AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

Agrekon

Agricultural Economics Research, Policy and Practice in Southern Africa

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ragr20

Identifying possible misspecification in South African soybean oil futures contracts

Andre Van der Vyver, Ulonka Barnard & Jean-Pierre Nordier

To cite this article: Andre Van der Vyver, Ulonka Barnard & Jean-Pierre Nordier (2022) Identifying possible misspecification in South African soybean oil futures contracts, Agrekon, 61:1, 80-93, DOI: 10.1080/03031853.2021.2006070

To link to this article: https://doi.org/10.1080/03031853.2021.2006070

Published online: 08 Mar 2022.

Submit your article to this journal

Article views: 82

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ragr20

AGREKON
2022, VOL. 61, NO. 1, 80–93
https://doi.org/10.1080/03031853.2021.2006070

Identifying possible misspecification in South African soybean oil futures contracts

Andre Van der Vyver, Ulonka Barnard and Jean-Pierre Nordier

UT Grain Management, Pretoria, Gauteng, South Africa

ABSTRACT
With the inclusion of a locally traded soybean oil futures contract, that is dual-listed and cash-settled of the Chicago Board of Trade futures contract, the South African Futures Exchange (SAFEX) aimed to provide local soybean crushing plants, the opportunity for managing their exposure toward the variation in soybean oil prices using effective hedging strategies. Which is only viable assuming adequate liquidity, that is currently lacking in these futures contracts. The soybean oil contract used for hedging local price exposure should also reflect local import parity and/or be correlated to local price movements. Therefore, with most soybean oil usually being imported from Argentina, one would expect SAFEX soybean oil futures contracts to reflect the cost of imported soybean oil from Argentina. Hence, the research study used the Engle–Granger (1987) cointegration approach, alongside a range of diagnostic tests to determine whether SAFEX soybean oil futures contracts, that is dual-listed and cash-settled of CBOT settlement values is a misspecification and whether or not SAFEX soybean oil futures contracts should rather be based on the Argentina free-on-board soybean oil prices which is a much better representation of South Africa's import parity and local industry prices.

ARTICLE HISTORY
Received 11 October 2020
Accepted 9 November 2021

KEYWORDS
Futures contract specifications; soybean oil; statistical analysis; Engle–Granger cointegration

1. Introduction

Soybean crushing[1] plants operate on a crush margin, which is the monetary difference between the combined sales value of mainly soybean meal and soybean oil and the cost of raw soybeans. However, given the high volatility in the prices of these three products, crushing plants normally secure these prices simultaneously. If not, they are vulnerable to the relative price variation between these three products.

Futures markets, such as the Johannesburg Stock Exchange (JSE) Commodities Derivatives Market (CDM) (previously known, and hereafter referred to, as the South African Futures Exchange (SAFEX)), provide futures contracts that can be used as a mechanism for securing these prices. Soybean crushing plants would usually buy soybean futures contracts, i.e., taking a long position in the futures market (known as a long anticipatory hedge), thereby offsetting the risk against higher soybean prices, whilst simultaneously selling soybean meal and soybean oil futures contracts (in a ratio aligned with production), thereby offsetting the risk against lower soybean meal and soybean oil prices (known as a short anticipatory hedge). This strategy allows soybean crushing plants to "lock-in" their crushing margin. The crushing plant will again close the positions, first the long soybean futures when it buys the physical stock, and second the short futures (meal and oil

CONTACT Andre Van der Vyver andre@grainmanagement.co.za UT Grain Management, 44 Herbert Baker Street Groenkloof, Pretoria, Gauteng 0181, South Africa

© 2022 Agricultural Economics Association of South Africa

contracts) when it sells the physical product (the sequence could be the opposite as long as the futures positions are replaced by physical positions). After having taken the appropriate positions in the futures market, soybean crushing plants would be ensured of an adequate crushing margin, covering the factory operational costs and their net profit, as long as they correctly offset the futures positions with physical positions. If prices in the futures and/ or physical market move, in-sync, the loss in one of the positions will always be offset by a similar gain in the other. But this is only viable given adequate (1) liquidity[2] within these futures contracts (which is not the case for SAFEX soybean oil futures contracts), and (2) the oil and meal futures contracts are adequately correlated to the local physical import parity price.

Furthermore, if South Africa is a net importer of the underlying commodity (as is the case with soybean oil) the CBOT[3] contract (as traded on SAFEX) futures' price normally represents the majority of the import cost[4] also known as the import parity cost. Therefore, with most soybean oil usually being imported from Argentina, one would expect SAFEX soybean oil futures contracts to reflect the cost of imported soybean oil from Argentina (which are significantly different at times through the season). The reason we import from Argentina and sometimes the EU, is obvious, import parity from this origin is the cheapest. South America produces the most soybeans (accounting for production in Brazil, Argentina, Paraguay and Uruguay). Furthermore, despite Argentinia's position as a leading soybean oil exporter, even their prices are largely derived from the CBOT (as the global price discovery mechanism) plus or minus fob premium and duties.

However, currently (2020), the SAFEX soybean oil futures contract is a CBOT contract that is dual-listed and cash-settled.[5] The research study seeks to determine whether this is a misspecification and whether or not SAFEX soybean oil futures contracts should rather be based on the Argentina free-on-board (fob)[6] soybean oil prices which is a much better representation of South Africa's import parity and local industry prices. If correct, it may also explain why market participants are reluctant to utilise SAFEX listed CBOT soybean oil futures contracts, explaining the low trading volumes and inadequate liquidity.

It could best be explained by the following analogy: Let's consider our local market Country Market-A. There is an instrument available in Country Market-B and there is an instrument available in Country Market-C. All three country markets, A, B and C forms part of the entire World Market-D. Market-A is too small to impact market-D. However, market-B and market-C both have significant impacts on market-D and as such also have pushback effects on each other. However, market-C > market-B, and such has a bigger effect on market-D. It is true that market-B supplies market-A; market-B > market-A. However, market-A is price wise affected by market-C, as is market-B. Hence, the risk of choosing market-B as opposed to market-C as a price reference, is that market-B's price can also be heavily influenced by its local domestic incidents which are not relevant to market-A or C and only partly affects market-D due to the fact that market-C can easily substitute market-B in market-D. Should market-A now have decided to participate with market-B and a local incident occurs in market-B, it will now have a severe unintended impact on the hedge of market-A. Therefore, making a qualitative decision should be best avoided. This research attempts to follow a quantitative approach.

Hence, the study used the Engle–Granger (1987) cointegration approach, alongside a range of diagnostic tests to evaluate the existence of adequate long- and short-run cointegration relationships amongst a linear combination of data variables underlying the current specifications of SAFEX soybean oil futures contracts versus that of an alternative linear combination of data variables that are cash-settled of Argentina fob prices (settlement values). Essentially evaluating its efficiency under Eugene Fama's semi-strong-form of market efficiency, in an attempt to identify possible mis-specification by referencing CBOT settlement values as opposed to Argentina settlement values that could ultimately lead to greater participation and improved liquidity

2. Literature review

Proponents of the Efficient Market Hypothesis (EMH) believe it constitutes the cornerstone of financial markets (Delcey 2019). That said, Fama (1965a) is known to have pioneered the term "efficient market" as a market where prices always "fully reflect" available information. However, Samuelson (1965) first analysed the role of current futures contract prices as an estimator for spot market prices in commodity markets, proposing the so-called martingale[7] definition of market efficiency. Samuelson argued for a stochastic process[8] (x_t) to be a martingale with respect to a sequence of information sets ($\emptyset_t$). That is, today's future price (x_t) would be the best-unbiased estimator of tomorrow's spot market price (x_{t+1}), given a sequence of information sets ($\emptyset_t$), where:

$$\boldsymbol{E}(\boldsymbol{x_t} \mid \emptyset_{\boldsymbol{t}}) = \boldsymbol{x_{t+1}} \tag{1}$$

Hereafter, Fama (1970) expanded his definition of an efficient market, suggesting three forms: (1) *Weak-form:* where historic information is a true reflection of current market prices; (2) *Semi-strong-form:* all publicly available information is fully reflected in current market prices; and the (3) *Strong-form:* all information (public & private) are fully reflected in current market prices.

But, according to LeRoy (1989), the theory of efficient markets is based on economic principles of competitive equilibrium. Hence, analysing efficient financial markets requires the assumption that comparative advantages exist from the differences in information that market participants hold (LeRoy 1989). Meaning, financial markets require participants to express opposite views that arise from their differences in available information, simultaneously quoting both bid and offer prices. For example, if it were universally known that the price of a security listed on a financial market is about to rise, all participants would bid to buy the security beforehand; hence, no offers to sell the security, causing no one to gain from their participation in the futures market once the higher price for the security actually materialises. Thus, the optimal condition would be that of a semi-strong form, since futures markets would cease to exist under the strong-form of market efficiency.

Most studies employ Ordinary Least Squares (OLS) regression analysis to investigate the efficiency of commodity futures markets, i.e., regressing spot market prices on previous futures prices (Goss 1981; Singh 2014). However, commodity prices usually appear to be integrated of order one, I(1). Meaning, commodity prices are non-stationary and need to be differenced to allow for stationarity. Why is this important? If a series is non-stationary, I(d) where $d > 0$, the use of OLS may produce spurious results (Granger and Rush 1974; Crowder and Hamed 1993; Wooldridge 2013). Therefore, Engle and Granger (1987) developed a far superior cointegration regression analysis (recognised as the Engle–Granger Cointegration technique) for testing market efficiency, given non-stationary variables.

Aulton, Ennew, and Rayner (1997) used this technique to evaluate the efficiency of futures markets for agricultural commodities in the United Kingdom (UK), specifically that of wheat, potatoes and pig meat. First, the researchers established the order of integration for each series (wheat, potato and pig meat prices), then proceeded to test for cointegration. Once the respective price series for wheat, potatoes and pig meat were found to be cointegrated the researchers used OLS to test for market efficiency. The results were varied, suggesting efficiency in wheat futures, but inefficiencies in potato and pig meat futures, with inefficiencies being correlated with relatively low volumes of trade (Aulton, Ennew, and Rayner 1997).

Wiseman, Darroch, and Ortmann (1999) also applied cointegration regression techniques for testing the efficiency of South Africa's white maize futures market. Analysing whether lagged futures prices for the July 1997 and July 1998 white maize futures contracts could be used to predict subsequent spot market prices. Their results showed no-cointegration amongst futures and spot market prices for the earlier and less liquid July 1997 white maize futures contract, but the existence of a long-run cointegrating relationship between futures and spot market prices for the July 1998 white maize futures contract with higher market liquidity. Hence, the increase in liquidity could improve market efficiency (Wiseman, Darroch, and Ortmann 1999).

McCullough and Strydom (2013) also used the Engle–Granger cointegration approach to re-evaluate the efficiency of the South African white maize futures market using near spot price data for longer time periods, which also pointed toward the existence of long-run co-integrating relationships and ultimately an efficient futures market for white maize in South Africa (McCullough and Strydom 2013). This suggests the use of spot price data for longer time periods under the Engle–Granger cointegration approach for evaluating the efficiency in commodity futures markets.

3. Theoretical framework

3.1 South African soybean oil futures contract specifications

As shown in Table 1, soybean oil futures, currently trading on SAFEX, are cash-settled of a CBOT settlement value, i.e., the underlying physical commodity (soybean oil) is not exchanged, only the monetary difference between the trade (willing buyer and seller dealt with each other on the exchange at a mutually agreed quoted price) and closing price. These prices are derived from CBOT soybean oil futures prices, while using the US$/ ZAR exchange rate quoted on the currency futures market to convert it from a US$ based to a ZAR based settlement value.

However, only 67 soybean oil futures contracts traded for July 2019 expire during the month of May 2019 (representing 1,655 MT of soybean oil), which is only 5% of the 32,878 MT of soybean oil produced in South Africa during the months June and July 2019. While 92,042 May 2019 soybean futures contracts (representing 46,02,100 MT of soybeans) traded during the same time period, which is 2300% of the 1,99,152 MT of soybeans crushed during the months of June and July 2019 (SAGIS 2019a, 2019c; Refinitiv 2020). Thereby highlighting the liquidity issue (as mentioned in Section 1.2) and the fact that domestic soybean crushing plants likely use international soybean oil futures for managing their price risk as opposed to SAFEX soybean oil futures.

3.2 Understanding the concept of cointegration

First, consider the difference between non-stationary and stationary time series data variables. Non-stationary time series data variables are those with varying mean, variances and covariances, trending upwards or downwards over time. Furthermore, non-stationary variables contain one or more unit roots (d), also known as the order of integration, where data variables are said to be integrated of order d, containing d number of unit roots, written as I(d), if after being differenced d times it becomes stationary. Therefore, stationary time series data variables do not contain a unit root, and are considered as integrated of order zero, I(0), having a constant mean, variance and covariances over time (Wooldridge 2013).

Now consider Murray's (1994) tale of the drunk and her dog. A drunk lady (Y_t) steps out of a bar with her dog (X_t). The lady follows a random walk, while her dog wanders aimlessly without a leash, moving further apart with the passing of time (both Y_t and X_t are non-stationary variables) where it becomes more and more difficult to estimate the distance between them. But if the drunk lady holds the dog on leash (ε_t) they could only move as far from each other as the leash allows them to. Therefore, cointegration is the event was the leash (ε_{tz}), a residual[9] in the long-run equilibrium relationship[10] between the drunk lady and her dog, is stationary.

Table 1. SAFEX soybean oil future contract specifications as set out by the JSE.

Underlying Instrument
"A soybean oil futures contract meeting all specifications as listed and traded on CBOT, a subsidiary of the CME Group Inc. The JSE reserves the right to amend the contract specifications including settlement methodology should these be amended by the reference exchange."
"The final settlement price for cash settlement of the contract will require two components, a CBOT settlement value and a Rand Dollar exchange rate which will be rounded to tow decimals."

Source: JSE (2019b).

This coincides with the formal definition by Granger (1981) for cointegration as the event were the residual term (ε_{tz}) in a long-run equilibrium relationship of two or more non-stationary data variables ($x_{t1}, x_{t2}, \ldots, x_{tz},$ *for all integers* $z \geq 1$*; and* Y_1) that are integrated of order one, I(1):

$$Y_1 = \alpha + \beta x_{t1} + \beta x_{t2} + \ldots + \beta x_{tz} + \varepsilon_{tz}, \quad t = 1, 2 \ldots, Z \text{ with } Y_1 \sim I(1) \;\&\; x_t \sim I(1) \tag{2}$$

is proofed to be stationary, $I(0)$, using Equation (3) under the null hypothesis of $\rho^* = 0 \sim$ non-stationary residuals (Wooldridge 2013).

$$\Delta\varepsilon_{tz} = \rho^*\varepsilon_{t-1} + \sum_{i=1}^{\rho-1} \rho_i^* \Delta\varepsilon_{t-i} + \omega_t, \qquad \text{where } \omega_t \text{ is } I(0) \tag{3}$$

4. Research data

Daily time series data for the period 25 June 2010 to 27 September 2019 was taken from a Thomson Reuters DataStream subscription, one of the world's leading and most reliable sources of financial data (Refinitiv 2020), in combination with daily time series data retrieved from the JSE's website (JSE 2019c). Therefore obtaining 2,313 observations for each of the following variables:

- Closing prices for spot month South African soybean oil futures (SA_OILS) quoted in ZAR/ MT.
- Fob closing prices for spot month CBOT soybean oil futures (US_OILS) quoted in USc/ lbs.
- Fob closing prices for spot month Argentina soybean oil futures (ARG_OILS) quoted in US$/ MT.
- Closing ZAR/ US$ (ZAR) exchange rate quotes as published by Refinitiv, on a Thomson Reuters DataStream subscription (Refinitiv 2020).

5. Research methodology

5.1 Unit root tests for stationarity

The Engle–Granger cointegration technique requires time series data variables to portray a common order of integration for establishing stationarity[11] amongst data variables. Since non-stationary time series data variables may produce spurious results, causing the validity of the standard assumptions to be questioned (Engle and Granger 1987; Wooldridge 2013). Therefore, the study will employ the most widely used formal unit root tests in economic literature, i.e., the Augmented Dickey–Fuller (ADF) and Phillips Perron (PP) tests for testing stationarity amongst the time series data variables (Wooldridge 2013). The ADF test is based on the null hypothesis that an underlying time series variable is integrated of order one, $I(1)$. That is, it contains a unit root and needs to be differenced once, after which it becomes stationary and tested under three possible structures: (1) A random walk with drift (a model referred to as "None"), Equation (4); (2) A random walk without a drift (a model referred to as "Intercept"), Equation (5); and (3) A deterministic trend with a drift (a model referred to as "Trend & Intercept"), Equation (6).

$$Y_t = \gamma + \gamma_{t-1} + \varepsilon_t, \; t = 1, 2, \ldots, \tag{4}$$

$$Y_t = \gamma_{t-1} + \varepsilon_t, \; t = 1, 2, \ldots, \tag{5}$$

$$Y_t = \alpha + \beta t + \varepsilon_t, \; t = 1, 2, \ldots, \tag{6}$$

Phillips and Perron (1988), developed the PP test, following the ADF test for stationarity. The PP test allows a wide range of weakly dependent and heterogeneously distributed data variables (potential regression bias as a result of omitted variables where the dependent variable is correlated with the residual term) to be tested for stationarity (Phillips and Perron 1988; Wooldridge 2013). It is evaluated against the same three structures as in Equations (4)–(6), using nonparametric methods to solve possible serial correlation amongst residual terms in the absence of lagged variables.

Tables 2–4 depict the ADF and PP unit root test results for all the time series data variables used in this study. Both the ADF and PP tests strongly agree that SA, US and Argentina soybean oil prices, including the ZAR/ US$ exchange rates are non-stationary in level form, after becoming stationary once differenced, hence I(1). However, SA soybean oil futures trading volumes seem to be stationary in both the level and differenced form, hence I(0). But cointegration tests require all data variables used to be of the same order of integration. Thus, the study will focus on I(1) data variables, excluding SA soybean oil futures trading volumes in the cointegration analysis.

5.2 Engle–Granger cointegration test

The Engle–Granger cointegration test is a relatively simple test that is used to identify the presence of a long-run co-integrating relationship amongst variables (Wooldridge 2013). This study will test whether the residual (ε_{USt}) from linear combinations of data variables underlying the specifications of South African soybean oil futures (Section 3.1) using CBOT fob soybean oil futures prices, Equation (7), and the residual (ε_{ARGt}), substituting CBOT with Argentina fob soybean oil futures prices, Equation (8), are stationary, I(0). Using a long-run relationship to evaluate the null hypothesis of no-cointegration, with the underlying *p*-values at the 1%, 5% and 10% levels of significance.

$$SA\_OILS_t = US\_OILS_t + ZAR_t + C_{US} + \varepsilon_{USt}, \quad t = 1, 2 \dots 2313 \tag{7}$$

$$SA\_OILS_t = ARG\_OILS_t + ZAR_t + C_{ARG} + \varepsilon_{ARGt}, \quad t = 1, 2 \dots 2313 \tag{8}$$

Table 5 together with Equation (9) depicts the OLS coefficients and test-statistics for the long-run relationship between SA_OILS, US_OILS and ZAR:

$$\begin{aligned} SA\_OILS_t &= 119.2496(US\_OILS_t) + 481.2402(ZAR_t) - 954.3332(C_{US}) + \varepsilon_{USt}, \\ t &= 1, 2 \dots 2313 \end{aligned} \tag{9}$$

The Adjusted R-squared (*Adj.* R^2) value of 0.3546 (Table 5), indicates that 35.46% of the variation in SA_OILS can be explained by the long-run relationship in Equation (9). However, according to econometric literature, an *Adj.* R^2 value below 0.7 is considered inconclusive for statistical inference (Wooldridge 2013).

Table 6 together with Equation (10) depicts the OLS coefficients and test-statistics for the long-run relationship between SA_OILS, ARG_OILS and ZAR:

$$\begin{aligned} SA\_OILS_t &= 4.368074(ARG\_OILS_t) + 348.7210(ZAR_t) - 1447.371(C_{US}) + \varepsilon_{USt}, \\ t &= 1, 2 \dots 2313 \end{aligned} \tag{10}$$

The *Adj.* R^2 value of 0.3712 (Table 6), indicates that 37.12% of the variation in SA_OILS can be explained by the long-run relationship in Equation (10).

Therefore, despite the use of statistical inference, we notice a slightly higher degree of variability in SA_OILS being explained by ARG_OILS (37.12%) as opposed to US_OILS (35.46%) in the long-run (Tables 5 and 6). Table 7 presents the Engle–Granger cointegration results for the long-run relationship between SA_OILS, US_OILS and ZAR, alongside a *p*-value of 0.0008 for the dependant variable SA_OILS. This is less than 0.10, 0.05 and 0.01, meaning we reject the null hypothesis of no-cointegration at a 1% level of significance.

Table 8 presents the Engle–Granger cointegration results for the long-run relationship between SA_OILS, ARG_OILS and ZAR, alongside a *p*-value of 0.0022 for the dependant variable SA_OILS. This

Table 2. ADF and PP unit root test results for SA, US and Argentina soybean oil prices in level and differenced form.

	Model	Augmented Dickey–Fuller (ADF) Lags	τ_T, τ_μ, τ	Phillips–Perron (PP) Band-width	PP	Conclusion
sa_oils	Trend & intercept	0	−3.01	6	−3.09	Non-stationary
	Intercept	0	−2.82*	6	−2.90**	
	None	0	−0.79	11	−0.79	
Δsa_oils	Trend & intercept	0	−46.97***	12	−46.97***	Stationary
	Intercept	0	−46.97***	12	−46.97***	
	None	0	−46.97***	12	−46.97***	
us_oils	Trend & intercept	0	−1.55	10	−1.57	Non-stationary
	Intercept	0	−1.43	11	−1.43	
	None	0	0.07	12	0.07	
Δus_oils	Trend & intercept	0	−47.40***	13	−47.40***	Stationary
	Intercept	0	−47.40***	12	−47.40***	
	None	0	−47.40***	12	−47.40***	
arg_oils	Trend & intercept	0	−1.64	1	−1.56	Non-stationary
	Intercept	0	−1.59	1	−1.56	
	None	0	−0.05	3	−0.023	
Δarg_oils	Trend & intercept	0	−50.61***	0	−50.61***	Stationary
	Intercept	0	−50.61***	0	−50.61***	
	None	0	−50.62***	0	−50.62***	

Source: IHS Markit (2017).
*Statistically significant at a 10% level.
**Statistically significant at a 5% level.
***Statistically significant at a 1% level.

Table 3. ADF and PP unit root test results for SA soybean oil futures trading volumes in level and differenced form.

	Model	Augmented Dickey–Fuller (ADF) Lags	τ_T, τ_μ, τ	Phillips–Perron (PP) Band-width	PP	Conclusion
sa_oils_volume	Trend & intercept	0	−46.28***	35	−51.38***	Stationary
	Intercept	0	−46.02***	35	−51.54***	
	None	0	−45.83***	35	−51.75***	
Δsa_oils_volume	Trend & intercept	10	−23.81***	30	−262.36***	Stationary
	Intercept	10	−23.76***	30	−262.39***	
	None	10	−23.75***	30	−262.44***	

Source: IHS Markit (2017).
*Statistically significant at a 10% level.
**Statistically significant at a 5% level.
***Statistically significant at a 1% level.

Table 4. ADF and PP unit root test results for ZAR/ US$ exchange rates in level and differenced form.

	Model	Augmented Dickey–Fuller (ADF) Lags	τ_T, τ_μ, τ	Phillips–Perron (PP) Band-width	PP	Conclusion
zar	Trend & intercept	0	−2.50	3	−2.50	Non-stationary
	Intercept	0	−0.90	1	−0.90	
	None	0	−1.57	2	−1.57	
Δzar	Trend & intercept	0	−47.47***	3	−47.47***	Stationary
	Intercept	0	−47.48***	3	−47.48***	
	None	0	−47.46***	1	−47.46***	

Source: IHS Markit (2017).
*Statistically significant at a 10% level.
**Statistically significant at a 5% level.
***Statistically significant at a 1% level.

Table 5. OLS long-run regression results using US_OILS.

Dependent Variable: SA_OILS
Method: Least Squares
Included observations: 2313 after adjustments

Variable	Coefficient	Std. Error	*t*-Statistic	Prob.
US_OILS	119.2496	3.665146	32.53612	0.0000
ZAR	481.2402	13.49457	35.66178	0.0000
C	−954.3332	290.4047	−3.286218	0.0010

R-squared	0.355186	Mean dependent var	9201.426
Adjusted R-squared	0.354628	S.D. dependent var	934.935
S.E. of regression	751.0805	Akaike info criterion	16.0822
Sum squared resid	1.30E+09	Schwarz criterion	16.08965
Log likelihood	−18596.06	Hannan–Quinn criter.	16.08492
F-statistic	636.2147	Durbin–Watson stat	0.043164
Prob(F-statistic)	0.000000		

Source: IHS Markit (2017).

is also less than 0.10, 0.05 and 0.01, hence rejecting the null hypothesis of no-cointegration at a 1% level of significance.

Thus, these results indicate that there exists a long-run co-integrating relationship between the variables as set out in Equations (7) and (8).

5.3 Estimating the error-correction-model (ECM)

We now estimate the Error-Correction Model's (ECM's), using the lagged residuals (ε_{USt-1} & ε_{ARGt-1}) from the long-run co-integrating relationships, together with the first difference forms (*D*) (Equations (11) and (12)):

$$D(SA\_OILS_t) = \varepsilon_{USt-1} + D(US\_OILS_t) + D(ZAR_t) + C_{US}, \quad t = 1, 2 \ldots 2313 \tag{11}$$

$$D(SA\_OILS_t) = \varepsilon_{ARGt-1} + D(ARG\_OILS_t) + D(ZAR_t) + C_{ARG}, \quad t = 1, 2 \ldots 2313 \tag{12}$$

Correcting any disequilibrium that might have occurred in the previous period ($t - 1$), outlying the short-run equilibrium relationship, where the coefficients indicate the speed of adjustment towards equilibrium. Table 9, together with Equation (13), represents the ECM and ultimately the

Table 6. OLS long-run regression results using ARG_OILS.

Dependent Variable: SA_OILS
Method: Least Squares
Included observations: 2313 after adjustments.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ARG_OILS	4.368074	0.128946	33.87526	0.0000
ZAR	348.7210	9.670804	36.05915	0.0000
C	1447.371	211.2464	6.851576	0.0000

R-squared	0.371772	Mean dependent var	9201.426
Adjusted R-squared	0.371228	S.D. dependent var	934.935
S.E. of regression	741.358	Akaike info criterion	16.05614
Sum squared resid	1.27E+09	Schwarz criterion	16.06359
Log likelihood	−18565.93	Hannan–Quinn criter.	16.05886
F-statistic	683.5045	Durbin–Watson stat	0.040192
Prob(F-statistic)	0.000000		

Source: IHS Markit (2017).

Table 7. Engle–Granger cointegration test results using US_OILS.

Series : SA_OILS US_OILS ZAR
Included observations: 2313 after adjustments
Null hypothesis: Series are not cointegrated
Cointegration equation deterministics: C
Automatic lags specification based on Schwarz criterion (maxlag=26)

Dependent	tau-statistic	Prob.*	z-statistic	Prob.*
SA_OILS	−4.991135	0.0008	−49.61058	0.0005
US_OILS	−4.763521	0.0020	−47.73169	0.0007
ZAR	−4.796583	0.0017	−48.53272	0.0006

Source: IHS Markit (2017).

Table 8. Engle–Granger cointegration test results using ARG_OILS.

Series: SA_OILS ARG_OILS ZAR
Included observations: 2313 after adjustments
Null hypothesis : Series are not cointegrated
Cointegration equation deterministics: C
Automatic lags specification based on Schwarz criterion (maxlag=26).

Dependent	tau-statistic	Prob.*	z-statistic	Prob.*
SA_OILS	−4.735696	0.0022	−45.46134	0.0012
ARG_OILS	−4.254700	0.0115	−36.59748	0.0073
ZAR	−4.192036	0.0140	−35.76034	0.0086

Source: IHS Markit (2017).

short-run dynamics between SA_OILS, US_OILS and ZAR.

$$\begin{aligned} D(SA\_OILS_t) = {} & -0.008352(\varepsilon_{USt-1}) - 3.899677(D(US\_OILS_t)) + 31.56401\,(D(ZAR_t)) \\ & - 1.280483\,, \\ t = {} & 1, 2 \ldots 2313 \end{aligned} \tag{13}$$

This ECM produced an *Adj*. R^2 value of 0.0019, which is far below the minimum requirement of 0.7, hence the study failed to produce a short-run equilibrium relationship amongst SA_OILS, US_OILS and ZAR, as set out in the soybean oil futures contract specifications as explained in Section 3.1. However, the speed of adjustment coefficient of 0.0083 (absolute value for the coefficient of the lagged residual term) in Table 9, seems to be statistically significant at a 5% level of significance, validating economic justification.

Table 10, together with Equation (14), represents the ECM and ultimately the short-run dynamics between SA_OILS, ARG_OILS and ZAR.

$$\begin{aligned} D(SA\_OILS_t) = {} & -0.014162(\varepsilon_{ARGt-1}) - 0.190757(D(ARG\_OILS_t)) + 34.91252\,(D(ZAR_t)) \\ & - 1.266553\,, \\ t = {} & 1, 2 \ldots 2313 \end{aligned} \tag{14}$$

Unfortunately, this ECM also produced an *Adj*. R^2 value (0.0059) far below the minimum requirement of 0.7. But this is higher than the *Adj*. R^2 value of the ECM using US_OILS (0.0019), with a speed of adjustment coefficient of 0.0083 (Table 9), as compared to the speed of adjustment coefficient of 0.014 (Table 10) for the ECM using ARG_OILS. Hence, these results suggest using ARG_OILS instead of US_OILS, since it appears to correct any disequilibrium that might have occurred in the previous

Table 9. US_OILS ECM results.

Dependent Variable: D(SA_OILS)
Method: Least Squares
Included observations: 2312 after adjustments.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESIDUAL_US_OILS(-1)	−0.008352	0.003705	−2.254464	0.0243
D(US_OILS)	−3.899677	5.617512	−0.694200	0.4876
D(ZAR)	31.56401	22.93198	1.376419	0.1688
C	−1.280483	2.761685	−0.463660	0.6429
R-squared	0.003196	Mean dependent var	−1.410900	
Adjusted R-squared	0.001901	S.D. dependent var	132.8556	
S.E. of regression	132.7293	Akaike info criterion	12.61623	
Sum squared resid	4.07E+07	Schwarz criterion	12.62617	
Log likelihood	−14580.36	Hannan–Quinn criter.	12.61985	
F-statistic	2.466814	Durbin–Watson stat	1.940992	
Prob(F-statistic)	0.060448			

Source: IHS Markit (2017).

period at a faster pace (Wooldridge 2013), proving greater efficiency under Eugene Fama's semi-strong form of the EMH (Section 2).

5.4 Diagnostic tests

5.4.1 Long-run equilibrium

With the use of various diagnostic tests we consider theoretical justifications for the interpretation of results. This includes a test by Jarque and Bera (1987), known as the Jarque–Bera test for normality, measuring the difference in kurtosis (the sharpness of a distribution's peak) and skewness of a time series variable to those from normally distributed random variables, under the null hypothesis of a symmetric and mesokurtic distributed series. Followed by White's test for heteroskedasticity, a langrage multiplier test, developed by White (1980), to ensure equality amongst variances under the null hypothesis of no heteroskedasticity. When $Var(\varepsilon|x)$ depends on x, the residual (ε) exhibits heteroskedasticity, which could invalidate standard inference procedures (Wooldridge 2013).

The study then employed the Breusch (1978) and Godfrey (1978) tests as one, known as the Breusch-Godfrey Lagrange Multiplier (LM) test for serial correlation (autocorrelation) under the null hypothesis of no autocorrelation up to the pth order. Autocorrelation exists where there is some correlation amongst residuals of different time periods, which could render spurious results (Wooldridge 2013). Finally, the study made use of Ramsey's RESET test, developed by Ramsey (1969), for identifying possible misspecification in terms of inclusion of irrelevant variables or the exclusion of relevant variables, while evaluating the correlation between the estimators and residual terms.

Table 11 depicts the diagnostic test results for the long-run equilibrium model using US_OILS, Equation (9). The Jarque–Bera test for normality produced a test statistic (p-value) of 0.200, hence we fail to reject the null hypothesis of symmetric and mesokurtic distributed series (normally distributed residuals) at a 1% (p-value > 0.01), 5% (p-value > 0.05) and 10% (p-value > 0.10) level of significance. Furthermore, White's test produced a p-value of 0.000, that rejects the null hypothesis at 10% level of significance, concluding the existence of heteroskedasticity (invalidating standard inference procedures). The Breusch-Godfrey's LM test also produced a p-value of 0.000, suggesting serial correlation amongst data variables. Finally, Ramsey's RESET test produced a p-value of 0.000, rejecting the null hypothesis of no misspecification at a 10% level of significance. Hence the long-run equilibrium relationship between SA_OILS, US_OILS and ZAR seems to exclude relevant time series data variables or include irrelevant time series data variables.

Table 12 depicts the diagnostic test results for the long-run equilibrium model using ARG_OILS, Equation (10). According to the Jarque–Bera test for normality, the residuals are not normally

Table 10. ARG_OILS ECM results.

Dependent Variable: D(SA_OILS
Method: Least Squares
Included observations: 2312 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESIDUAL_ARG_OILS(-1)	−0.014162	0.003732	−3.794943	0.0002
D(ARG_OILS)	−0.190757	0.233914	−0.815503	0.4149
D(ZAR)	34.91252	22.89167	1.525119	0.1274
C	−1.266553	2.756138	−0.459539	0.6459
R-squared	0.007174	Mean dependent var	−1.410900	
Adjusted R-squared	0.005884	S.D. dependent var	132.8556	
S.E. of regression	132.4642	Akaike info criterion	12.61223	
Sum squared resid	4.05E+07	Schwarz criterion	12.62217	
Log likelihood	−14575.74	Hannan–Quinn criter.	12.61585	
F-statistic	5.559213	Durbin–Watson stat	1.936767	
Prob(F-statistic)	0.000845			

Source: IHS Markit (2017).

distributed across time. White's test reveals the presence of heteroskedasticity amongst variables, invalidating standard inference procedures. The Breusch-Godfrey LM test also reveals the existence of serial correlation amongst data variables and most importantly Ramsey's RESET test suggests possible misspecification amongst these time series data variables. Therefore, the long-run equilibrium relationship between SA_OILS, ARG_OILS and ZAR also seem to exclude relevant time series data variables or include irrelevant time series data variables.

5.4.2 Short-run equilibrium

Table 13 depicts the diagnostic test results for the short-run equilibrium model using US_OILS, Equation (11). The Jarque–Bera test for normality produced a test statistic (*p*-value) of 0.000, which rejects the null hypothesis of normally distributed residuals. Furthermore, White's test produced a *p*-value of 0.042, that rejects the null hypothesis at the 10% levels of significance, concluding the existence of heteroskedasticity in some moderate form. The Breusch-Godfrey's LM test produced a *p*-value of 0.150, hence we fail to reject the null hypothesis of no serial correlation amongst the time series data variables at 10%, 5% and 1% level of significance. Finally, Ramsey's RESET test produced a *p*-value of 0.401, which is greater than 0.10, hence we also fail to reject the null hypothesis of no misspecification amongst the time series data variables at a 10%, 5% and 1% level of significance. Therefore, the JSE contract specifications for South African soybean oil futures contracts does not seem to be mis specified in the short-run.

Table 14 depicts the diagnostic test results for the short-run equilibrium model using ARG_OILS, Equation (12). According to the Jarque–Bera test for normality the residuals are not normally distributed across time. White's test produced a *p*-value of 0.023, that rejects the null hypothesis at 10% level of significance, also concluding the existence of heteroskedasticity in some moderate form. The Breusch-Godfrey's LM test produced a *p*-value of 0.123, hence we fail to reject the null hypothesis of no serial correlation amongst the time series data variables at 10%, 5% and 1% level of significance. Finally, Ramsey's RESET test produced a *p*-value of 0.032, which is greater than 0.01 but smaller than 0.05. Hence, we accept the null hypothesis of no misspecification at a 1% level of

Table 11. US_OILS long-run equilibrium diagnostic test results.

Test	Test-Statistic	*p*-Value	Conclusion
Jarque–Bera	3.14	0.200	Residuals are normally distributed
White	535.98	0.000	Heteroskedasticity of a high degree
Breusch–Godfrey	2214.88	0.000	Serial correlation up to 1 lag
Ramsey reset	370.60	0.000	Possible misspecification

Source: IHS Markit (2017).

Table 12. ARG_OILS long-run equilibrium diagnostic test results.

Test	Test-statistic	*p*-Value	Conclusion
Jarque–Bera	43.98	0.000	Residuals are not normally distributed
White	168.84	0.000	Heteroskedasticity of a high degree
Breusch–Godfrey	2220.97	0.000	Serial correlation up to 1 lag
Ramsey reset	62.95	0.000	Possible misspecification

Source: IHS Markit (2017).

Table 13. US_OILS short-run equilibrium diagnostic test results.

Test	Test-statistic	*p*-Value	Conclusion
Jarque–Bera	233820.30	0.000	Residuals are not normally distributed
White	17.48	0.042	Heteroskedasticity of some moderate form
Breusch–Godfrey	2.07	0.150	No serial correlation up to 1 lag
Ramsey reset	10.53	0.401	No Misspecification

Source: IHS Markit (2017).

Table 14. ARG_OILS short-run equilibrium diagnostic test results.

Test	Test-statistic	*p*-Value	Conclusion
Jarque–Bera	231394.00	0.000	Residuals are not normally distributed
White	19.23	0.023	Heteroskedasticity of some moderate form
Breusch–Godfrey	2.38	0.123	No serial correlation up to 1 lag
Ramsey reset	4.04	0.032	Possible misspecification

Source: IHS Markit (2017).

significance but reject the null hypothesis at a 5% and 10% level of significance. Further supporting the statement for South African soybean oil futures not being misspecified in the short-run.

6. Conclusion and recommendations

As mentioned in Section 3.1, there is very little participation in SAFEX soybean oil futures contracts that, according to its contract specifications, are cash-settled of CBOT settlement values, while the majority of the country's soybean oil originates from Argentina. Hence, the study used the Engle–Granger (1987) cointegration approach to evaluate the existence of adequate long and short-run cointegration relationships amongst a linear combination of data variables underlying the current specifications of SAFEX soybean oil futures versus that of an alternative linear combination of data variables that are cash-settled of Argentinian settlement values. Essentially evaluating its efficiency under Eugene Fama's semi-strong-form of market efficiency (where all publicly available information is fully reflected in current market prices, Section 2), attempting to identify possible misspecification by referencing CBOT settlement values as opposed to Argentinian settlement values that could ultimately lead to greater participation and improved liquidity.

The study did however fail to produce statistically significant long- and short-run equilibrium relationships between CBOT and Argentina settlement values, despite cointegrating relationships. Hence, it appears that there isn't a significant difference in the accuracy of using Argentinian settlement values as opposed to that of CBOT for estimating subsequent spot market prices for SAFEX soybean oil futures contracts, that usually represents the price at which soybean crushers or importers would sell soybean oil in South Africa. The study then employed a range of diagnostic tests for a theoretical justification of these results. Identifying possible misspecification amongst the variables for both CBOT and Argentinian cash settlement values in the long-run, as-well-as Argentinian cash settlement values in the short-run, which was not the case for the CBOT settlement values in the short-run. Thus, pointing towards short-run market efficiency in SAFEX soybean oil futures contracts referencing CBOT settlement values.

In conclusion, SAFEX soybean oil futures contracts that are based on CBOT settlement values do not incorporate all the information used by market participants in forming a prediction of subsequent spot market prices in the long-run. It does however incorporate sufficient information for such practices in the short-run, attracting speculators[12] who hope to profit from short-term price variations in the absence of hedgers (typically soybean crushers) that seek to employ effective long-term hedging strategies. The implication is that South African soybean crushers would most likely continue using more liquid international markets for securing their crushing margin until the country becomes self-sustainable, meeting local demand with local production through an increase in both soybean production and crushing capacity. While simultaneously catering for more players within the soybean crushing industry that would ultimately improve the accuracy and transparency in the price formulation for domestic soybean oil futures in the long-run. South Africa will probably remain a net importer of soybean oil for a while, however, with the huge increases in soybean production in 2021, the playing ground is changing rapidly. This could lead the way towards more liquid and self-efficient local physically settled SAFEX soybean oil futures contracts, given the collective participation amongst the majority of these players. If so, an unintended consequence could be that liquidity in the cash-settled contract may decline even further which possibly could lead to its delisting.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes

1. Soybean crushing plants use soybeans (100%) to produce soybean meal (80%), soybean oil (18%) and low protein soybean hulls (2%), through a process known as "crushing".
2. Liquidity can be considered as the most important constituent for successfully creating an agricultural futures contract, implying the existence of willing buyers and sellers should a participant wish to buy/ sell a futures contract (Van der Vyver 1994).
3. Today formally known as the United States (US) CME Group and previously known as the Chicago Board of Trade (CBOT).
4. In practice, the US fob value for soybean oil accurately reflects the CBOT soybean oil futures contract. To determine the SA import parity price, shipping and offloading costs, as well as local transport costs and import taxes, are added to the US fob value. The added-on costs are normally stable, however, the CBOT or US fob value, converted to Rand, is very volatile and it is this value that participants would like to secure. However, industry views differ and some specialists are of the opinion that a cash settlement contract will never work for the very reason that it does not take into account freight and premiums.
5. Overseas agricultural futures contracts that are dual listed on SAFEX are always cash settled (compared to local contracts that are physically settled), meaning the underlying physical commodity is not exchanged, only the monetary difference between the trade and closing price.
6. Free-on-board (fob) refers to the underlying commodity's price, i.e. assuming delivery without the charge (freight) to ship from country of origin.
7. A betting strategy is used to ensure a favourable outcome with an arbitrarily high probability.
8. This is a sequence of random variables indexed by time. Meaning, if past conditions were different, we would fail to generate similar variables over the same time period.
9. Residual is the difference between the actual and estimated value of a variable.
10. A linear combination of two or more time series data variables: $x_{t1}, x_{t2, \dots,} x_{tz,}$ for all integers $z \geq 1$; and Y_1 .
11. Time series data is considered stationary if for every group of time indices $1 \leq T_1 < T_2 < \dots < T_n$, the joint distribution of data variables $(x_{t1,} x_{t2, \dots,} x_{tn})$ is similar to the joint distribution of $(x_{t1+z,} x_{t2+z, \dots,} x_{tn+z})$ for all integers $z \geq 1$, meaning their mean, variance and covariances are constant over time.
12. Speculators are those participants that seek to gain from short-term price movements in future markets, buying and selling futures contracts, within a relatively short time frame. Hoping to earn a monetary profit from their endeavours (SAIFM 2017).

References

Aulton, A.J., C.T. Ennew, and A.J. Rayner. 1997. Efficiency tests of futures markets for UK agricultural commodities. *Journal of Agricultural Economics* 48, no. 3: 498–524.

Breusch, T.S. 1978. Testing for autocorrelation in dynamic linear models. *Australian Economic Papers* 17, no. 31: 334–355.

Crowder, W.J., and A. Hamed. 1993. A cointegration test for oil futures market efficiency. *Journal of Futures Market* 13, no. 1: 933–941.

Delcey, T. 2019. Samuelson vs Fama on the efficient market hypothesis: The point of view of expertise. *Economia* 9, no. 1: 37–58.

Engle, R.F., and C.W.J. Granger. 1987. Co-integration and error correction: representation estimation and testing. *Econometrica* 55, no. 2: 251–67.

Fama, E.F. 1965a. The behaviour of stock market prices. *The Journal of Business* 38, no. 1: 34–105.

Fama, E.F. 1970. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance* 25, no. 2: 383–417.

Godfrey, L.G. 1978. Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica* 46, no. 6: 1293–1301.

Goss, B.A. 1981. The forward pricing function of the London metal exchange. *Applied Economics* 13, no. 1: 133–150.

Granger, C.W.J. 1981. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* 16, no. 1: 121–130.

Granger, C.W.J., and M. Rush. 1974. Spurious regressions in econometrics. *Journal of Econometrics* 2, no. 1: 111–120.

IHS Markit. 2017. Eviews (Version 10) [Computer software]. California: HIS Global Inc.

Jarque, C.M., and A.K. Bera. 1987. A test for normality of observations and regression residuals. *International Statistical Review* 55, no. 1: 163–172.

Johannesburg Stock Exchange (JSE). 2019b. Cash settled agricultural contract specifications. https://www.jse.co.za/content/JSEContractSpecificationItems/CBOTSpecs.pdf (Accessed January 15, 2019).

Johannesburg Stock Exchange (JSE). 2019c. Agriculture stats. https://www.jse.co.za/downloadable-files?RequestNode=/Safex/agriculture.stats (Accessed September 28, 2019).

LeRoy, S. 1989. Efficient capital markets and martingales. *Journal of Economic Literature* 27, no. 4: 1583–1621.

McCullough, K., and B. Strydom. 2013. The efficiency of the South African white maize futures market. *Agrekon* 53, no. 3: 18–33.

Murray, M.P. 1994. A drunk and her dog: an illustration of cointegration and error correction. *The American Statistician* 48, no. 1: 37–39.

Phillips, P.C.B., and P. Perron. 1988. Testing for a unit root in time series regression. *Econometrica* 75, no. 2: 335–346.

Ramsey, J.B. 1969. Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society* 31, no. 2: 350–371.

Refinitiv. 2020. *Thomson Reuters Datastream*. Subscription service (Accessed February 1, 2020).

Samuelson, P.A. 1965. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review* 6, no. 2: 41–49.

Singh, A. 2014. Commodity futures market efficiency and related issues: a Review of existing literature. *Asian Journal of Business and Economics* 4, no. 4: 3–13.

South African Grain Information Service (SAGIS). 2019a. *Supply and demand: Monthly since* 1997. http://www.sagis.org.za/sd_marketing_year.html (Accessed October 10, 2019).

South African Grain Information Service (SAGIS). 2019c. *Monthly products*. https://www.sagis.org.za/products.html (Accessed October 2, 2019).

South African Institute of Financial Markets. 2017. The derivatives market. AP Faure, October 2017.

Van der Vyver, A. 1994. Futures Contracts for Agricultural Commodities: is the time ripe for South Africa. *Agrekon* 33, no. 1: 50–56.

White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, no. 1: 817–838.

Wiseman, J.A., M.A.G. Darroch, and G.F. Ortmann. 1999. Testing the efficiency of the South African futures market for white maize. *Agrekon* 38, no. 3: 321–335.

Wooldridge, J.M. 2013. *Introductory econometrics: A modern approach*, 5th ed. Mason, OH: South-Western, Cengage Learning.