

Jaap Bos, Maarten van der Molen

**A Bitter Brew? Futures
Speculation and Commodity
Prices**

RM/12/044

METEOR

Maastricht University School of Business and Economics
Maastricht Research School of Economics
of Technology and Organization

P.O. Box 616
NL - 6200 MD Maastricht
The Netherlands

A Bitter Brew? Futures Speculation and Commodity Prices

Jaap W.B. Bos

Maastricht University School of Business and Economics

`j.bos@maastrichtuniversity.nl`

Maarten van der Molen

Rabobank Netherlands

`m.t.molen@rn.rabobank.nl`

We introduce a new approach to measuring the possible impact of futures speculation on spot commodity prices. We advocate the use of a non-parametric, highly flexible empirical model for measuring this impact, in order to account for possible non-linearity in the transmission from futures to spot market. Empirical results for the coffee market show that most of the changes in spot prices can be attributed to shifts in demand and - in particular - supply. Nevertheless, speculation is an important part of the coffee price generation process. The effect of speculation on the price of coffee is indeed spiky, which explains why traditional, mean-variance based methods have failed to identify this. However, it is also significant, both statistically and economically. An extensive robustness analysis confirms the validity of our results, and - within the limitations posed by the data - we have been able to establish causality.

We thank Mette Asmild, Judith Beugels, Peter Bogetoft, Omar Hassib, Koert Jansen, Piet Keizer, Robert Merin, Mark Sanders, Anna Wisniewska and seminar participants at Maastricht University School of Business and Economics, Groningen University, University of Gent, Copenhagen Business School, Sveriges Riksbank and the German Bundesbank as well as participants of the XII European Workshop on Productivity and Efficiency Analysis in Verona for helpful comments. Thanks to Paulo Rodrigues for many insightful discussions regarding this paper, and to Caroline Studdert for excellent editorial assistance. We are grateful to Triodos Bank Netherlands for facilitating the data collection supporting this research, as well as for providing excellent insights into commodities finance and fair trade. This research would not have been possible without the data work and extremely helpful advice regarding all matters related to coffee of Jos Algra of Twin-Novotrade. Also, we thank Toby Quanttrill and Dave Goodyear of the Fairtrade Foundation for providing us with detailed fair trade prices. Finally, Maarten van de Molen gratefully acknowledges the support of Triodos Bank and the Utrecht University Science Shop. The usual disclaimer applies. The views expressed in this article are not necessarily shared by Triodos Bank Netherlands, Rabobank Netherlands, Twin-Novotrade, the Fairtrade Foundation and the Utrecht University Science Shop.

I. Introduction

Much of the economic literature dwells on the notion that as a market clears, demand equals supply, and an asset, physical good or derivative is traded at the market clearing price. But how much does that price indeed reflect the underlying fundamentals, demand, supply and their constituents, such as productivity, purchasing power, the price of raw materials and energy? And how much does it matter whether the market clearing price reflects those fundamentals?

In the last decade, the latter question has gained in importance as rising food prices have led to riots in India, Mexico and Yemen, a boycott of tomatoes in Argentina during the presidential elections, a one-day boycott of pasta organized by Italian shops and a price freeze on certain basic foods by the Russian government (Kingsbury, 2007). The consequences of rising food prices are particularly worrying for third world countries, where people spend between 60 and 80 percent of their income on food, compared with 10 to 20 percent for people in industrialized countries (Wahl, 2008). The World Bank forecasts that between 73 and 105 million people will fall below the absolute poverty line in the near future due to food crises as the result of rising prices (World Bank, 2008).

But what has produced these high prices? Is it, for example, a downward shift in supply as a result of soil exhaustion and climate change? Is it an upward shift in demand as a result of increasing wealth in developing countries? Has the price elasticity of supply changed as a result of changes in production techniques? Have consumers become less price elastic?

This paper presents an analysis of the development of the price of coffee. Although coffee is not the largest commodity in the commodity futures index, it is traded globally, and its production, consumption and inventories are well documented. The paper's contribution consists of asking a very important additional question: has commodity futures speculation contributed to rising prices? We present a simple model, where the price of a commodity is the result of demand and supply as well as factors that can shift demand and supply. In an extension of the model, producers can try to earn a profit, via arbitrage between spot and futures markets. There is an arbitrage opportunity if there is 'enough' production: with too much production, arbitrage gains from selling in the future are not feasible for many producers, whereas with too little production, producers cannot credibly commit to trading on the futures market, as they risk violating their short-run budget constraint. We demonstrate how our model is in fact similar to a production model where the price is the output, and inputs include demand and supply, specified in such a way that increases in each input can be expected to increase the price.

This model can be operationalized using a non-parametric production model, incorporating inefficiency in production. Known advantages of this modeling approach include the fact that it does not require normally distributed variables, nor does it require us to specify a production function, but it does allow us to measure deviations from the 'optimal' output (here, the price) without the ex-ante imposition of a distribution on these deviations. As a result, we can measure the possible effect of speculation in a flexible, relatively agnostic manner. And since our model is an efficiency model, this effect is measured as a decrease in inefficiency.

We apply the model using a single commodity, coffee, chosen for the reasons given above. Using a detailed data set on coffee production, consumption and related factors, we investigate whether speculative futures market positions have an effect on the coffee price. Subsequently, we execute an elaborate set of robustness tests to establish causality, minimize measurement problems, test the sensitivity to model assumptions and measure policy implications.

Our results show that futures speculation is indeed part of the coffee price generation process. The effect of speculation on the coffee price is spiky, but statistically and economically significant. We find some evidence of substitution effects with oil, the most important commodity in commodity index futures, and evidence of market impact as a result of sharp increases in index futures investment by institutional investors in the last decade.

The paper is set out as follows. In Section II, we present a short overview of the key developments in the literature on this topic, and summarize the most important developments in the coffee market. Section III describes our theoretical model, empirical setup and the formal derivations of our hypothesis. In Section IV, we introduce our data. Section V gives our results, and Section VI our conclusions.

II. Context

In this section, we describe the context in which our analysis takes place. We first present a brief review of the literature on the relationship between futures and spot commodity prices. Subsequently, we discuss developments in the coffee market.

A. The possible (or impossible) effect of futures on spot prices

For most commodities, the rise in prices has coincided with the increasing popularity of commodity index futures, as institutional investors increased

their exposure to commodities following the deregulation in important commodity derivatives markets which began in 2000. As yet unresolved is the issue of whether there is also a causal relationship between the two events, and if so, in which direction.

The latter issue gained momentum in 2008, when hedge fund manager Michael Masters claimed in testimony before the US Senate that commodity index futures investments are partly to blame for the commodity price increases. Masters later urged the Senate to vote in favor of the new Wall Street Transparency and Accountability Act (Masters, 2008). Meanwhile, Irwin and Sanders (2010) published the preliminary results of a study commissioned by the OECD's Food, Agriculture and Fisheries Department on "The Impact of Index and Swap Funds on Commodity Futures Markets," concluding that there was no such relationship.¹

The literature on the possible (or impossible) relationship between spot and futures prices of commodities has a long history. Stein (1961) developed a model where spot and futures prices are determined simultaneously, as a result of changes in (excess) supply or changes in price expectations. Garbade and Silber (1983) showed that storage and transaction costs constrain the supply of arbitrage services, as a result of which futures contracts provide imperfect risk transfer in the short run. They find that futures markets tend to lead cash markets.

Turnovsky (1983), relying on the fact that the futures price is a weighted average of the current and expected future spot price, showed that the impact of the futures market on the long-run spot price and its variance depends on the underlying parameters of supply and inventory demand. Related to this, Fama and French (1987) studied two different models of commodity futures prices. With their first model, they focused on the importance of convenience yields (cost of carry) for explaining the basis (the difference between future and spot prices). With their second model, the futures price was modeled as a function of the forecast of the spot price plus an expected premium.

Deaton and Laroque (1992, 1996) further contributed by enhancing the rational expectations competitive storage model. They showed that since a commodities market as a whole cannot carry negative inventories, the commodity prices predicted by the model are characterized by a high degree of autocorrelation, with occasional extreme spikes (Deaton and Laroque, 1992). Subsequently, they explored the non-linearities in the model further and introduced risk-neutral commodity speculators, whose introduction increased the autocorrelation in spot prices (Deaton and Laroque, 1996).

¹Other studies with overviews of the literature include Acharya et al. (2011), CFTC (2006), Wahl (2008), Irwin and Sanders (2011), Irwin et al. (2009), Tang and Xiong (2010) and Meijerink et al. (2011).

Currently, the idea that increased futures price volatility can result in spot price volatility is largely unquestioned (Irwin and Sanders, 2011; Irwin et al., 2009). Not generally accepted however is the possibility that futures prices may also affect the level of spot prices, despite early evidence pointing in this direction. Chan (1992) showed that there is an asymmetric lead-lag relationship between futures and stocks. When there is co-movement among stocks, a futures index tends to lead the cash index, since the former is able to react faster to new information than the latter. Subrahmanyam (1991) demonstrated theoretically that while the introduction of an index does not have an effect on price volatility, it does strengthen the relationship between the price of the index and the prices of the underlying assets in both directions. Booth et al. (1999) showed empirically that index futures dominate stock indexes and index options, using German DAX data. The low transaction costs of index futures explain their high information share.

Pindyck (1993) formalized the relationship between futures and spot commodity prices, emphasizing the role of changes in the expected future convenience yield. Given that, in Pindyck's model, the spot price is the present value of future convenience yields, the channel introduced here can be seen as a second order effect, where changes in current and future demand, besides having a direct effect on the price, change expected future convenience yields (and futures prices), thereby further affecting spot prices. Hochman et al. (2011) presented evidence in line with Pindyck (1993), finding that inventory adjustments contribute significantly to commodity spot price increases, after controlling for other factors such as economic growth, bio-fuel expansion, exchange rate fluctuations, and energy price inflation.

Proving that causality runs in one or possibly both directions can be very difficult. Irwin and Sanders (2010) and Irwin et al. (2009) based their analysis on the most commonly used methods: vector auto regression (VAR) models and Granger causality tests. They found no evidence that higher futures prices have led to higher spot prices. Advantages of these methods are that they make full use of the time series available for spot and futures prices, and allow for statistical inference. But there are also disadvantages. First, lack of evidence of Granger causality may be the result of the non-linearities Deaton and Laroque (1996) explored. Second, any hypothesis test based on these methods concerns the *average* impact of one price on the other, and is relatively insensitive to short periods in which this impact may be more intense, as suggested by Garbade and Silber (1983) and Chan (1992), among others. Third, an implicit assumption in these tests is that the possible impact of speculation is independent of other factors that may affect the price, such as inventories.

A different approach was taken by Korniotis (2009), who tried to exploit a quasi-natural experiment that occurred because futures contracts exist for

some industrial metals and not for others. He found that co-movement of spot prices between both groups has stayed constant despite a surge in futures speculation.² Bryant et al. (2006) took a slightly different approach and applied a Causal Inference (CI) algorithm. He found that Keynes' normal backwardation theory, which says that futures speculators can earn a premium as hedgers pass on risk through futures market, was rejected. Interestingly, any effect of speculation on spot price volatility was also rejected.

Recent studies have focused on three additional elements that may explain a causal link between futures and spot market. The first element is presented by Tang and Xiong (2010), Singleton (2011), Henderson et al. (2012) and others, who examined the changes in the micro-structure of futures markets. Singleton (2011) focused on crude oil, and found evidence of an "economically and statistically significant effect of investment flows on futures prices."³ He found that the index positions and "managed-money spread positions" affected crude oil futures prices the most. Tang and Xiong (2010) found that the rise of commodity futures index investment has contributed to the increase in the correlation between commodity spot prices in the US, reflecting the financialization of these markets. He contrasted this with developments in China, where commodity markets are driven much more by fundamentals and this increase in correlation has not occurred. Finally, Henderson et al. (2012), using a new data set of Commodity-Linked Notes, found that hedging trades executed by the issuers of these notes have resulted in a significant price impact on futures prices.

The second element was presented by Acharya et al. (2011), who applied the limits-to-arbitrage theory to commodity markets. In their model, capital-constrained speculators face hedging commodity producers in commodity futures markets. Hedging costs depend positively on the demand for hedging by producers and negatively on the risk capacity of speculators. As institutional investors' (speculative) demand for commodity index futures increases, the commodity futures risk premium is no longer simply determined by the convenience yield. They therefore concluded that "limits to financial arbitrage generate limits to hedging by producers, and affect both asset and goods prices." The same limits-to-arbitrage argument has also been used by Hong and Yogo (2012), who showed that as a result, futures prices have become less

²Korniotis (2009) also tried to examine the relationship between speculation and spot prices using a storage model applied to industrial metals. He finds no evidence of hoarding, as inventories tend to drop as spot prices increased. Also, the S&P Goldman-Sachs Commodity Index returns are not correlated with metal prices.

³Singleton controls for US and other stock market returns, "balance-sheet flexibility of large financial institutions, open interest, the future/spot basis, and lagged returns on oil futures."

informative about the real economy, including spot prices.⁴

The third element was presented by Frankel (2008), who focused on the impact of real interest rates on real prices of mineral and agricultural commodities. In light of the low real interest rates of the last years, his argument goes as follows: low interest rates lead to increased demand for storable commodities, and may also increase the supply of these commodities. At the same time, the cost of inventory drops and treasury bills become a less attractive investment. The result is an increase in spot commodity prices.

In summary, there are enough arguments theoretically speaking as to why the futures price may go up as a result of a speculative influx, as well as to why this may have an effect on the spot price. In Section III, we discuss the question of how we can ascertain whether the latter is indeed the case. For now, we conclude that supply and demand in the spot market, factors that can shift that supply and demand, and the market microstructure of the futures market can all have an effect on the spot price. Each of these elements will therefore appear in our theoretical and empirical analysis.

B. Developments in the coffee market

In 2009/2010, the last year of our sample, coffee accounted for USD15.4 billion in exports, making it the most traded tropical agricultural commodity (International Coffee Organization, 2012). In 2010, coffee was produced in approximately 70 countries, exceeding 10 percent of total export earnings in seven of those countries (International Coffee Organization, 2012). At the same time, consumption was 133.9 million bags, having increased by approximately 1.2 percent since the early 1980s (International Coffee Organization, 2012).⁵ Futures contracts for Arabica coffee, which is the prime bean in most coffee blends, are traded on the New York Mercantile Exchange.

The coffee market has for a long time been regulated by the International Coffee Agreement (ICA). Due to overproduction in the period from 1954 to 1956, Latin American producers decided to start controlling their production to stabilize the price. In 1958, all Latin American countries signed the Latin American Agreement (LAA), with the objective of stabilizing prices. However, because a large portion of coffee production was in Africa, the LAA did not work. In 1962, a large number of producing countries both from Africa and Latin America, as well as consuming countries, signed the first ICA. They agreed on a price target and export quotas for each producing country (Davi-ron and Ponte, 2005).

⁴Hong and Yogo (2012) showed that open interest, which is not (or less) hampered by limits to arbitrage is a much more accurate signal.

⁵One bag = 60kg.

According to Daviron (1996), four factors contributed to the success of the regime. First, the participation of consuming countries in the ICA was important. The consuming countries did not undermine the agreement by buying coffee from non-members. Second, in the producing countries, governments had control over exports. Because the volumes that these governments controlled were relatively large compared with the market, governments could effectively regulate the market. Third, Brazil's acceptance of production cuts was essential to create upward pressure on the price, as Brazil is the largest producer of coffee. And fourth, the members agreed to import part of each other's production when production was too high, thus sharing the burden of overproduction.

However, the ICA was eventually undermined by free-riding and dissention about quotas. Furthermore, more and more non-member countries started or increased production, and subsequently traded at lower prices. When the consumption of coffee in the United States rose, importers started to buy from non-member countries. Due to Cold War politics, relations between Latin America and the United States changed. The United States no longer saw the political left in Brazil as a threat and the quota system prevented the United States from punishing other Latin American countries (Bates, 1997).

When the ICA collapsed in 1989, market power shifted from the producing countries' coffee agencies to private trading companies. This was partly due to the shift in control over inventory stocks. While the ICA was functioning, inventory had mostly been held in producing countries. As a consequence of its breakdown however, inventory was moved to consuming countries. In the period after the breakdown of the ICA, prices dropped enormously. Up to 1989, the ICA had provided the producers of coffee with stable and minimum prices. After 1989, the futures (and options) markets of New York, for Arabica, and London, for Robusta, took off (Rutten and Youssef, 2007).

When we compare the indicator price of the International Coffee Organization (ICO) five years before the breakdown and five years after, we observe that prices dropped from \$1.34 per pound (1984-1988) to \$0.77 (1990-1994). This led to the bankruptcy of many public coffee agencies in producing countries. In 1993, the coffee producing countries established the Association of Coffee Producer Countries (ACPC) to try to regain control over the supply of coffee and export flows. Akiyama and Varangis (1990) provided an assessment of the impact of the ICA on producing countries, and found that the ICA contributed to a reduction of risk, but mainly benefited large producers in terms of returns. Local effects, however, differ significantly since there was evidence of oligopsony power in some producing countries (Lopez and You, 1993).

Since the breakdown of the ICA, a liberalization process has taken place in producing countries and not all major producing countries have been willing

to join the ACPC. The withdrawal of member countries from the ACPC in 1998/1999 plus the fact that Brazil exceeded its quota by six million bags of coffee resulted in the emergence of a new plan that came into operation in October 2000. This plan was intended to solve the problem of overproduction, which had already been a market feature for a couple of years. It was agreed by the member countries that if the ICO price indicator was below \$95 cents on a 15 day average basis, supply would be reduced by 20 per cent. As well as the member countries, non-member countries promised to cooperate with the plan. But the main weakness of the plan was that it did not include any provision for destroying stocks, so the problem of overproduction was not resolved. And the price indicator provided by the ICO fluctuated between \$0.69 in May 2000 and \$0.43-0.54 in the 2002 and 2003 years. In 2004, there was some recovery, with a price of \$0.62 (Daviron and Ponte, 2005). Meanwhile, with the rise of fair trade and eco labels, coffee growers have tried to benefit from consumers' willingness-to-pay for speciality coffees (Loureiro and Lotade, 2005). In terms of its magnitude, however, fair trade production is still relatively small.

With overall demand for coffee on the rise, interest has turned to production. Coffee supply is characterized by long lags (Wickens and Greenfield, 1973; Parikh, 1979), with cooperatives generally being less scale-efficient than investor-owned producers (Mosheim, 2002). With coffee production contributing significantly to local economies, Ninan and Sathyapalan (2005) found, for India, that biodiversity conservation has very high opportunity costs. Hein and Gatzweiler (2006), however, found that Arabica coffee production itself also benefits from the genetic diversity and resources contained in highland forests (in Ethiopia), providing a strong argument for the need to halt deforestation.

Wollni and Zeller (2007) found, for Costa Rica, that producers that have decided to supply the specialty coffee niche have been able to increase their prices; these producers are usually in cooperatives. In many cases, however, producers are not able to benefit fully from increases in world coffee prices, as described by Fafchamps and Hill (2005) and Fafchamps and Hill (2008). Based on data gathered in Uganda, Fafchamps and Hill showed how asymmetric information problems plague small producers, who as a result lose out to the rest of the supply chain: when international prices rose, growers received only a small fraction of the price increase.

In summary, for most of the past twenty years, coffee production has been characterized by a fairly large amount of competition. As coffee consumption rose, so did coffee production, albeit gradually. As yet largely unclear is the composition of international coffee prices. As a result, relatively little is known about the consequences for coffee growers of changes in these international

prices.

III. Theory

In this section, we model the price of coffee, using an approach that is akin to models of price discovery (Gansler, 2009), where the coffee price is the output of a process and demand, supply, demand and supply shifters and futures market equivalents all function as potential inputs.

This way of modeling has two advantages. First, it allows us to reformulate the debate regarding the possible impact of future market speculation on spot prices clearly as a question regarding omitted variables. What our modeling process amounts to is asking whether future market speculation is indeed an omitted variable in the coffee price generation process. The second advantage is that, in demonstrating that this process can be viewed in much the same way as a traditional production process, we can use an existing, non-parametric production approach to investigate the possible impact of futures market speculation without imposing much structure on the transformation function and while allowing for possible 'peak' effects.

In the rest of this section, we first introduce a basic model for the short-run price determination of coffee. Subsequently, we extend the model by introducing a futures market. After demonstrating how excluding relevant inputs into the coffee price generation process leads to increasing underestimation of the price, we then proceed with the development of an empirical framework to see whether and to what extent speculation has been part of this process. After proving that the results from our empirical analysis can indeed be used to test our model's key predictions, we introduce and discuss three hypothesis tests that will establish the role futures market speculation plays in pushing up coffee prices.

A. Basic model

To model the coffee price generation process, we build on the model of short-run price determination for produce commodities of Sexton and Zhang (1996). Sexton and Zhang (1996) are interested in the relative bargaining power of producers and retailers, in the context of highly perishable produce commodities.⁶ In our analysis, we do not - to all intents and purposes - distinguish between producers and retailers. Instead, we allow for storage of a commodity and the

⁶The empirical analysis in Sexton and Zhang (1996) focuses on the California iceberg lettuce market.

possibility, but by no means the certainty, that futures market developments affect short-run supply in the spot market. Except where noted otherwise, we closely follow Sexton and Zhang (1996), except in notation.⁷ The model developed here is typical for commodities, in that supply shocks are important drivers of the price, and demand is assumed to be relatively inelastic.

At any given time t , the total production of coffee is Q_t . The total production is assumed to be the result of past, long-run investment decisions, and as a result is independent of the current market price. At time t , supply is S_t .

The market value of a bag of coffee, P_t , is the result of supply and demand

$$P_t = R(S_t, \mathbf{D}_t) \quad (1)$$

where S_t is total coffee produced at t and shipped to the market at t , and \mathbf{D}_t is a vector of factors that can shift demand. Furthermore, by definition $\partial P / \partial S < 0$ and $\partial P / \partial \mathbf{D} > 0$.⁸

Since coffee can be stored, $S_t \leq Q_t$, and farmers can decide to sell at a later date. Farmers can hold a nonnegative amount of inventory I_t .⁹ The derived demand for coffee faced in the spot market is therefore determined by subtracting from the total demand all elements that can lower current supply, including (those affecting) inventory levels. Put differently, if producers face positive opportunity costs, O_t , to supply in the spot market, they reduce S_t . Per bag, these opportunity costs can be expressed as $O_t = O(S_t, \mathbf{W}_t)$, where \mathbf{W}_t represents a vector of factors that negatively affect current supply. It is assumed that $\partial O_t / \partial S_t \leq 0$. Inverse spot market demand can then be written as

$$P_t = R(S_t, \mathbf{D}_t) - O(S_t, \mathbf{W}_t) \quad (2)$$

Alternatively, in equilibrium, we can write spot market supply as

$$S_t = S(P_t, \mathbf{D}_t, \mathbf{W}_t) \quad (3)$$

In perfect competition, the price of a bag of coffee will equal the marginal cost of producing coffee C'_t . The intersection of $S(\cdot)$ and C'_t determines total long-run supply Q^* . Whenever $Q_t \geq Q^*$, $S_t = Q^*$ and, in the absence of scarcity, the spot price is determined solely by marginal cost, $P_t = C'_t$. Put differently, if there is enough supply (relative to demand), the world coffee price is determined by supply and demand and equals the (relatively stable)

⁷In Sexton and Zhang (1996), explicit attention is paid to the existence of multiple markets. Here, we abstract from this distinction without loss of generalization.

⁸The latter condition is by assumption, and we shall therefore include each element in \mathbf{D}_t either as such or as its inverse, to ensure $\partial P / \partial \mathbf{D} > 0$.

⁹Of course, with positive inventories it is also possible that supply (including sales of inventories) is bigger than production. This does not affect our model or our empirical analysis.

marginal cost C'_t .

As a result, in this case we can write the price of coffee as $(P_t \cdot \phi_t^D) + (P_t \cdot (1 - \phi_t^D))$, where

$$\phi_t^D = \frac{C'_t}{P} \quad (4)$$

measures the extent to which the price reflects marginal costs. In a perfectly competitive market, in the simplest case, $P_t = C'_t$, and $\phi_t^D = 1$. Otherwise, $\phi_t^D < 1$.

In practice, of course, demand and supply shifters can affect the price in the short run. Whether this is the case, depends on the realizations of Q_t , C'_t , \mathbf{D}_t and \mathbf{W}_t . If we let ρ_t denote the probability of $P_t = C'_t$, then

$$\begin{aligned} P_t &= C'_t && \text{with } prob = \rho_t, \\ P_t &= G(S_t, \mathbf{D}_t, \mathbf{W}_t) && \text{with } prob = 1 - \rho_t. \end{aligned} \quad (5)$$

So, the expected or fundamental price of coffee at time t can be expressed as

$$E(P_t) = \tilde{P}_t = \rho_t [C'_t] + (1 - \rho_t) [G(S_t, \mathbf{D}_t, \mathbf{W}_t)]. \quad (6)$$

If $Q_t \geq Q^* \forall Q_t$, then $P_t \equiv C'_t$, $\rho_t \equiv 1$, and supply and demand are the only factors influencing the price at any time. If, however, this is not the case, then ignoring situations where $\rho_t > 0$ leads to an underestimation of the price by $1 - \phi_t^{D,W}$, where

$$\phi_t^{D,W} = \frac{\rho_t [C'_t]}{\rho_t [C'_t] + (1 - \rho_t) [G(S_t, \mathbf{D}_t, \mathbf{W}_t)]} \quad (7)$$

and $0 \leq \phi_t^{D,W} \leq 1$.¹⁰

Intuitively, as long as there is abundant production ($Q_t \geq Q^*$), the price is not driven up from its 'fundamental' value, where it is determined by supply and demand, and where $P_t = C'_t$. If that is not the case, demand and - in particular - supply shocks can (temporarily) shift the demand and/or supply curve and thereby drive up the price, \tilde{P}_t .

B. Futures market

Old text:

¹⁰Note that we speak of an *underestimation*, since both \mathbf{D}_t and \mathbf{W}_t have been defined in such a way that an increase in either results in an increase in demand, respectively an decrease in supply, i.e., an increase in the price.

We now consider the existence of a futures market, where farmers can sell their products and guarantee future prices by engaging in a futures contract. We identify two types of traders on this futures market (CFTC, 2008). First, there are those that have an underlying position on the spot market, the so-called 'commercial' traders. Although these traders can of course both increase and decrease their price risk exposure, we shall refer to them as hedgers, in line with the literature. Also in line with the literature, the 'non-commercial' traders in this futures market will be referred to as speculators.

For $Q_t \geq Q^*$, the spot market price is again determined by (spot) supply and demand, because production is so high that demand and supply shocks (changes in \mathbf{D}_t and \mathbf{W}_t) cannot raise the price above marginal costs C'_t .

However, for $Q_t < Q^*$, there exists a per unit surplus, U_t :

$$\begin{aligned} U_t &= U \left(S_t, \mathbf{D}_t, \mathbf{W}_t, C'_t, \mathbf{D}_t^F, \mathbf{W}_t^F \right), \\ &= H \left(S_t, \mathbf{D}_t, \mathbf{W}_t, \mathbf{D}_t^F, \mathbf{W}_t^F \right) - C'_t. \end{aligned} \quad (8)$$

When $Q_t < Q^*$, $X_t > 0$. Given \mathbf{D}_t , \mathbf{W}_t , and C'_t , U_t is determined by the level of (spot) supply, S_t , and factors that shift demand \mathbf{D}_t^F and supply \mathbf{W}_t^F in the futures market. U_t is decreasing in S , for all $S \in [0, Q_t)$.

Following Deaton and Laroque (1992, 1996), given real interest rate r and deterioration rate δ , the expected value of a bag of coffee in the next period is $\left(\frac{1-\delta}{1+r}\right)$ times the expected next-period price. The best estimate of that price is assumed to be the futures price, F_{t+n} , where $n \geq 1$. Without loss of generalization, we assume here that $n = 1$. Since, by definition, $\partial F_{t+n} / \partial \mathbf{D}_t^F > 0$ and $\partial F_{t+n} / \partial \mathbf{W}_t^F > 0$, U_t is non-decreasing in \mathbf{D}_t^F and \mathbf{W}_t^F . Hence, demand and supply shifts in the futures market, besides pushing up the futures price, can reflect future scarcity, which *may* drive up spot prices as well.¹¹

As a result, $H(\cdot) \geq G(\cdot)$ and the spot price cannot easily be determined. Put simply, the surplus can be earned both on the spot and the futures market. As S_t is decreased, and more is sold on the futures market, the resulting scarcity on the spot market raises the surplus U_t .¹²

Therefore, as Q_t decreases, this puts pressure on the price in two ways: the available surplus increases, and - since the commodity can be stored and thereby future supply becomes less certain - the futures price is more likely

¹¹The latter can result when the opportunity cost of selling on the spot market is positive, i.e., if $P < \left(\frac{1-\delta}{1+r}\right) F_{t+n}$.

¹²Keep in mind, however, that the extent to which the farmer can benefit in this manner depends on the total amount produced; the more coffee is harvested, the less credible is a reduction of spot supply. In this context, think of the impact of competition among farmers: as total available harvest per farmer increases, committing to collusion becomes less credible.

to increase, and the farmer has more reason to sell on the futures market. As long as $Q_t \geq Q^*$, it is still the marginal cost C'_t that determines the price, \tilde{P}_t . For $Q_t < Q^*$, the effect now depends on the share, γ_t , of the available surplus U_t that is earned on the spot market, where γ_t lies within the unit interval and of course $\partial\gamma/\partial S_t < 0$.

In addition, as is the case in the spot market, factors that shift demand \mathbf{D}_t^F and supply \mathbf{W}_t^F in the futures market drive up the futures price, and thereby $\partial\gamma_t/\partial\mathbf{D}_t^F > 0$ and $\partial\gamma_t/\partial\mathbf{W}_t^F > 0$, under the assumption that $\partial F_{t+n}/\partial\mathbf{D}_t^F > 0$ and $\partial F_{t+n}/\partial\mathbf{W}_t^F > 0$, as was the case for the spot market.¹³ Hence, $\gamma_t = \gamma(S_t, \mathbf{D}_t^F, \mathbf{W}_t^F)$, where the latter include demand and supply shifters for the hedgers.

We can therefore now write¹⁴

$$\begin{aligned} P_t &= C'_t && \text{with } prob = \rho_t \\ P_t &= C'_t + \gamma(\cdot) [H(\cdot) - C'_t] && \text{with } prob = 1 - \rho_t, \end{aligned} \quad (9)$$

where $\gamma(\cdot) = \gamma(S_t, \mathbf{D}_t^F, \mathbf{W}_t^F)$ and $H(\cdot) = H(S_t, \mathbf{D}_t, \mathbf{W}_t, \mathbf{D}_t^F, \mathbf{W}_t^F)$.

Consequently, \tilde{P}_t can now be expressed as

$$E(P_t) = \rho_t [C'_t] + (1 - \rho_t) [C'_t + \gamma(\cdot) [H(\cdot) - C'_t]]$$

where $\gamma(\cdot) = \gamma(S_t, \mathbf{D}_t^F, \mathbf{W}_t^F)$ and $H(\cdot) = H(S_t, \mathbf{D}_t, \mathbf{W}_t, \mathbf{D}_t^F, \mathbf{W}_t^F)$.

From our description of the spot market price setting process above, it is clear that - in the presence of storable commodities - the existence of futures market demand and supply shocks can increase the likelihood of a price increase, by driving up $H(\cdot)$ and γ . Therefore, ignoring situations where $\rho_t > 0$ and $\gamma_t > 0$ leads to an underestimation of the price by $1 - \phi_t^{\mathbf{D}, \mathbf{W}, \mathbf{F}}$, where

$$\phi_t^{\mathbf{D}, \mathbf{W}, \mathbf{F}} = \frac{\rho_t [C'_t]}{\rho_t [C'_t] + (1 - \rho_t) [C'_t + \gamma(\cdot) [H(\cdot) - C'_t]]} \quad (10)$$

and $\mathbf{F}_t = (\mathbf{D}_t^F, \mathbf{W}_t^F)$.

Put differently, in the presence of a futures market, there is further potential for an increase in the spot price, although we by no means assume that the latter is the case.

As shown in Section II, the debate regarding rising commodity prices does not focus on the commercial futures market positions referred to above, but rather on the non-commercial, or speculative positions, i.e., relating to those parties that have no underlying position in the spot market. As a final ex-

¹³Note that we assume that r and δ are stable and thereby do not affect γ_t .

¹⁴In their analysis, Sexton and Zhang (1996) parametrize γ , which in their case represents the farmers' share.

tension, therefore, we include \mathbf{D}_t^V and \mathbf{W}_t^V , the demand and supply shifters for these non-commercials. Wrongly excluding this second set results in an underestimation of the price by $1 - \phi_t^{\mathbf{D}, \mathbf{W}, \mathbf{F}, \mathbf{V}}$, where $\mathbf{V}_t = (\mathbf{D}_t^V, \mathbf{W}_t^V)$,

$$\phi_t^{\mathbf{D}, \mathbf{W}, \mathbf{F}, \mathbf{V}} = \frac{\rho_t [C_t']}{\rho_t [C_t'] + (1 - \rho_t) [C_t' + \gamma(\cdot) [\tilde{H}(\cdot) - C_t']]} \quad (11)$$

and $\tilde{H}(\cdot) = \tilde{H}(S_t, \mathbf{D}_t, \mathbf{W}_t, \mathbf{D}_t^F, \mathbf{W}_t^F, \mathbf{D}_t^V, \mathbf{W}_t^V)$. Since, as before, U_t is non-decreasing in \mathbf{D}_t^V and \mathbf{W}_t^V , $\tilde{H}(\cdot) \geq H(\cdot)$.

Summing up, we can conclude from our analysis thus far that

$$1 \geq \phi_t^{\mathbf{D}, \mathbf{W}, \mathbf{F}, \mathbf{V}} \geq \phi_t^{\mathbf{D}, \mathbf{W}, \mathbf{F}} \geq \phi_t^{\mathbf{D}, \mathbf{W}} \geq \phi_t^{\mathbf{D}} \geq 0. \quad (12)$$

As a result, if demand and supply suffice to explain the price of coffee, $\phi_t^{\mathbf{D}, \mathbf{W}, \mathbf{F}, \mathbf{V}} = \phi_t^{\mathbf{D}, \mathbf{W}, \mathbf{F}} = \phi_t^{\mathbf{D}, \mathbf{W}} = \phi_t^{\mathbf{D}} = 1$, and there is no underestimation of the price. On the other hand, if speculative positions on the futures market are an essential input in the spot price generation process, then $\phi_t^{\mathbf{D}, \mathbf{W}, \mathbf{F}, \mathbf{V}} > \phi_t^{\mathbf{D}, \mathbf{W}, \mathbf{F}}$, and *not* including these positions leads to an underestimation of the price. Put differently, futures market speculative positions then inflate the coffee price by $(\phi_t^{\mathbf{D}, \mathbf{W}, \mathbf{F}, \mathbf{V}} - \phi_t^{\mathbf{D}, \mathbf{W}, \mathbf{F}}) \times \text{price}$.

C. Empirical framework

Implicit in the description of the coffee price production process so far is the idea that the market *transforms* the inputs (demand, supply, etc.) into the output, i.e., the coffee price. The actual process through which the price is then established is one of *price discovery* (Gansler, 2009). And the coffee price is, eventually, established on the coffee market, the market on which physical quantities of the commodity are bought and sold.

Assuming we can identify the inputs that are relevant to the market, we can start to identify the production possibility set (PPS) within which the market operates (Thanassoulis et al., 2008). As the PPS contains all feasible combinations of input and output vectors, the location within the PPS tells us the extent to which the inputs, at a given point in time, 'produce' the price.

Following Thanassoulis et al. (2008), we can describe the PPS T as:

$$T = \{(x, y) \in \mathfrak{R}_+^{m+s} | x \text{ can produce } y\}, \quad (13)$$

where T contains all feasible combinations of inputs $x \in \mathfrak{R}_+^m$ and output $y \in \mathfrak{R}_+^s$, and both are assumed to be non-negative. In addition, to ensure that an increase in an input x leads to an increase in output y , we include those

'inputs' that otherwise would decrease output by their inverse values. Hence, for example, we include supply as $1/S$.

Naturally, what we are interested in is the set of *efficient* combinations, so we can measure the inefficiency of the other combinations. Therefore, adding time subscripts t , we can use the following basic DEA model (Charnes et al., 1978, 1985):

$$\begin{aligned}
 & \max_{\theta, \lambda} \theta_t \\
 & \text{Subject to } -\theta_t y_t + Y\lambda \geq 0 \\
 & \quad x_t - X\lambda \geq 0 \\
 & \quad N1'\lambda = 1 \\
 & \quad \lambda \geq 1,
 \end{aligned} \tag{14}$$

where $N1$ is a vector of ones, $1 < \theta_t < \infty$, and $\theta - 1$ is the proportional increase in the coffee price (the output y) that can be produced at time t with input quantities (the market variables) held constant. Inefficiency, χ_t , is defined as $1 - \left(\frac{1}{\theta_t}\right)$, and varies between zero and one. Inefficiency is measured against a convex hull of intersecting planes which tightly envelope the data. Since the weights λ can sum to a value greater than or equal to one, the model is characterized by a 'variable returns to scale' (VRS) technology, which ensures that an inefficient combination of inputs and output is benchmarked against combinations of similar magnitude.

In line with the theoretical framework, we can now use the model defined by equation (14) to test whether speculative positions in the futures market are an essential component of the input set that produces the coffee price. First, we consider $\{\mathbf{D}\}$, an input set that consists purely of demand (D) and inverse supply ($1/S$), and we estimate $\chi_t^{\mathbf{D}}$. Second, we consider $\{\mathbf{D}, \mathbf{W}\}$, an input set that consists of demand, supply and a set of variables describing supply shifters (W). With this input set we estimate $\chi_t^{\mathbf{D}, \mathbf{W}}$. Third, we consider $\{\mathbf{D}, \mathbf{W}, \mathbf{F}\}$, an input set that consists of demand, supply and market condition variables, as well as the long and inverse short position of commercial futures market participants, the hedgers (F). This input set allows us to estimate $\chi_t^{\mathbf{D}, \mathbf{W}, \mathbf{F}}$. Fourth, we consider $\{\mathbf{D}, \mathbf{W}, \mathbf{F}, \mathbf{V}\}$, an extension of $\{\mathbf{D}, \mathbf{W}, \mathbf{F}\}$ to include the long and inverse short position of index funds, the speculators (V). Now we can estimate $\chi_t^{\mathbf{D}, \mathbf{W}, \mathbf{F}, \mathbf{V}}$, the inefficiency that is present in the price generation process once all inputs describe above have been included.

In line with our theoretical predictions in equation (12), we can now write

$$0 \leq \chi_t^{\mathbf{D}, \mathbf{W}, \mathbf{F}, \mathbf{V}} \leq \chi_t^{\mathbf{D}, \mathbf{W}, \mathbf{F}} \leq \chi_t^{\mathbf{D}, \mathbf{W}} \leq \chi_t^{\mathbf{D}} \leq 1. \tag{15}$$

Finally, we consider whether our coffee price production model satisfies the usual properties of a production function (Chambers, 1988). First, each of our inputs satisfies a non-negativity constraint, following from the definition of our inputs. Second, weak essentiality requires that the production of output is impossible without employing at least one input, which follows from our inclusion of supply as an input. Third, output (i.e., the coffee price) should be non-decreasing in each input, in the absence of any congestion problems. In fact, we have ensured the monotonicity of our production function by defining each of our inputs x such that if $\mathbf{x}^0 \geq \mathbf{x}^1$, then $f(\mathbf{x}^0) \geq f(\mathbf{x}^1)$. The most difficult property to satisfy in the context of the coffee price production model is concavity, which requires that all marginal products are non-decreasing. Therefore, in Section V.C, we pay special attention to robustness tests related to this property. For now, we conclude that our model appears to satisfy most of the standard properties and proceed to relate the empirical DEA model to the theoretical framework.

D. Relation between model and empirical framework

To see why the equation (15) must hold, we note here that we can write

$$\chi_t^{\mathbf{D},\mathbf{W},\mathbf{F}} = 1 - \frac{\hat{P}_t^{\mathbf{D},\mathbf{W},\mathbf{F}}}{P_t}, \quad (16)$$

where $\hat{P}_t^{\mathbf{D},\mathbf{W},\mathbf{F}}$ is the predicted ('inefficient') price from employing the model using demand, supply, demand and supply shifters, as well as hedgers' positions, and P_t is the actual price of coffee at time t . Likewise, we can write

$$\chi_t^{\mathbf{D},\mathbf{W},\mathbf{F},\mathbf{V}} = 1 - \frac{\hat{P}_t^{\mathbf{D},\mathbf{W},\mathbf{F},\mathbf{V}}}{P_t}, \quad (17)$$

where $\hat{P}_t^{\mathbf{D},\mathbf{W},\mathbf{F},\mathbf{V}}$ is the predicted price from employing the model using all inputs, including speculative futures market demand and supply.

In our theoretical model, we derived the *efficiency* of the pricing process, denoted as ϕ . Now we can also rewrite $\phi_t^{\mathbf{D},\mathbf{W},\mathbf{F}}$ as

$$\phi_t^{\mathbf{D},\mathbf{W},\mathbf{F}} = \frac{\rho_t \hat{P}_t^{\mathbf{D}}}{\hat{P}_t^{\mathbf{D},\mathbf{W},\mathbf{F}}} = \frac{\rho_t (1 - \chi_t^{\mathbf{D}}) P_t}{(1 - \chi_t^{\mathbf{D},\mathbf{W},\mathbf{F}}) P_t}, \quad (18)$$

and we can write $\phi_t^{\mathbf{D},\mathbf{W},\mathbf{F},\mathbf{V}}$ as

$$\phi_t^{\mathbf{D},\mathbf{W},\mathbf{F},\mathbf{V}} = \frac{\rho_t \hat{P}_t^{\mathbf{D}}}{\hat{P}_t^{\mathbf{D},\mathbf{W},\mathbf{F},\mathbf{V}}} = \frac{\rho_t (1 - \chi_t^{\mathbf{D}}) P_t}{(1 - \chi_t^{\mathbf{D},\mathbf{W},\mathbf{F},\mathbf{V}}) P_t}. \quad (19)$$

So for $\phi_t^{\mathbf{D},\mathbf{W},\mathbf{F}} \leq \phi_t^{\mathbf{D},\mathbf{W},\mathbf{F},\mathbf{V}}$, we can write

$$\rho_t \frac{(1 - \chi_t^{\mathbf{D}}) P_t}{(1 - \chi_t^{\mathbf{D},\mathbf{W},\mathbf{F}}) P_t} \leq \rho_t \frac{(1 - \chi_t^{\mathbf{D}}) P_t}{(1 - \chi_t^{\mathbf{D},\mathbf{W},\mathbf{F},\mathbf{V}}) P_t}, \quad (20)$$

or

$$(1 - \chi_t^{\mathbf{D},\mathbf{W},\mathbf{F}}) P_t \geq (1 - \chi_t^{\mathbf{D},\mathbf{W},\mathbf{F},\mathbf{V}}) P_t. \quad (21)$$

Therefore testing whether $\phi_t^{\mathbf{D},\mathbf{W},\mathbf{F}} \leq \phi_t^{\mathbf{D},\mathbf{W},\mathbf{F},\mathbf{V}}$ amounts to testing whether

$$\chi_t^{\mathbf{D},\mathbf{W},\mathbf{F}} \geq \chi_t^{\mathbf{D},\mathbf{W},\mathbf{F},\mathbf{V}}. \quad (22)$$

The same logic applies to the other inefficiency scores.

E. Hypotheses

Formally, in case index fund speculation is *not* part of the coffee price generation process, we can formulate three hypotheses to test whether or not index fund speculation has contributed to the coffee price. The first hypothesis can be formulated as follows:

Hypothesis 1

$$H_0^1: \chi_t^{\mathbf{D}}, \chi_t^{\mathbf{D},\mathbf{W}}, \chi_t^{\mathbf{D},\mathbf{W},\mathbf{F}} = 0;$$

$$H_a^1: \chi_t^{\mathbf{D}}, \chi_t^{\mathbf{D},\mathbf{W}}, \chi_t^{\mathbf{D},\mathbf{W},\mathbf{F}} > 0.$$

According to this hypothesis, the futures market positions of speculators are *not* part of the coffee price generation process. As a result, there is no inefficiency left once we have selected the correct input set ($\{\mathbf{D}\}$, $\{\mathbf{D}, \mathbf{W}\}$ or $\{\mathbf{D}, \mathbf{W}, \mathbf{F}\}$).

Of course, there may be other, possibly omitted variables that affect the coffee price. We therefore also test the following hypothesis:

Hypothesis 2

$$H_0^2: \chi_t^{\mathbf{D},\mathbf{W},\mathbf{F}} - \chi_t^{\mathbf{D},\mathbf{W},\mathbf{F},\mathbf{V}} = 0;$$

$$H_a^2: \chi_t^{\mathbf{D},\mathbf{W},\mathbf{F}} - \chi_t^{\mathbf{D},\mathbf{W},\mathbf{F},\mathbf{V}} > 0.$$

According to this hypothesis, adding the long and inverse short position of speculators to the coffee price input set does not result in a decrease in inefficiency. Rejection of this hypothesis would indicate that, at least sometimes, speculators' positions explain part of the coffee price.

Clearly, rejection of hypotheses 1 and 2 is evidence in favor of speculators playing a role in the coffee price generation process. Conditional on this rejection, we propose a third and more restrictive hypothesis:

Hypothesis 3

$$H_0^3: \chi_t^{\text{D,W,F,V}} > 0;$$

$$H_a^3: \chi_t^{\text{D,W,F,V}} = 0.$$

Rejection of this third hypothesis, in addition to rejection of hypotheses 1 and 2, means that once we include speculators' future market positions, we have fully accounted for any inefficiencies in the coffee price generation process. Therefore, rejection of hypothesis 3 provides very strong evidence that index fund speculation is part of the coffee price generation process.

Before we discuss our data, we address one last important consideration regarding our hypothesis tests. As our discussion of the literature and the coffee market has shown, *if* index fund speculation has had an impact on the coffee price, there is no reason to assume that this impact can be observed *on average*. After all, the discussion so far has pointed out (a) that there may be many other aspects (possibly covered by our other inputs) that can cause coffee price jumps, and (b) it may be the case that the effect of index fund speculation on the coffee price is part of a complex interaction of factors, including possible price impact, inventory changes, and substitution effects with other commodities.

When testing our hypotheses, we shall take this into account in several ways. First, in addition to testing whether each hypothesis can be rejected on average, we provide more detailed, distribution-based tests. Second, in addition to tests of the three hypotheses outlined above, we shall relate our results to the developments in the coffee market, both in a formal and in a more descriptive manner. Third, we shall discuss a series of robustness tests aimed at challenging some of the assumptions made so far.

IV. Data

Our analysis focuses on explaining the price of Arabica coffee. Arabica is the prime, broadly traded coffee bean. Its producers try to sell at a relatively large margin, yet it is also a necessary ingredient in blends. Unfortunately for

the purpose of our research, there is no 'world' spot price of coffee: coffee is traded spot on many markets, and available data on spot prices are scarce and difficult to compare as they may refer to what is charged by either producers or retailers (Algra, 2009).

To measure the spot price of coffee, we therefore follow the literature (Irwin and Sanders, 2010), and rely on the fact that the futures price converges to the spot price as the futures contract reaches its expiration date (Gorton et al., 2007). Hence, we can closely approximate the spot price by tracking the price of Arabica at the New York futures market. Here, one to three month futures are traded on coffee to be delivered in one month to three years time. Futures are quoted in US dollar cents per pound, and each contract applies to 37,500 pounds of Arabica coffee (ICE, 2009). When the futures contract is very close to expiration, it occasionally shows extreme fluctuations (Gorton et al., 2007). We therefore make use of the strong convexity of the convergence path (Gorton et al., 2007) and measure the coffee price as the average of the price of the nearest contract.¹⁵ Over the sample period 1989 - 2008, the average monthly price was a little over 102 dollar cents per pound. The price fluctuated between less than 45 and more than 260 dollar cents per pound. Figure 2(a) shows that following a decline from 1998 up to 2002, coffee prices have since then showed a steady rise.

The inputs in $\{\mathbf{D}\}$ are the demand and inverse supply of coffee. Annual data on coffee production are provided by the Foreign Agricultural Service of the United States Department of Agriculture (FAS-USDA). Production is expressed in thousands of bags of 60kg and over the sample period, Arabica production ranges between 61,389 and 84,903 production units. The production of Arabica shows a steady increase over the course of our sample period (Foreign Agricultural Service of the United States Department of Agriculture, 2009).¹⁶ Since most coffee consumption consists of blended coffee, we also include the production of Robusta coffee beans. Data on Robusta production are also supplied by the FAS-USDA. Robusta production ranges between 27,146 and 49,117 units over the course of our sample period, with an even steeper increase over time.¹⁷ As most coffee is consumed as a blend, coffee consumption is recorded for Arabica and Robusta jointly.¹⁸ Consumption is also expressed

¹⁵As a robustness test, we also test using the last price on the second-nearest contract, since the latter is less susceptible to these fluctuations (see Table 4).

¹⁶We observe that these numbers may be underestimated, since they are provided by the International Coffee Organization (ICO), which may have an interest in underestimating production in order to put upward pressure on the price of coffee imports.

¹⁷In order to match data on prices, production and consumption, we have also run our empirical analysis after converting production numbers to a monthly frequency, by simple imputation. Results are robust.

¹⁸Again, there may be bias in these numbers, as the International Coffee Organization (ICO) may have an interest in overestimating consumption to drive prices upward.

in thousands of bags (of 60kg) and ranges between 90,500 and 128,000 units (Osorio, 2009). Figures 2(a) and 2(c) show that, in 2004/2005 and in 2006/2007, prices increased despite the fact that supply was higher than demand.

Now we turn to the extended input set $\{\mathbf{D}, \mathbf{W}\}$, which includes \mathbf{W} , a number of supply shifters. The first two are the inventories of coffee importing countries (GCA, 2009b,c,a) and the inventories of coffee exporting countries (FAS-USDA, 2009). Changes in these inventories may help clear the market. Indeed, as expected, Figure 3d shows that reductions in local inventories of coffee exporting countries have coincided with the increases in coffee prices in the last decade. In the 1990s, however, reductions in these inventories did not coincide with coffee price increases. In fact, after the collapse of the International Coffee Agreement (ICA) and the ending of the quota system in 1989, inventories of coffee were held by importing countries instead of producing countries, as shown in Figure 2(d). This diminished the ability of coffee exporting countries to drive up prices.¹⁹

The next supply shifter is the occurrence of frost and droughts (Dijk et al., 1998). During our sample period, coffee production was seriously affected by both. In 1994, production dropped due to frost in Brazil, the largest coffee producer, as the frost caused serious damage to coffee trees.²⁰ Our sample period also includes two periods of drought, in 1994 and 1999 (see Figure 2(c)). The ICO estimated that 40 percent of these years' crops in Brazil was lost because of the drought (International Coffee Organization, 2008).

Since our sample period covers 20 years, supply shifters must include factors that affect productivity. Here, we include four such factors. First, we include the yield per hectare in hectograms per hectare (for both Arabica and Robusta, combined), as provided by the statistical bureau of the Food and Agriculture Organization of the United Nations (FAOSTAT) (Food and Agriculture Organization of the United Nations, 2007).²¹ From the same source, we then also use the crop productivity and a broad measure of agricultural productivity. For each measure, we use the production weighted average of the 17 largest

¹⁹In addition, the quantity of coffee produced decreased, which reduced the inventories and put an upward pressure on the coffee price as well. In 1997, large roasters launched a new system of supply, called supplier managed inventories. This meant that they reduced their inventories to cut costs and outsourced the stock management to the large traders, thus increasing stock mobility and putting downward pressure on prices. Due to supply problems in countries such as Colombia, Indonesia and Cote d'Ivoire and the late harvest in Brazil that year, the roasters were not be able to replenish their inventories timely, which led to acute shortages in the coffee market and triggered panic buying.

²⁰The fact that Robusta is less vulnerable to frost was in the reason for the introduction of Robusta in Brazil after the frost period of 1976 (International Coffee Organization, 2008).

²¹As the data provided by FAOSTAT only run until 2007, we use the numbers for 2007 for 2008 as well.

Table 1: Descriptive statistics⁽¹⁾

Variable	Description	Source	Note 1/x	Mean	Std. Dev.	
☞ Coffee price	Arabica beans, bag of 60kg	New York market	Futures (2)	n.a.	102	39
☐ Production of Arabica beans	1,000 bags of 60kg, monthly	FAS-USDA (2009)	yes	70,728	6,035	
☐ Production of Robusta beans	1,000 bags of 60kg, monthly	FAS-USDA (2009)	yes	36,316	7,794	
☐ World coffee consumption	1,000 bags of 60kg, monthly	Osorio (2009)	no	106,720	11,344	
☐ Inventory of importers	1,000 bags of 60kg, annual	GCA (2009b,c,a)	no	5,142	2,039	
☐ Inventory of exporters	1,000 bags of 60kg, annual	FAS-USDA (2009)	no	31,067	9,667	
☐ Frosts and droughts	1994 and 1999	ICO (2008)	(7)	n.a.	n.a.	
☐ Yield per hectare	hectogram per hectare, annual	FAOSTAT (2007)	(4)	yes	6,506	821
☐ Crop productivity	weighted average, annual	FAOSTAT (2007)	(6)	yes	101	6
☐ Agricultural productivity	weighted average, annual	FAOSTAT (2007)	yes	100	8	
☐ Price of fertilizer	monthly	Thomson Datastream	no	124	34	
☐ Dollar exchange rate	against a currency basket	Thomson Datastream	(5)	no	93	10
☐ Trading volume	trading volume of commercials	CFTC (2010)	(3)	no	11,373	5,433
☐ Open interest and commitment	open interest of commercials	CFTC (2010)	no	63,668	38,136	
☐ Speculative long position	long position of non-commercials	CFTC (2010)	no	15,604	1,446	
☐ Speculative short position	short position of non-commercials	CFTC (2010)	yes	10,562	9,797	

Notes: (1) = 240 observations, from January 1989 until December 2008; (2) = per bag of 60kg; (3) = first expiring future; (4) = hectare =2.471 acres, as the data provided by FAOSTAT only run until 2007, we use the numbers for 2007 for 2008 as well, for both Arabica and Robusta, combined; (5) = the Thomson Datastream USD Index measures the performance of the US Dollar against a basket of currencies: EUR, JPY, GBP, CAD, CHF and SEK; (6) = index covering the 17 largest coffee producing countries, with weights based on share in coffee production, base year 1999=100; (7) = Frost and droughts are included with a value of 2 (1) in the period in which they (do not) occur.

coffee producing countries. During our sample period, productivity mostly declined, putting pressure on prices. As a final measure, we include the price of fertilizer, as the latter constitutes a large part of the production costs of coffee are determined by the price of fertilizer. Over time, fertilizer became more expensive, thus pushing up prices. A recent study by the International Coffee Organization (2009) states that in the 2007 and 2008, fertilizer represented between 16.2% and 23.2% of the total costs in Brazil, 23.7% in Columbia and 12.3% in Costa Rica.²²

Most coffee is traded in US dollars. As a result, exchange rate fluctuations can influence the price setting process, both directly and through their effect on inventories. We include an exchange rate index of the US currency against a basket of currencies taken from Thomson Datastream.²³

Next, we turn to the input set $\{\mathbf{D}, \mathbf{W}, \mathbf{F}\}$, which includes demand and supply from so-called 'commercial' traders. Over our sample period, the market for coffee futures has become increasingly popular among all kinds of traders (Crown, 1997). We measure the total trade volume, the number of futures traded in the market corresponding to the nearest expiring future (i.e., the future we use to measure the spot price of coffee). If the trade volume increases, the liquidity in the market will increase and so the accuracy of the prediction of the price will increase (CFTC, 2008). Also, we employ a more direct measure of liquidity in the futures market, in the form of the open interest and commitment of traders. Open interest relates to the number of outstanding contracts in the futures market corresponding with the price of the futures used in our dataset.²⁴ Both variables show a strong increase, in particular in the second half of our sample period.

Finally, we turn to input set $\{\mathbf{D}, \mathbf{W}, \mathbf{F}, \mathbf{V}\}$, including speculative positions. The position held by noncommercial parties in the coffee futures market is significant. In 1997, Crown (1997) estimates that the market share of speculators was around 70 percent. To account for the position of speculators we include two variables. First, we include the speculative long position, measured as the long positions held by non-commercial market participants (CFTC, 2008). Second, we include the inverse of the speculative short position, to ensure monotonicity for this input. The net position of speculators is shown in Figure 2(e).

²²The price of fertilizer has a base date of November 1 1950, and is published monthly by Thomson Datastream. An important ingredient in fertilizers is oil. In addition, oil is also included in most commodity indexes, together with coffee. These commodity indexes are used by speculators as a major guideline for determining their investment strategy in commodity futures markets. We also ran our analyses including the price of oil as an additional input variable in the pricing process. Results do not change.

²³Alternatively, we considered the exchange rate of the US dollar against the Brazilian Real. However, since the Real was only introduced again in 1994, it cannot be used for our entire sample period.

²⁴These positions can be taken by both commercial and non-commercial parties.

This shows that long positions exceeded short positions, as expected given the presence of (long-only) index funds, most of the time, but not always. Also, net positions increased significantly after 2000.

It is difficult to address the full impact of noncommercial traders. The positions of the different types of traders, published by the CFTC, do not necessarily reflect the true positions of traders, speculators and index speculators. Some Wall Street banks are exempted from the CFTC's speculative position limits. So when noncommercial traders enter into a commodity index swap with these Wall Street banks, they bypass speculative position limits. In this way they have unlimited access to commodities futures markets. It is said that index decorators use this mechanism in 85-90% of their transactions. Also, noncommercial traders are sometimes wrongly categorized in the CFTC's classification scheme (Masters, 2008). In our robustness section, we shall therefore return to this issue and use a data set that is available for a much smaller sample period to identify more precisely different types of speculators.

Summing up, we have now defined our four input sets. Each of the input sets is of course nested in the next. In the next section, we return to our empirical results.

V. Results

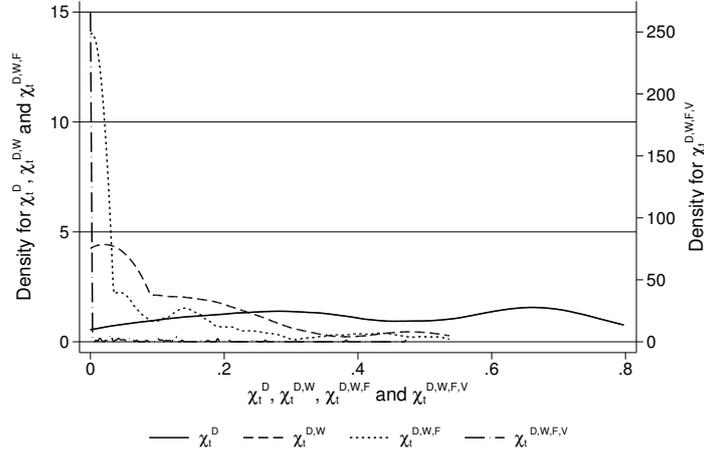
First, we describe our results. This is followed by a description of our hypothesis tests. Next, we provide a more in-depth analysis of the elements that have contributed to rising coffee prices. Finally, we describe an extensive series of robustness tests aimed at validating our results.

A. Estimation results

Figure 1 shows the inefficiency distributions. From the figure, we observe that the inefficiency distributions become more and more skewed as we add inputs, with lower average inefficiency scores.

Table 2 contains the inefficiency scores resulting from estimating our model with each of the four input sets, respectively. A number of observations can be made from the table. First, and as expected, the mean inefficiency drops as we move from the basic input set containing only demand and supply to the most elaborate input set, including speculative long and short positions. One important reason for this, to which we will return in section C, why this is the case concerns the issue of dimensionality in DEA models: the more variables

Figure 1: Distributions of inefficiency scores



are added, the more likely it is that an observation will be efficient. However, it is worth noting that dimensionality is unlikely to play a role here since the total number of observations (240) greatly outnumbers the total number of variables, even in the most elaborate case (with 15 inputs). We observe that supply-side developments, as captured by W , result in a marked decrease in inefficiency of more than 30%. The inclusion of commercial traders' (hedgers') positions further reduces inefficiency by 4.5%. Speculative positions contribute by slicing another 4.4% off overall inefficiency, on average.

Table 2: Descriptive statistics

inefficiency (input sets)	mean	std. dev.	min.	max.
χ_t^D (supply and demand)	0.417	0.236	0.000	0.797
$\chi_t^{D,W}$ (+ supply and demand shifters)	0.107	0.133	0.000	0.537
$\chi_t^{D,W,F}$ (+ hedgers' positions)	0.062	0.117	0.000	0.537
$\chi_t^{D,W,F,V}$ (+ speculators' positions)	0.024	0.067	0.000	0.472

Second, we observe a similar pattern when we look at the standard deviation of inefficiency. The standard deviation is reduced by almost 44%, when we introduce supply and demand shifters. Subsequently introducing commercial futures market positions only leads to a further reduction by 12%. However, introducing speculative positions lowers the standard deviation of inefficiency

further by almost 43%.

Finally, although the minimum inefficiency is equal to zero in all cases, the maximum inefficiency drops as we add inputs. Interestingly, adding the futures market positions of commercial participants does not lower the maximum inefficiency.

B. Hypothesis tests

Table 3 gives an overview of our main hypothesis tests. As explained in Section III.E, we shall test each of our hypotheses in a number of ways. First, we use a standard t-test, assuming unequal variances.²⁵ The t-tests provide the most direct tests of our hypotheses. Unfortunately, the t-test is most effective in case of normally distributed variables, something that obviously does not apply here as is clear from Figure 1.

Table 3: Hypothesis tests

Hypothesis		t-test	KW	KS
$H_0^1: \chi_t^D = 0$	(demand and supply)	0.0000	0.0001	0.0000
$H_0^1: \chi_t^{D,W} = 0$	(+ demand and supply shifters)	0.0000	0.0001	0.0000
$H_0^1: \chi_t^{D,W,F} = 0$	(+ hedgers' positions)	0.0000	0.0001	0.0000
$H_0^2: \chi_t^{D,W,F} - \chi_t^{D,W,F,V} = 0$	(+ speculators' positions)	0.0000	0.0001	0.0000
$H_0^3: \chi_t^{D,W,F,V} > 0$	(+ speculators' positions)	0.0000	0.0001	0.0000

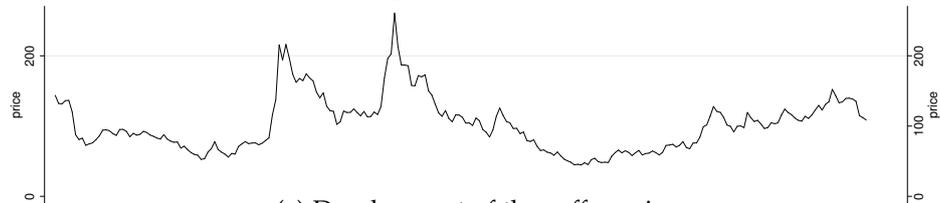
Notes: KW= Kruskal-Wallis rank test; KS = Kolmogorov-Smirnov test.

Therefore, we also avail of two alternative tests. To test for differences in efficiency ranks, we apply a Kruskal-Wallis rank test. To operationalize this test, we generate counterfactual inefficiency distributions as applicable. For example in testing Hypothesis 1, we generate a series of inefficiency scores that are always equal to zero. To compare distributions, and test for significant differences among distributions, we make use of a Kolmogorov-Smirnov test and the counterfactual inefficiency scores we generated for the rank test (Fiorio, 2004).

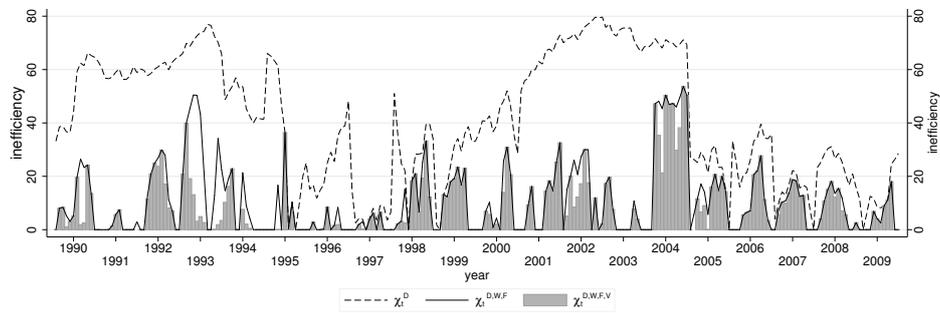
From Table 3, we observe that we can reject the null for the first hypothesis: for each of the first three input sets, the inefficiency that remains in our model

²⁵We use a Welch correction.

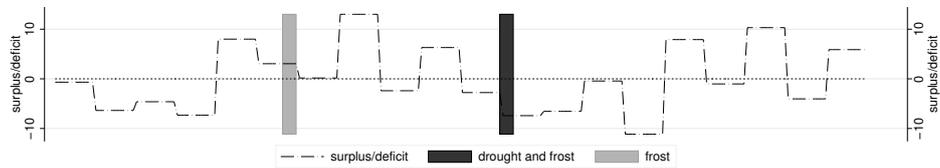
Figure 2: Results



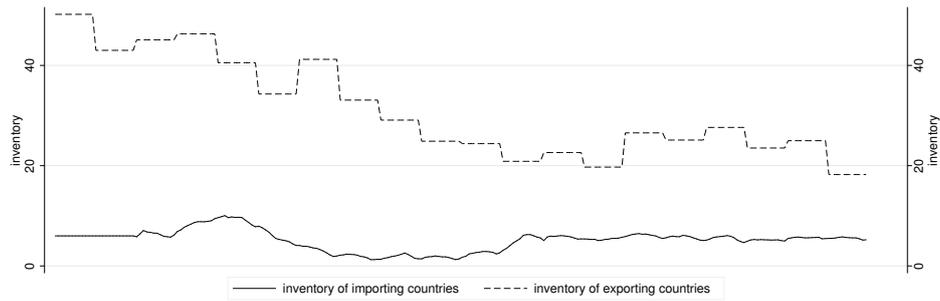
(a) Development of the coffee price



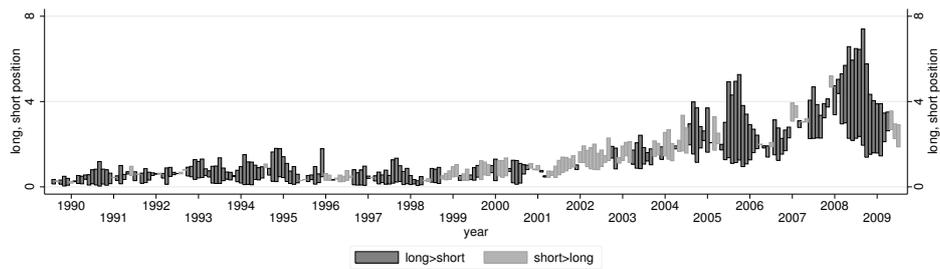
(b) Explaining the coffee price



(c) Surplus/deficit supply



(d) Inventories of coffee



(e) Long and short positions by speculators

differs significantly from zero. The null for the second hypothesis is also rejected: the decrease in efficiency that results from adding speculators' long and short positions to the input set differs significantly from zero. Finally, the null for the third hypothesis is also rejected: the inefficiency that remains once we add speculators' positions is not significantly different from zero.

The strength of our results is reflected graphically in Figure 2(b), which shows the development of inefficiency scores over the sample period. Clearly, the most significant decrease in inefficiency comes from adding supply and demand shifters. With the exception of a few occasions in the early 1990s, however, almost all of the remaining inefficiency is gone once we include speculators' futures market positions.

Summing up, our results so far show clear evidence that speculative trading in the futures market serves as an input in the spot coffee price. The effect of speculation on the spot price is spiky, which may explain why other, mean-variance based analyses have had a hard time measuring it. In the next subsection, we explore our results further in order to establish that they are robust, reflect causality and are economically significant.

C. Further analysis

In order to find out whether our results are robust, we carry out a long series of additional tests. Table 4 contains an overview of these additional analyses, as well as - when applicable - how they affect our hypothesis tests. The appendix that accompanies this paper contains a complete and detailed overview of the robustness tests.

We start, in panel [A], by tackling a number of measurement issues. First, we change the way we measure the spot price, and find that our hypothesis tests are qualitative and quantitatively unchanged. When we subsequently impute those variables that are measured annually, we again find that test results are unaffected. Next, we make use of more detailed data, collected since 2007 by the CFTC, regarding the composition of speculative long and short positions. When we focus specifically on index funds, we find that when the net position of index funds increases, $(\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}) \times \text{price}$ decreases, in line with our second hypothesis. We then do the same using the Disaggregated Commitments of Traders reports, collected by the CFTC since 2006, and we find a strong negative correlation between inefficiency and the position of swap dealers, who have the least rapport with the underlying fundamentals in the spot market. Our final two robustness tests in panel [A] concern the oil price. Since oil is not only an input in the production of many agricultural commodities, including coffee, but also the main commodity in commodity

index futures, we have so far refrained from including the oil price in our input sets. As a robustness test, we now include it, and find not only that this does not affect our tests results, but also that the substitution effect that we discuss later on in this subsection is still present.

In panel [B], we tackle the assumptions underlying our model. As explained earlier, we apply a production model to a price discovery setting, thereby implicitly assuming that the usual characteristics of an input-output relationship hold. Now, we relax the assumptions underlying our model in order to find out how this affects our hypothesis test results. We start with the assumption of free disposability of inputs and convexity. In order to relax these assumptions, we re-run our analysis using a Full Disposable Hull (FDH) model, instead of a DEA model. Our test results are still valid, with the exception of Hypothesis 3, which indeed seems to suffer from the dimensionality issues that sometimes plague FDH results (Bogetoft and Otto, 2011). Our next two robustness tests concern the efficiency measure itself, which is assumed to be radial and output-oriented. Relaxing both assumptions, respectively using an additive measure and non-oriented efficiency, does not have much of an effect. The same holds when we include, to the extent possible, inputs as ratios (e.g. demand divided by supply) in order to reduce dimensionality issues. Using a non-increasing returns to scale model and dropping most of the demand and supply shifters also has no effect. In line with our theoretical model, we find some evidence that an increase in the futures price Granger-causes higher inventories for importers, and that a good harvest results in more inefficiency.

In panel [C], we take a closer look at the way in which we test our hypotheses. We start by considering the fact that our inefficiency distributions are highly skewed, and test each of our hypothesis on the tails of each distribution, using a Kolmogorov-Smirnov test. As expected, this approach does not work for Hypothesis 2, which - given that it concerns the differences between two inefficiency distributions - does not suffer from skewness problems. For the other hypothesis tests, results are robust. In line with Irwin and Sanders (2010) and others, we find evidence that increased speculation increases the volatility of spot prices. Next, we randomize the speculative long and short positions, preserving the overall distributions of each, but distributing each position randomly over the sample period. If speculation does not *cause* the decrease in efficiency, then we expect our null hypothesis not to be rejected, which is indeed the case (at the 1% level). Our next robustness test concerns the rise of index fund speculation: graphical evidence suggests that $(\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}) \times \text{price}$, the part of the price which, according to our model, can be attributed to speculation, does not start to co-move with long futures positions until after 2000, i.e., when interest fund specula-

Table 4: Robustness tests

	Issue	Additional analysis	H_0^1	H_0^2	H_0^3
[A] Measurement	(1) Spot price measurement	Use second most nearby futures contract	✓	✓	✓
	(2) Frequency	Imputation of annual variables	✓	✓	✓
	(3) Index speculators	New variable for speculators	n.a.	n.a.	n.a.
	(4) Managed funds	New variable for managed funds	n.a.	n.a.	n.a.
	(5) Oil price as an input	Include oil price and re-run hypotheses tests	✓	✓	✓
	(6) Oil price as an input	Include oil price and re-run [E.1] (substitution effect)	n.a.	n.a.	n.a.
[B] Assumptions	(1) Free disposability and convexity	FDH estimations	✓	✓	X
	(2) Radial expansion	Additive efficiency measure	✓	✓	✓
	(3) Output orientation	Non-oriented efficiency measure	✓	X	✓
	(4) Scaling and dimensionality	Apply ratios	✓	✓	✓
	(5) Returns to scale	non-increasing	✓	✓	✓
	(6) Dimensionality	Drop most demand-/supply-shifters	✓	✓	✓
	(7) Hoarding	Inventories increase as a result of future price?	n.a.	n.a.	n.a.
	(8) Good harvest	χ is high when there is a good harvest	n.a.	n.a.	n.a.
[C] Hypotheses	(1) Tail effects	Kolmogorov-Smirnov tests on tails	✓	X	✓
	(2) Volatility test	Welch test on std. dev of inefficiencies	n.a.	n.a.	n.a.
	(3) Spurious correlation	randomizing speculative long and short position	X	X	X
	(4) Demand/supply shift	Market development futures market	n.a.	n.a.	n.a.
	(5) Fit using $\{D, W, F, V\}$	Figure C5.a	n.a.	n.a.	n.a.
[D] Causality	(1) Significance	Bootstrapped $\chi_t^{D,W,F} - \chi_t^{D,W,F,V}$, Figure A.1c	n.a.	n.a.	n.a.
	(2) VAR analysis	Impulse response of coffee price	n.a.	n.a.	n.a.
	(3) VAR analysis	Variance of impulse response of coffee price	n.a.	n.a.	n.a.
	(4) VAR analysis	Normality tests using coffee price	n.a.	n.a.	n.a.
	(5) VAR analysis	Impulse response of $(\chi_t^{D,W,F} - \chi_t^{D,W,F,V}) \times$ coffee price	n.a.	n.a.	n.a.
	(6) VAR analysis	Variance of impulse response of $\chi_t^{D,W,F} - \chi_t^{D,W,F,V}$	n.a.	n.a.	n.a.
	(7) VAR analysis	Normality tests using $(\chi_t^{D,W,F} - \chi_t^{D,W,F,V}) \times$ price	n.a.	n.a.	n.a.
	(8) Causality	Granger causality tests	n.a.	n.a.	n.a.
	(9) Lags	6 month lag in DEA model (maximum between har- vests)	✓	✓	✓
	(10) Lags	Impulse response for 1 year (cycle of min. 2 harvests)	n.a.	n.a.	n.a.
[E] Policy	(1) Substitution effects	With S&P and price of oil, Figure E.1	n.a.	n.a.	n.a.
	(2) Substitution effects	Changes in price of oil, Table ??	n.a.	n.a.	n.a.
	(3) Market impact	Figure E.3	n.a.	n.a.	n.a.
	(4) Impact on Fair trade	Figure E.4	n.a.	n.a.	n.a.
	(5) Consumer surplus	Figure E.5	n.a.	n.a.	n.a.

Significance tested at 1% level; Based on t-test.

tion started taking off. A one percent increase in these long positions reduces $(\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}) \times \text{price}$ by 0.42%. Finally, we compare the actual price of coffee with the price that, according to our results, would prevail without futures speculation, and show that the model tracks the actual price remarkably well.

In panel [D], we try to tackle the issue of causality. We start by presenting the bootstrapped distributions of $\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}$, in order to verify that measurement error does not affect our tests for hypothesis 2. Next, we follow the approach taken by Irwin and Sanders (2010), and estimate VAR models using the spot price of coffee and the long futures positions. Our initial results are in line with Irwin and Sanders (2010) and Irwin et al. (2009): the cumulative impulse response of the spot price to a one standard deviation change in the long futures position is zero, with a very large confidence interval. However, variance decomposition of that same impulse response shows that the variance of the spot price does react. When we repeat the VAR analysis, instead using the part of the spot price that, according to our model, is explained by speculation, the results change. The impulse response is now positive, albeit still not significant, and the variance is no longer affected. More interesting are the Granger causality tests from both VAR analyses: whereas initially, neither spot nor futures position Granger-cause one another, the futures position does Granger-cause the part of the spot price that is explained by speculation. Further robustness analyses reveal that our hypothesis test results are robust for the inclusion of all inputs with a lag (of six months), in line with our model where inventory adjustments should play an important role. Interestingly, the six month cumulative impulse response of the part of the spot price that is explained by speculation is now not only positive, but also significantly different from zero. Summing up, the results in panel [D] provide additional evidence that speculation is not only an input in our model, but has indeed caused part of the changes in the spot price.

Finally, in Panel [E] we investigate a number of policy-related issues. We find evidence of a substitution effect: as the relative returns on oil compared with other commodities in the index drop, commodity futures index funds shift more of their investments into other commodities, strengthening the speculative channel we identify in this paper. Further analysis shows that this effect is particularly likely to occur in the presence of steep oil price increases. Together, these results contribute to the complex relationship between oil prices and the prices of other commodities, such as coffee. An increase in the price of oil can increase the spot price of commodities, since oil is an input in the production of most commodities. However, a decline in the price of oil does not necessarily have the opposite effect: as a result of the substitution effect that is briefly described above, such a decline may actually result in

an increase in other commodities' futures prices, thereby pushing spot prices up instead of reducing them. Next, we find evidence in support of the market impact of long speculators in the coffee futures market. Fair trade has not however been affected, because the speculation has mostly influenced spot prices when they were well below the minimum guaranteed fair trade price. Finally, using some basic assumptions, we estimate that the loss in consumer surplus as a result of speculation averages approximately 20%, but varies widely over time. IN summary, these additional results provide further evidence of the impact speculation has had on both producers and consumers of coffee.

VI. Conclusion

In this paper, we have introduced a new approach to measuring the possible impact of futures speculation on spot commodity prices. We advocate the use of a non-parametric, highly flexible empirical model for measuring this impact, in order to account for possible non-linearity in the transmission from futures to spot market.

Empirical results for the coffee market show that most of the changes in spot prices can be attributed to shifts in demand and - in particular - supply. Failure to account properly for these inputs in the coffee price generation process may lead to serious overestimation of the effects of speculation. Nevertheless, speculation is an important part of the coffee price generation process. The effect of speculation on the coffee price is indeed spiky, which explains why traditional, mean-variance based methods have failed to identify it. However, it is also significant, both statistically and economically. An extensive robustness analysis confirms the validity of our results, and, within the limitations posed by the data, we are able to establish causality.

It is important to bear in mind the complexity of the process that is described in this paper and captured by our empirical analysis. Contrary to many other studies, we find that futures speculation does have an impact on spot prices. However, we also argue and establish empirically that this impact depends crucially on other aspects, including inventories, the futures market microstructure, harvest sizes, the price elasticity of demand and many other factors that are included in our analysis. On occasion, as described in this paper, these factors align to create a 'perfect storm' that temporarily pushes up the spot price of coffee. Although futures speculation is an important factor in the mix, our results imply that we should not expect miracles from the futures trading caps for speculators currently being proposed in many legislatures. Rather, our results suggest that limiting the negative consequences of futures

speculation without harming the liquidity that speculators bring to the futures market will require a broader set of measures.

What do our results imply for other commodities? Compared with most other commodities, coffee is widely traded and produced by many countries, and the natural requirements for coffee production (elevation, a dry season) mean that climate change may affect it more than is the case with many other commodities. All of these elements suggest that speculation can be expected to be at least as important as an input for other commodities. However, coffee is also relatively easy to store and importing countries' inventories are at a low level, suggesting the pass-through from futures to spot market is relatively easy. Ultimately, therefore, a definitive answer to this question is left open for further research.

References

- Acharya, V. V., Lochstoer, L. A., and Ramadorai, T. (2011). Limits to arbitrage and hedging: Evidence from commodity markets. *National Bureau of Economic Research Working Paper Series*, No. 16875:–.
- Akiyama, T. and Varangis, P. N. (1990). The impact of the international coffee agreement on producing countries. *The World Bank Economic Review*, 4(2):157–173.
- Algra, J. (2009). Coffee and new trends in commodities? mimeo, Novotrade.
- Bates, R. H., . (1997). *Markets and States in Tropical Africa: The Political Basis of Agricultural Policies*. "Markets and States in Tropical Africa: The Political Basis of Agricultural Policies, Berkeley.
- Bogetoft, P. and Otto, L. (2011). *Benchmarking with DEA, SFA, and R*. Springer.
- Booth, G. G., So, R. W., and Tse, Y. (1999). Price discovery in the german equity index derivatives markets. *J. Fut. Mark.*, 19(6):619–643.
- Bryant, H. L., Bessler, D. A., and Haigh, M. S. (2006). Causality in futures markets. *Journal of Futures Markets*, 26(11):1039–1057.
- CFTC (2008). Staff report on commodity swap dealers & index traders with commission recommendations. Internet site: http://www.cftc.gov/stellent/groups/public/@newsroom/documents/file/cftcstaff_reportonswapdealers09.pdf (accessed November 1, 2010), Commodity Futures Trading Commission, Washington, DC.

- CFTC (2010). Commitment of active traders report. Technical report, Commodity Futures Trading Commission. Last accessed November 1, 2011.
- CFTC (December 5, 2006). Commission actions in response to the "comprehensive review of the commitments of traders reporting program" (June 21, 2006). EPTD discussion paper 57, Commodity Futures Trading Commission.
- Chambers, R. G. (1988). *Applied Production Analysis: A Dual Approach*. Cambridge University Press, Massachusetts, USA.
- Chan, K. (1992). A further analysis of the lead-lag relationship between the cash market and stock index futures market. *The Review of Financial Studies*, 5(1):123–152.
- Charnes, A., Cooper, W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6):429–444.
- Charnes, A., Cooper, W. W., Golany, B., Seiford, L., and Stutz, J. (1985). Foundations of Data Envelopment Analysis for Pareto-Koopmans efficient empirical production functions. *Journal of Econometrics*, 30:91–107.
- Cherchye, L., Kuosmanen, T., and Post, T. (2000). What is the economic meaning of FDH? a reply to Thrall. *Journal of Productivity Analysis*, 13(3):263–267.
- Crown, T. (1997). Coffee futures price behavior and fund investment. *International Coffee Organization Coffee Newsletter*, 4.
- Daviron, B. (1996). The rise and fall of governmental power on the international coffee market. In Griffon, M. and Guillaumont, P., editors, *Economics of Agricultural Policies in Developing Countries*, pages –. Editions de la Revue Française d'Économie.
- Daviron, B. and Ponte, S. (2005). *The Coffee Paradox: Global Markets, Commodity Trade and the Elusive Promise of Development*. Zed Books, London.
- Deaton, A. and Laroque, G. (1992). On the behaviour of commodity prices. *The Review of Economic Studies*, 59(1):1–23.
- Deaton, A. and Laroque, G. (1996). Competitive storage and commodity price dynamics. *The Journal of Political Economy*, 104(5):896–923.
- Dehn, J. (2000). Commodity price uncertainty and shocks: implications for economic growth. Working Paper Series 2000-10, Center for the Study of African Economies, Oxford University.

- Dijk, J. B. v., Doesburg, D. H. M., Heijbroek, A. M. A., Wazir, M. R. I. A., and Wolff, G. S. M. d. (1998). *The World Coffee Market*. Rabobank International, Utrecht.
- Fafchamps, M. and Hill, R. V. (2005). Selling at the farmgate or traveling to market. *American Journal of Agricultural Economics*, 87(3):717–734.
- Fafchamps, M. and Hill, R. V. (2008). Price transmission and trader entry in domestic commodity markets. *Economic Development & Cultural Change*, 56(4):729–766.
- Fama, E. F. and French, K. R. (1987). Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage. *The Journal of Business*, 60(1):55–73.
- Fiorio, C. V. (2004). Confidence intervals for kernel density estimation. *Stata Journal*, 4(2):168–179.
- Food and Agriculture Organization of the United Nations (2007). <http://faostat.fao.org/site/567/desktopdefault.aspx?pageid=567#anchor>.
- Foreign Agricultural Service of the United States Department of Agriculture (2009). http://www.fas.usda.gov/psdonline/psddownload.aspxfilepsd_coffee_csv.zip.
- Frankel, J. A. (2006). The effect of monetary policy on real commodity prices. *National Bureau of Economic Research Working Paper Series*, No. 12713:–.
- Frankel, J. A. (2008). The effect of monetary policy on real commodity prices. In *NBER Chapters*, pages 291–333. National Bureau of Economic Research, Inc.
- Gansler, G. (2009). Statement by chairman Gary Gensler on speculative position limits and enhanced transparency initiative. July 7.
- Garbade, K. D. and Silber, W. L. (1983). Price movements and price discovery in futures and cash markets. *The Review of Economics and Statistics*, 65(2):289–297.
- GCA (2009a). http://www.greencoffeeassociation.org/images/uploads/resources/professional_resources-warehouses_stocks_current.pdf.
- GCA (2009b). http://www.greencoffeeassociation.org/professionalresources/warehouse_informationfiles.

- GCA (2009c). http://www.greencoffeeassociation.org/images/uploads/resources/professional_resources-warehouse_stocks_historical.pdf.
- Gorton, G., Hayashi, F., and Rouwenhorst, K. G. (2007). The fundamentals of commodity futures returns. Yale ICF Working Paper No. 07-08, Yale University.
- Hein, L. and Gatzweiler, F. (2006). The economic value of coffee (*coffea arabica*) genetic resources. *Ecological Economics*, 60(1):176–185.
- Henderson, B. J., Pearson, N. D., and Wang, L. (2012). New evidence on the financialization of commodity markets. Technical report, Available at SSRN: <http://ssrn.com/abstract=1990828> or <http://dx.doi.org/10.2139/ssrn.1990828>.
- Hochman, G., Rajagopal, D., Timilsina, G., and Zilberman, D. (2011). The role of inventory adjustments in quantifying factors causing food price inflation. Policy Research Working Paper 5744, World Bank.
- Hong, H. and Yogo, M. (2012). What does futures market interest tell us about the macroeconomy and asset prices? *Journal of Financial Economics*, 105(3):473–490.
- ICE (2009). https://www.theice.com/report_center_docs.jsp, file coffee_prices.
- International Coffee Organization (2008). http://www.ico.org/frosts_droughts.asp.
- International Coffee Organization (2009). Study on fertilizer prices.
- International Coffee Organization (Accessed July 6, 2012). World coffee trade. http://www.ico.org/trade_e.asp.
- Irwin, S. H. and Sanders, D. R. (2010). The impact of index and swap funds on commodity futures markets: Preliminary results. *OECD Food, Agriculture and Fisheries Working Papers*, 27.
- Irwin, S. H. and Sanders, D. R. (2011). Index funds, financialization, and commodity futures markets. *Applied Economic Perspectives and Policy*, 33(1):1–31.
- Irwin, S. H., Sanders, D. R., and Merrin, R. P. (2009). Devil or angel, the role of speculation in the recent commodity price boom (and bust). *Journal of Agricultural and Applied Economics*, 41(2):377–391.

- Kingsbury, K. (2007). After the oil crises, a food crisis? Technical report, Time.
- Korniotis, G. M. (2009). Does speculation affect spot price levels? the case of metals with and without futures markets. Finance and Economics Discussion Series 2009-29, Federal Reserve Board.
- Lopez, R. A. and You, Z. (1993). Determinants of oligopsony power: The Haitian coffee case. *Journal of Development Economics*, 41(2):275–284.
- Loureiro, M. L. and Lotade, J. (2005). Do fair trade and eco-labels in coffee wake up the consumer conscience? *Ecological Economics*, 53(1):129–138.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, New York, 2 edition.
- Masters, M. W. (2008). Testimony of Michael W. Masters before the Committee on Homeland Security and Government Affairs of the United States Senate. May 20.
- Meijerink, G., van Berkum, S., Shutes, K., and Solano, G. (2011). Price and prejudice: Why are food prices so high? Report 2011-035, Landbouw-Economisch Instituut (Agricultural Economics Institute foundation) Wageningen University. ISBN 978-90-8615-518-7.
- Mosheim, R. (2002). Organizational type and efficiency in the Costa Rican coffee processing sector. *Journal of Comparative Economics*, 30(2):296–316.
- Ninan, K. and Sathyapalan, J. (2005). The economics of biodiversity conservation: a study of a coffee growing region in the western ghats of India. *Ecological Economics*, 55(1):61–72.
- Osorio, N. (2009). Coffee market outlook. Technical report, ICO/UNCTAD, Geneva.
- Parikh, A. (1979). Estimation of supply functions for coffee. *Applied Economics*, 11(1):43–54.
- Park, B. U., Simar, L., and Weiner, C. (2000). The fdh estimator for productivity efficiency scores: Asymptotic properties. *Econometric Theory*, 16(6):855–877.
- Pindyck, R. S. (1993). The present value model of rational commodity pricing. *The Economic Journal*, 103(418):511–530.
- Rutten, L. and Youssef, F. (2007). Market-based price risk management. Technical report, International Institute for Sustainable Development.

- Sexton, R. J. and Zhang, M. (1996). A model of price determination for fresh produce with application to California iceberg lettuce. *American Journal of Agricultural Economics*, 78(4):924–934.
- Simar, L. and Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science*, 44(1):49–61.
- Singleton, K. J. (2011). Investor flows and the 2008 boom/bust in oil prices. Available at SSRN: <http://ssrn.com/abstract=1793449> or <http://dx.doi.org/10.2139/ssrn.1793449>, Stanford University.
- Stein, J. L. (1961). The simultaneous determination of spot and futures prices. *The American Economic Review*, 51(5):1012–1025.
- Subrahmanyam, A. (1991). A theory of trading in stock index futures. *The Review of Financial Studies*, 4(1):17–51.
- Swed, F. S. and Eisenhart, C. (1943). Tables for testing randomness of grouping in a sequence of alternatives. *The Annals of Mathematical Statistics*, 14(1):66–87.
- Tang, K. and Xiong, W. (2010). Index investment and financialization of commodities. *National Bureau of Economic Research Working Paper Series*, No. 16385:–.
- Thanassoulis, E., Portela, M. C. S., and Despić (2008). Data envelopment analysis: The mathematical programming approach to efficiency analysis. In Fried, H. O., Knox Lovell, C. A., and Schmidt, S. S., editors, *The Measurement of Productive Efficiency and Productivity Growth*, pages 251–420, New York, USA. Oxford University Press.
- Turnovsky, S. J. (1983). The determination of spot and futures prices with storable commodities. *Econometrica*, 51(5):1363–1387.
- Wahl, P. (2008). Speculation undermines the right of food. Mimeo.
- Wickens, M. R. and Greenfield, J. N. (1973). The econometrics of agricultural supply: An application to the world coffee market. *The Review of Economics and Statistics*, 55(4):433–440.
- Wollni, M. and Zeller, M. (2007). Do farmers benefit from participating in specialty markets and cooperatives? the case of coffee marketing in Costa Rica. *Agricultural Economics*, 37(2-3):243–248.

World Bank (2008). Double jeopardy: Responding to high food and fuel prices. Technical report, World Bank.

Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics*, 14(1):45–52.

Appendix of Futures Speculation and Commodity Prices

A.1. Spot price measurement

As explained in our description of the data, we measure the spot price of coffee by taking the futures price just before expiration, thus relying on the fact that the latter converges to the former. In practice, however, the futures price occasionally shows some wide swings in the days before expiration, before settling at the spot price (Gorton et al., 2007). In order to rule out the possibility that such swings affect our findings, we repeat our analysis here with a different measure for the spot price: instead, we take the second-nearest futures price. This price is generally much less susceptible to the types of swings we just described.

Figure A.1: Price development

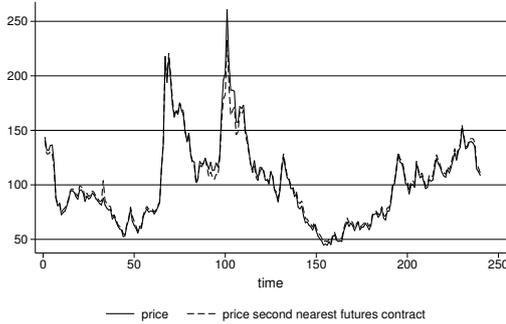


Table A.1: Tests with alternative price

Hypothesis	t-test	KW	KS
$H_0^1: \chi_t^D = 0$	0.0000	0.0001	0.0000
$H_0^1: \chi_t^{D,W} = 0$	0.0000	0.0001	0.0000
$H_0^1: \chi_t^{D,W,F} = 0$	0.0000	0.0001	0.0000
$H_0^2: \chi_t^{D,W,F} - \chi_t^{D,W,F,V} = 0$	0.0000	0.0001	0.0000
$H_0^3: \chi_t^{D,W,F,V} > 0$	0.0000	0.0001	0.0000

Notes: KW= Kruskal-Wallis rank test; KS = Kolmogorov-Smirnov test.

Figure A.1 compares our original measure for the price with the new measure, and shows that the resulting prices are very, very similar indeed. Table A.1 repeats our hypotheses tests, and finds that results do not change when we use the alternative measurement of the spot price.

A.2. Frequency

As is clear from our description of the data, our analysis is characterized by a mismatch of the frequency at which we observe variables. This type of mismatch is highly typical of a study that combines high-frequency financial data with lower-frequency production data (Zhou, 1996).

In order to assess whether this mismatch affects our results, we repeat our analysis with imputed values for variables with a lower frequency, and interpolate these variables. The variables that are interpolated in this way are the production of both Arabica and Robusta, total coffee consumption, inventories held by importers and exporters and agricultural productivity. Not

shown here are additional interpolations controlling for seasonality as results do not change materially.

Figure A.2: χ with imputation

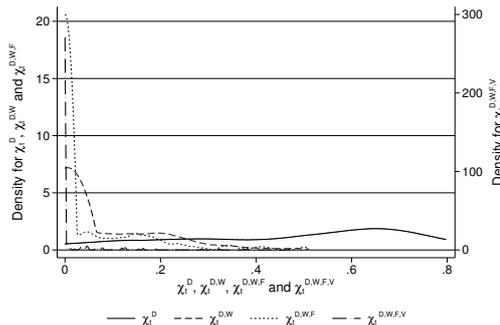


Table A.2: Tests with imputation

Hypothesis	t-test	KW	KS
$H_0^1: \chi_t^D = 0$	0.0000	0.0001	0.0000
$H_0^1: \chi_t^{D,W} = 0$	0.0000	0.0001	0.0000
$H_0^1: \chi_t^{D,W,F} = 0$	0.0000	0.0001	0.0000
$H_0^2: \chi_t^{D,W,F} - \chi_t^{D,W,F,V} = 0$	0.0000	0.0001	0.0000
$H_0^3: \chi_t^{D,W,F,V} > 0$	0.0000	0.0001	0.0000

Notes: KW= Kruskal-Wallis rank test; KS = Kolmogorov-Smirnov test.

Figure A.2 shows the resulting inefficiency distributions. As is to be expected, inefficiency distributions become more skewed. However, from Table A.2 we learn that none of the hypotheses tests is affected by this change.

A.3. Measurement of index funds

In our analysis, we can accurately distinguish between commercial and non-commercial futures market positions. However, part of our argument regarding the impact of the latter on spot prices hinges on the particular behavior of a subset of non-commercial traders: index funds. It is these funds that are most likely to invest long-long, rolling over positions, largely ignoring fundamentals and generating substantial increases in - particularly - long futures positions. Since 2007, the CFTC has collected more detailed data on non-commercial traders, on a quarterly basis. Unfortunately, the resulting series is too short to rerun our analysis. We therefore present a more modest analysis.

Figure A.3: χ and index funds

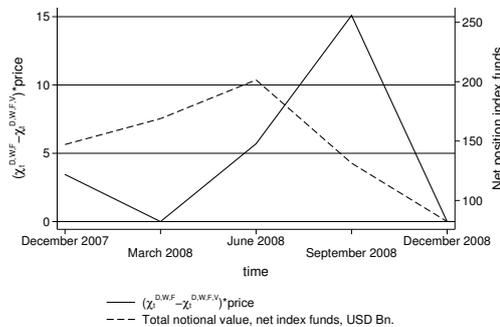


Table A.3: Correlations

	$\Delta \times \text{price}$	net	long	short	equiv
$\Delta \times \text{price}$	1.0000				
net	0.0853	1.0000			
long	0.1590	0.9879*	1.0000		
short	-0.2349*	-0.7826*	-0.8541*	1.0000	
equiv	0.1720	0.2503	0.1435	0.0292	1.0000

Notes: $\Delta \times \text{price} = (\chi_t^{D,W} - \chi_t^{D,W,F,V}) \times \text{price}$; net, long, short refer to notional positions; equiv refers to futures equivalent, net position; *= significant at least at 10% level.

Figure A.3 shows the development of the net position of index funds, com-

pared with the development of $(\chi_t^{D,W} - \chi_t^{D,W,F,V}) \times \text{price}$.²⁶ In line with our second hypothesis, we expect the latter to be high, when the former is also high. Indeed, this is what the Figure shows. The graph is similar when we use only long positions. In Table A.3, we correlate $(\chi_t^{D,W} - \chi_t^{D,W,F,V}) \times \text{price}$ with index funds' positions. As expected, the correlation is positive/negative with the long/short positions of index funds. Not surprisingly, the correlation with the net position of index funds is also positive. Finally, the correlation with the futures equivalent position including delta-adjusted options - estimates of the contracts that would have been closed if there had not been offsets - is also positive. Not surprisingly, given the low number of observations, correlations are for the most part not significant.

A.4. Managed funds

From 2006, the CFTC has published the Commitments of Traders (COT) report, which contains a breakdown of open interest positions above the minimum reporting levels of the CFTC. These data are available on a higher weekly frequency. As these are all positions, including those of commercial traders, we relate them to $\chi^{D,W}$.

Figure A.4 plots the long positions of commercial traders, the long positions of swap funds²⁷ and $\chi^{D,W}$. In line with our first hypothesis, we expect to see no relation with the latter and a negative relation with the former positions. This is in fact more or less what we observe.

Figure A.4: χ and managed funds

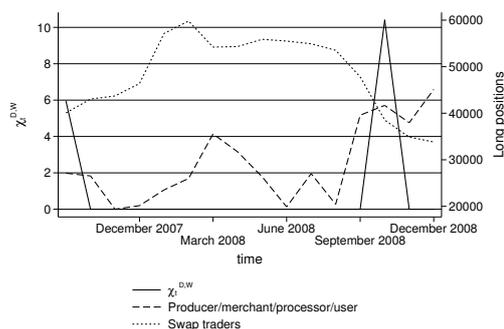


Table A.4: Correlations

	$\chi^{D,W}$	prod	funds	swap	other
$\chi^{D,W}$	1.0000				
prod	0.2640	1.0000			
funds	0.4463*	-0.2384	1.0000		
swap	-0.5756*	-0.1201	-0.00913	1.0000	
other	-0.4592*	0.0121	-0.8076*	0.0603	1.0000

Notes: prod, funds, swap and other refer to change in long positions of producers/merchants/processors/users, managed funds, swap dealers and other reportables, respectively; * = significant at least at 10% level.

²⁶We compare with the inefficiency excluding all futures positions, in order to avoid measurement of commercial versus non-commercials.

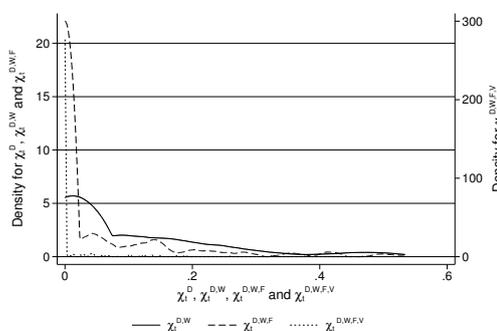
²⁷In the Disaggregated Commitments of Traders (COT) reports, there are swap dealers that are not involved in index trading. These traders are not included as index traders in the Commodity Index Trader (CIT) supplement that the CFTC publishes since January 2007, and that was used in the previous subsection. Institutional investors that do not operate via a swap dealer are classified here as managed money or as other reportables.

Table A.4 provides somewhat more robust evidence. In this table, we correlate $\chi^{D,W}$ with the long positions of all different types of investors in the Disaggregated Commitments of Traders reports. Correlation with the long positions of 'producers' is positive, but not significant. Correlation with managed funds is also positive. The negative correlation is strongest, as expected, for swap dealers, who have the least rapport with the underlying fundamentals in the spot market.

A.5. Oil price as an input: hypotheses tests

In our analyses in the paper, we do not include the oil price as a direct input. One reason for not doing so, is the fact that other variables (the dollar exchange rate, price of fertilizer) already capture quite some part of the effects of oil price changes. Also, in our policy section *E*, we test for substitution effects: the mechanics of index funds *can* imply that low oil prices (or small oil price changes) create the conditions for coffee futures market positions to affect the spot price of coffee - in the absence of momentum in the oil futures market.

Figure A.5: χ with oil price as input Table A.5: Tests with oil price as input



Hypothesis	t-test	KW	KS
$H_0^1: \chi_t^D = 0$	n.a.	n.a.	n.a.
$H_0^1: \chi_t^{D,W} = 0$	0.0000	0.0001	0.0000
$H_0^1: \chi_t^{D,W,F} = 0$	0.0000	0.0001	0.0000
$H_0^2: \chi_t^{D,W,F} - \chi_t^{D,W,F,V} = 0$	0.0000	0.0001	0.0000
$H_0^3: \chi_t^{D,W,F,V} > 0$	0.0000	0.0001	0.0000

Notes: KW= Kruskal-Wallis rank test; KS = Kolmogorov-Smirnov test.

In Figure A.5, we show the inefficiency distributions after including the oil price as an extra input in our data envelopment analysis model. As expected, with one extra input, the distributions are somewhat more skewed. Hypotheses tests in Table A.5 confirm the results from the paper for all hypotheses.

A.6. Oil price as an input: substitution effect still there?

In a later section in this Appendix, we formally investigate the above substitution effect. We find that inefficiency indeed is high when the oil price is low, and vice versa. There is, however, no correlation with the S&P500, suggesting that there is indeed some substitution within the index.

In Figure A.6, we repeat this analysis, by estimating the kernel density of $\chi_t^{\mathbf{D},\mathbf{W},\mathbf{F}} - \chi_t^{\mathbf{D},\mathbf{W},\mathbf{F},\mathbf{V}}$, conditional on the price of oil and the S&P500, respectively, *after* including the oil price as an input in the efficiency estimate. Again, the graph reflects a substitution effect. The correlation analysis in Table A.6 suggests the same.

Figure A.6: Substitution with oil price as input

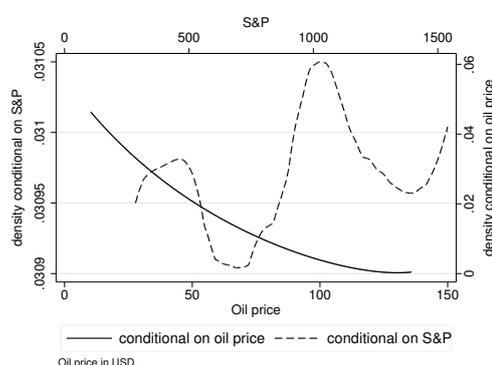


Table A.6: Correlations

	Δ	oil price	S&P500	$\chi_t^{\mathbf{D},\mathbf{W},\mathbf{F}}$	$\chi_t^{\mathbf{D},\mathbf{W},\mathbf{F},\mathbf{V}}$
Δ	1.0000				
oil price	-0.0097	1.0000			
S&P500	0.0585	0.5596*	1.0000		
$\chi_t^{\mathbf{D},\mathbf{W},\mathbf{F}}$	0.8064*	-0.0863	-0.0730	1.0000	
$\chi_t^{\mathbf{D},\mathbf{W},\mathbf{F},\mathbf{V}}$	0.1510*	-0.1327*	-0.1920*	0.7064*	1.0000

Notes: $\Delta = \chi_t^{\mathbf{D},\mathbf{W},\mathbf{F}} - \chi_t^{\mathbf{D},\mathbf{W},\mathbf{F},\mathbf{V}}$; * = significant at least at 10% level.

B.1. Free disposability of inputs and outputs

In a traditional production analysis, true production possibilities are assumed to be convex. Since the observed input-output combinations are assumed to be a representative sample from these true possibilities, the same holds for these combinations. In the context of our analysis, convexity can reasonably be questioned. To do so, we estimate our production model using a Free Disposable Hull (FDH) model.

FDH relies only on the assumption that observed production possibilities satisfy free disposability: it is always possible to waste inputs. With this waste, FDH can in fact be considered as an approximation of the traditional production analysis, with congestion.²⁸ With this adjustment, the resulting efficiency scores can still be interpreted as Pareto-Koopmans or Debreu-Farrell measures of technical efficiency (Cherchye et al., 2000).

In light of our approach, applying FDH does have one drawback: it is somewhat more sensitive to dimensionality issues, which can lead to a bias in inefficiency scores (Park et al., 2000). Figure B.1 indeed shows more skewed inefficiency distributions, with the exception of $\chi_t^{\mathbf{D},\mathbf{W},\mathbf{F},\mathbf{V}}$. In Table B.1, only the results for our third hypothesis test change significantly, demonstrating that with free disposability, our input set $\{\mathbf{D}, \mathbf{W}, \mathbf{F}, \mathbf{V}\}$ no longer reduces inefficiency to zero. The Kolmogorov-Smirnov test for our second hypothesis

²⁸In light of our coffee price model, congestion can be seen as late price discovery.

suggests that we can no longer reject the null that $\chi_t^{D,W,F} - \chi_t^{D,W,F,V} = 0$. However, the difference is significant again when evaluated above mean or median (i.e. in the tail).

Figure B.1: Distributions with FDH

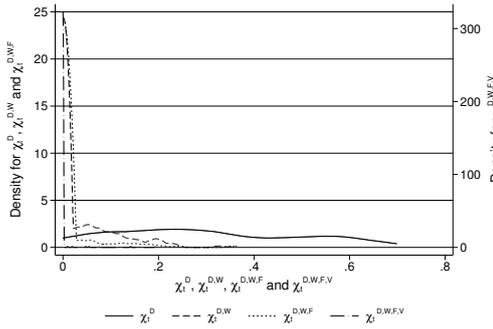


Table B.1: Hypotheses tests with FDH

Hypothesis	t-test	KW	KS
$H_0^1: \chi_t^D = 0$	0.0000	0.0001	0.0000
$H_0^1: \chi_t^{D,W} = 0$	0.0000	0.0001	0.0000
$H_0^1: \chi_t^{D,W,F} = 0$	0.0000	0.0001	0.0000
$H_0^2: \chi_t^{D,W,F} - \chi_t^{D,W,F,V} = 0$	0.0004	0.0001	0.3930
$H_0^3: \chi_t^{D,W,F,V} > 0$	0.0088	0.4758	0.9940

Notes: KW= Kruskal-Wallis rank test; KS = Kolmogorov-Smirnov test.

B.2. Radial expansion

In economic analyses, we are typically not only interested in the direction of effects, but certainly also in their magnitude. In our DEA model, we rely on a radial measure of inefficiency: our Debreu-Farrell measure quantifies the resulting relative efficiency. But what is the magnitude of the maximum absolute decreases in prices? In order to find out, we re-estimate our model, this time using an additive inefficiency measure that quantifies the maximum slack in the price of coffee.

Of course, here too, the introduction of a new methodology has its drawbacks: the resulting Koopmans inefficiency is no longer invariant with respect to units of measurement.

Figure B.2: Distributions with additive in- efficiency

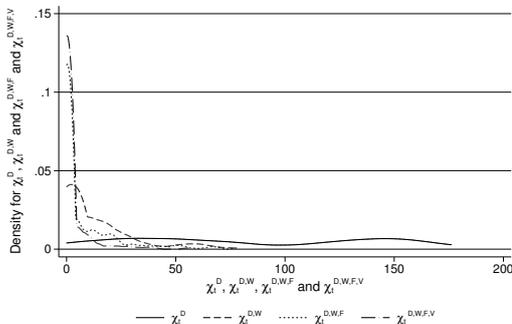


Table B.2: Tests with additive inefficiency

Hypothesis	t-test	KW	KS
$H_0^1: \chi_t^D = 0$	0.0000	0.0001	0.0000
$H_0^1: \chi_t^{D,W} = 0$	0.0000	0.0001	0.0000
$H_0^1: \chi_t^{D,W,F} = 0$	0.0000	0.0001	0.0000
$H_0^2: \chi_t^{D,W,F} - \chi_t^{D,W,F,V} = 0$	0.0000	0.0001	0.0000
$H_0^3: \chi_t^{D,W,F,V} > 0$	0.0000	0.0001	0.0000

Notes: KW= Kruskal-Wallis rank test; KS = Kolmogorov-Smirnov test.

We recall from our descriptive statistics that the price of coffee is on average approximately 102 dollar cents per pound. In Figure B.2 we show the distri-

butions of the additive inefficiency scores, now comparable to the price. Table B.2 is testament to the fact that hypotheses test results remained unaffected, even for hypothesis 3.

B.3. Output orientation

Our additive inefficiency measure in the previous paragraphs shares one important feature with the inefficiency measures introduced in the paper: both are output (i.e., price) oriented. But what if in reality not everything is aimed at increasing prices? After all, some of our 'inputs' can surely not be said not to aim to maximize prices; think for example of consumers.

To see how robust our results are to relaxing the pure focus on output orientation, we can estimate the same additive efficiency as in the previous subsection, but now in a non-oriented manner: effectively what we then do is to weight reductions in inputs and increases in output, using the weights resulting from the non-oriented Pareto-Koopmans DEA model (Charnes et al., 1985).

Figure B.3: Distributions with non-oriented inefficiency

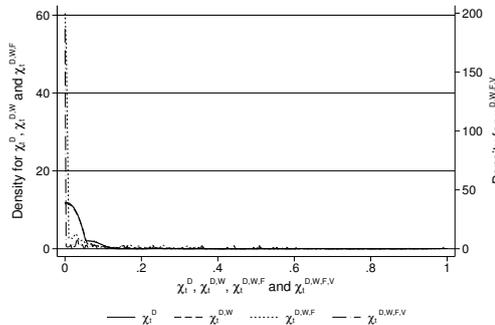


Table B.3: Hypotheses tests with non-oriented inefficiency

Hypothesis	t-test	KW	KS
$H_0^1: \chi_i^D = 0$	0.0006	0.0001	0.0020
$H_0^1: \chi_i^{D,W} = 0$	0.0001	0.0001	0.0000
$H_0^1: \chi_i^{D,W,F} = 0$	0.0000	0.0001	0.0000
$H_0^2: \chi_i^{D,W,F} - \chi_i^{D,W,F,V} = 0$	0.0000	0.0001	0.0000
$H_0^3: \chi_i^{D,W,F,V} > 0$	0.0000	0.0001	0.0000

Notes: KW= Kruskal-Wallis rank test; KS = Kolmogorov-Smirnov test.

Figure B.3 shows the resulting inefficiency distributions, which are very similar to what we saw in Figure B.2. The same holds true for the hypothesis test results in Table B.3.

B.4. Scaling and additivity

In estimating our DEA model, we assume that inputs are additive and can be scaled: a slightly lower set of inputs results in a slightly lower output, i.e., price. These assumptions manifest themselves in our choice of a constant returns to scale DEA model (Bogetoft and Otto, 2011), a choice that we address in the next subsection.

For now, we test the importance of these assumptions in a very simple, rather intuitive manner: a number of our inputs can in fact be argued not to affect our output through their absolute size, but rather through their relative size compared with another input. The most obvious examples are demand and supply. Other examples include the long and short positions of speculators.

Likewise, other inputs can be argued to affect the output (price) in the same manner. Examples include inventories of importers and exporters, yield, crop and agricultural productivity, and the total trading volume and open interest/commitment of hedgers.

Therefore, as a robustness test, we include the first set of inputs as ratios, and the second set of inputs as sums, thus reducing the potential for dimensionality issues, somewhat relaxing the scaling assumption, at the expense of increasing our reliance on the additivity assumption.

Figure B.4: Distributions with ratios and sums

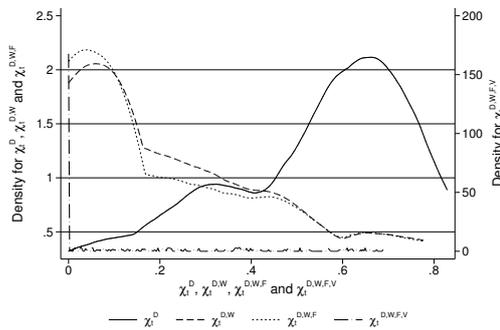


Table B.4: Tests with ratios and sums

Hypothesis	t-test	KW	KS
$H_0^1: \chi_t^D = 0$	0.0000	0.0001	0.0000
$H_0^1: \chi_t^{D,W} = 0$	0.0000	0.0001	0.0000
$H_0^1: \chi_t^{D,W,F} = 0$	0.0000	0.0001	0.0000
$H_0^2: \chi_t^{D,W,F} - \chi_t^{D,W,F,V} = 0$	0.0000	0.0001	0.0000
$H_0^3: \chi_t^{D,W,F,V} > 0$	0.0000	0.0001	0.0000

Notes: KW= Kruskal-Wallis rank test; KS = Kolmogorov-Smirnov test.

Figure B.4 shows the impact, at first glance rather profound, on the inefficiency distributions. It is important to note, however, that one inefficiency distribution remains largely unaffected: $\chi_t^{D,W,F,V}$. Table B.4 shows that hypothesis test results still remain unaffected.

B.5. Returns to scale

In the previous subsection, we took a simple, but somewhat unusual approach to relaxing assumptions regarding scaling. Here, we proceed with a more customary approach: we relax our constraint on the curvature of the efficient price frontier, and estimate a non-increasing returns to scale (NIRS) DEA model.

Inefficiency distributions in Figure B.5 tell a similar story as in Figure B.4: with the exception of $\chi_t^{D,W,F,V}$, average inefficiency increases significantly, as more density is shifted towards higher inefficiency values. Table B.5 shows that hypothesis test results are unaffected.

Figure B.5: Distributions with non-increasing returns to scale

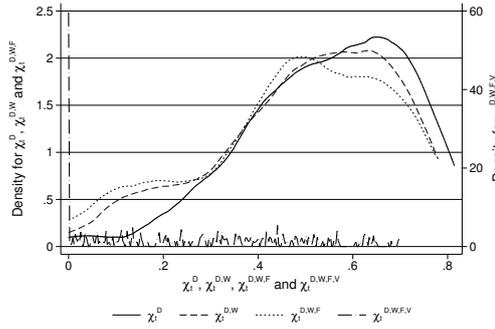


Table B.5: Tests with non-increasing returns to scale

Hypothesis	t-test	KW	KS
$H_0^1: \chi_t^D = 0$	0.0000	0.0001	0.0000
$H_0^1: \chi_t^{D,W} = 0$	0.0000	0.0001	0.0000
$H_0^1: \chi_t^{D,W,F} = 0$	0.0000	0.0001	0.0000
$H_0^2: \chi_t^{D,W,F} - \chi_t^{D,W,F,V} = 0$	0.0000	0.0001	0.0000
$H_0^3: \chi_t^{D,W,F,V} > 0$	0.0000	0.0001	0.0000

Notes: KW= Kruskal-Wallis rank test; KS = Kolmogorov-Smirnov test.

B.6. Dimensionality

In our theory section, as well as in our empirical analysis, we have emphasized both demand and supply, and factors that *shift* demand and supply. We think including inputs that reflect the latter enhances the validity of our model's results. However, our choice here can also be questioned, on two grounds. First, in so doing, we have extended the set of inputs, thereby possibly increasing dimensionality issues: with too many variables, 'efficiency-by-default' is a serious possibility. Second, we have included factors that are strictly speaking supposed to reflect changes in (the productivity of) inputs in an input-output analysis.

Therefore, as a robustness test, we now estimate our model with fewer control variables. Of the variables included in **W**, we retain only our measures of frost and drought, as well as the inventories of importers and exporters.

Figure B.6: Distributions with fewer inputs

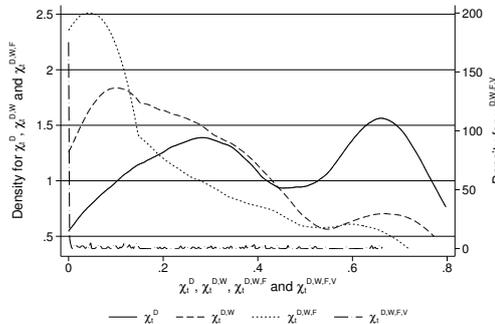


Table B.6: Tests with fewer inputs

Hypothesis	t-test	KW	KS
$H_0^1: \chi_t^D = 0$	0.0000	0.0001	0.0000
$H_0^1: \chi_t^{D,W} = 0$	0.0001	0.0001	0.0000
$H_0^1: \chi_t^{D,W,F} = 0$	0.0000	0.0001	0.0000
$H_0^2: \chi_t^{D,W,F} - \chi_t^{D,W,F,V} = 0$	0.0000	0.0001	0.0000
$H_0^3: \chi_t^{D,W,F,V} > 0$	0.0000	0.0001	0.0000

Notes: KW= Kruskal-Wallis rank test; KS = Kolmogorov-Smirnov test.

Figure B.6 contains the resulting inefficiency distributions. Clearly, the omitted variables account for an important part of the inefficiency. However, as is

clear from the distribution of $\chi_t^{D,W,F,V}$, including the speculative positions results in a very large reduction of that inefficiency. The latter is confirmed by H_0^3 in Table B.6, which also shows that the other hypothesis results remain intact.

B.7. Hoarding and future price developments

One of the more subtle implications of our theoretical model is the notion that one way in which future price developments may have an impact on the spot market and spot price is through hoarding: as the futures price increases, building up an inventory becomes more attractive, *ceteris paribus*.

The argument here is intricate: who exactly benefits from building up the inventory? Is it the farmer, who can afford, given a good harvest, to sell less on the spot market in order to receive the higher futures price? Or is it an intermediary, who either does the same, or tries to guard herself against such an increase by bumping up her inventory. A cautious preliminary answer is given in Figure B.7 and Table B.7. As the Figure shows, inventories of exporters have been shrinking since the early 1990s. Possible additional explanations for these declines include high price variability (Dehn, 2000) and low interest rates (Frankel, 2006), both acting to reduce the collateral value of inventories.

Figure B.7: Hoarding?

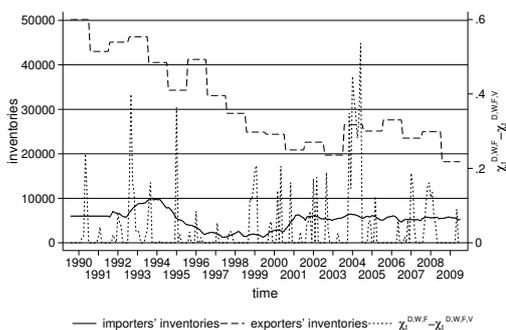


Table B.7: Granger causality hoarding

left-hand side	right-hand side	χ^2	df	Prob $>\chi^2$
exporters	futures price	4.6304	6	0.592
futures price	exporters	4.3863	6	0.625
importers	futures price	6.6043	6	0.359
futures price	importers	15.993	6	0.014

Notes: exporters = exporters' inventory; importers = importers' inventory; futures price of 3rd closest contract, measured at beginning of month.

Table B.7 reports results from a Granger causality test, conducted after running a straightforward VAR model (with a lag structure of up to 6 months). From the Table, we learn that the futures price indeed Granger-causes a change in inventories (the impulse response, now shown, is positive) for importers. For exporters, no effect is found.

B.8. Good harvest

Another subtle implication from our model is the impact of production on the link between the futures market and the spot price. The argument here is as follows: when producers are faced with poor harvests, they simply cannot afford *not* to sell spot. And after a big harvest, producers have a strong incentive *not* to sell spot, but cannot expect to gain much by doing this.

Therefore, the model suggests that the link between the futures market and the spot price is particularly strong when there is a relatively good harvest.

Figure B.8 shows the development of both $\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}$ and the difference, at any time, between production and consumption of coffee (plotted here using a kernel density estimator). In line with the model, we expect increases in inefficiency to occur when there is excess production. We note however that, in our empirical setup, an increase in production itself is expected to lower the price (since we include $1/\text{production}$ as an input).

Figure B.8: Excess production and inefficiency

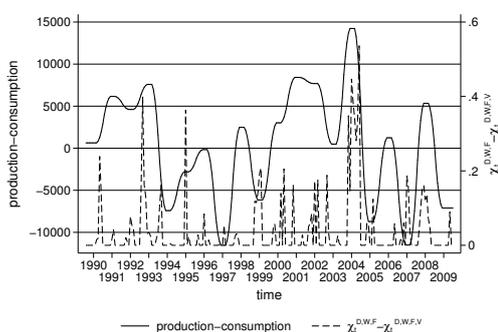


Table B.8: Formal tests

$\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}$	mean	std. dev.	obs.
production-consumption <= med.	0.0256	0.0606	120
production-consumption > med.	0.0519	0.1039	120
Hypothesis	t-test	KW	KS
$H_0^2: \Delta \chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}} = 0$	0.0172	0.0112	n.a.

Notes: med.=median; $\Delta \chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}$ = discrete difference between value for both groups; p-values shown for both tests; KW= Kruskal-Wallis rank test; KS = Kolmogorov-Smirnov test.

To test whether the effect is still significant, we split the distribution of the difference between production and consumption at the median. We then test whether $\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}$ is significantly higher when 'excess production' is above the median. The results from this test are included in Table B.8 and confirm that inefficiency is indeed significantly higher.²⁹

C.1. Tail effects

When we introduced our hypothesis tests, we emphasized the importance of not solely testing for an *average* effect of speculation on the spot price. As a result, we used both parametric (student t) and non-parametric (Krusal-Wallis, Kolmogorov-Smirnov) tests for our hypotheses.

²⁹Note that we cannot perform a Kolmogorov-Smirnov (KS) test here, since the split at the median means that the distributions are no longer continuous over time.

For robustness purposes, we now take things one step further, by testing whether our hypotheses hold up once we ignore the more extreme values further out in the tail of the inefficiency distributions.

In most of the cases presented here, the bulk of the density of inefficiency is located relatively close to zero. But what about the most extreme inefficiencies? Figure C.1 illustrates this, by plotting all inefficiency distributions when χ_t^D is above its median.

Figure C.1: Tail effects

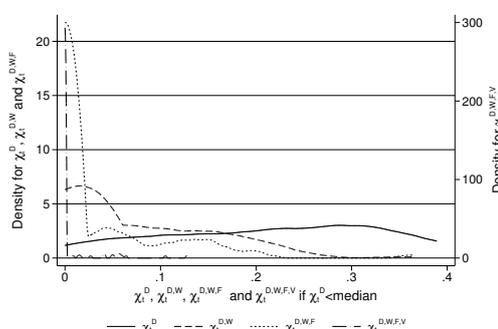


Table C.1: Hypotheses tests without the tail

Hypothesis	5th	25th	50th
$H_0^1: \chi_t^D = 0$	0.0000	0.0000	0.0000
$H_0^1: \chi_t^{D,W} = 0$	0.0000	0.0000	0.0000
$H_0^1: \chi_t^{D,W,F} = 0$	0.0000	0.0000	0.0000
$H_0^2: \chi_t^{D,W,F} - \chi_t^{D,W,F,V} = 0$	0.0000	0.0000	0.0000
$H_0^3: \chi_t^{D,W,F,V} > 0$	0.0000	0.0000	0.0000

Notes: KS = Kolmogorov-Smirnov test; 5th (25th, 50th) means that 5th (25th, 50th) highest percentile of the inefficiency distributions is ignored.

For H_0^2 , we base the cut-off on $\chi_t^{D,W,F}$.

In Table C.1, we test each of our hypotheses after ignoring the inefficiencies above the 5th, 25th and 50th highest percentile respectively, using Kolmogorov-Smirnov tests. As is evident from the Table, the hypothesis tests are highly robust to the exclusion of extreme values. Note that this is particularly true for our second hypothesis, where we test the difference between inefficiency *without* and *with* the inclusion of speculators' positions.

C.2. Volatility tests

Our paper focuses on the possible effect of futures market speculation on the level of spot prices. But what about spot price volatility? Empirical evidence with respect to volatility increases is much more prevalent (Irwin and Sanders, 2010).

In order to test whether our approach confirms that futures market speculation has increased spot price volatility, we perform two sets of additional tests. In the first set, reported in Table C2.a, we perform a Welch test on the equality of variances on two prices: the original price of coffee, and the price of coffee we derive by excluding the effect of speculation. The test indeed confirms that we cannot reject the hypothesis that speculation has increased the volatility of the price of coffee.

In the second set, reported in Table C2.b, we test for volatility differences between each of our inefficiency measures. The interpretation of results is

straightforward: with each extension of the set of inputs in our model, the volatility of the resulting inefficiency score drops.

Table C2.a: Variance test on price

Descriptives		
	price	price _{no spec}
std. dev.	38.507	39.742
Tests		
	<i>f</i>	<i>Pr(F > f)</i>
price - price _{no spec}	0.939	0.000

Notes: std. dev. = standard deviation; F-test for variance; price_{no spec} = 1 - ($\chi_t^{D,W,F} - \chi_t^{D,W,F,V}$) × price.

Table C2.b: Variance test on inefficiency

Descriptives				
	χ_t^D	$\chi_t^{D,W}$	$\chi_t^{D,W,F}$	$\chi_t^{D,W,F,V}$
std. dev.	0.236	0.133	0.117	0.067
Tests				
	<i>f</i>			<i>Pr(F > f)</i>
$\chi_t^D - \chi_t^{D,W}$	3.138			0.000
$\chi_t^{D,W} - \chi_t^{D,W,F}$	1.312			0.018
$\chi_t^{D,W,F} - \chi_t^{D,W,F,V}$	3.040			0.000

Notes: std. dev. = standard deviation; F-test for variance.

C.3. Spurious correlation

Some of the robustness tests we have carried out so far have been aimed at ensuring that we are not falling into a dimensionality trap which would result in lower inefficiency scores purely because of *adding* more inputs, rather than because of the *impact* of those inputs on the price.

Here, we continue along those lines, but in a somewhat different manner: we randomize the positions of futures market participants, by redistributing the actual positions in a random manner across our sample period. In doing so, we explore whether our inefficiency scores have been the result of spurious correlation.

Figure C.3: Distributions with randomized futures positions

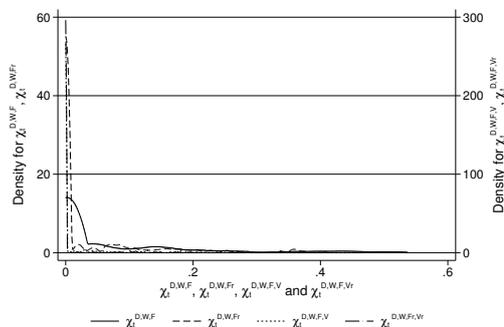


Table C.3: Hypothesis tests with randomized futures positions

Descriptives	mean	std. dev.	obs.
$\chi_t^{D,W,F}$	0.0622	0.0075	240
$\chi_t^{D,W,Fr}$	0.0367	0.0052	240
$\chi_t^{D,W,F,V}$	0.0235	0.0043	240
$\chi_t^{D,W,Fr,Vr}$	0.0194	0.0041	240
Hypothesis	t-test	KW	KS
$\chi_t^{D,W,F} = \chi_t^{D,W,Fr}$	0.0000	0.0001	0.0000
$\chi_t^{D,W,F,V} = \chi_t^{D,W,Fr,Vr}$	0.1734	0.0277	0.0000

Notes; KW= Kruskal-Wallis rank test; KS = Kolmogorov-Smirnov test.

Figure C.3 shows the resulting inefficiency distributions, both for $\chi_t^{D,W,F}$ and $\chi_t^{D,W,F,V}$ - the original inefficiency scores - and $\chi_t^{D,W,Fr}$ and $\chi_t^{D,W,Fr,Vr}$ - with randomized futures positions respectively for hedgers, and for hedgers and speculators.

In Table C.3, we formally test whether randomizing futures positions has an effect on the inefficiency scores. If futures positions are indeed strategic, we may expect inefficiency scores with randomized positions to be significantly lower. In the Table, we learn that average inefficiency indeed drops once we randomize futures positions, both for $\chi_t^{\text{D,W,F}}$ and for $\chi_t^{\text{D,W,Fr,V}}$. The difference however, is only significant for $\chi_t^{\text{D,W,F}}$: put simply, these results show that speculators' futures market positions are indeed *not* spuriously correlated with spot prices.

C.4. Likelihood of demand/supply shift

An important ingredient in our theoretical model is the inclusion of inputs that cause a demand or supply shift, thus affecting the spot price of coffee. Assuming there is a channel from the futures market to the spot market, a demand or supply shift in the former market still does not have to affect the price in the latter. To see why, consider a simultaneous increase in demand and supply of coffee futures: the result is an increase in the equilibrium number of futures traded, but the price of coffee futures would be unaffected, as would be the spot price of coffee.

As it happens, most commodity futures markets have gone through big structural changes over the past twenty years (Masters, 2008). Not only have volumes increased tremendously, the composition of short and - in particular - long positions has changed as well. In the beginning of our sample period, more than two thirds of long futures positions for most commodities were held by commercial parties. By the end of our sample period, these parties held less than one third in many markets, the remainder of the demand for long futures coming from non-commercial participants (Masters, 2008).

We can use the fact that commodity futures index trading did not become very popular until the beginning of this century, as an identification mechanism, to find out more about the relationship between futures and spot market developments.

In Figures C4.a and C4.b, we plot long and short positions in the futures market for the first part of our sample period (1990-2000) and the last part (2001-2009), respectively. The most important change, going from the former graph to the latter, is the rise of futures index trading, resulting mainly in the increase in the demand for long futures. From Figure C4.a, we observe no clear pattern. Things change when we turn to Figure C4.b, where we observe a coincidence between the increase in long futures held by noncommercial participants and $\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,Fr,V}}$, the part of the coffee price that is attributed to speculation.

Figure C4.a: 1990-2000

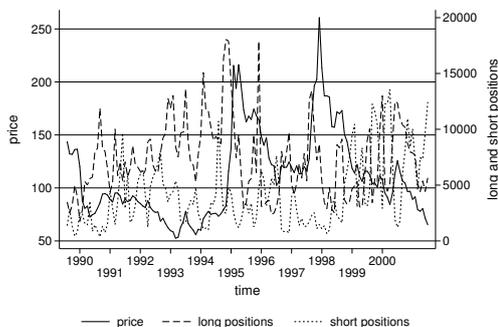
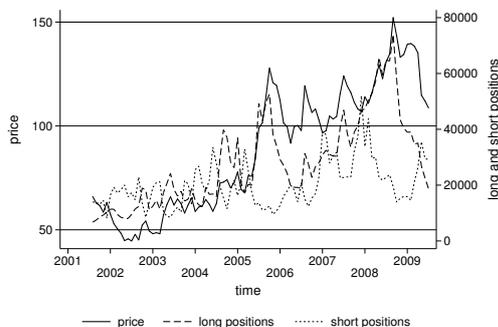


Figure C4.b: 2000-2009



To get a feel for the economic significance of the pass-through from futures to spot market, we regressed $\chi_t^{D,W,F} - \chi_t^{D,W,F,V}$ on the (3-month lagged) long and short positions of non-commercials, in a simple regression. Both positions have a significant coefficient, which - as expected - is positive for short and negative for long positions. Adjusted R^2 for the regression is 0.04. Considering the elasticities, we find that a one percent increase in the long position decreases $\chi_t^{D,W,F} - \chi_t^{D,W,F,V}$ by 0.42%. Likewise, an increase in the short position of non-commercials of one percent results in a 0.58% increase in $\chi_t^{D,W,F} - \chi_t^{D,W,F,V}$. Since the latter is still expressed as a percentage of the price, we can interpret both elasticities in a fairly straightforward manner.

C.5. Model fit

The final item on our list of robustness checks related to our hypothesis tests concerns the fit of our model. Results from testing our third hypothesis in particular suggest that our analysis does not suffer from an omitted variable problem: once we include noncommercial futures market positions, the spot price of coffee is 'fully' explained.

Since we use a non-parametric model and emphasize the 'spikiness' of the main effect we are interested in, using a least squares measure of fit (such as R^2) is not obvious. As a simple robustness check, we nevertheless regress the actual price of coffee on the price with speculation. The resulting R_{adj}^2 is 0.99, and the coefficient for the latter variable is of the order of 0.9676 with a t-value of 126.99.

In Figures C5.a and C5.b, we attempt to demonstrate the fit of our model in two ways. First, in Figure C5.a, we plot the actual price of coffee and the price of coffee 'with the remaining inefficiency' in our model, after including the position of speculators. Visual inspection confirms the regression results. Figure C5.b contains the actual price as well as the 'predicted' price, *without*

speculation, and it tells a different story.

Figure C5.a: Price with speculation

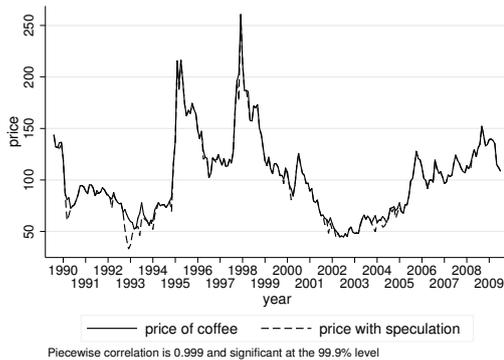
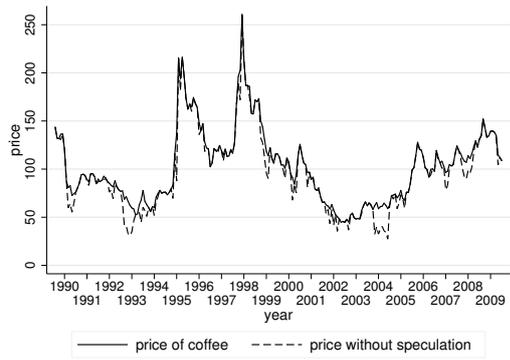


Figure C5.b: Price without speculation



On the one hand, it reinforces what is already clear from Figure 3b, in the paper, that the effect of speculation on the price of coffee is spiky. On the other hand, it demonstrates what others (Irwin and Sanders, 2010) have previously stated before: speculation is not the most important driver of spot prices.

D.1. Significance

Although we have tried hard to present a convincing case when it comes to our modeling strategy, we realize that our choice of a non-parametric model has - at least - one downside: it makes it more difficult to compare our results with the rest of the literature. Therefore, in this part of the Appendix, we carry out a series of robustness tests aimed at confronting our results with those obtained by others. In particular, we shall emphasize the importance of establishing *causality*, by reverting to the same types of analysis as those used by Irwin et al. (2009).

Figure D1.a: $\chi_t^{D,W,F,V}$

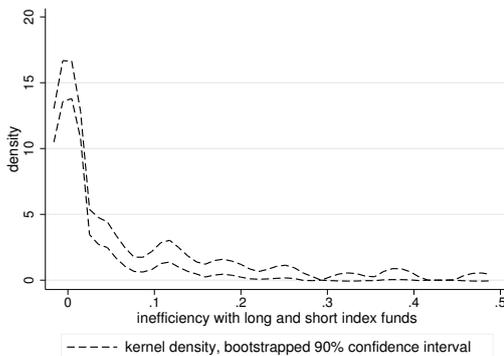
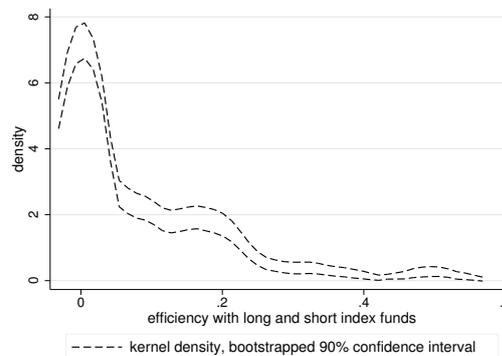


Figure D1.b: $\chi_t^{D,W,F} - \chi_t^{D,W,F,V}$



Before we do so, however, we pause for a moment to consider the importance of (the lack of) stochastic properties of our modeling approach. As stated in the paper, all of our hypothesis tests, as well as the rest of our analyses, are carried out on bootstrapped inefficiency scores, following Simar and Wilson (1998).

But what if differences in inefficiency scores *after* adding speculative long and short positions reflect changes in the underlying data generating process? In order to find out whether our results are robust to this, in Figure D1.a we start by plotting $\chi_t^{\text{D,W,F,V}}$ after another round of bootstrapping (this time with a more traditional bootstrap, instead of the Simar and Wilson (1998) procedure already applied for the initial estimations). Instead of plotting the (bootstrapped) mean inefficiency scores, in this figure we report the upper and lower confidence bounds for the 90% confidence interval. What we hope to see is a small confidence interval that does not vary much in size along the horizontal axis (i.e., as inefficiency increases). This is indeed what Figure D1.a shows.

Subsequently, and more interestingly, we apply the same line of thinking to $\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}$, in Figure D1.b. Again, the confidence interval is small and does not vary much in size as inefficiency increases, suggesting that our results are rather robust.

D.2. VAR analysis: impulse response of coffee price

In Irwin and Sanders (2010) and Irwin et al. (2009), two key tools are employed to establish causality. First, the authors make use of vector auto regression (VAR) models to study impulse responses from increases in - particularly - long non-commercial futures positions. Second, and related, the authors make use of Granger causality tests. As explained in the body of our paper, both types of methods are not without their drawbacks, particularly in this specific context. In the following subsections of this Appendix, we shall nevertheless repeat their analyses. We do so, however, with a slight twist: instead of using only the spot price of coffee as a unit of analysis, we also - in addition - use the part of the spot price that, according to our results, is the result of speculation.

We start in Figure D.2 and Table D.2 with a simple VAR analysis. Table D.2 contains the key statistics from the analysis. Figure D.2 shows the impulse response to the spot price of coffee from a one standard deviation increase in the long position of speculators. As in Irwin and Sanders (2010) and Irwin et al. (2009), the Figure shows that there is, on average, no impact from speculation on the (spot) price.

Figure D.2: Impulse response of coffee price

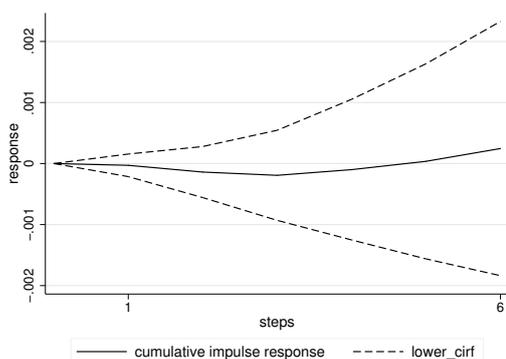


Table D.2: VAR results

Sample	7-240	N	234
Log likelihood	-3159.412	HQIC	21.6758
FPE	2.29e+09	SBIC	21.8873
Equation	K	R ²	P> χ^2
long speculators	13	0.9126	0.0000
price	13	0.9220	0.0000

Notes: FPE= final prediction error; N = number of observations (lag order = 6); K= number of parameters.

In our paper, we argue that this result is basically a 'statistical artifact'. To see whether that is indeed the case, we now proceed to explore these results a bit further.

D.3. VAR analysis: variance of impulse response of coffee price

We start doing so in Figure D.3 and Table D.3 with a variance decomposition of the impulse response in Figure D.2: after all, our argument has been that, just because there is no evidence of an *average* effect, this does not mean that there is not, on occasion, an extreme (or spiky) effect of speculation on the price.

Figure D.3 shows the result graphically. In essence, this figure can be interpreted in much the same way as the previous one, with one important difference: instead of considering the first moment of the distributions of variables of interest (price, long positions), we now consider the second moment. Indeed, the figure shows that there is a positive effect. In simple terms, although based on Figure D.2 there is no evidence based on a VAR analysis of an *average* effect, based on Figure D.3 we cannot rule out that there is occasionally an extreme effect. Importantly, both figures result from the *same* analysis.

How can we tell whether our interpretation of Figure D.3 does not in itself present a statistical artifact? To see whether there is evidence of this, we run a fairly simple test, presented in Table D.3, following Lütkepohl (2005). The Table contains, for a given lag length, LR test statistics compared to a VAR model with one less lag. The largest drop in final prediction error and sharpest decrease in AIC occur when we add the first lag, suggesting that the optimal number of lags for our model is indeed one. The likelihood ratio test, however,

puts the optimal number of lags at three. In either case, it does not appear that our interpretation of Figure D.3 is the result of an insufficient lag length.

Figure D.3: Variance of impulse response of coffee price

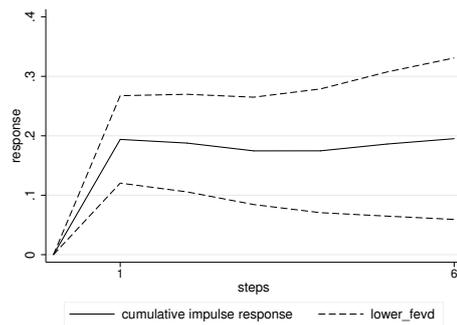


Table D.3: Selection-order criteria

lag	LR	df	p	FPE	AIC
0	n.a.	n.a.	n.a.	3.2e+11	26.4597
1	1172.2	4	0.000	2.2e+09*	21.4844*
2	5.0	4	0.285	2.2e+09	21.4971
3	9.6*	4	0.049	2.2e+09	21.4905
4	3.6	4	0.466	2.2e+09	21.5094
5	3.0	4	0.554	2.3e+09	21.5306
6	7.5	4	0.113	2.3e+09	21.5329

Notes: FPE= final prediction error.

D.4. VAR analysis: normality tests

How appropriate are VAR estimations when applied to spot prices and long futures market positions in the first place? In order to find out, we consider the distributional properties of both variables. After all, VAR models use least squares to minimize error, and as a result work best when variables considered are normally distributed.

Figure D.4: Long speculators and price

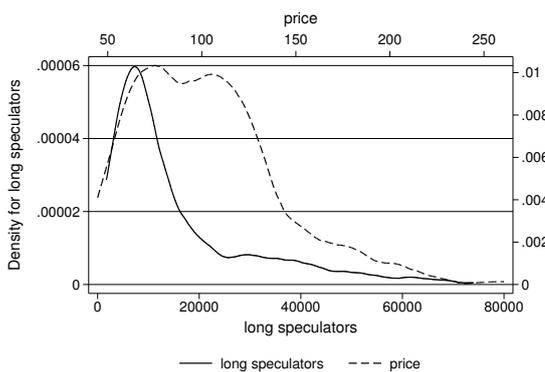


Table D.4: Normality tests

equation	χ^2	df	Prob > χ^2
Jarque-Bera test			
long speculators	152.287	2	0.00000
price	2729.691	2	0.00000
Skewness test			
long speculators	9.484	1	0.0021
price	180.438	1	0.0000
Kurtosis test			
long speculators	142.803	1	0.0000
price	2549.253	1	0.0000

Figure D.4 shows the distributions of both non-commercial long positions (the 'speculators') and the spot price of coffee. As is confirmed by the normality tests in Table D.7, both variables are skewed. In addition, the spot price of coffee in particular is characterized by a high degree of kurtosis. As a result, the mean of the two variables is not as informative as it would be if they were normally distributed, as is also clear from the confidence interval in Figure D.2.

D.5. Impulse response of $(\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}) \times \text{coffee price}$

The previous robustness tests have shown that although a standard VAR test yields the same result (no response of the coffee price to change in long speculators' position) as in previous research, a variance decomposition based on the same VAR estimates suggests that there may be an incidental response. We therefore now proceed by *repeating* the VAR estimates, this time replacing the main variable of interest, the coffee price, with its 'speculative' equivalent, i.e., $(\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}) \times \text{coffee price}$. Based on our model and empirical results, we expect two outcomes. First, we expect a positive impulse response to changes in long speculators' position, although the confidence interval may still be relatively large. Second, we expect no effect on the variance of $(\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}) \times \text{coffee price}$; after all, the latter already internalizes the spiky nature of the speculators' impact.

Figure D.5: Impulse response

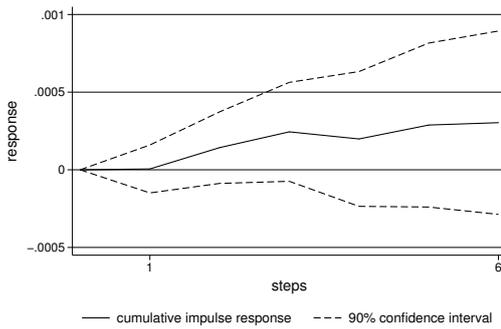


Table D.5: VAR results

Sample	7-240	N	234
Log likelihood	-3042.996	HQIC	20.6808
FPE	8.45e+08	SBIC	20.8923
Equation	K	R ²	P>χ ²
Long speculators	13	0.9130	0.0000
price	13	0.2265	0.0000

Notes: FPE= final prediction error; N = number of observations (lag order = 6); K= number of parameters.

Figure D.5 and Table D.5 relate to our first expectation. Indeed, we observe that - although the confidence interval is still relatively large - there is now a positive response to increased speculation. With a reasonable fit for the equation on which the impulse response is based, an R_{adj}^2 of 22.65, these results suggest that even with a mean-variance based technique such as a VAR model, there is some evidence that speculation affects the coffee price.

D.6. Variance of impulse response of $(\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}) \times \text{price}$

Figure D.6 and Table D.6 address the second expectation raised above. Indeed, these results show that that is no significant effect of long speculators on the variance of the 'speculative' part of the coffee price.

Figure D.6: Variance of impulse response

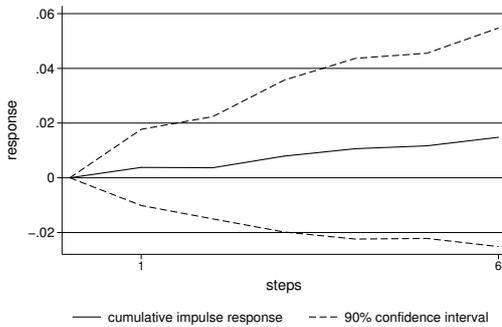


Table D.6: Selection-order criteria

lag	LR	df	p	FPE	AIC
0	n.a.	n.a.	n.a.	1.0e+10	23.0352
1	608.4*	4	0.000	7.9e+08	20.4692
2	8.5	4	0.076	7.9e+08	20.4672
3	8.9	4	0.064	7.8e+08*	20.4635*
4	0.9	4	0.923	8.1e+08	20.4937
5	4.1	4	0.387	8.2e+08	20.5102
6	1.5	4	0.822	8.5e+08	20.5379

Notes: FPE= final prediction error.

D.7. VAR analysis, Normality tests using $\chi_t^{D,W,F} - \chi_t^{D,W,F,V} \times price$

Naturally, given our earlier remarks regarding the relevance of the (normality of) distributions in VAR models, we now repeat the analysis for $\chi_t^{D,W,F} - \chi_t^{D,W,F,V} \times price$ and the long position of speculators.

Figure D.7: Long speculators and $(\chi_t^{D,W,F} - \chi_t^{D,W,F,V}) \times price$

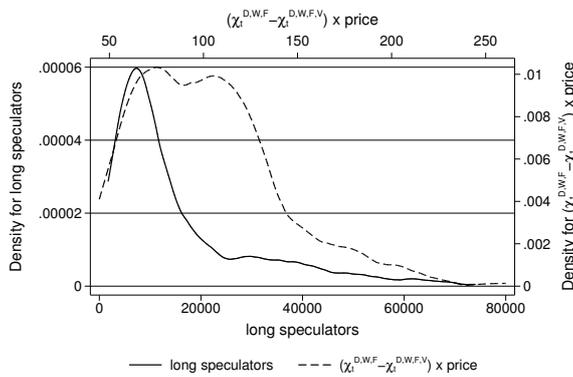


Table D.7: Normality tests

equation	χ^2	df	Prob > χ^2
Jarque-Bera test			
long speculators	128.917	2	0.00000
price	2835.113	2	0.00000
Skewness test			
long speculators	8.634	1	0.0033
price	301.668	1	0.0000
Kurtosis test			
long speculators	120.283	1	0.0000
price	2533.445	1	0.0000

Notes: FPE= final prediction error.

Results are similar to what we have found before: both variables are skewed, and the 'speculative' part of the coffee price is also characterized by a high degree of kurtosis.

D.8. Causality: Granger causality tests

The second tool employed by Irwin and Sanders (2010), Irwin et al. (2009) and others to establish causality, is a Granger causality test. The long position of speculators is said to Granger-cause the coffee price if past values of the

latter significantly contribute to predicting the former, *given* past values of the former.

In Figures D8.a and D8.b, we run Granger causality tests. In Figure D8.a, we replicate along the lines of Irwin and Sanders (2010). Whereas Irwin and Sanders (2010) use net future positions, we use only the long positions. Results are not qualitatively affected by this, however. Another difference compared with Irwin and Sanders (2010), is the fact that we estimate a VAR model for the coffee market, whereas Irwin and Sanders (2010) estimate a SUR model across all commodities markets. The reason for their choice of a SUR model is probably the fact that they focus on index speculators. For that group, data are available for a much shorter time period, requiring them to rely more on cross-sectional variation. In our analysis, index speculators are the most likely to affect the spot price, but they by no means constitute the only group of futures speculators that can have an effect. We therefore focus on the total positions taken by non-commercial parties, and explore in detail the time variation both in the coffee price and in those positions.

Table D8.a: Granger causality

excluded	χ^2	df	Prob > χ^2
<i>long speculators</i>			
price	8.6194	6	0.196
<i>price</i>			
long speculators	3.2355	6	0.779

Table D8.b: Granger causality, $\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}$

excluded	χ^2	df	Prob > χ^2
$(\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}})$			
price	5.4409	6	0.489
<i>price</i>			
$(\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}})$	11.809	6	0.066

Despite these differences, our initial Granger causality results are similar to those obtained by Irwin and Sanders (2010): we cannot reject the null hypothesis that the long position of speculators has no effect on the coffee price. Interestingly enough, we do not find evidence of Granger causality in the opposite direction either (Irwin and Sanders (2010) do not report these results).

What happens when we replicate these tests in line with our own approach? In fact, this is not straightforward: since the long position of speculators is an input in the model, establishing Granger causality from that position to the 'speculative' part of the coffee price can be merely a statistical artifact. Instead, therefore, we opt for something different. We test for Granger causality between $\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}$ and the coffee price. In so doing, we rely on the fact that the only difference between the two inefficiency measures is the positions taken by speculators. By doing this, we therefore minimize a common pitfall in Granger causality tests: the existence of an unobserved third variable which affects both the price and the speculators' positions. Also, this test is more closely related to what Irwin and Sanders (2010) aim to do, since we now pick up the effects of the net position of speculators.

The results for this second set of Granger causality tests are summarized in Table D8.b. First, we observe that we cannot reject the null hypothesis that the coffee price has no effect on $\chi_t^{D,W,F} - \chi_t^{D,W,F,V}$. However, with a p-value of 6.6%, Granger causality in the opposite direction is much more likely.

In the last set of robustness analyses, we have tried to the best of our abilities to replicate tests from previous papers. In doing so, we have tried to find a balance between staying as close as possible to those papers, and staying as true as possible to our modeling approach here. At the very least, our results indicate that even with more traditional approaches such as those replicated here, concluding that there is no evidence of a causal link between futures market speculation and spot prices is far from straightforward.

D.9. Lags, 6 month lag in DEA model (maximum between harvests)

Before we turn to some policy implications, we take the opportunity to conduct two more robustness tests. Both tests address the issue of 'simultaneity', and in doing so corroborate our results. In fact, in what follows here, we are not solely aiming to further our analysis in terms of causality; we are also trying to stay closer to our theoretical model, where speculators' positions can take some time to affect prices through inventory levels.

Figure D.9: Distributions with lags

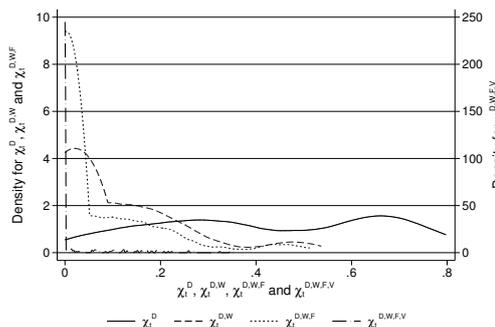


Table D.9: Hypotheses tests with lag

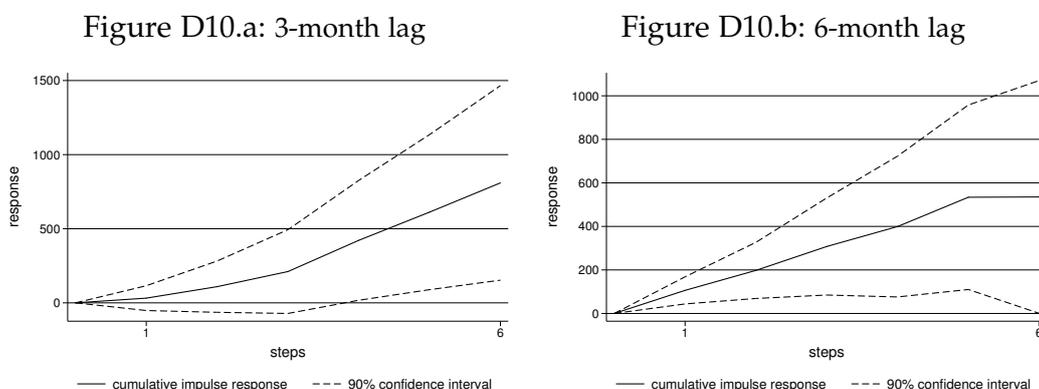
Hypothesis	t-test	KW	KS
$H_0^1: \chi_t^D = 0$	0.0000	0.0001	0.0000
$H_0^1: \chi_t^{D,W} = 0$	0.0000	0.0001	0.0000
$H_0^1: \chi_t^{D,W,F} = 0$	0.0000	0.0001	0.0000
$H_0^2: \chi_t^{D,W,F} - \chi_t^{D,W,F,V} = 0$	0.0000	0.0001	0.0000
$H_0^3: \chi_t^{D,W,F,V} > 0$	0.0000	0.0001	0.0000

Notes: KW= Kruskal-Wallis rank test; KS = Kolmogorov-Smirnov test.

In Figure D.9 and Table D.9, we repeat our earlier analysis, but this time include a 6-month (= minimum of one harvest) lag for all inputs, including speculators' long and short positions. From the figure, we observe that including a 6-month lag does not change the inefficiency distributions much. In Table D.9, we see that all hypotheses test results remain unaltered.

D.10. Lags, Impulse response for 3 and 6 months

Now that we have established that including inputs with a 6-month lag does not change our conclusions, we ask whether a lagged effect alters the impact of an increase in the long speculators' position. To do so, in Figures D10.a and D10.b we consider the impulse response from an increase of that same position on the lagged values of $(\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}) \times \text{price}$. We include the latter in our VAR model with both a 3 and a 6 month lag, respectively, to ensure that we cover at least one harvest cycle (coffee can be harvested up to 4 times a year).



If anything, we expect to see a more pronounced effect, compared with Figure D.5, since our model suggests that inventory effects play a role. The latter may take some time to materialize (Deaton and Laroque, 1996). In Figure D10.a, we observe that the impulse response is now indeed significantly different from zero after a few steps (= months). In Figure D10.b, the response is immediately significantly different from zero.

E.1. Substitution effects: with S&P and price of oil

Now that we have thoroughly investigated causality, it is time to turn towards the policy consequences that emanate from our paper.

In our conclusion, we argue that it is worthwhile to reconsider the preeminence of oil in most commodity indices, given the destabilizing effects oil can have on other commodity prices. In particular, we argue that there may be a substitution effect: as the relative returns on oil drop compared with other commodities in the index, commodity futures index funds shift more of their investments into other commodities, strengthening the speculative channel we identify in this paper.

In Figure E.1 and Table E.1 we investigate this substitution effect, albeit in a rather rudimentary manner. In Table E.1, we calculate the simple correlations

between $\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}$, the price of oil and the S&P500. If the substitution effect exists, we can expect a negative correlation with $\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}$. Indeed, this is the case for oil, although the correlation is not significant.

Figure E.1: Substitution effects

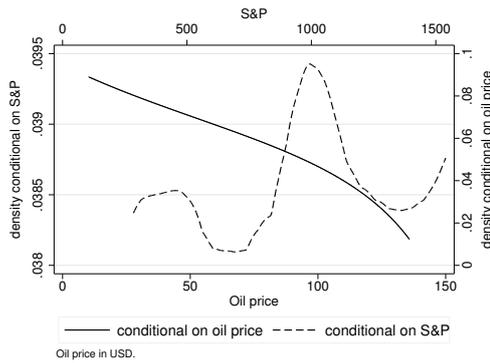


Table E.1: Correlations

	$\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}$	oil price	S&P
$\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}$	1.0000		
oil price	-0.0367	1.0000	
S&P	0.0667	0.5596	***1.0000

Notes: Pairwise correlation; * * * = significant at 1% level.

In Figure E.1, we plot the kernel density of $\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}$, conditional on the price of oil and the S&P500, respectively. If the substitution effect exists, we should see a downward sloping curve: $\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}$ is highest, when the alternative asset is at its lowest price. Indeed, this is the case for oil. For the S&P500, there is no clear pattern (as expected).

E.2. Substitution effects: changes in price of oil

The analysis we presented above regarding the substitution effect, we considered the *level* of the price of oil. In our description of the substitution effect, however, we focused on *changes* of the oil price. More specifically, we argued that what might mitigate the relationship between speculation and the coffee price was a sharp increase in the price(s) of other commodities in the index.

We now turn to this in Tables E2.a and E2.b. In Table E2.a, we split our sample into two parts: one part where the oil price is increasing, and one part where it is not. We then compare $\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}$ for both groups, using a simple t-test, with a Welch correction, to allow for unequal variances. Also, we conduct a Kruskal-Wallis rank test. In Table E2.b, we do the same, but now for the second order effect, i.e., the increase in the increase of the oil price.

If the substitution effect exists, it should show up in Table E2.b in particular. It is indeed there that we find a significant difference in the average values of $\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}$: when the oil price is increasing rapidly, $\chi_t^{\text{D,W,F}} - \chi_t^{\text{D,W,F,V}}$ is significantly lower than when the oil prices in not increasing rapidly.³⁰

³⁰Note though that the p-value from the Kruskal-Wallis rank test is still relatively high, at 0.089.

Summing up, from our analyses here and above, we conclude that there is some evidence in favor of a substitution effect.

Table E2.a: Δ oil price

		$\chi_t^{D,W,F} - \chi_t^{D,W,F,V}$		
	Obs	Mean	Std. Err.	Std. Dev.
Δ oil price ≤ 0	103	0.035	0.008	0.082
Δ oil price > 0	137	0.042	0.008	0.089
$t = -0.607$		Welch's DF = 229.604		
Ha: difference < 0		$p = 0.272$		
Ha: difference $\neq 0$		$p = 0.544$		
Ha: difference > 0		$p = 0.728$		
$\chi^2 = 0.843$		$p = 0.358$		

Δ oil price = percentage change of the coffee price; Welch's DF = Welch's degrees of freedom (testing with unequal variances); χ^2 results from Kruskal-Wallis rank test, with ties.

Table E2.b: Δ^2 oil price

		$\chi_t^{D,W,F} - \chi_t^{D,W,F,V}$		
	Obs	Mean	Std. Err.	Std. Dev.
Δ^2 oil price ≤ 0	168	0.043	0.007	0.091
Δ^2 oil price > 0	72	0.028	0.009	0.074
$t = 3.410$		Welch's DF = 118.871		
Ha: difference < 0		$p = 1.000$		
Ha: difference $\neq 0$		$p = 0.001$		
Ha: difference > 0		$p = 0.000$		
$\chi^2 = 2.893$		$p = 0.089$		

Δ^2 oil price = percentage change in Δ oil price; Welch's DF = Welch's degrees of freedom (testing with unequal variances); χ^2 results from Kruskal-Wallis rank test, with ties.

E.3. Market impact

In our introduction, we mentioned the 'limits to arbitrage' as an explanation of why speculators might drive up the futures price. In line with Acharya et al. (2011), we argued that the change in the micro structure of commodities futures markets resulting from a large influx of - particularly - long speculators, could have an effect on futures prices, and thereby on spot prices.

Here, we take this argument a step further. In particular, we investigate the possibility of market impact (Fafchamps and Hill, 2008). In Figure E.3, we plot the total volume held by long speculators as well as the ratio of long speculators to total futures trading. The figure shows that from 2002 onwards, increases in the total market volume coincide with the influx of long speculators.

If there is market impact, then according to our model (i) the effect on the price is expected to persist for some time, and (ii) the impact should be reflected in $\chi_t^{D,W,F}$ and the price, but not in $\chi_t^{D,W,F,V}$.

In Table E.3, we therefore test for momentum (Swed and Eisenhart, 1943). We use a so-called run-test, to see whether observations of inefficiency scores and the price, respectively, are serially independent. To do so, we apply a threshold level and count how many 'runs' are above or below this threshold. For the inefficiency scores, the threshold level is zero, i.e., we effectively test how many times consecutively inefficiency is higher than zero. For the price, in line with the standard test procedure, we use the median as the threshold

level. For all variables, it is the case that positive/negative serial correlation will result in a low/high number of runs.

Figure E.3: Evidence of market impact?

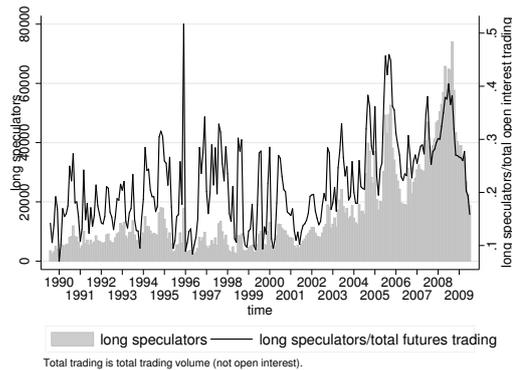


Table E.3: Testing for momentum

Variable	≤ 0	> 0	N(runs)	Prob $> z $
χ_t^D	2	238	5	0.880
$\chi_t^{D,W}$	5	235	9	0.000
$\chi_t^{D,W,F}$	9	231	11	0.000
$\chi_t^{D,W,F,V}$	52	188	44	0.000
price	120	120	11	0.000

Notes: non-parametric run-test, with threshold(0), except for the price, where the median is used as the threshold.

From Table E.3, we observe that the number of runs is below what can be expected (based on complete serial independence) for all variables.³¹ However, the number of runs is exactly four times as high for $\chi_t^{D,W,F,V}$ as for $\chi_t^{D,W,F}$ and the price. Summing up, positive serial correlation is much less likely once we control for futures speculators. Of course, this is indirect evidence of market impact. More robust statistical analyses would require the availability of longer disaggregated series for futures market participants (cf., Section A.3).

E.4. Impact on Fair trade

One of the key motivations for many who criticize 'food speculation' is its impact not just on consumers, but also on producers. For example, has speculation affected fair trade, by making the latter less attractive to coffee producers?

Fair trade coffee producers receive a guaranteed minimum price, which has remained largely unaltered in the last 20 years. If the world price is higher than this fair trade price, the latter is adjusted, and the producer receives a small markup. Figure E.4 shows the development of the price, the fair trade price and the price without speculation. For speculation to have a negative effect on fair trade, we would expect to see that it is (partly) responsible for pushing the actual price above the guaranteed minimum fair trade price. As is clear from the figure, this is seldom the case, with the exception of the late 1990s and the late 2000s.

³¹Note that all p-values in the table are for two-sided tests. Since we are interested in the one-sided p-value, we should divide all values by two. Also, values equal to threshold are treated as if they are below the threshold.

Figure E.4: Fair trade

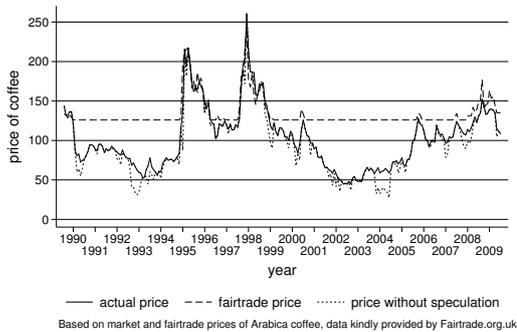


Table E.4: Relation with fair trade

	$\chi_t^{D,W,F} \times \text{price}$			
	Obs	Mean	Std. Err.	Std. Dev.
fair trade \leq price	23	3.503	1.226	5.882
fair trade $>$ price	217	5.882	0.610	8.981
$t = 0.916$	Welch's DF = 35.080			
Ha: difference < 0	$p = 0.111$			
Ha: difference $\neq 0$	$p = 0.222$			
Ha: difference > 0	$p = 0.889$			
$\chi^2 = 9.073$	$p = 0.003$			

Notes: fair trade = fair trade price; Welch's DF = Welch's degrees of freedom (testing with unequal variances); χ^2 results from Kruskal-Wallis rank test, with ties.

In Table E.4, we look at $\chi_t^{D,W,F} \times \text{price}$, i.e., the part of the price that can be attributed to speculation. We then construct a dummy that takes on a value of one/zero, when the fair trade price is above/below or equal to the world price. Upon testing whether $\chi_t^{D,W,F} \times \text{price}$ is significantly higher when the world price has broken through the minimum fair trade price, we find no significant differences.

E.5. Consumer surplus

Is there a cost to consumers as a result of speculation? To answer that question, we estimate the loss of consumer surplus by considering the effect of $eff_{contr-efl\&s}$ on the coffee price. Assuming a constant price elasticity of demand, and a non-zero price elasticity of supply, we can calculate a rough estimate of the loss in consumer surplus as follows.

First, we estimate the price elasticity of demand by regressing the log of the price on the log of demand. Next, we can calculate the percentage change in the price as a result of speculation, and after multiplying by the total demand, we obtain the percentage change in demand, and the demand that would have existed *without* speculation.

Having now obtained both the change in price and the change in demand as a result of speculation, we can calculate the adjustment in consumer surplus. Comparing that to the original consumer surplus, we can calculate the percentage loss of consumer surplus.

Figure E.5: Consumer surplus loss

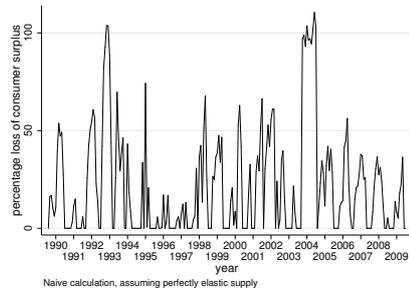


Table E.5: Input for calculations

Variable	Mean	Std. Dev.
price	102.040	38.507
$\text{price} \times (1 - \chi_t^{\text{D,W,F}})$	92.751	40.541
consumption	106,720	11,344
$\text{consumption, price} \times (1 - \chi_t^{\text{D,W,F}})$	106,823	11,344
original consumer surplus	5,434,398	2,063,625
adjustment in consumer surplus	994,976	1,147,148
% loss in consumer surplus	21.482	26.740

Notes: nobs=240; estimate of price elasticity of demand is -0.00897.

Key numbers from these calculations are reported in Table E.5. Average loss of consumer surplus is more than 20 percent. Figure E.5 depicts the estimated consumer surplus losses over time, showing that there is no clear pattern.