work on coffee market price discovery. While Holmes and Otero (2020) assess the informational efficiency of coffee futures as predictors of spot prices, and Bohl, Gross and Souza (2019) document bidirectional information transmission between Brazilian and U.S. futures markets, we provide the first direct quantification of price discovery shares between a major African physical auction and global futures contracts. Second, we exploit granular data on individual coffee grades (`AA`, `C`, `T`) from the NCE auction, allowing us to move beyond aggregate spot indices and examine heterogeneity in price discovery across quality tiers. This disaggregation is crucial, as different grades may embody distinct information about supply shocks, quality differentials, and local demand conditions. Third, we develop a theoretical VECM extension that accommodates multiple spot grades with potentially heterogeneous adjustment speeds to the futures-spot equilibrium, providing a flexible framework for testing hypotheses about the structure of price formation.

Our main findings are as follows. The spot market at the Nairobi Coffee Exchange dominates the price discovery process, with an aggregate information share of approximately 83%. This result holds robustly across alternative measures, with the Gonzalo-Granger component share indicating a 88% contribution from the spot market. Among individual grades, we find that the *lowest-quality* grade (`T`) contributes most to the Hasbrouck share (39%), consistent with its role in conveying long-run supply-demand fundamentals. The component share further reveals that grade `T` also contributes the largest weight in the permanent component (42%), suggesting that it adjusts *relatively* most slowly to disequilibrium and thus embodies a greater share of permanent information. Hypothesis tests confirm that both futures and spot prices adjust significantly to deviations from the long-run equilibrium, indicating bidirectional information flow, but the magnitude and variance of spot innovations ultimately drive the common trend.

The remainder of the paper proceeds as follows. Section 2 provides institutional background on the determination of international coffee prices and the operation of the Nairobi Coffee Exchange. Section 3 develops the theoretical framework, beginning with a discrete version of the Gibson and Schwartz (1990) model and deriving testable VECM representations with associated price discovery measures. Section 4 presents the empirical analysis, including unit root tests, cointegration results, computation of information and component shares, hypothesis tests on adjustment speeds, and forecast error variance decompositions and impulse responses, and a discussion of the results. Section 5 concludes.

2 International Coffee Prices and the NCE

In commodity markets, price discovery often involves a complex interaction between a benchmark futures contract and multiple spot grades. The coffee market trades a liquid futures contract (e.g., ICE Coffee C) alongside numerous physical grades (e.g., Colombian Mild, Brazilian Natural, Robusta), each with its own unique supply and demand dynamics. However, day-to-day physical coffee prices are determined by supply and demand. Price setting criteria are mostly quality (origin), and availability (supply). While each parcel of coffee is unique with regard to its characteristics, flavor and quality; by grouping more or less comparable types of coffee together, average prices can be calculated and traded. Indicator prices, published daily by the International Coffee Organization (ICO) in London, represent and track the four main types of coffee available in the international market: Colombian mild Arabicas, Other mild Arabicas, Brazilian and other natural Arabicas, and Robustas. These indicator prices represent spot or cash prices, quoted in the market for coffee that is more or less immediately available (or within a reasonable time-span). The four categories enable the ICO to calculate market prices for these four broad groups and so monitor price developments for each. In addition, using an agreed formula, the ICO publishes a Daily Composite Indicator Price that combines these four into a single price representing "all coffee". This represents the best indication of a current "international price for coffee".

Futures prices reflect the estimated future supply and demand for a defined average quality of coffee (e.g. Arabica coffee futures prices in New York, Robusta coffee futures prices in London). In these markets, forward trading is used to offset price risk in the green coffee market where different qualities of coffee are traded. Traders therefore link individual prices with the futures price by establishing a price difference, the differential. The differential takes into account (i) differences between an individual coffee and the standard quality on which the futures market is based and, (ii) the supply. For example, by combining the New York or London futures price and the differential, one usually obtains the FOB (free on board) price for a particular type of green coffee. This enables the market to simply quote, for example, "Quality X from Country Y for October shipment at New York December plus 5" (US cts/lb). Traders and importers know the cost of shipping coffee from each origin to Europe, the United States, Japan or wherever, and so can easily transform "plus 5" into a price "landed final destination".

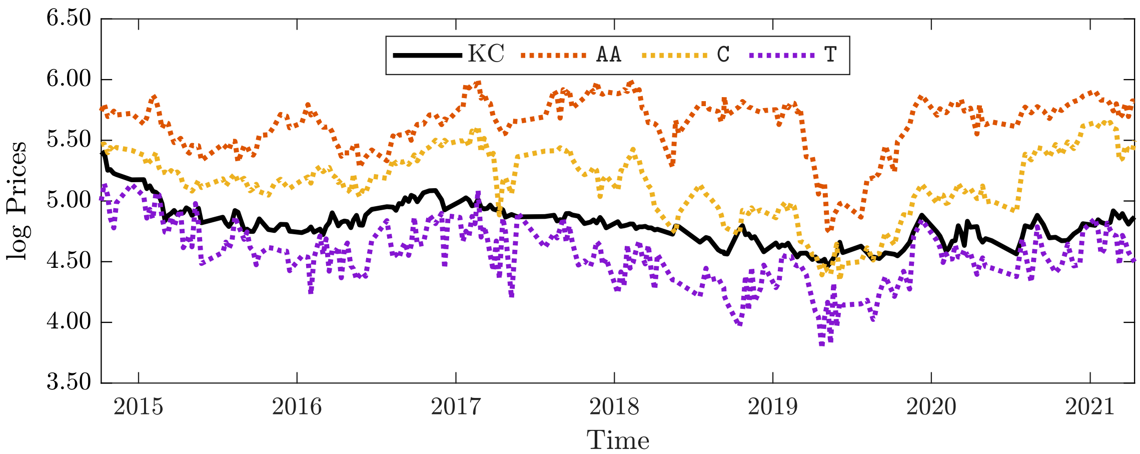

Figure 1: NCE Auction Spot Prices & KC Futures Contract Price

The bold dotted lines represent the three NCE spot prices (grades AA, C and T), while the solid line is the nearest maturity KC futures price.

The Nairobi Coffee Exchange (NCE) is the physical market of coffees from at least five different African countries. While the output sold at the market is not large compared to producers such as Brazil and Ethiopia, the "Colombian Milds" sold at the auction have a 48–

54% weight in the calculation of international coffee prices by the ICO.[1] The NCE coffee also have a price premium of 20 US$ on prices of similar grade coffee traded at the International Coffee Exchange.[2] A look at the time series of prices from the NCE and the average prices published by the International Coffee Organization suggests co-movement or integration, with NCE prices showing a constant premium as shown in Figure 1.

Table 1: Arabica Futures Contracts Specifications

Symbol	KC
Contract Size	37,500 pounds
Price Quotation	US Cents
Min. Price Fluctuation	0.05 cent/lb. ($18.75 per contract)
Settlement	Physical Delivery

The time series analysed come from two databases. Spot prices are from the physical auction at the Nairobi Coffee Exchange (NCE). A detailed description of how these data are generated through the auction mechanism is given in Appendix A.3. The futures price are nearest maturity prices of Arabica contracts traded at the ICE. The Arabica prices are from the International Coffee Organization (ICO), representing the New York cash price; ex-dock in US cents per pound. The futures prices are sourced from the CHRIS database hosted at the data provider Quandl[3]. The continuous futures contract chains together a series of individual futures contracts that provide a long-term price history that is suitable for our analysis. Table 1 gives details of the contract specifications. The time series are of weekly frequency from April 2014 to October 2021. Figure 1 shows the time variation of the prices we analyze.

3 Error Correction and Price Discovery

We flesh out how the error correction mechanism can arise from a popular model of commodity price dynamics due to Gibson and Schwartz (1990). We express their continuous time assumptions about the dynamics of the spot price and the unobserved "*convenience yield*" in discrete time, and approximate the latter from the arbitrage free expression of the relationship between futures and spot prices. This leads to a co-integration relationship between futures and spot prices similar to, but distinct from the present value models of Pindyck (1993) and Campbel and Shiller (1987). We start with the simple case of a single futures price with one spot market, then extend the model to account for multiple spot prices as will be the case when considering the different coffee grades auctioned at the NCE. We then give a causal interpretation of the model and discuss the information share measures arising from the theoretical

[1] See Annex 1 in http://www.ico.org/documents/icc-105-17e-rules-indicator-prices-final.pdf

[2] See e.g. Prices Paid to Growers from the International Coffee Exchange Prices Database http://www.ico.org/coffee_prices.asp

[3] https://www.quandl.com/data/CHRIS-Wiki-Continuous-Futures. Quandl Code: 'CHRIS/ICE_KC1'

framework.

3.1 Error Correction

Using standard no arbitrage arguments, the futures price at date t, for a unit of a commodity to be delivered at date $T \geq t+1$, may be expressed as:

$$F(t,T) = S(t)e^{(r-c)(T-t)} \tag{1}$$

where $S(t)$ is the commodity spot price, r is the continuously compounded interest rate and c is the "convenience yield". Gibson and Schwartz (1990) modeled the spot price as a random walk with a drift and the convenience yield as a mean reverting (Ornstein–Uhlenbeck) process, respectively:

$$d\log S = \left(\mu - \frac{1}{2}\sigma_s^2\right)dt + \sigma_s dW_s \;\text{ and }\; dc = k_c[\lambda - c]dt + \sigma_c dW_c,$$

where W_s, W_c are standard Brownian motions with mean 0, variance $\sqrt{dt}$ and *correlation* $dW_s dW_c = \rho dt$, so that spot price shocks are positively correlated with convenience yield shocks. Letting $s_t = \log S(t)$, the spot price can be represented in discrete time as the process

$$s_t = k_s + s_{t-1} + e_{st}, \tag{2}$$

where $k_s = \left(\mu - \frac{1}{2}\sigma_s^2\right)$ is the drift and e_{st} is an i.i.d normal variable with zero mean and variance σ_s^2. The convenience yield can similarly be represented as the discrete time process

$$c_t = \lambda k_c + (1-k_c)c_{t-1} + e_{ct},$$

with the error process following a normal distribution with zero mean and variance σ_c^2. For a slowly mean reverting process, k_c is small, so $\lambda k_c \approx 0$ and $(1-k_c) = \phi$ is close to 1. We can therefore write the convenience yield process as the simple first order auto regression:

$$c_t = \phi c_{t-1} + e_{ct}. \tag{3}$$

The two equations imply that the log-futures price consists of a permanent component, the random walk with a drift part, that is, the commodity spot price and a stationary AR(1) component, the convenience yield. Taking logarithms of (1) with $T = t+1$, and writing $\log F(t+1,t) = f_t^{t+1}$, the unobserved convenience yield can be expressed as the "*basis*"[4] plus an interest adjustment: $-c_t = f_t^{t+1} - s_t - r$. Assuming that the interest rates between any two

[4] The basis is generally defined as the contemporaneous difference between the nearest maturity futures price and the spot (cash) price.

time periods are not changing too quickly so that $r((t+1)-t) \approx \phi r(t-(t-1))$, then (3) can be written as

$$f_t^{t+1} = k_s + (1-\phi)s_{t-1} + \phi f_{t-1}^t - e_{ct} + e_{st}. \tag{4}$$

Dropping the superscripts and stacking (2) and (4) into the price vector $p_t = (s_t, f_t)'$, we obtain the VAR(1) representation: $p_t = c + Ap_{t-1} + u_t$, with $c = k_s(1,1)'$, and A, u_t, and the error covariance matrix $\text{Var}(u_t)$, given by:

$$A = \begin{bmatrix} 1 & 0 \\ 1-\phi & \phi \end{bmatrix}, \quad u_t = \begin{bmatrix} e_{st} \\ e_{st} - e_{ct} \end{bmatrix}, \quad \Omega = \begin{bmatrix} \sigma_s^2 & \sigma_s^2 - \rho\sigma_s\sigma_c \\ \sigma_s^2 - \rho\sigma_s\sigma_c & \sigma_f^2 \end{bmatrix},$$

where $\sigma_f^2 = \sigma_s^2 + \sigma_c^2 - 2\rho\sigma_s\sigma_c$. The system is cointegrated with cointegrating vector $\beta = (-1,1)'$ and error correction term $z_t = s_t - f_t$, which follows an AR(1) process: $z_t = \phi z_{t-1} + e_{ct}$.

3.2 Multiple Prices

We now present a discrete time "futures first" model, where the futures price reacts to lagged spot prices and its own lag, while spot prices follow random walks with drifts. The model yields a natural VECM representation, from which we derive permanent–transitory decompositions and compute price discovery metrics: Hasbrouck's (1995) Information Share and Gonzalo and Granger's (1995) Component Share.

As before, let $f_t \equiv f_t^{t+1}$ denote the log price of the benchmark futures contract for delivery at $t+1$, and let s_{it} denote the log spot price of grade $i = 1, \ldots, N$. The structural equations are:

$$\begin{aligned} f_t &= k_s + (1-\phi)\bar{s}_{t-1} + \phi f_{t-1} - e_{ct} + e_{st}, \\ s_{it} &= k_s + s_{i,t-1} + e_{ist}, \qquad i = 1, \ldots, N, \end{aligned} \tag{5}$$

where $\bar{s}_{t-1} = \frac{1}{N}\sum_{i=1}^N s_{i,t-1}$ is the average lagged spot price. The shocks are distributed as $e_{st} \sim N(0,\sigma_s^2)$, $e_{ct} \sim N(0,\sigma_c^2)$, $e_{ist} \sim N(0,\sigma_{si}^2)$, for $i = 1, \ldots, N$, with $\text{Cov}(e_{st}, e_{ct}) = \rho\sigma_s\sigma_c > 0$, and all other correlations equal to zero, so individual spot price innovations are uncorrelated with the futures price innovation by assumption. The parameter $\phi \in (0,1)$ governs the persistence of the futures price deviation from the long–run equilibrium with the spot prices. To write (5) in VECM form, let $\mathbf{1}_N$ be an $N \times 1$ vector of ones and define the stacked price vector $\mathbf{p}_t = (f_t, \mathbf{s}_t)'$, where $\mathbf{s}_t = (s_{1t}, \ldots, s_{Nt})$. Then the cointegrating vector is $\beta = (-1, \frac{1}{N}\mathbf{1}_N')'$ with cointegrating residual $z_t = \beta'\mathbf{p}_t = \bar{s}_t - f_t$, which gives the VECM representation of (5) as

$$\Delta\mathbf{p}_t = \mu + \alpha z_{t-1} + \mathbf{u}_t, \tag{6}$$

where $\mu = k_s \mathbf{1}_{N+1}$, $\alpha = (1-\phi, 0, \ldots, 0)'$, $\mathbf{u}_t = (e_{st} - e_{ct}, e_{1st}, \ldots e_{Nst})'$ and the covariance matrix of the innovations Ω is equal to

$$\Omega = \begin{bmatrix} \sigma_f^2 & \mathbf{0}_{1\times N} \\ \mathbf{0}_{N\times 1} & \mathrm{diag}(\sigma_{s1}^2, \ldots, \sigma_{sN}^2) \end{bmatrix}. \tag{7}$$

Following the Granger representation theorem, the levels prices vector can be expressed as

$$\mathbf{p}_t = \mathbf{C} \sum_{j=1}^{t} \mathbf{u}_j + \mathbf{C}^*(L)\mathbf{u}_t + \mathbf{p}_0 + \mu t, \tag{8}$$

where $\mathbf{C}$ is the long-run impact matrix. For the VECM (5), the orthogonal complements are $\alpha'_\perp = \beta'_\perp = [\frac{1}{N}\mathbf{1}_N, \mathbf{I}_N]$, where $\mathbf{I}_N$ in an $N \times N$ identity matrix such that $\alpha'_\perp \alpha = 0$ and $\beta'_\perp \beta = 0$. The long-run impact matrix is

$$\mathbf{C} = \beta_\perp (\alpha'_\perp \beta_\perp)^{-1} \alpha'_\perp = \begin{bmatrix} 0 & \frac{1}{N}\mathbf{1}'_N \\ \mathbf{0}_{N\times 1} & \mathbf{I}_N \end{bmatrix}.$$

Thus, the permanent component (common stochastic trend) is driven solely by the *average of the spot-price innovations*:

$$\tau_t = \alpha'_\perp \sum_{j=1}^{t} \mathbf{u}_j = \sum_{j=1}^{t} \bar{e}_{j\cdot}, \qquad \bar{e}_{j\cdot} = \frac{1}{N} \sum_{i=1}^{N} e_{ijt}.$$

This result implies that in this "futures–first" specification, the futures price is error-correcting to the spot average, so the long-run trend is determined by spot-market shocks. This has implications for the information share of a given market. Because the permanent innovation is $\bar{e}_{j\cdot}$, its variance is $\sigma_\tau^2 = \frac{1}{N^2}\sum_{i=1}^{N} \sigma_{si}^2$ (assuming independent spot shocks). Hasbrouck's (1995) information share **(IS)** for market i is the proportion of σ_τ^2 attributable to that market's innovation. Given the block-diagonal structure of Ω, the Cholesky factorization $\Omega = LL'$ yields a lower-triangular L. The long-run impact of orthogonalized shocks is $\mathbf{C}L$. Because only the spot-market block of $\mathbf{C}$ is non-zero, *only spot markets contribute to price discovery* in this model; the futures market's IS is zero. The IS for spot grade i is

$$IS_i = \frac{\sigma_{si}^2/N^2}{\sigma_\tau^2} = \frac{\sigma_{si}^2}{\sum_{j=1}^{N} \sigma_{sj}^2}. \tag{9}$$

Thus, the Hasbrouck share for each spot grade is simply its innovation variance relative to the total spot-innovation variance. The Gonzalo and Granger Component Share (**CS**) measure is based on the orthogonal complement of the adjustment coefficients. Here $\alpha_\perp = [\mathbf{0}'_{N\times 1}, \mathbf{I}_N]'$, so the permanent component weights are $(0, 1/N, \ldots, 1/N)$. Normalizing to sum to one, the *component share* for the futures market is 0, and for each spot grade is $1/N$. This equal weighting arises because all spots adjust identically (no spot-specific error correction).

The "futures-first" VECM for a multi-grade commodity market implies that price discovery, as measured by the IS and CS, is entirely attributable to the spot markets when futures prices error-correct to the spot average and spot prices follow random walks. For markets like coffee, where the futures contract is highly liquid, the model can be relaxed to allow for two-way feedback, yielding richer price-discovery patterns. We present such a model in the next section.

3.3 Spot-Specific Adjustment Speeds

We extend the previous specification by allowing each spot grade to adjust to the futures?spot equilibrium at its own speed. Let $f_t \equiv f_t^{t+1}$ denote the log futures price and s_{it} the log spot price of grade $i = 1,\ldots,N$. The structural system is

$$\begin{aligned} f_t &= k_s + (1-\phi)\bar{s}_{t-1} + \phi f_{t-1} - e_{ct} + e_{st}, \\ s_{it} &= k_s + s_{i,t-1} + \alpha_{si}(\bar{s}_{t-1} - f_{t-1}) + e_{ist}, \quad i = 1,\ldots,N, \end{aligned} \tag{10}$$

where $\bar{s}_{t-1} = \frac{1}{N}\sum_{j=1}^{N} s_{j,t-1}$. The parameters α_{si} capture grade-specific speeds of adjustment toward the long-run equilibrium. Stacking prices as $\mathbf{p}_t = (f_t, \mathbf{s}_t)'$, the system admits the VECM representation (6), with $\alpha = (1-\phi, \alpha_{s1}, \ldots, \alpha_{sN})'$ and the innovation covariance matrix $\Omega = \mathrm{Var}(\mathbf{u}_t)$ given by (7). Since the system contains $k = N+1$ variables and the cointegration rank is one, there are $k-1 = N$ common stochastic trends.

3.3.1 Gonzalo-Granger Component Shares (CS)

The levels equation admits the permanent transitory decomposition (8) with the long-run impact matrix $\mathbf{C} = \beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'$, where the orthogonal complements satisfy $\beta'\beta_\perp = 0$ and $\alpha'\alpha_\perp = 0$. A convenient choice is $\beta_\perp = (0_{1\times N}, I_N)'$ so that the permanent shocks are linear combinations of spot-market innovations. The vector $\alpha_\perp$ spans the null space of α'. Since α has dimension $(N+1)\times 1$, $\alpha_\perp$ is $(N+1)\times N$ and satisfies

$$(1-\phi)a_0 + \sum_{i=1}^{N} \alpha_{si} a_i = 0,$$

for an arbitrary $\alpha_\perp = (a_0, a_1, \ldots, a_N)'$. Any basis of this null space yields a valid permanent component.

Following Gonzalo and Granger (1995), the permanent component is defined as $g_t = \alpha_\perp' \mathbf{p}_t$, which is driven exclusively by permanent shocks. The component shares (**CS**) measure each market's weight in the permanent component. In the single-cointegration case, a convenient normalized vector orthogonal to α is[5]

$$\tilde{\alpha}_\perp = \begin{pmatrix} 1 \\ -\frac{\alpha_{s1}}{1-\phi} \\ \vdots \\ -\frac{\alpha_{sN}}{1-\phi} \end{pmatrix}. \tag{11}$$

Normalizing so that the weights sum to one yields the Gonzalo-Granger component shares:

$$CS_j = \frac{\tilde{\alpha}_{\perp,j}}{\mathbf{1}'\tilde{\alpha}_\perp}. \tag{12}$$

Markets that adjust more rapidly to disequilibrium (large $|\alpha_{si}|$) receive smaller weights in the permanent component.

3.3.2 Hasbrouck Information Shares (IS)

While the component share depends only on adjustment coefficients, Hasbrouck's (1995) information share (**IS**) measures each market's contribution to the variance of innovations to the common efficient price. Assume the innovation covariance matrix is block–diagonal such that $\Omega = \mathrm{diag}(\sigma_f^2, \Sigma_s)$ where the spot block allows correlation between the spot prices:

$$\Sigma_s = \begin{pmatrix} \sigma_{s1}^2 & \rho_{12}\sigma_{s1}\sigma_{s2} & \cdots \\ \rho_{12}\sigma_{s1}\sigma_{s2} & \sigma_{s2}^2 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}.$$

Let the normalized null-vector of α be given by (11), and define the futures innovation $e_{ft} = e_{st} - e_{ct}$ and $\gamma_i = -\alpha_{si}/(1-\phi)$, then permanent innovations satisfy

$$\tau_t = \tilde{\alpha}_\perp' \mathbf{u}_t = e_{ft} + \sum_{i=1}^{N} \gamma_i e_{sit},$$

Since futures-spot covariances are zero by assumption, the variance of the permanent innovation is[6]

$$\sigma_\tau^2 = \sigma_f^2 + \gamma' \Sigma_s \gamma.$$

[5] The vector $\alpha_\perp$ is defined only up to multiplication by a non-zero scalar, since it spans the null space of α'. Because Gonzalo–Granger component shares are obtained by normalizing $\alpha_\perp$ so that its elements sum to unity, any scalar rescaling leaves the shares invariant. Hence alternative normalizations of $\alpha_\perp$ produce identical component share estimates.

[6] With non-zero covariances, $\sigma_\tau^2 = \sigma_f^2 + 2\sum_{i=1}^{N} \gamma_i \mathrm{Cov}(e_{ft}, e_{sit}) + \sum_{i=1}^{N}\sum_{j=1}^{N} \gamma_i \gamma_j \Sigma_{s,ij}$.

Given shocks are already orthogonal between the futures price and individual spots, we obtain the futures and spot grade i information shares

$$IS_f = \frac{\sigma_f^2}{\sigma_f^2 + \gamma' \Sigma_s \gamma} \quad \text{and} \quad IS_{si} = \frac{\sum_{j=1}^{N} \gamma_i \gamma_j \Sigma_{s,ij}}{\sigma_f^2 + \gamma' \Sigma_s \gamma}.$$

In the special case with independent spot shocks, i.e. when $\Sigma_s = \text{diag}(\sigma_{si}^2)$, spot information shares are given by

$$IS_{si} = \frac{\alpha_{si}^2 \sigma_{si}^2}{(1-\phi)^2 \sigma_f^2 + \sum_{j=1}^{N} \alpha_{sj}^2 \sigma_{sj}^2}. \tag{13}$$

Because the Cholesky factor depends on ordering, upper and lower bounds are obtained by permuting the ordering of variables.

3.3.3 Special Cases and Economic Interpretation

The heterogeneous-adjustment model component shares (12) yields several insights. If $\alpha_{si} = 0$ for all i, spot prices are weakly exogenous and futures fully adjust, so the permanent component is determined by spot shocks. Larger $|\alpha_{si}|$ implies faster spot adjustment and therefore smaller component shares. Finally, a larger $(1-\phi)$ implies stronger futures adjustment and therefore a smaller futures role in price discovery.

Information shares given by (13) depend additionally on innovation variances and contemporaneous correlations. This has several implications for price discovery. When there is no spot adjustment ($\alpha_{si} = 0$), then all price discovery occurs in the forward market. Secondly, we can have a single dominant *spot grade*. If $\sigma_{s1}^2 \gg \sigma_{sj}^2$ for $j \neq 1$, and α_{s1} is moderate, that spot market dominates price discovery. Larger α_{si} implies that spot market i adjusts quickly to disequilibrium so contributes less to the permanent component. On the other hand, a smaller σ_{si}^2, means there is more precise information in market i, so it will have a larger weight in permanent component. A larger $1-\phi$ implies that the futures price adjusts more and the forward market has less contribution to price discovery.

For coffee, different grades (Colombian Mild, Brazilian Natural, Robusta or $\{$`AA, AB,` $\dots\}$) likely have different adjustment speeds and innovation variances. Higher-quality grades may adjust slower if they have more stable demand, while less liquid grades have higher volatility or correlate more closely with futures. The extended model allows testing whether the ICE futures contract leads price discovery, if specific grades or origin markets (e.g., Colombian) dominate price formation, and whether adjustment speeds relate to market liquidity or quality differentials. The framework is particularly relevant for commodities like coffee, where multiple physical grades coexist with a benchmark futures contract, and where price discovery patterns may vary across quality tiers and regional markets. The framework therefore permits disentangling two distinct channels of price discovery: adjustment dynamics (component shares) and innovation variance contributions (information shares).

4 Empirical Analysis

We implement the empirical estimation and testing procedure in several steps. We first perform standard stationarity tests. We estimate the VECM under the maintained hypothesis of rank one cointegration. Adjustment coefficients α and innovation covariance matrix Ω are obtained via maximum likelihood. Component shares are computed using the normalized null vector of α as in Gonzalo and Granger (1995). Information shares are computed using the Hasbrouck (1995) methodology based on the estimated long-run impact matrix and the Cholesky factorization of Ω. Upper and lower bounds are reported under alternative orderings to account for contemporaneous correlation. Finally, we test the hypotheses listed in Table 2 below.[7]

Table 2: Hypotheses Tests

Null Hypothesis	Interpretation
$H_{01} : \alpha_{s1} = \alpha_{s2} = \cdots = \alpha_{sN} = 0$	No spot adjustment.
$H_{02} : \alpha_{s1} = \alpha_{s2} = \cdots = \alpha_{sN}$	Equal adjustment speeds.
$H_{03} : \phi = 1$	No futures adjustment.

4.1 Estimation Results

We estimate the vector error correction model specified by eq. (6) with the cointegrating vector restricted to $\beta = (-1, \frac{1}{N}\mathbf{1}_N)'$ as derived in Section 3.2. The system includes the log of the ICE Coffee C futures price and the log of three representative spot grades from the Nairobi Coffee Exchange (`AA`, `C`, and `T`).[8] The sample consists of weekly observations from

[7]We need to compute the covariance matrix of the adjustment speeds α to test the hypotheses above. To do so, consider a k-dimensional VECM(ℓ) with cointegration rank r and innovation covariance matrix Σ_u is the . Following Johansen's (1991) maximum likelihood framework, define $Z_{0t} = \Delta p_t$, $Z_{1t} = p_{t-1}$, $Z_{2t} = \left(\Delta p'_{t-1}, \ldots, \Delta p'_{t-\ell+1}\right)'$. Partial out the short-run dynamics by projecting Z_{0t} and Z_{1t} onto Z_{2t} and denote the resulting residuals by $R_{0t} = Z_{0t} - \hat{\Pi}_2 Z_{2t}$, $R_{1t} = Z_{1t} - \hat{\Pi}_1 Z_{2t}$. The concentrated system can then be written as $R_{0t} = \alpha\beta' R_{1t} + u_t$.. Let $W_t = \beta' R_{1t} \in \mathbb{R}^r$, and stack observations over $t = 1, \ldots, T$ to obtain $R_0 = W\alpha' + U$, where R_0 is $T \times k$, W is $T \times r$, and U collects the innovations. Conditional on β, estimation of α reduces to a multivariate regression with common regressors W.

The OLS estimator is $\hat{\alpha}' = (W'W)^{-1} W' R_0$. Vectorizing and using standard properties of the Kronecker product, $\mathrm{vec}(\hat{\alpha}) = \mathrm{vec}(\alpha) + \left((W'W)^{-1} \otimes I_k\right)\mathrm{vec}(U)$. Since $\mathrm{Var}\big(\mathrm{vec}(U)\big) = \Sigma_u \otimes I_T$, it follows that the asymptotic covariance matrix of the adjustment coefficients (conditional on β) is

$$\mathrm{Var}\big(\mathrm{vec}(\hat{\alpha})\big) = \Sigma_u \otimes (W'W)^{-1} = \Sigma_u \otimes \left(\beta' R_1' R_1 \beta\right)^{-1}.$$

In empirical implementation, Σ_u is replaced by its consistent estimator $\hat{\Sigma}_u = T^{-1}\sum_{t=1}^{T} \hat{u}_t \hat{u}_t'$, and R_1 is constructed from the residuals of the partial regressions described above. The resulting matrix has dimension $(kr) \times (kr)$,and standard errors for the elements of α are obtained as the square roots of its diagonal entries. This expression corresponds to the covariance of $\hat{\alpha}$ conditional on the estimated cointegration space.

[8]The NCE sells many different green coffee *grades*: `AA,AB, C, HE, MH, ML, PB,T, TT, UG1, UG2, UG3`. The three grades analyzed are the most consistently traded during the sample period; including additional grades would reduce the number of usable observations due to missing auctions.

April 2014 to October 2021, yielding $T = 394$ observations after accounting for lags.[9] The VECM(1) is estimated using Johansen's (1991) maximum likelihood procedure with no constant in the cointegrating space (model H2).

4.1.1 Unit Root and Cointegration Tests

Before estimating the vector error correction model, we first examine the time series properties of the four price series: the ICE Coffee C futures price (`KC`) and the three Nairobi spot grades (`AA`, `C`, `T`). All series are expressed in natural logarithms. The augmented Dickey-Fuller (ADF) test is applied to each series to assess the presence of a unit root. The results, reported in Table 3, indicate that for all four series the null hypothesis of a unit root cannot be rejected at conventional significance levels, as the p-values exceed 0.05. This confirms that each price series is integrated of order one, I(1), a necessary condition for cointegration analysis.

Table 3: Augmented Dickey-Fuller Unit Root Tests

Series	ADF Statistic	p-value
`KC` (futures)	-0.7542	0.3748
`AA`	-0.0733	0.6241
`C`	-0.1027	0.6133
`T`	-0.4894	0.4717

Notes: The test regression includes a constant; lag length selected by AIC.

Given the I(1) property of the individual series, we proceed to test for cointegration using Johansen's (2002) trace test. The test is applied to a vector autoregressive model of lag order 1 (selected by information criteria) with an unrestricted constant in the cointegrating space (model H2 in `MATLAB's Econometrics Toolbox`). The results, summarized in Table 4, strongly reject the null hypothesis of no cointegration ($r = 0$) with a trace statistic of 64.79 ($p = 0.001$). In contrast, the null of at most one cointegrating relationship ($r \leq 1$) cannot be rejected ($p = 0.2465$). Consequently, we conclude that there exists a single cointegrating vector among the four price series, consistent with the theoretical prediction of a long-run equilibrium linking the futures price and the average of the spot grades.

Table 4: Johansen Cointegration Rank Test (Trace)

r	Hypothesis	Trace Statistic	5% Critical Value	p-value
0	$r = 0$	64.79	40.18	0.0010
1	$r \leq 1$	18.37	24.27	0.2465
2	$r \leq 2$	6.65	12.32	0.4162
3	$r \leq 3$	0.14	4.13	0.7840

[9] The exact number of weeks with simultaneous trading in both markets is 394.

Having established the presence of one cointegrating relation, we estimate the VECM with the cointegrating vector restricted to $\beta = (-1, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})'$ as derived in Section 3.2. The estimated adjustment coefficients (loading matrix α) are $\hat{\alpha} = (0.0407, -0.0684, -0.0884, -0.1441)'$. The coefficient for the futures equation is positive (0.0407), indicating that when the average spot price exceeds the futures price (i.e., a negative basis), the futures price tends to rise in the next period, restoring equilibrium. All three spot adjustment coefficients are negative, ranging from -0.0684 for grade AA to -0.1441 for grade T, implying that spot prices fall when they are above the long-run equilibrium relative to futures. The magnitudes suggest that grade T adjusts most quickly to disequilibrium, while grade AA adjusts more slowly. The statistical significance of these coefficients is confirmed by Wald tests reported in Section 4.1.3. Finally, the residual covariance matrix $\hat{\Omega}$ from the VECM estimation is:

$$\hat{\Omega} = \begin{bmatrix} 0.0028 & 0.0011 & 0.0009 & 0.0016 \\ 0.0011 & 0.0087 & 0.0033 & 0.0020 \\ 0.0009 & 0.0033 & 0.0075 & 0.0052 \\ 0.0016 & 0.0020 & 0.0052 & 0.0250 \end{bmatrix}.$$

The relatively low off-diagonal elements suggest modest contemporaneous correlations among the innovations, with the futures innovations exhibiting the smallest variance. This covariance structure is used in the subsequent computation of Hasbrouck information shares and Gonzalo-Granger component shares.

4.1.2 Hasbrouck Information Shares

Table 5 reports the Hasbrouck's (1995) information shares, computed as the mid-point (mean), upper and lower bounds obtained by ordering the futures price first and last in the Cholesky factorization of the residual covariance matrix (Baillie et al., 2002). The bounds – 95% confidence intervals – are relatively narrow, indicating that the estimates are not highly sensitive to the ordering of the variables. The futures market contributes approximately 17% of the price discovery, while the three spot grades together account for the remaining 83%. Among the spot grades, the *highest-quality* grade (AA) has the *lowest* share (20%), followed by C (24%) and T (19.8%). This ranking is consistent with the idea that *lower-quality* beans may convey more information about long-run supply and demand conditions, possibly because they are more actively traded and/or because their price differentials relative to the futures contract are more volatile. Given that grade T has the largest α_{si} and σ^2_{si}, it dominates the information shares as predicted by (13). The combined spot market share of 83% shows that the spot markets contributes more to price discovery in aggregate than the forward market.

Table 5: Information and Component Shares

		Information Share			Component Share
Market	Symbol	Lower	Upper	Mid-Point	Value
Futures (ICE Coffee C)	`KC`	0.0935	0.2449	0.1692	0.1191
Spot Grade	`AA`	0.1158	0.2860	0.2009	0.2004
	`C`	0.1496	0.3284	0.2390	0.2588
	`T`	0.2992	0.4826	0.3909	0.4218

4.1.3 Gonzalo-Granger Component Shares

While the Hasbrouck information shares measure a market's contribution to the variance of the common efficient price innovations, the Gonzalo and Granger (1995) component share (CS) offers a complementary perspective by focusing on the composition of the common factor itself (de Jong, 2002). As Baillie et al. (2002) explain, the CS model measures each market's contribution to the common factor, where the contribution is defined as a function of the markets' error correction coefficients. It decomposes the permanent component into a linear combination of the observed prices, with the weights reflecting each market's share in the price discovery process.

Figure 2 presents the Gonzalo-Granger component shares for the ICE Coffee C futures contract and the three Nairobi spot grades reported in the last column of Table 5. The results reveal that the futures market has a component share of 12%, which is notably lower than its Hasbrouck information share of 17%. Conversely, the combined spot market component share is 88%, higher than the 83% observed in the Hasbrouck measure. This divergence between the two metrics is a well-documented phenomenon in the price discovery literature. As de Jong (2002) demonstrates, the Gonzalo-Granger measure works only with the "weight" that the innovation of market i has in the increment of the common factor, while the Hasbrouck measure additionally incorporates the variance of the innovations in each market's price. The two models provide similar results only when the residuals are uncorrelated; when substantive correlation exists, they typically yield different results. The lower futures share in the GG metric suggests that when considering only the adjustment coefficients (the error correction process), the futures market plays a *less* substantial role than when innovation variances are taken into account.

Among the spot grades, `T` exhibits the largest component share (42%), more than double that of `AA` (14.55%) as shown in Figure 2 above. This ranking is consistent with the Hasbrouck shares, where `T` was the dominant spot grade. The larger share for `T` implies that it has the smallest *relative* adjustment coefficient among the spot grades – that is, it adjusts most slowly to the equilibrium, thereby contributing more to the permanent component. This interpretation is consistent with the permanent-transitory decomposition framework where markets that are

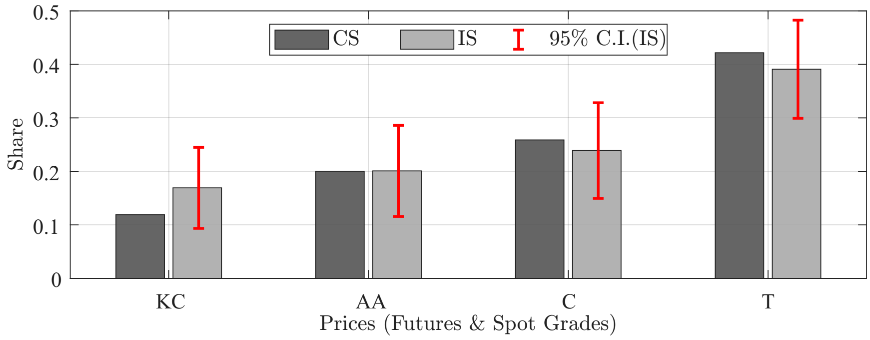

Figure 2: Component and Information Shares

slower to correct deviations bear a larger weight in the common factor (de Jong, 2002).

Although spot grade `T` has the largest adjustment coefficient in magnitude (-0.1441), implying it reacts most strongly to deviations from equilibrium, it receives the largest component share. At first glance, the result appears paradoxical since we should expect a rapidly adjusting market to play a smaller role in price discovery. However, GG shares (12) depend on relative adjustment speeds:

$$CS_{si} \propto \frac{|\alpha_{si}|}{1-\phi}.$$

Although spot markets adjust strongly in absolute terms, the futures market adjusts very weakly (0.0407). Consequently, the relative adjustment ratios are large: $\frac{0.1441}{0.0407} \approx 3.54$. Thus, in relative terms, spot markets absorb most of the equilibrium correction burden. The permanent component is therefore primarily determined by spot innovations. Since the futures adjustment coefficient is small, the relative adjustment ratios are large, leading spot markets to dominate the permanent component.

4.1.4 Hypothesis Tests on Adjustment Speeds

To further characterize the dynamics of price adjustment, we test three hypotheses on the adjustment coefficients α using Wald tests based on the estimated covariance matrix of the parameters. Because the cointegrating vector is fixed, the adjustment coefficients are identified and the Wald statistics are asymptotically χ^2 distributed. Table 6 summarizes the results.

Table 6: Wald Tests on Adjustment Coefficients

Hypothesis	Description	Wald Stat.	d.f.	p-value
H_{01}: $\alpha_{s1} = \alpha_{s2} = \alpha_{s3} = 0$	All spot adjustments are zero	39.3195	3	0.0000
H_{02}: $\alpha_{s1} = \alpha_{s2} = \alpha_{s3}$	Equal adjustment speeds across spot grades	5.0608	2	0.0796
H_{03}: $\alpha_f = 0$	No futures adjustment	15.3792	1	0.0001

The strong rejection of H_{01} (p-val < 0.01) indicates that spot prices do adjust to deviations from the long-run equilibrium. This confirms that the Nairobi spot market is not passively following the futures price; instead, it actively contributes to the error-correction process. The point estimates of the spot adjustment coefficients (not shown) are all positive and statistically significant, implying that when the average spot price is above the futures price (i.e., the basis is negative), spot prices tend to fall in the next period, consistent with a return to equilibrium. Hypothesis H_{02}, which posits equal adjustment speeds for the three spot grades, is not rejected at the 5% level (p-val $= 0.08$). Although the p-value is marginally above 0.05, the result suggests that the speed at which different coffee grades revert to the equilibrium with the futures price is not statistically distinguishable. In economic terms, this could reflect a common market microstructure or similar exposure to the same fundamental shocks. All three grades are auctioned in the same weekly session, and buyers and sellers observe the same order flow; therefore, it is plausible that their responses to disequilibrium are similar. Finally, the rejection of H_{03} (p-val $= 0.0001$) demonstrates that the futures price also adjusts to the cointegrating relationship. The futures adjustment coefficient is negative and significant, meaning that when the average spot price exceeds the futures price, the futures price rises in the subsequent period. Hence, the price discovery process is bidirectional: while the spot market as a whole has a larger information share, the futures market does respond to information originating in the physical auction.

4.1.5 Forecast Error and Impulse Responses

Figures 3 and 4 give the forecast error variance decomposition and impulse response functions from the estimated VECM(1).The solid, dotted, dashed, and dashed-dotted lines represent, respectively, the variance contribution of a one-standard deviation shock to variable $j = \{\texttt{KC}, \texttt{AA}, \texttt{C}, \texttt{T}\}$, to the forecast variance 40 periods into the future. For each of the spot prices, shocks to futures price `KC` has negligible effect on to the forecast error. However, shocks to the spot prices contribute up to 20% of the forecast error variance of the futures price `KC`, rising to almost 30% for grades `C` and `T`. The impulse response functions show a similar pattern. Negative shocks to spot prices lead to a positive response of the futures price, suggesting the market interprets a fall in the usually high price grades `AA`, `C` and `T` as a negative supply shock in quality coffee, leading to a rise in the forward prices.

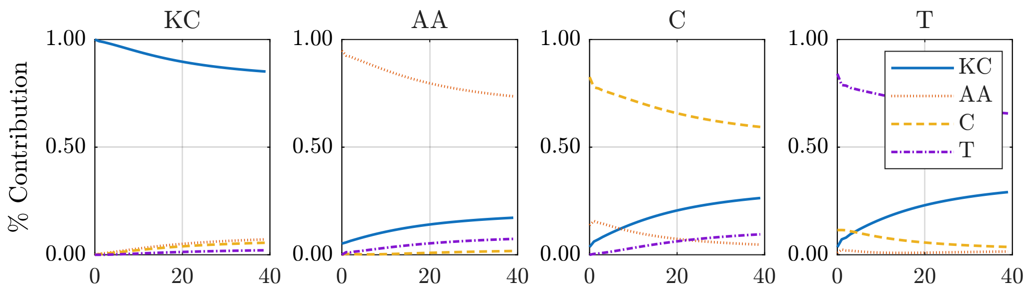

Figure 3: Forecast Error Variance Decomposition

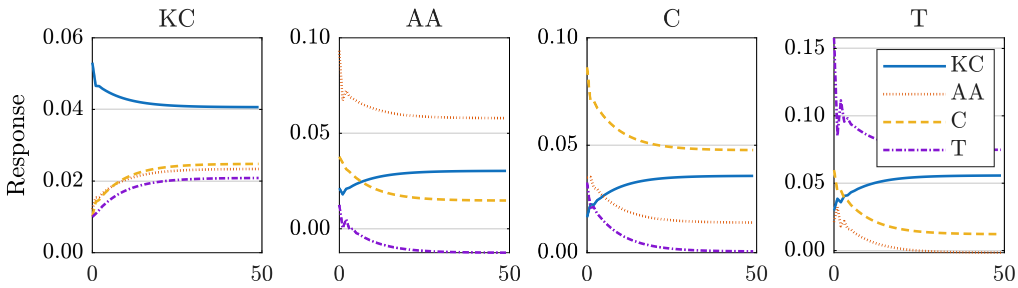

Figure 4: Impulse Response Functions

4.2 Discussion

The finding that the physical market dominates price discovery aligns with the theoretical model of Figuerola-Ferretti and Gonzalo (2010), who show that in commodity markets the convenience yield can induce a leader-follower relationship. In their framework, when the convenience yield is highly persistent and the spot price is driven by supply-demand fundamentals specific to the production region, the spot market can become the primary locus of price discovery. Our results for Kenyan coffee, a product with distinct quality attributes and a concentrated auction mechanism, provide empirical support for this prediction. These findings also complement the work of Bohl et al. (2019), who document bidirectional information transmission between Brazilian and U.S. coffee futures markets. While their focus is on two futures contracts, our analysis extends the picture by including the physical spot market in a major producing country. The substantial information share of the Nairobi auction suggests that local supply-demand conditions – reflected in the prices of different grades – are incorporated into the global price through the futures market's adjustment to the cointegrating relationship. Holmes and Otero (2020) find that coffee futures prices are unbiased predictors of future spot prices, but they do not quantify the relative contribution of each market to the price discovery process. Our information share estimates fill this gap by attributing a larger role to the spot market.

The component share results reinforce the finding that price discovery in the Kenyan coffee market is a shared process between the global futures contract and the local physical auction. While the spot market collectively dominates with a 88% share, the futures market's 12% contribution is modest, but still gives it a role as the global benchmark. The heterogeneity across spot grades, with T emerging as the most influential in the GG metric, suggests that different grades play distinct roles in the price discovery process. T may represent a "bellwether" grade that is most representative of the broader market conditions, adjusting slowly to temporary dislocations and thus embodying more of the permanent information. These findings resonate with previous applications of the GG methodology in commodity markets. Baillie et al. (2002) note that the GG model focuses on the error correction process, involving only permanent shocks that result in disequilibrium. In our context, the disequilibrium arises because the Nairobi auction and the ICE futures market process news at different rates. The modest GG share for the futures market indicates that it is a passive reflector of spot prices that mildly contributes to the error correction process. This bidirectional adjustment was confirmed by our Wald tests, which rejected the hypothesis of no futures adjustment ($p - val = 0.0001$).

The GG results also provide a useful robustness check on the Hasbrouck findings. The fact that both metrics point to spot market dominance (83-88%) lends confidence to the conclusion that the Nairobi Coffee Exchange is the primary locus of price discovery for Kenyan coffee. However, the IS metric's higher futures share serves as an important reminder that the futures market's role should not be understated. As Theissen (2002) demonstrates in the context of floor versus screen trading, using multiple price discovery measures can provide a more complete picture of how information is incorporated across markets. The heterogeneity across spot grades, with `T` emerging as the most influential, highlights the importance of disaggregating the spot market by quality grade rather than treating it as a homogeneous entity. These findings underscore the complementary nature of the two price discovery metrics and the value of employing both in empirical analysis.

Taken together, these results imply that for the Kenyan coffee market, price discovery occurs primarily in the spot market, but the ICE futures contract retains a substantial role. The significant adjustment of both futures and spot prices to the equilibrium error indicates a healthy two-way feedback that helps maintain the long-run relationship. The finding that spot grades adjust at similar speeds is consistent with a well-integrated local market where quality differentials do not lead to heterogeneous dynamic responses to the global benchmark. This pattern resonates with the model of Section 3.3, which allows for grade-specific adjustment speeds but does not require them to differ significantly when the underlying information flow is common across grades.

5 Conclusion

This paper has investigated the locus of price discovery in the global coffee market, focusing on the interaction between the physical spot auction at the Nairobi Coffee Exchange (NCE) and the financial futures contracts traded on the International Coffee Exchange (ICE). Utilizing a vector error correction model and two complementary measures of price discovery, the Hasbrouck information share and the Gonzalo and Granger component share, we provide robust evidence that the physical spot market plays a dominant role in price formation. Our key finding is that the NCE spot market accounts for approximately 83% of price discovery as measured by the Hasbrouck information share, with the remaining 17% attributable to the ICE futures contract. This result challenges the conventional wisdom that highly liquid futures markets naturally lead price discovery, particularly for globally traded commodities.

Hypothesis tests confirm that price discovery is a bidirectional process: both futures and spot prices adjust significantly to deviations from the long-run equilibrium. The futures market's adjustment coefficient is negative and highly significant, meaning that when spot prices exceed futures, the futures price rises in the subsequent period, incorporating information from the physical auction. Conversely, spot prices adjust positively to disequilibrium, falling when they are above the futures price. This two-way feedback ensures that the long-run cointegrating relationship between the markets is maintained, despite the geographical and institutional separation of the trading venues.

Our findings have important implications for market participants and policymakers. For coffee producers and exporters in East Africa, the results underscore the importance of the NCE auction as a mechanism for price discovery and suggest that participation in this physical market provides valuable information not fully reflected in global futures prices. For traders and risk managers, the results highlight the need to monitor physical auction outcomes in producing regions, as these contain incremental information relevant to futures price dynamics. For policymakers concerned with market development, our analysis supports investments in transparent, efficient physical auction mechanisms as a means of enhancing price discovery and potentially improving the bargaining position of producers in global value chains.

Several avenues for future research emerge from this study. First, extending the analysis to other major coffee-producing regions – such as Brazil, Colombia, and Vietnam – would help establish whether the dominance of physical markets in price discovery is unique to East Africa or a more general phenomenon. Second, incorporating additional futures contracts, such as the ICE Robusta contract, would allow for a more complete picture of the global price discovery network. Third, exploring the role of speculative activity, hedging pressure, and order flow in the futures market could shed light on the mechanisms through which information from physical auctions is transmitted to financial markets. Finally, the development of higher-frequency data from physical auctions would enable more precise measurement of information transmission speeds and the dynamics of intraday price discovery. Despite these opportunities for further investigation, the current study provides compelling evidence that for Kenyan coffee, the physical market, not the financial one, is where prices are truly discovered.

References

Baillie, R. T., Booth, G. G., Tse, Y. and Zabotina, T.: 2002, Price discovery and common factor models, *Journal of Financial Markets* **5**(3), 309–321. (cited on page 2, 14, 15, 18)

Bohl, M. T., Gross, C. and Souza, W.: 2019, The role of emerging economies in the global price formation process of commodities: Evidence from Brazilian and U.S. coffee markets, *International Review of Economics & Finance* **60**, 203–215.
URL: *https://www.sciencedirect.com/science/article/pii/S1059056018303228* (cited on page 2, 18)

Campbel, J. Y. and Shiller, R. J.: 1987, Cointegration and tests of present value models, *Journal of Polictical Economy* **95**, 1062 – 88. (cited on page 5)

de Jong, F.: 2002, Measures of contributions to price discovery: a comparison, *Journal of Financial Markets* **5**(3), 323–327. Price Discovery.
URL: *https://www.sciencedirect.com/science/article/pii/S1386418102000289* (cited on page 15, 16)

Figuerola-Ferretti, I. and Gonzalo, J.: 2010, Modelling and measuring price discovery in commodity markets, *Journal of Econometrics* **158**(1), 95–107. Twenty Years of Cointegration.
URL: *https://www.sciencedirect.com/science/article/pii/S0304407610000552* (cited on page 2, 18)

Gibson, R. and Schwartz, E.: 1990, Stochastic convenience yield and the pricing of oil contingent claims, *Journal of Finance* **45**(3), 959 – 76. (cited on page 3, 5, 6)

Gonzalo, J. and Granger, C.: 1995, Estimation of common long-memory components in cointegrated systems, *Journal of Business & Economic Statistics* **13**(1), 27–35. (cited on page 2, 7, 8, 9, 12, 15)

Harris, F. H., McInish, T. H. and Wood, R. A.: 2002, Common factor components versus information shares: a reply, *Journal of Financial Markets* **5**(3), 341–348. Price Discovery.
URL: *https://www.sciencedirect.com/science/article/pii/S1386418102000307* (cited on page)

Hasbrouck, J.: 1995, One Security, Many Markets: Determining the Contributions to Price Discovery, *The Journal of Finance* **50**(4), 1175–1199.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1995.tb04054.x* (cited on page 2, 7, 8, 10, 12, 14)

Holmes, M. J. and Otero, J.: 2020, A tale of two coffees? Analysing interaction and futures market efficiency, *Studies in Economics and Finance* . (cited on page 2, 18)

Johansen, S.: 1991, Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models, *Econometrica: Journal of the Econometric Society* pp. 1551–1580. (cited on page 12, 13)

Johansen, S.: 2002, A small sample correction for the test of cointegrating rank in the vector autoregressive model, *Econometrica* **70**(5), 1929–1961. (cited on page 13)

Pindyck, R. S.: 1993, The Present Value Model of Rational Commodity Pricing, *The Economic Journal* **103**(418), 511–530.
URL: *https://doi.org/10.2307/2234529* (cited on page 5)

Theissen, E.: 2002, Price discovery in floor and screen trading systems, *Journal of Empirical Finance* **9**(4), 455–474. (cited on page 19)

Appendix

A Nairobi Coffee Exchanges Auction

This section gives a description of the organization of trading activities in the Nairobi Coffee Exchange (NCE). The exchange is an auction where buyers and sellers meet to trade pre-submitted, graded and "tasted" lots of coffee. Both the buyers and sellers are registered members of the exchange. Auctions are held once every week, on Wednesdays at the NCE Auction hall. Over the last five years, the NCE has held its auction on almost every week. In the next subsections, I give a brief summary of the auction mechanism and the organization of activities on a typical auction day.

A.1 The Auction Mechanism

The NCE operates a clock auction mechanism designed for the exchange by Aucsys,[10] a Belgian electronic trade systems developer. The NCE operates a "Dutch auction", where the seller selects a start price from which the price moves down until a buyer intervenes. A central display panel, as shown in Figure A.1, enables the buyers to identify the coffee on offer: lot number, description, production area, quality, proposed start price (usually augmented by a small random ± value), the amount of each step by which the opening price will reduce (optional), buyer's name, plus information on upcoming lots, current auction status etc.

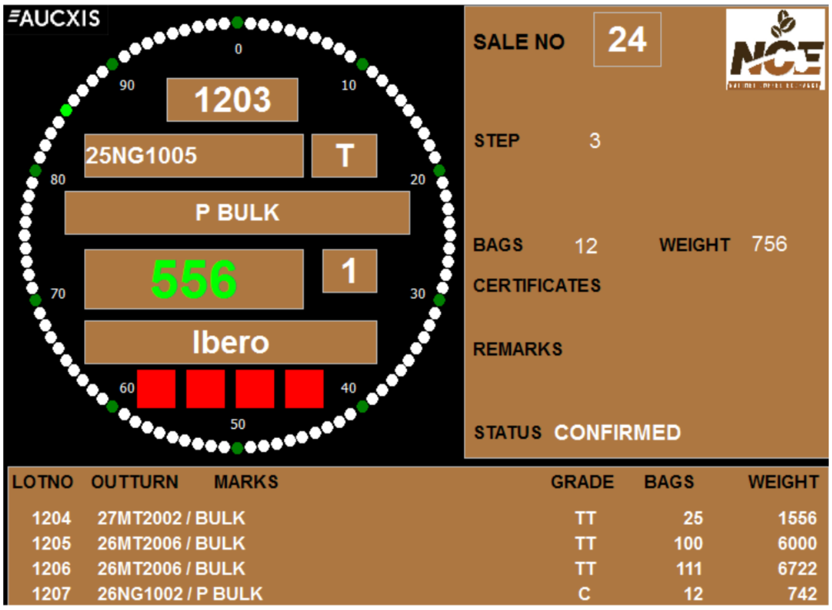

Figure A.1: Electronic Display Panel

To begin the sale the auctioneer enters a price and a pre-determined clock step. The price immediately starts to go down and, when a buyer is interested, s/he pushes one of three bid

[10] https://www.aucxis.com/en/news/aucxis-modernises-auction-system-nairobi-coffee-exchange

buttons, depending on the amount by which he wishes to increase the price. Then, as further bids are made, the value increases throughout the bidding time already determined by the auctioneer and displayed. When the time is up, the lot is assigned to the highest bidder. As soon as a lot is sold, the following lot moves from the "next lots" display to become the current lot. Figure A.2 shows the auction hall during a break, when sellers(auctioneers) change, with the display panel above the next sellers who are preparing to start auctioning their lots. The downward slopping desks are each allocated to a registered buyer, with buttons allowing the buyer register interest.

Figure A.2: Auction Hall

A.2 Buying and Selling Process

Before a typical auction day, lots of coffee are delivered to a common warehouse where "warrants" are issued by the "warehouse-men" in conjunction with the sellers or "marketing agents". After the a lot has been delivered to the warehouse with an accompanying sale catalogue, the sellers draw and present representative samples to the trade sample room. Figure B.3a represents typical samples from different lots displayed in the sample room. Buyers are then free to pick samples, at a fee equal to the average price over the previous month, for roasting and tasting.

On the auction day, buyers take seats in their pre-allocated positions/desks in the auction hall. Figure B.3b shows a typical rows of buyers or their agents participating in or observing the bidding process. Sellers control the auction process, choosing their minimum prices, the speed of price changes and sequence of selling their inventory. For each sale (of an individual lot from a single buyer), buyers are free to enter the bidding process or skip and wait for the

(a) Sample Room (b) Buyers at the Auction Hall

Figure B.3: Buying Process

next sale. Once the seller initiates the auctioning process by starting the clock timer, the price begins falling. By pressing a set of buttons available at each buyer's desk,[11] a buyer indicates interest in purchase, resulting into a reversal of the displayed price level which starts rising. If multiple buyers are pressing the buy button at the same time, the price rises faster, only stopping when there is one buyer left. The price at the last button press is the sale's price and the last remaining bidder is the winner of the lot. Once the auction has ended, sellers prepare invoices for the respective winning bidders and remit coffee warrants to the traders after payments have been made.

A.3 NCE Auction Data

This section gives a description of the data available for analysis from the NCE.

A.3.1 Seller Side Data

Sellers information at the NCE is available through their catalogues. Every week, before the auction, sellers submit a detailed description of the lots of coffee they will have on sale. The catalogues includes standard identifying information such as seller name and date of intended auction/sale. Since the product is sold in lots, with each lot potentially being from a different region, harvest area and/or mill, among other characteristics, the sales are organized per lot, each of which presents product of different quality/grade and quantity. For each lot, a unique ID is issued by the "warehouse-men" and a descriptive "mark" added (and also recorded on 2.5kg the sample bags). Finally grade, quantities on offer and packaging information is included. Table 7 shows a sample catalogue from available to buyers before the auction.

[11] The NCE has 80 buyer buttons: See http://www.aucxis.gr/images/p-fisf/20130125-Reference_list-E.PDF

Table 7: Typical Seller Catalogue

SALE No. N.C.E. 1
THE NAIROBI COFFEE EXCHANGE
THROUGH THIKA COFFEE MARKETING LIMITED
WILL OFFER BY AUCTION
1,045 Bags On Tuesday 2nd OCTOBER 2018
at 9.00 a.m.
AT THE EXCHANGE HALL
Wakulima House
NAIROBI
1,045 Bags of Kenya Coffee
Prompt Date 9TH OCTOMBER 2018

LOT	MARK	GRADE	BAGS	PKT	WEIGHT	SALE	SEASON
601	41TK0025/MUKURWE-ESTATE	T	11	8	668	1	2018/19
602	48TK0065/POINT-MZURI	T	16	59	1019	1	2018/19
603	51TK2004/T/BULK	T	15	40	940	1	2018/19
604	51TK2005/T/BULK	T	13	13	793	1	2018/19

A.3.2 Buyer Side Data

Buyers have access to the the sellers' catalogues before the auction. Auctions are held every Wednesday, from 9 AM and may last from 6 to 8 hours, depending on the number of lots on offer and how active the bidding is during the sale of each individual bid. There are 80 different "buyer desks" where bids are submitted through a set of buttons available at each desk. The following information is collected at every sale: a transaction number, lot number, season, marks, grade, sale number (of the season), bags and weight bought, buyer code, price, seat number (of buyer in auction hall), agent code (marketing agent or seller identifier) and time. Table 8 shows an extract of the transactions listings file showing information that is relevant to the analysis. Note that the lot numbers, marks ad weight match those of Table 7 as the transactions correspond to the catalogue of the same marketing agent (21/Thika Coffee Mills).

Table 8: Transaction Listing, Sale 1 of 2nd October 2018

TRCN.	LOT	MARKS	WEIGHT	BUYER	PRICE	SEAT	AGENT	TIME
21192	601	41TK0025	668	160	64	40	21	10:42:14
21193	602	48TK0065	1019	160	59	40	21	10:42:44
21194	603	51TK2004	940	74	89	74	21	10:43:21
21195	604	51TK2005	793	74	88	74	21	10:43:43
21196	605	51TK2006	1036	74	97	74	21	10:44:13

University of Warsaw
Faculty of Economic Sciences
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl
ISSN 2957-0506