MSc Economics
Econometrics project

6th May 2023
Achour Omar
ID: 13818207

Analysis of wheat commodity prices:

Can a farmer infer on the price of his crop simply by observing his own farm and the neighbouring fields?

Abstract:

Wheat is a staple constituent in dietary intake of most regions in the world. In the context of a surge in prices, major producers' conflict and consumer apprehension, this study seeks to investigate the effect of commodity prices on wheat, mainly its substitutes, and those involved in its production process, oil and fertilizers. Indices are constructed to reflect these components, a cointegration analysis is conducted and a model is drafted. In doing so, this study aims to verify the no-correlation hypothesis of oil and wheat; whether the random walk hypothesis applies in this case; and the perfect market hypothesis of all information being incorporated into the price.

Table of Contents

1. Introduction

1.1. Overview

Wheat has held a historical importance from the start of human civilisation and early settlements. To this day it is one of the major crops grown for its complex carbohydrates, descent levels of protein and other nutrients.

In modern times, Wheat was introduced by the first English colonists to North America and Australia and quickly became the main cash crop of farmers due the increasing demand from settlers. The United States, Canada, Australia are thus one of the major producers today along with traditional producers in the Mediterranean region (Italy, Turkey), the Indian subcontinent and South America (Brazil, Argentina).

Its flour is used for many forms of bread, pasta and animal feed. According to the United Nations, it accounts for 20% of all caloric consumption by humans on earth. (NAWG, 2022)

There are six basic classes, Hard Red Winter (HRW), Soft Red Winter (SRW), Hard Red Spring (HRS), Hard White (HW), Soft White (SW), and Durum wheat. (USDA, 1994)

Hard Red Winter, as the name suggests is planted in the winter months (typically around November) and harvested in the summer season (June-July) and due to its high protein content is the most dominant wheat variety as it makes up about 40% of total U.S. wheat production. It is grown in the Central Plains in regions like Kansas, North Dakota and Minnesota. Wheat is a water-efficient and is seldom irrigated and often left to rainfall. (USDA, 1994)

The production process starts with ploughing the land and sowing seeds, traditionally done by human with equines, it is now, and for some time, a machine-based process with tractors and various motor-based machines powered by Diesel engines. Applying fertilizers for the development and growth of the crop involves mainly primarily fertilizers like Ammonium Nitrate, Urea, Muriate of Potash and Diammonium Phosphate.

1.2. Demand Side factors

Wheat is highly reflective of demographic shifts in the population, because each new-born is an additional consumer, increases in population could drive wheat prices.

Substitutes of wheat constitute a direct threat to wheat consumption in some sense, as oats and barley can be used as flour bases for bread for example, rice the major rival of wheat is the dominant grain in many Asian countries, and equivalently corn in the Americas, is increasingly present in “healthier” diets in Western countries.

Increasing demand for biofuel as a substitute to oil and gas has seen a significant increase in the last decade with policy shifts towards sustainable sources of energy.

Although wheat is predominantly made for human consumption, it is also used as animal feed and therefore shifts to plant-based diets and move from cattle meat, the process of which accounts for 14.5% of global greenhouse gas emissions can impacts demand. (UC Davis, 2019)

1.3. Supply Side Factors

From the producer's point of view, commodities involved in the production process are major indicators. Energy prices of oil, natural gas, and/or coal used as inputs for fertilizers' production or transportation costs. (Zeneli, 2022)

They are used to power tractors and almost every machine used in ploughing, sowing, treatment and harvesting to milling and post-production processes.

Fertilizers' prices of Nitrate, Urea and Potash are significant inputs involved in the production cycle. Increases in these commodities should be reflected in the price of the crop.

Another important factor is weather and climate conditions, with climate change making headlines, increases in average temperatures or decreases in average precipitation could reduce grains yields and drive prices up by laws of demand and supply. It is important to note that, wheat is a relatively robust crop, that resists extreme temperatures and is not as gourmand in water as other substitutes.

1.4. Other factors

As it acts as a sort of buffer, from embargoes, bad weather conditions and potential geopolitical shifts and risks of famine, wheat stock is a primary concern for wheat consuming nations, and their respective government and is often subsidized reducing production costs on the accounts level. And thus, shifting production and demand factors towards desired policy. (Duke, 2014)

This can weaken the predicting power of supply and demand factors and distort any market hypothesis assumption results.

Major financial crises, wars and pandemics as observed in the last few decades can have an influence on wheat price fluctuations. The case of Ukraine and Russia, major world producers of fertilizers, oil and wheat and the blockade imposed by Russia on Ukrainian exports, has raised concerns for food and energy supplies as many countries rely on these imports. Russia and Ukraine have in recent years accounted for about one-quarter of global exports of wheat. (World Bank, 2022)

Consumers' craze for food and amenities stock, noticed during the first pandemic in early 2020 wave could constitute, perhaps momentary, demand shock to explore along with disruption in supply of production inputs. The financial crisis of 2008, the 2000 Dotcom bubble, the 1984 financial crisis and the layoff that ensued could influence demand as well as increased commodity prices influence supply. The first and second Oil shocks of 1971-72 and 1979-80 are directly related to oil price spikes and therefore can influence wheat prices. Which begs the question:

Do Wheat prices depend on the price of their substitutes and that of commodities involved in their production, mainly oil and fertilizers?

To put it simply,

-Can a farmer infer on the price of his crop simply by observing his own farm and the neighbouring fields?

The observation, in this context is that of prices, of commodities involved in the production process and (oil, fertilizers) and substitutes (corn, barley, rice). These will constitute the supply and demand factors in this study. Neighbouring fields in this context are farms that produce substitutes of Wheat.

2.Literature Review

2.1. Review

Due to its importance as a major source of nutrition, wheat and more broadly grain commodities have been the subject of research for some time. A study by Harvard researchers, (Frankel and Rose, 2010) on the determinants of agricultural commodities involves inventories, uncertainty, speculation, economic growth and expected inflation, the study sought to invoke both macroeconomic variables and microeconomic foundations with adverse selection/asymmetric information playing a part, behavioural aspects are present to reflect the effect of market participants (speculation, storage..) in situations where signals of changing economic environment are present. Surprisingly the interest rate has insignificant effect. This study is rich in determining factors and constitutes a reference for readers. It is however beyond the scope of this study.

Another study by (Nazlioglu and Soytas, 2012) examines the relationship between oil and commodity prices using a panel cointegration analysis, the results show evidence of oil prices impact on agricultural commodities, the study included exchange rates. This comes as a contradiction to traditional literature findings of neutrality between the two and could be attributed to the econometric method used, as panel cointegration and Granger causality based on VAR and VECM models involve more than one dependant variable and allows for linear relations to exist between the variables in question. Time series analysis may differ in findings where causality exist jointly but is insignificant individually.

A more recent study, (Zeneli, 2022) makes the link between energy and grains markets in the context of surging prices, testing for structural breaks along the way. Cointegration was found among the variables but no Granger causality.

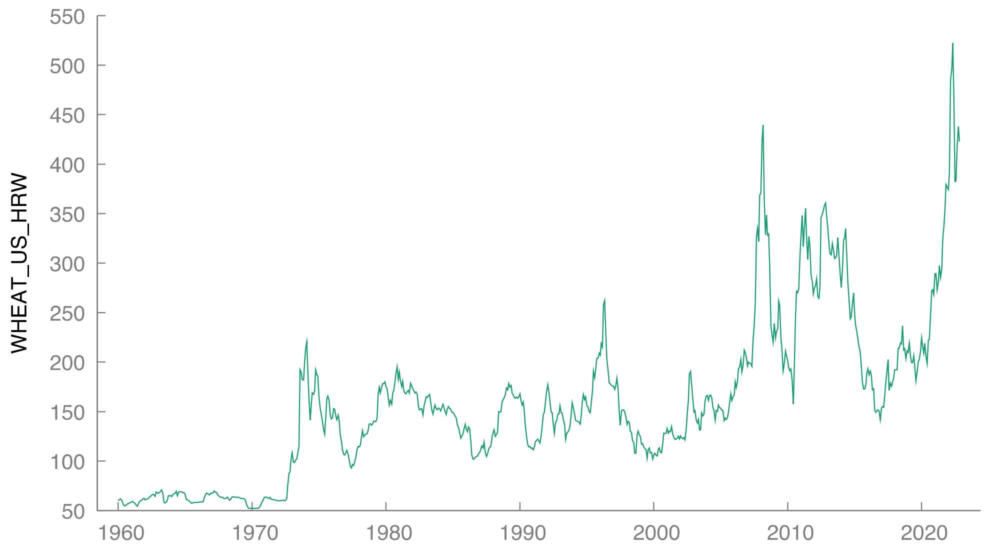
The World Bank, Commodity Markets Outlook April 2022 edition, evaluates the pandemic, the Russia-Ukraine war and the incumbent recession and their impact on commodities. Results of a comparison of energy price shocks with previous occurrences finds that higher prices of energy increased the production costs of other commodities, the study concludes with a retrospective on the preceding crises and the eventual return to equilibrium, tax breaks further aggravated the situation rather than adjusting supply and demand imbalances.

2.2. Hypothesis

The study seeks to examine the no-correlation hypothesis of US HRW Wheat and Oil. It investigates any relationship with respect to fertilizers and evaluates major substitutes like Rice, Maize corn and Barley. Another axis of the study is to establish whether wheat prices follow a random walk or are determined by some autoregressive process. A third axis would be to evaluate whether prices contain all available information or supplementary instruments are necessary thus inferring on the perfect markets hypothesis.

3.Data

3.1. Layout



Source: World Bank Commodity Price 2022, Gretl

Figure 1: US HRW Wheat prices from 1960 to 2022

The dataset consists of monthly prices of major commodities from the World Bank Pink sheet. Across all, It contains monthly average spot price indices from January 1960 to December 2022 for:

-Crude oil in US Dollars per Barrel, a barrel represents approximately 159 litres. For UK Brent, Dubai Fateh (1985-) and Saudi Light (1960-1984), and US West Texas Intermediate (1982-). An equal weight average is also available under (Crude Petro), which we will use in our analysis as the missing values in WTI will be replaced by the existing Dubai and Brent values until they appear in the average after 1982.

-Grains in US Dollars per metric tonne, of Barley (missing values from September 2020 onwards) and Maize corn both from the US, Rice Thai 5% broken White rice (Rice 05), is the rice index with no missing values is used in the analysis.

In terms of substitution, Wheat is the most consumed grain followed closely by Rice and Corn, -as all three constitute more than half the calories consumed by human beings (Awika, 2011). Barley is fourth, Oats and Sorghum follow.

Although being the most consumed, wheat is second in production to corn, that is used for other purposes as well. Sorghum is very small in terms of trade volume it is discarded in this analysis; oats data is not present.

Only maize and rice datasets are fully descriptive of the period and thus we need to create a synthetic model using time dummy variable to accommodate barley in the substitutes.

-Fertilizers provide the primary nutrients for the crop: Nitrogen (N), Phosphorus (P, or P2O5) and Potassium (K, or K2O). Nitrogen supports vegetative growth. Phosphorus improves roots and flowering. Potassium strengthens resistance to environmental assaults, from extreme temperatures to pest attacks.

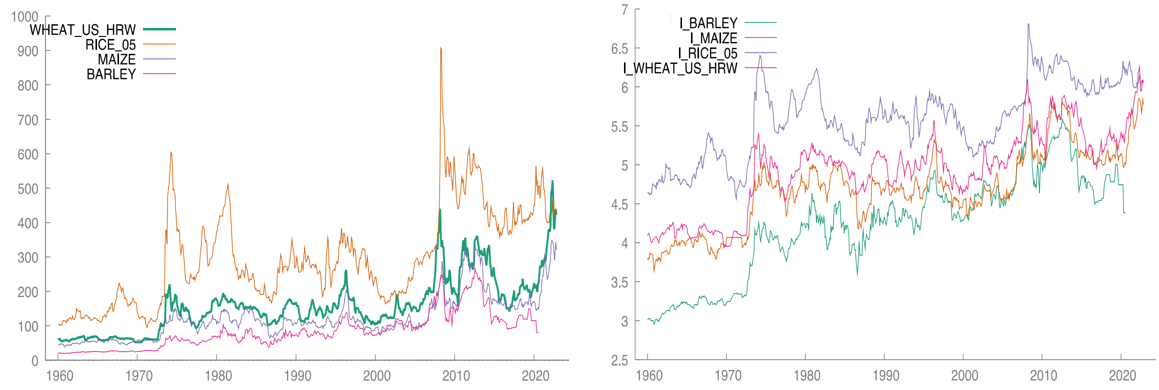
Wheat requires the application of around 80:40:40 NPK kg/Ha, the average yield per acre of wheat is 40.2 Bushels/acre or around 2,703 kg/Ha. (US Dept of Agriculture, 2013)

Intuitively, 2 units of N, 1 of P and 1 of K correspond to 33.8 units of Wheat.

The dataset contains the following fertilizers, their percentage composition in NPK is added between brackets in percentage for reference. Diammonium Phosphate (DAP) (18,46, 0) (only available from 1967 onwards), Triple Super Phosphate (TSP) (0, 46,0) is classified as hazardous, Urea (46, 0, 0) , Potassium Monochloride (Potash) (0,0,52), Phosphate Rock is insoluble and is not used directly as fertilizer. These prices are given in kilograms per metric tonne.

We construct two indices based on demand (substitutes of wheat) and supply (fertilizers) The weights are based on external data of world grain production (Statista, 2023) and the practical NPK formula used in crop treatment. This constitutes the only external information to the model so far and is assumed available to the farmer.

3.2. Description



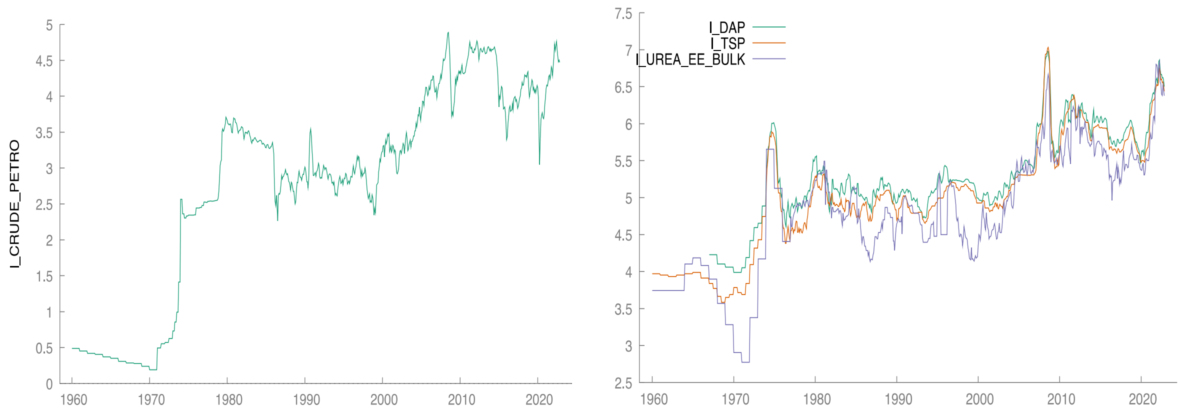
Source: World Bank Commodity Price 2022, Gretl

Figure 2: Grain substitutes prices per Bushel (left) and log-prices per Bushel(right)

The data plot of grain prices seems to fit within the financial crises' framework, suggesting spikes in prices around the same time as financial crises namely the first and second Oil shocks of 1971-72 and 1979-80, the 1984 financial crisis, the 2000 Dotcom bubble, the financial crisis of 2008 and most recently, the Covid pandemic and Ukraine War.

The data seems to revert back to its fluctuations after each crisis. Which could suggest structural breaks within the periods. The log-prices suggest an over-the-period permanent increase which might signal a trend.

Overall, the strongest result is the apparent high correlation between substitutes typical of grains markets and suggests some degree of elasticity between them.



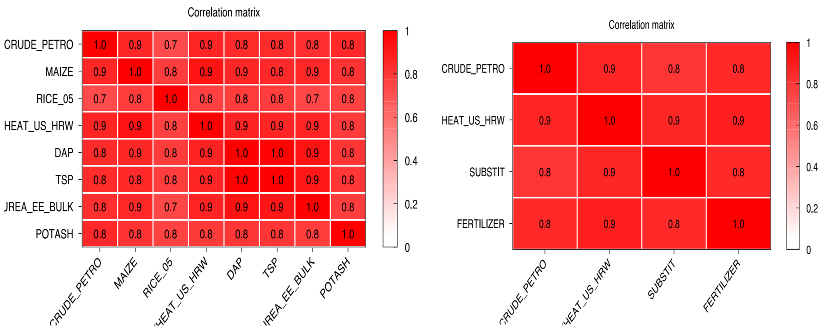
Source: World Bank Commodity Price 2022, Gretl

Figure 3: Oil index log-price per Barrel (left) Fertilizers prices per Tonne (right)

The plots of the petroleum index (in logs), fits the crises story of the two 1970's oil shocks the 1990 Asian crisis, the mortgage-backed securities crisis and visibly the

2020 lockdown resulting in less overall traffic and thus less demand for fuel, the Ukraine war has pushed it close to its highest peak of pre-2008.

Similarly, Fertilizers have much in common when it comes to their trend, with recent spikes in 2008 and 2022 and one in mid-1970s Although its seems that prices have been flat of regulated before that period,



Source: Author's work, Gretl

Figure 4: Correlation matrix of the determinants of the model

The correlation matrix above confirms the observation of high correlation between the variables, it is unchanged when the commodities are grouped into indices substitutes (consisting of Rice and Maize) Fertilizers (consisting of TSP, DSP, Urea and Potash). This type of correlation makes the interpretation the model more difficult and creates an overfitting problem. It is relevant to mention that no perfect collinearity exists and that high correlation should be managed partly by choosing the adequate model, the fit however would be extremely high and therefore less truthful to the goodness (of fit) of the model.

4.Model

4.1. Indices construction

We build two indices to describe supply and demand. A fertilizers index, inferring on the supply side, along with the already available oil index referred to above, is a weighted average of TSP, Urea and Potash according to their nutrient content (in Phosphorus, Nitrogen and Potassium) and the relative amount necessary to produce a unit of output C.

	Nutritional Content	**In Nutrient**	**Need for C output**		**Amount in Kg**	**Percent weight**
Urea	46%	N	2 N	2/0.46	4.34	51.48%
TSP	46%	P	1 P	1/0.46	2.17	25.74%
Potash	52%	K	1 K	1/0.52	1.92	22.77%
Total					8.43	100%

Source: Author's Work

Table 1: Weights for Fertilizer index

The index is therefore

$$Fert = W_{Urea} * UREA + W_{TSP} * TSP + W_{Potash} * Potash$$

One unit of this index could be thought of as a pre-packed unit of fertilizer input.

A substitutes index, is built to reflect the demand side, consisting of Corn, Rice and Barley. As data for Barley is unavailable from August 2020 onwards, it is synthetically accounted for using time dummy variables: The index is a weighted average of the three crops by volume of world production for the period up to August 2020 and then of only Rice and Corn in the subsequent months. The index weights are obviously made to equal one in each period.

	World Production In Million metric Tonnes	Total percentage share	Total share of Substitutes index SUB3	Total share of restricted Substitutes Index SUB2
		All	Excl. Wheat	Excl. Wheat and Barley
Corn	1151.36	44.48%	63.79%	69.6%
Wheat	783.4	30.26%	-	-
Rice	502.98	19.43%	27.86%	30.40%
Barley	150.48	5.81%	8.33%	-
Total	2588.22	100%	100%	100%

Source: Author's work

Table 2: Weight of respective indices of Substitutes

We use a dummy variable for the period (1960-01 to 2020-08), D_b to account for the presence of data for Barley. Our final idex is therefore the Sub_3 index if barley data is present and Sub_2 index if it's not.

Where:

$$Sub_3 = W_{3,Corn} * Corn + W_{3,Rice} * Rice + W_{3,Barley} * Barley$$

And

$$Sub_2 = W_{2,Corn} * Corn + W_{2,Rice} * Rice$$

The subscripts 2 and 3 refer to the number of elements included in the index.

The index is therefore:

$$Sub_d = D_b * Sub_3 + (1 - D_b) * Sub_2$$

One unit of this index could be thought of a pre-packed unit of other grains output, a blend of other grains that can be used for the same purpose and wheat.

4.2. Diagnostics and cointegration analysis

Consider the linear model

$$W = \alpha * Sub + \beta * Fert + \gamma * Oil + \varepsilon$$

W is the price of wheat; *Sub* (S) is the final substitutes index and *Fert* (F) the fertilizer index and ε refers to the error term. We shall use log-prices for all variables from now on.

We estimate first using OLS for each equation, see Appendix A,

At first sight, the model (Models 1 and 2) seems to fit the data with high adjusted R-squared and near zero p-values for both the demand and supply models.

Rho suggests the presence of a unit root for which we tested and indeed the ADF test could not reject this hypothesis. The model should be estimated in first differences instead. The Durbin-Watson test statistic below 2 suggests positive autocorrelation among the residuals we should therefore add lagged variables (two are suggested by the ARCH test in appendix). We also perform the Breusch-Pagan and White tests confirming heteroskedastic error terms, robust errors should be used to curb this issue, While a suggestion of non-linearity is rejected weakly (at the 5% but not at the 10% level) by the Ramsey RESET test for the demand equation (substitutes) and accepted for the supply equation this could be induced by the oil regressor for which we could add its squared terms. A test for the normality of residuals in the first suggest normally distributed errors in the substitute model and evidence against it for the fertilizers/oil model.

All three regressors and the dependant variable are integrated of degree one, as confirmed by the ADF test. We proceed to estimate in first differences.

As there in no individual cointegration between wheat and the substitutes, wheat and oil or wheat and fertilizers. We attempt a cointegrating test for all combination of two and three, but the four variables together result in the trace test confirming the existence of cointegrating relations (H1: r>2 then H0 r=2), where depending on lag order (3 and 6 or 12) there are two or three cointegrating relations.

From 6 lags onwards three cointegrating relations are clearly apparent while in the case of two, an argument could be made that the non-zero coefficients for Oil is weak and fails testing for a null of zero. We stick to three cointegrating vectors each with the substitutes index as follow

$$Z_i: \left(1 * i - \gamma_{i,s} * Sub\right) \mid i \in \{W, Fert, Oil\}$$

Where $\gamma_{i,s}$ stands for the normalised cointegrating vector coefficient of wheat, fertilizers and oil indices with respect to the substitutes index. This desirable property provides us with three restrictions for each variable, in the case of wheat, one normalising, for wheat variable (unit coefficient) and two zero coefficients for Oil and fertilizers in the cointegrating relation.

As we have three relations, no cointegrating relation exists between each variable and other except with the substitutes index in this specification, the latter cointegrates with all three variables individually including wheat which constitutes the three cointegrating relations, this is simply a matter of specification (ordering) and has no underlying implication for the index.

4.3. Econometric model

A Vector Autoregressive model (VECM) is therefore best to account for the cointegrating relationships as well as the estimation in first differences, the model uses Maximum likelihood estimation and therefore relies on distributional assumption and the fact that data is independent and identically distributed (i.i.d).

A model for wheat prices could be:

$$\Delta W_t = \alpha_w + \sum_{I \in \{W,F,O\}} \beta_{w,i} * \left(I_t - \gamma_{i,S} * S_t\right) + \sum_{I \in \{W,F,O,S\}} \sum_{j=1}^{L-1} \delta_{I,j} \Delta I_{t-j} + \theta_w * t + \varepsilon_w$$

Where the subscript w refers to the estimates in the equation of wheat. α is a constant, $\beta_{w,i}$ coefficient of the cointegrating relations (1, -$\gamma_{i,S}$) for (W,S), (F,S) and (O, S), L refers to the number of lags, ε_w is an error term to which the ML assumptions apply, the unrestricted time trend doesn't make it to the model as it fails a relevance test.

$$\Delta W_t = \alpha_w + +\beta_{w,f} * Z_{(w,s)} + \beta_{w,f} * Z_{(f,s)} + \beta_{w,o} * Z_{(o.s)} + \sum_{j=1}^{l-1} \delta_w \Delta W_{t-j} + \sum_{j=1}^{l-1} \delta_f \Delta F_{t-j} + \sum_{j=1}^{l-1} \delta_o \Delta O_{t-j} + \sum_{j=1}^{l-1} \delta_s \Delta S_{t-j} + \varepsilon_w$$

This model is extremely cumbersome, as with twelve lags it estimates 44 differenced lags, 3 cointegrating relations, a constant and a trend.

Starting with a preliminary model (Model 4), we refine the VECM to exclude irrelevant regressors and improve the fit. Note, the issue of upward bias is dealt with by including lags and that of unit roots by differencing. First, all unnecessary lags are excluded, and lag order is kept to two which would yield first difference first order lags (Model 6).

$$\Delta W_t = \alpha_w + \sum_{I \in \{W,F,O\}} \beta_{w,i} * Z_I + \sum_{I \in \{W,F,O,S\}} \delta_w * I_{t-1} + \varepsilon_w$$

Dropping the w subscript for ease, Model 6 is:

$$\Delta W_t = \alpha + \beta_s * Z_{w,s} + \beta_f * Z_{f,s} + \beta_o * Z_{o,s} + \delta_w \Delta W_{t-1} + \delta_f \Delta F_{t-1} + \delta_o \Delta O_{t-1} + \delta_s \Delta S_{t-1} + \varepsilon_w$$

5. Findings

5.1. Model results

The estimates are:

VECM system, lag order 2
Cointegration rank = 3
Case 3: Unrestricted constant
beta (cointegrating vectors)

l_Wheat	1.0000	0.0000	0.0000
l_Fert_idx	0.0000	1.0000	0.0000
l_Oil	0.0000	0.0000	1.0000
l_Sub_idx	-1.0992	-1.6867	-2.9296

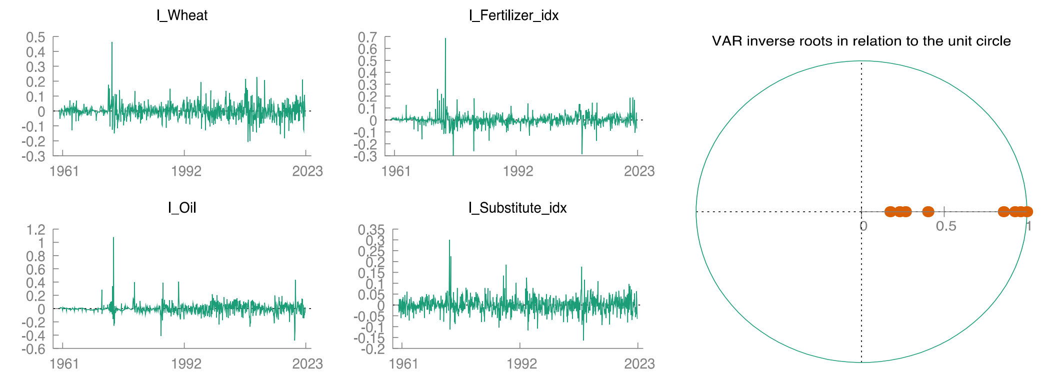
Equation 1: d_l_Wheat

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.136166	0.0396156	3.437	0.0006	***
d_l_Wheat_1	0.266080	0.0383748	6.934	<0.0001	***
d_l_Fertilizer_idx_1	0.0224998	0.0382453	0.5883	0.5565	
d_l_Oil_1	0.0306351	0.0237816	1.288	0.1981	
d_l_Substitute_idx_1	0.120258	0.0513025	2.344	0.0193	**
EC1	−0.0734183	0.0158755	−4.625	<0.0001	***
EC2	0.0204216	0.00969591	2.106	0.0355	**
EC3	0.00873095	0.00403435	2.164	0.0308	**

Source: Author's work, Gretl

Model 6: VECM regression with differenced first lag (2 lags) of wheat, substitutes, fertilizers and oil indices in log-prices.

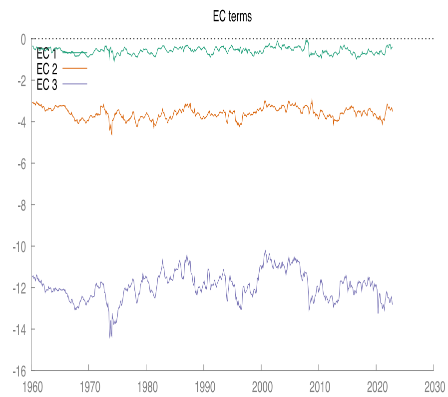
The fertilizers and oil indices are statistically insignificant. Error correction terms have small coefficients and the dominant regressors are first-differenced wheat lag and the first-differenced substitutes index. The fit is very low typical of first differenced data of integrated of degree one variables.



Source: Author's work, Gretl

Figure 5: Residuals in VECM (left) inverse roots in a regular VAR (right)

Our four constituents are all I(1) variables, so a regular VAR has four unit root one for each (Figure 5) The VECM, estimates as an error deviation from a stationary process in first-differences (Figure 6). The fact the variables are I(1), unit-roots suggest a random walk governing the movement of the prices, the attempt at estimating in first differences is usually expected to yield low adjusted R-squared and poorer fits. Nonetheless, we try to improve the fit.



Source: Author's work, Gretl

Figure 6: Error Correction terms fitted values from VECM

Surprisingly enough, both coefficients for lagged difference for fertilizers and oil are statistically irrelevant, while their level values are still contained in the error correction term. The fit does not improve and is still very low. In this case the VECM collapses to:

$$\Delta W_t = \alpha_w + \sum_{I \in \{W,F,O\}} \beta_{w,i} * Z_I + \delta_w \Delta W_{t-1} + \delta_s \Delta S_{t-1} + \varepsilon_w$$

Or,

$$\Delta W_t = \alpha + \beta_s * Z_{w,s} + \beta_f * Z_{f,s} + \beta_o * Z_{o,s} + \delta_w \Delta W_{t-1} + +\delta_s \Delta S_{t-1} + \varepsilon_w$$

Saving the Zs from the VECM from Model 6, we simply regress ΔW_t using a VAR, on the three Z_i's along with ΔW_{t-1} -we add ΔW_{t-2} - and ΔS_{t-1}.

This yields Model 7, only the regressors with statistical significance at 1% are kept, this model has so far the best fit (yet a low one). ΔW_{t-2} and $EC2$ are statistically insignificant, whereas for EC1 and EC3, the null that they are zero is rejected, even though they have small coefficients relative to those of ΔW_{t-1} and ΔS_{t-1}. A case could be made to exclude them as their impact is low.

VAR system, lag order 1

Equation 1: d_l_Wheat

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.104928	0.0360026	2.914	0.0037	***
d_l_Wheat_1	0.242110	0.0347223	6.973	<0.0001	***
d_l_Substitute_idx	0.445349	0.0457108	9.743	<0.0001	***
EC1	−0.0860853	0.0148221	−5.808	<0.0001	***
EC3	0.0129156	0.00336507	3.838	0.0001	***

Source: Author's work, Gretl

Model 7: VAR(2) regression of first difference wheat on two lags, FD substitutes index in log-prices and the EC terms from the VECM in Model 6

5.2. Model adjustment

Eliminating the error terms, results in a simple first difference autoregression of wheat and the substitutes in log-prices. Again, both oil and fertilizers do no make it, as well as lagged substitutes. This simple model stands for comparison, is slightly more biased downwards due to omitted variables, has poorer fit and does not account for the effect of the supply variables. Its simplicity however is prized and a case could be made to exclude the error terms as their amplitude is likely not to result in any significant impact. The second lag is picked up again by the VAR, in the absence of the error correction terms:

VAR system, lag order 2
Equation 1: d_l_Wheat

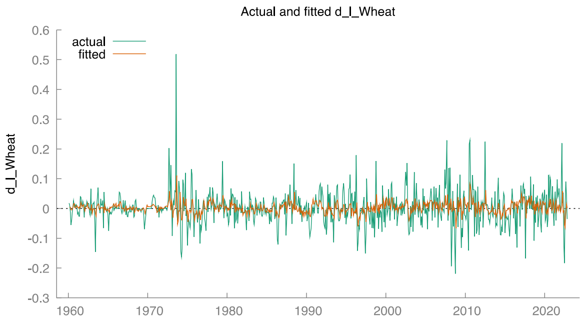
	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.00125798	0.00187218	0.6719	0.5018	
d_l_Wheat_1	0.214462	0.0460768	4.654	<0.0001	***
d_l_Wheat_2	−0.0985516	0.0405692	−2.429	0.0154	**
d_l_Substitute_idx	0.420835	0.0841995	4.998	<0.0001	***

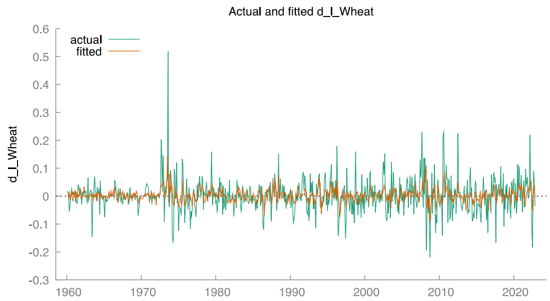
Source: Author's work, Gretl

Model 9: VAR of first difference wheat on two lags and first differenced substitutes index in log-prices

This VAR, discards cointegration for a simpler model including first-difference lags and the differenced substitutes index all significant at the 1% level. There is no bias, no autocorrelation. For the case of heteroskedastic robust standard errors, the second lag is at 1.54%, significant at the 5% level, slightly above the 1% confidence interval. The coefficients for the supply variables are on the edge of a 10% (within) confidence interval and do not make it to this model.

To the naked eye the models (in orange) mimic the fluctuation in first differences in actual data albeit at a lower amplitude.





Source: Author's work, Gretl

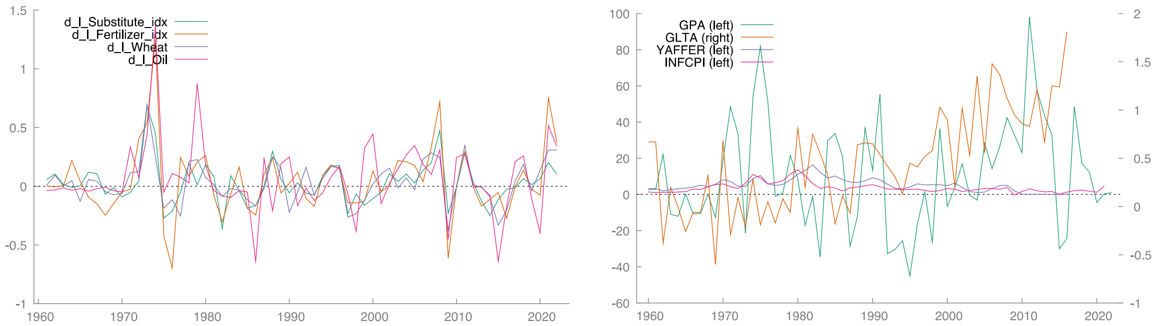
Figure 7: VECM Model 7 with EC terms (left) and VAR Model 9 without EC terms (right)

The information criteria are so close that a subjective choice will be made, that of opting for VAR Model 9 for its simplicity, despite the VECM error correction terms being statistically significant their amplitude is small, with a low fit overcrowding the model for less than 3 percentage points of fit is perhaps not as lucrative. The model is kept, but we shall continue with the VAR Model 9.

5.3. Extensions of the model

As a way to verify our findings, we attempt two extensions of this model, the detailed analysis is included in Appendix A.

First, we explore structural breaks using the Chow test, Only the first Oil shock however, is statistically significant and is included as a period dummy in the Chow augmented regression, this apart from overcrowding the model does not improve the fit and is discarded.



Source: Author's work, Gretl

Figure 9: Yearly replica of the model components (left) and the new instrument (right)

Second, and most important, we attempt a first departure from this dataset. We look for factors that could influence demand and supply. Using yearly historical data for interest rates, inflation, population growth, precipitation and land temperature anomalies, a detailed discussion of these proxies is in the appendix. Resampling our data to yearly values, treating regressors as endogenous and using these instrument in an over-identified VAR. The results collapse to an autoregression of second order in in first differences, as the instruments are made redundant.
This is a strong point for the perfect markets hypothesis as all outside information is incorporated in the price. Note that, the substitutes index, is also lost in this calibration. We return to Model 9 in monthly data to discuss our findings:

6. Synthesis

The VAR's first equation for differenced log-price wheat is:

$$\Delta W_t = 0.214462 \times \Delta W_{t-1} - 0.0985516 \times \Delta W_{t-2} + 0.420835 \times \Delta Sub_t + \varepsilon_t$$

First difference lags' standard errors of roughly 0.04 each or around one fifth and one half of the magnitude of the first and second lags infer on the existence of some stability with regards to the variance of wheat prices. In other words, prices do not skyrocket to extremely high prices (in the range of double and ten-fold for example) as the commodity markets are extremely regulated and more often than not are subject to interventions by higher authorities. More importantly wheat is a vital commodity and its price is subsidized for the consumers in many areas of the world, it is so essential in peoples' diets that extreme increases could spark serious uprisals, volatility is always thus met some form of regulation and prices stay within a limited range. In this monthly-data model, a change in the previous month is likely to increase the change in this period by one fifth or 0.21 log-USD per Bushel. The period before that is negatively correlated and impacts by -0.09 log-USD per Bushel this could simply be the result of cyclical behaviour of differenced unit-root variables and is somewhat reminiscent of signal-like graphs. This adjustment creates a stabilizing factor within the equation and the stationarity around the mean.

A more important finding is the substitutes index, this is a logical link within the model and is reflective of microeconomic theory. As changes in the price of substitute (Maize, Rice and Barley) increase so does the change in wheat prices: a unit increase in the index (weighted average by quantity) results in a 0.4 log-USD per Bushel increase.

As preferences in diets are usually stable throughout the period, the proportion of quantity demanded and thus produced of each grain would stay the same, this justification for the index serves to describe the relationship between wheat and its substitutes. Higher substitutes prices lead to higher wheat prices as the production factors are somewhat similar in each. This variable is twice as volatile as wheat lags but due to its proportionally higher amplitude results in the same t-ratio as the first difference first lag.

Only 15.66% -a little less than a sixth- of the fluctuation in the data is explained by this model.

The attempt to instrument it with external variables known to influence prices in general (inflation, interest rates), crops production specifically (Precipitation and temperature anomalies) and demand (population growth) has not been fruitful none of these have been statistically significant and the model reverts back to an Autoregressive process of second order. In comparison with the monthly data, even the substitutes index has been dropped. The coefficients for the first and second lags are almost equal and of opposite sign. This is suggestive of one coefficient for both in differences or a second differenced regressor of log-wheat prices in first lag. This feature is not present in monthly data and we shall focus our finding on the latter.

7. Conclusion

The project outlines a specific setting where only commodity price data is available and tries to extract a model of inference on wheat prices, relaxing this assumption later, by adding external data did not improve it.

Going back to the farmer's theoretical question, of whether he can predict the price of wheat by observing the price of its substitutes (the neighbouring fields) and his production process (fertilizers, oil). The answer, for the production process is negative since fertilizer prices and oil prices are statistically insignificant. This is in line with the no-correlation hypothesis of wheat and oil prices.

Observing substitutes however, does help the farmer infer on the price of his crop and so does remembering the previous prices. This is not a memoryless process and thus keeping count of past prices does indeed improve the farmer's insight on the price of his crop. These two results are indeed the most important takeaways from the model.

The farmer should have in mind that the predictive capability of this model is weak and that this framework should not be used for predicting purposes as only a fraction -one sixth- is explained by it. An easy suggestion would be that other information is out there unknown that could explain the model, this information would have to be to the exception of inflation, interest rates, population growth, precipitation and land temperature anomalies since none of them are statistically significant.

However, in the absence of wheat data, substitutes are highly useful and prices of the synthetic product could be used to infer on wheat prices, this proximity could be used as a control method to evaluate policy regarding wheat price regulation and vice-versa.

The results conveyed by the model discard the hypothesis that prices follow a random walk, even by eliminating the second lag and the substitutes index the coefficient is 0.25 and not unity, it is not a Markov process.

All available information, as shown by the extension of the model, is completely contained within the first two lags and the substitutes index (in the case of monthly data). Instruments were redundant and no further improvement resulted from adding them. This is in line with the efficient market hypothesis that all available information is reflected in the price and that no arbitrage gain is possible.

In all, the change in wheat prices is not completely random nor is it related to usual macroeconomic variables. It contains all available information and contain memory of up to two periods, its behaviour can be mimicked to a small extent using its substitutes.

Wheat prices are also heavily regulated and subject to government interventions grants, price caps, subventions, tax breaks, import and export barriers these factors are likely to affect price movement and distort our finding.

The consumption random walk hypothesis, could be used to justify the large hidden element in the model, this micro-founded hypothesis suggesting prices, as tools rather than targets, adjust to equate demand and supply. This could be the closing stage of our analysis and the opening of the reader's.

Self-reflective critique:

Throughout the analysis, the coefficients of the indices are assumed as given and static, , this assumes that the proportion of grains output is known, which is external to the model. This is indeed a simplification that could be replaced by cointegration coefficients in the case of the fertilizer index and the substitutes.

To keep the model simple the cointegration vectors were dropped from the final model, since the error correction terms amplitudes was very small (0.01<) and did not improve the fit significantly. This does not make them completely insignificant but rather less relevant.

The use of ML estimation, with no distributional assumption is also a drawback from this analysis, since the data is not normally distributed, only by the law of large numbers, can one make the case for a long enough period of monthly data that the assumption could be held although very cautiously.

An important factor, that would have been in line with microeconomic theory would be grain production and consumption historical data used in a market-clearing framework where prices adjust to reflect demand and supply fluctuation, since this data has not been available, this type of analysis was not possible and only relative proportion of current grain production is used for the index assuming some monotonicity in consumer preference.

Futures contracts data were not available for the majority of the period, these would have been used as complementary regressors reflecting the speculative behaviour of the market/ consumers, the effect of expectations of increased prices on current prices would have painted a more complete picture of exchange traded commodities as it is a driving factor.

References

Balabanoff, S. (1995). Oil futures prices and stock management A cointegration analysis. *Energy Economics*, 17(3), pp.205–210. doi:https://doi.org/10.1016/0140-9883(95)00024-o.

Enghiad, A., Ufer, D., Countryman, A.M. and Thilmany, D.D. (2017). An Overview of Global Wheat Market Fundamentals in an Era of Climate Concerns. *International Journal of Agronomy*, 2017, pp.1–15. doi:https://doi.org/10.1155/2017/3931897.

Frankel, J.A. and Rose, A.K. (2010). Determinants of Agricultural and Mineral Commodity Prices. *HKS Faculty Research Working Paper Series*. [online] Available at: http://nrs.harvard.edu/urn-3:HUL.InstRepos:4450126.

FRED (2023). *Effective Federal Funds Rate*. [online] Stlouisfed.org. Available at: https://fred.stlouisfed.org/series/FEDFUNDS.

Macrotrends LLC (2020). *U.S. Inflation Rate 1960-2020*. [online] Macrotrends.net. Available at: https://www.macrotrends.net/countries/USA/united-states/inflation-rate-cpi.

Mir B., A.B., Brooks, N.L. and Mc Elroy, R.G. (2000). *Characteristics of US Wheat farming*
[online] U.S. Department of Agriculture: Resource Economics Division, Economic Research Service. Available at: https://www.ers.usda.gov/webdocs/publications/47116/51556_sb968b.pdf?v=0 Statistical Bulletin No. 968.

National Centers for Environmental Information. (n.d.). *Global Surface Temperature Anomalies | National Centers for Environmental Information (NCEI)*. [online] Available at: https://www.ncei.noaa.gov/access/monitoring/global-temperature-anomalies/anomalies Global monthly land temperatures

National Oceanic and Atmospheric Administration. (2022). *Global precipitation anomaly*. [online] Available at: https://www.noaa.gov via ourworldindata.org.

Nazlioglu, S. and Soytas, U. (2012). Oil price, agricultural commodity prices, and the dollar: A panel cointegration and causality analysis. *Energy Economics*, 34(4), pp.1098–1104. doi:https://doi.org/10.1016/j.eneco.2011.09.008.

Shahbandeh, M. (2023). *Grain production worldwide 2022/23, by type*. [online] Statista. Available at: https://www.statista.com/statistics/263977/world-grain-production-by-type/.

Symeonidis, L., Prokopczuk, M., Brooks, C. and Lazar, E. (2012). Futures basis, inventory and commodity price volatility: An empirical analysis. *Economic Modelling*, 29(6), pp.2651–2663. doi:https://doi.org/10.1016/j.econmod.2012.07.016.

World Bank (2023). *Inflation, consumer prices (annual %) - United States | Data*. [online] data.worldbank.org. Available at: https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG?locations=US.

World Bank (2021). *Population, Total | Data.* [online] Worldbank.org. Available at: https://data.worldbank.org/indicator/SP.POP.TOTL.

World Bank (2022a). *SPECIAL FOCUS The Impact of the War in Ukraine on Commodity Markets*. [online] Available at: https://openknowledge.worldbank.org/server/api/core/bitstreams/c7f65ff2-9e31-5b2a-9ceb-ed37900e8223/content.

World Bank. (2022b). *World Bank Commodity Price Data (The Pink Sheet)*. [online] Available at: https://thedocs.worldbank.org/en/doc/5d903e848db1d1b83e0ec8f744e55570-0350012021/related/CMO-Historical-Data-Monthly.xlsx Monthly prices.

World Bank. (2023). *World Bank Commodity Price Data (The Pink Sheet)*. [online] Available at: https://www.worldbank.org/en/research/commodity-markets Annual Prices .

Zeneli, F. (2022). Energy and Grains Prices Cointegration and Causality Linkage. *Commodities*, 1(2), pp.167–180. doi:https://doi.org/10.3390/commodities1020011.

Appendices:

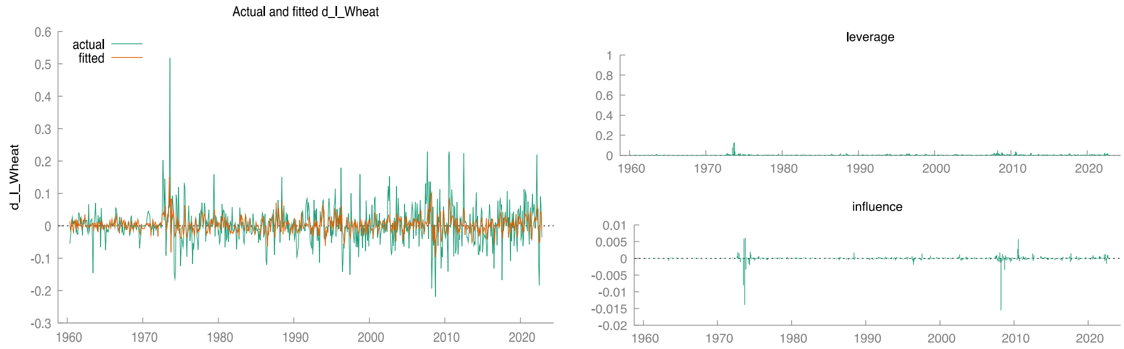
Appendix A : Extensions of the model

5.3.1. Extension 1: Exploring structural breaks

Perhaps the runs in financial crises could explain some of the fluctuations, we test for structural breaks using an OLS regression with the same variables as in Model 9, there is however an interesting contrast when estimating with or without robust errors,

Using the Chow test, we investigate the presence of financial crises we have described earlier, and try to estimate period specific coefficients that would improve the data. We therefore test for structural break, in each period of global financial crisis. Seven global crises are included, the two oil shocks crises, the banking crisis of ’84, the dotcom bubble, the great recession, the covid crisis and the Ukraine crisis is estimated separately as well. The test requires heteroskedastic error terms.

For the model estimated with robust errors, Model 10A, there are structural breaks for every financial crisis to the exception of the banking crisis of 1984 and the Dotcom bubble. This model has the required specification and would yield better results. Only the first Oil shock however, is statistically significant and is included as a period dummy in the Chow augmented regression, this apart from overcrowding the model does not improve the fit.



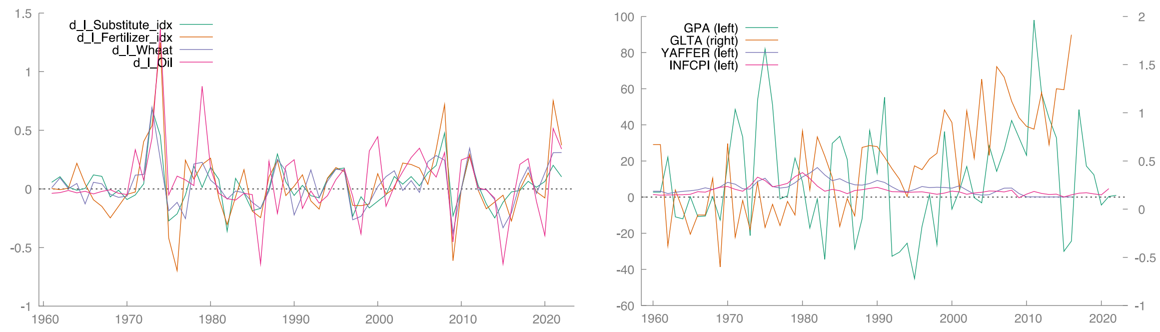
Source: Author’s work, Gretl

Figure 8: OLS fitted values (left) and influential values (right)

The analysis with information exclusive to this dataset now comes to an end. The models do not encompass all there is to know to predict or describe the wheat price variable. We shall now look at external data, and try to use instruments to improve the fit of the model.

Returning to our theoretical question, the wheat farmer only observing the surrounding fields cannot fully estimate the price of his crop just by observing other fields. In other words, he can only infer on one-sixth of the change in wheat prices, and thus the majority is left untapped.
Despite the production factors being irrelevant, the substitutes and thus the demand factors are of significance, the issue is that there is important information lying outside the spectrum of commodity prices alone. We shall look at them now:

5.3.2. Extension 2: Exploring additional macroeconomic data



Source: Author's work, Gretl

Figure 9: Yearly replica of the model components (left) and the new instrument (right)

An interesting proxy for the state of the economy in general, and the supply side specifically, is the interest rate, higher rates broadly imply less borrowing, less input and less production. We shall use the Federal Reserve of St. Louis Federal Funds Monthly Effective Rate in Percentage -Not Seasonally Adjusted-. Interest rates are the main tool for monetary policy and are usually reflective of high and low stress periods.
Inflation, is also a measure of the propensity to consume and is a proxy for increased demand unmet with supply it can also be thought of as an aggregate price index that likely has an influence on individual prices. As wheat is a prominent consumption good it is a constituent part of the CPI index and thus its exogeneity is questionable.

Another important factor is the weather, in the context of a global climate change examining deviations in global temperatures from the average monthly temperatures could explain partly the losses incurred to crops as a result of drought, extreme weather and such. We shall use the global land temperature anomalies dataset from the National Centre for Environmental Information of monthly anomalies with respect to the century average (1901-2000), the data represents deviation in Celsius (or Fahrenheit) from the average temperature. As wheat is grown globally, global average land temperature is used.
Precipitation, as water is an essential factor in the production of crops, and for any living creature for that matter, drought -and to a lesser extent, in the case of wheat, floods- can damage the growth cycle if not annihilate it. Low precipitation have been frequent in the last decades and more often than not damage has been reported. We use the Global precipitation anomaly data from National Oceanic & Atmospheric Administration of yearly deviations from a moving average with respect to the base period (1901-2000).
High temperatures are usually accompanied by low precipitation and correlation would be high.

Population and population growth are important demand factor, as each person needs to feed, an increase in population should results directly in an increase in consumption and a surge in prices, as land is restrained. We use population growth as the percentage change in world population total by year data from the World Bank.

As most of the instruments are yearly, we resample the data for yearly prices. We use yearly moving average for the monthly interest rates picking year end data for each year simply resulting in yearly average interest rates federal fund effective rate in percentages (YAFFER). Global land temperature anomalies in Celcius (GLTA) and population growth in percentages (POPGR).

We resample our data and repeat the VAR for yearly data instead. This should be a reference point to compare with the instrumented model.
The instrument dataset contains missing values from 2016 (GLTA, YAFFER) onwards 2021(INFCPI) and only POPTOT and GPA are fully descriptive of the period.

The model in yearly data, collapses to first-difference autoregression model of second degree. Neither fertilizers, substitutes nor oil make it to this yearly replica of the monthly model. This would serve as our benchmark.

```
VAR system, lag order 2

Equation 1: d_l_Wheat

              coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const         0.0287822   0.0227134     1.267    0.2102
  d_l_Wheat_1   0.379112    0.126409      2.999    0.0040   ***
  d_l_Wheat_2  -0.335948    0.128993     -2.604    0.0117   **
```

Source: Author's work, Gretl

Model 12: VAR of first difference wheat (yearly) instrumented with Precipitation, temperature, interest rates, inflation and population growth.

This VAR model, is a first difference of log wheat on its two lags, oil, substitutes, and fertilizers indices instrumented with the Global precipitation anomaly (deviation), Global Land temperature anomaly, Yearly average Federal fund interest rates, Inflation index and log population growth. These datum have been tested for a unit root and have been included as I(0) to match the differenced data. There is also no evidence of cointegration between them.

The instrument are redundant and all fail significance tests, except for the inflation index that is barely insignificant at the 10% level. (10%<10.93%)

Despite the overcrowding of the model, the fit has not changed. The instruments have failed to add any additional information and the model collapses to a VAR(2) just as in the monthly data.

Appendix B: Models

Model 1: OLS, using observations 1960:01-2022:11 (T = 755)
Dependent variable: l_Wheat

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	−0.207873	0.0605838	−3.431	0.0006	***
l_Substitute_idx	1.02510	0.0120125	85.34	<0.0001	***

Mean dependent var	4.938761	S.D. dependent var	0.515774
Sum squared resid	18.79699	S.E. of regression	0.157996
R-squared	0.906287	Adjusted R-squared	0.906163
F(1, 753)	7282.203	P-value(F)	0.000000
Log-likelihood	322.8168	Akaike criterion	−641.6335
Schwarz criterion	−632.3801	Hannan-Quinn	−638.0691
rho	0.930382	Durbin-Watson	0.140969

White's test for heteroskedasticity -
Null hypothesis: heteroskedasticity not present
Test statistic: LM = 9.73707
with p-value = P(Chi-square(2) > 9.73707) = 0.00768461

Breusch-Pagan test for heteroskedasticity -
Null hypothesis: heteroskedasticity not present
Test statistic: LM = 8.52763
with p-value = P(Chi-square(1) > 8.52763) = 0.00349795

Test for normality of residual -
Null hypothesis: error is normally distributed
Test statistic: Chi-square(2) = 1.95335
with p-value = 0.37656

RESET test for specification -
Null hypothesis: specification is adequate
Test statistic: F(2, 751) = 8.47358
with p-value = P(F(2, 751) > 8.47358) = 0.000229551

Model 1: Preliminary OLS regression of wheat log-price on the substitutes index

Model 2: OLS, using observations 1960:01-2022:11 (T = 755)
Dependent variable: l_Wheat

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	2.46131	0.0708652	34.73	<0.0001	***
l_Fertilizer_idx	0.444478	0.0200068	22.22	<0.0001	***
l_Oil	0.111086	0.0109024	10.19	<0.0001	***

Mean dependent var	4.938761	S.D. dependent var	0.515774
Sum squared resid	19.80474	S.E. of regression	0.162284
R-squared	0.901263	Adjusted R-squared	0.901001
F(2, 752)	3432.106	P-value(F)	0.000000
Log-likelihood	303.1021	Akaike criterion	−600.2042
Schwarz criterion	−586.3240	Hannan-Quinn	−594.8575
rho	0.926391	Durbin-Watson	0.149161

Breusch-Pagan test for heteroskedasticity -
Null hypothesis: heteroskedasticity not present
Test statistic: LM = 25.8954
with p-value = P(Chi-square(2) > 25.8954) = 2.38168e-06

RESET test for specification -
Null hypothesis: specification is adequate
Test statistic: F(2, 750) = 10.6042
with p-value = P(F(2, 750) > 10.6042) = 2.87454e-05

Test for normality of residual -
Null hypothesis: error is normally distributed
Test statistic: Chi-square(2) = 37.4988
with p-value = 7.19835e-09

Model 2: Preliminary OLS regression of wheat log-price on the fertilizers index and oil index log-price

VECM system, lag order 6
Maximum likelihood estimates, observations 1960:07-2022:11 (T = 749)
Cointegration rank = 2

Case 3: Unrestricted constant
beta (cointegrating vectors, standard errors in parentheses)

l_Wheat	1.0000	0.0000
	(0.0000)	(0.0000)
l_Fertilizer_idx	0.0000	1.0000
	(0.0000)	(0.0000)
l_Oil	-0.079342	-0.093279
	(0.028022)	(0.064299)
l_Substitute_idx	-0.86017	-1.3938
	(0.083429)	(0.19144)

alpha (adjustment vectors)

l_Wheat	-0.078774	0.027566
l_Fertilizer_idx	0.045554	-0.032882
l_Oil	0.015770	-0.031156
l_Substitute_idx	0.037043	0.019027

Log-likelihood = 4555.2453
Determinant of covariance matrix = 6.131191e-11
AIC = -11.8965
BIC = -11.2799
HQC = -11.6589

Equation 1: d_l_Wheat

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.0983859	0.0266477	3.692	0.0002	***
d_l_Wheat_1	0.290040	0.0396273	7.319	<0.0001	***
d_l_Wheat_2	−0.0157177	0.0407634	−0.3856	0.6999	
d_l_Wheat_3	0.0259785	0.0404654	0.6420	0.5211	
d_l_Wheat_4	0.0255371	0.0400339	0.6379	0.5237	
d_l_Wheat_5	0.0730585	0.0400596	1.824	0.0686	*
d_l_Fertilizer_idx_1	0.0540197	0.0408315	1.323	0.1863	
d_l_Fertilizer_idx_2	−0.108996	0.0413483	−2.636	0.0086	***
d_l_Fertilizer_idx_3	0.0175587	0.0412648	0.4255	0.6706	
d_l_Fertilizer_idx_4	−0.0991372	0.0408011	−2.430	0.0154	**
d_l_Fertilizer_idx_5	0.120563	0.0393551	3.063	0.0023	***
d_l_Oil_1	0.0231243	0.0244393	0.9462	0.3444	
d_l_Oil_2	0.0181369	0.0250784	0.7232	0.4698	
d_l_Oil_3	−0.0228913	0.0252426	−0.9069	0.3648	
d_l_Oil_4	0.0436855	0.0250731	1.742	0.0819	*
d_l_Oil_5	0.00249753	0.0245949	0.1015	0.9191	
d_l_Substitute_idx_1	0.133352	0.0538902	2.475	0.0136	**
d_l_Substitute_idx_2	−0.0686695	0.0567873	−1.209	0.2270	
d_l_Substitute_idx_3	−0.0132361	0.0575102	−0.2302	0.8180	
d_l_Substitute_idx_4	0.0778049	0.0573679	1.356	0.1754	
d_l_Substitute_idx_5	−0.140909	0.0553586	−2.545	0.0111	**
EC1	−0.0787737	0.0180645	−4.361	<0.0001	***
EC2	0.0275659	0.00980659	2.811	0.0051	***

Mean dependent var	0.002711	S.D. dependent var	0.057606
Sum squared resid	2.142423	S.E. of regression	0.054361
R-squared	0.136877	Adjusted R-squared	0.109495
rho	−0.008996	Durbin-Watson	2.017879

Model 3: Preliminary VECM regression of wheat, substitutes, fertilizers and oil indices in log-prices. (Two cointegrating relations)

VECM system, lag order 12

Maximum likelihood estimates, observations 1961:01-2022:11 (T = 743)
Cointegration rank = 3
Case 3: Unrestricted constant
beta (cointegrating vectors, standard errors in parentheses)

l_Wheat	1.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)
l_Fert_idx	0.0000	1.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)
l_Oil	0.0000	0.0000	1.0000
	(0.0000)	(0.0000)	(0.0000)
l_Sub_idx	-1.1121	-1.7251	-3.2734
	(0.051372)	(0.10636)	(0.29148)

alpha (adjustment vectors)

l_Wheat	-0.10194	0.015739	0.015327
l_Fertilizer_idx	0.051411	-0.042120	0.0084605
l_Oil	0.0069496	-0.0099809	-0.0099572
l_Substitute_idx	0.014511	0.019091	0.00070410

Log-likelihood = 4614.0593
Determinant of covariance matrix = 4.7438496e-11
AIC = -11.8925
BIC = -10.6762
HQC = -11.4236

Equation 1: d_l_Wheat

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.203899	0.0491888	4.145	<0.0001	***
d_l_Wheat_1	0.287116	0.0417407	6.879	<0.0001	***
d_l_Wheat_2	0.00959747	0.0430484	0.2229	0.8236	
d_l_Wheat_3	0.0499865	0.0427793	1.168	0.2430	
d_l_Wheat_4	0.0400618	0.0426977	0.9383	0.3484	
d_l_Wheat_5	0.102753	0.0421762	2.436	0.0151	**
d_l_Wheat_6	0.0515275	0.0432342	1.192	0.2337	
d_l_Wheat_7	−0.0488888	0.0433297	−1.128	0.2596	
d_l_Wheat_8	0.0141348	0.0433100	0.3264	0.7442	
d_l_Wheat_9	0.0537403	0.0433460	1.240	0.2155	
d_l_Wheat_10	0.0704154	0.0429429	1.640	0.1015	
d_l_Wheat_11	0.0607014	0.0427508	1.420	0.1561	
d_l_Fertilizer_idx_1	0.0349460	0.0427414	0.8176	0.4139	
d_l_Fertilizer_idx_2	−0.0767752	0.0433719	−1.770	0.0771	*
d_l_Fertilizer_idx_3	−0.00619287	0.0434103	−0.1427	0.8866	
d_l_Fertilizer_idx_4	−0.0859449	0.0434301	−1.979	0.0482	**
d_l_Fertilizer_idx_5	0.0988820	0.0431740	2.290	0.0223	**
d_l_Fertilizer_idx_6	0.0431431	0.0423222	1.019	0.3084	
d_l_Fertilizer_idx_7	0.0441634	0.0429087	1.029	0.3037	
d_l_Fertilizer_idx_8	−0.0375015	0.0430777	−0.8706	0.3843	
d_l_Fertilizer_idx_9	0.0752875	0.0428438	1.757	0.0793	*
d_l_Fertilizer_idx_10	−0.00278691	0.0425377	−0.06552	0.9478	
d_l_Fertilizer_idx_11	0.0106437	0.0411447	0.2587	0.7960	
d_l_Oil_1	0.0247753	0.0254286	0.9743	0.3302	
d_l_Oil_2	0.0206549	0.0260178	0.7939	0.4275	
d_l_Oil_3	−0.0304542	0.0261151	−1.166	0.2440	
d_l_Oil_4	0.0255747	0.0263185	0.9717	0.3315	
d_l_Oil_5	−0.00989747	0.0263428	−0.3757	0.7072	
d_l_Oil_6	−0.0428893	0.0262497	−1.634	0.1027	
d_l_Oil_7	−0.00323344	0.0257962	−0.1253	0.9003	
d_l_Oil_8	0.0134372	0.0257868	0.5211	0.6025	
d_l_Oil_9	−0.00183368	0.0257993	−0.07107	0.9434	
d_l_Oil_10	−0.0594711	0.0254471	−2.337	0.0197	**
d_l_Oil_11	0.00215352	0.0250154	0.08609	0.9314	

d_l_Substitute_idx_1	0.131424	0.0560526	2.345	0.0193	**
d_l_Substitute_idx_2	−0.0783097	0.0587573	−1.333	0.1830	
d_l_Substitute_idx_3	−0.0133439	0.0596098	−0.2239	0.8229	
d_l_Substitute_idx_4	0.0614258	0.0594445	1.033	0.3018	
d_l_Substitute_idx_5	−0.103459	0.0591277	−1.750	0.0806	*
d_l_Substitute_idx_6	−0.0925479	0.0586898	−1.577	0.1153	
d_l_Substitute_idx_7	0.0114534	0.0598480	0.1914	0.8483	
d_l_Substitute_idx_8	−0.0317762	0.0598426	−0.5310	0.5956	
d_l_Substitute_idx_9	0.0872568	0.0597880	1.459	0.1449	
d_l_Substitute_idx_10	−0.0156305	0.0596697	−0.2620	0.7934	
d_l_Substitute_idx_11	0.0682680	0.0578628	1.180	0.2385	
EC1	−0.101939	0.0215662	−4.727	<0.0001	***
EC2	0.0157387	0.0116222	1.354	0.1761	
EC3	0.0153270	0.00460209	3.330	0.0009	***

Mean dependent var	0.002672	S.D. dependent var	0.057827
Sum squared resid	2.021636	S.E. of regression	0.053934
R-squared	0.185225	Adjusted R-squared	0.130125
rho	0.003777	Durbin-Watson	1.992268

Model 4: Preliminary VECM regression of wheat, substitutes, fertilizers and oil indices in log-prices. (Three cointegrating relations)

VECM system, lag order 3
Maximum likelihood estimates, observations 1960:04-2022:11 (T = 752)
Cointegration rank = 3
Case 5: Unrestricted trend and constant
beta (cointegrating vectors, standard errors in parentheses)

l_Wheat	1.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)
l_Fert_idx	0.0000	1.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)
l_Oil	0.0000	0.0000	1.0000
	(0.0000)	(0.0000)	(0.0000)
l_Sub_idx	-1.1604	-1.3684	-3.3195
	(0.10340)	(0.14813)	(0.71925)

alpha (adjustment vectors)

l_Wheat	-0.066821	0.016541	0.010258
l_Fertilizer_idx	0.055113	-0.046101	0.00054184
l_Oil	0.076072	-0.043090	-0.014312
l_Substitute_idx	0.037485	0.011615	-0.0010365

Log-likelihood = 4523.5601
Determinant of covariance matrix = 7.0019081e-11
AIC = -11.8818
BIC = -11.5376
HQC = -11.7492

Equation 1: d_l_Wheat

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.122627	0.0422874	2.900	0.0038	***
d_l_Wheat_1	0.269264	0.0388249	6.935	<0.0001	***
d_l_Wheat_2	−0.0215412	0.0398120	−0.5411	0.5886	
d_l_Fertilizer_idx_1	0.0506068	0.0401239	1.261	0.2076	
d_l_Fertilizer_idx_2	−0.0754741	0.0394761	−1.912	0.0563	*
d_l_Oil_1	0.0231009	0.0242643	0.9521	0.3414	
d_l_Oil_2	0.00886086	0.0243091	0.3645	0.7156	
d_l_Substitute_idx_1	0.144770	0.0535038	2.706	0.0070	***
d_l_Substitute_idx_2	−0.0447812	0.0546504	−0.8194	0.4128	
time	−1.30915e-05	1.20586e-05	−1.086	0.2780	
EC1	−0.0668212	0.0166715	−4.008	<0.0001	***
EC2	0.0165413	0.0108200	1.529	0.1267	
EC3	0.0102579	0.00368366	2.785	0.0055	***

Mean dependent var	0.002558	S.D. dependent var	0.057552
Sum squared resid	2.226044	S.E. of regression	0.054884
R-squared	0.105117	Adjusted R-squared	0.090586
rho	0.001856	Durbin-Watson	1.995797

Model 5: Preliminary VECM regression up to 3 lags of wheat, substitutes, fertilizers and oil indices in log-prices. (Three cointegrating relations)

VECM system, lag order 2
Maximum likelihood estimates, observations 1960:03-2022:11 (T = 753)
Cointegration rank = 3
Case 3: Unrestricted constant
beta (cointegrating vectors, standard errors in parentheses)

l_Wheat	1.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)
l_Fert_idx	0.0000	1.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)
l_Oil	0.0000	0.0000	1.0000
	(0.0000)	(0.0000)	(0.0000)
l_Sub_idx	-1.0992	-1.6867	-2.9296
	(0.054191)	(0.098528)	(0.35064)

alpha (adjustment vectors)

l_Wheat	-0.073418	0.020422	0.0087310
l_Fertilizer_idx	0.051125	-0.040712	0.0033032
l_Oil	0.072174	-0.032010	-0.017117
l_Substitute_idx	0.040749	0.013211	-0.0024962

Log-likelihood = 4505.6083
Determinant of covariance matrix = 7.4621274e-11
AIC = -11.8715
BIC = -11.6504
HQC = -11.7863

Equation 1: d_l_Wheat

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.136166	0.0396156	3.437	0.0006	***
d_l_Wheat_1	0.266080	0.0383748	6.934	<0.0001	***
d_l_Fertilizer_idx_1	0.0224998	0.0382453	0.5883	0.5565	
d_l_Oil_1	0.0306351	0.0237816	1.288	0.1981	
d_l_Substitute_idx_1	0.120258	0.0513025	2.344	0.0193	**
EC1	−0.0734183	0.0158755	−4.625	<0.0001	***
EC2	0.0204216	0.00969591	2.106	0.0355	**
EC3	0.00873095	0.00403435	2.164	0.0308	**

Mean dependent var	0.002571	S.D. dependent var	0.057515
Sum squared resid	2.238018	S.E. of regression	0.054809
R-squared	0.100336	Adjusted R-squared	0.091883
rho	0.013020	Durbin-Watson	1.973347

Model 6: VECM regression with differenced first lag (2 lags) of wheat, substitutes, fertilizers and oil indices in log-prices.

Model 7: OLS, using observations 1960:03-2022:11 (T = 753)
Dependent variable: d_l_Wheat

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.135317	0.0396353	3.414	0.0007	***
d_l_Substitute_idx_1	0.126960	0.0501225	2.533	0.0115	**
EC1	−0.0719077	0.0158471	−4.538	<0.0001	***
EC2	0.0195341	0.00961097	2.032	0.0425	**
EC3	0.00883489	0.00401747	2.199	0.0282	**
d_l_Wheat_1	0.268541	0.0383123	7.009	<0.0001	***

Mean dependent var	0.002571	S.D. dependent var	0.057515
Sum squared resid	2.246644	S.E. of regression	0.054841
R-squared	0.096868	Adjusted R-squared	0.090823
F(5, 747)	16.02440	P-value(F)	5.07e-15
Log-likelihood	1120.747	Akaike criterion	−2229.493
Schwarz criterion	−2201.749	Hannan-Quinn	−2218.805
rho	0.011440	Durbin's h	NA

Model 7: OLS regression of first difference wheat on FD substitutes indices in log-prices and the EC terms from the VECM in Model 6 , (fertilizers and oil are statistically insignificant)

Model 11: OLS, using observations 1960:03-2022:11 (T = 753)
Dependent variable: d_l_Wheat

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.00109042	0.00193795	0.5627	0.5738	
d_l_Substitute_idx	0.405680	0.0459654	8.826	<0.0001	***
d_l_Wheat_1	0.191855	0.0343111	5.592	<0.0001	***

Mean dependent var	0.002571	S.D. dependent var	0.057515
Sum squared resid	2.111697	S.E. of regression	0.053062
R-squared	0.151116	Adjusted R-squared	0.148853
F(2, 750)	66.75661	P-value(F)	2.08e-27
Log-likelihood	1144.069	Akaike criterion	−2282.138
Schwarz criterion	−2268.266	Hannan-Quinn	−2276.794
rho	0.027347	Durbin's h	2.227130

White's test for heteroskedasticity -
Null hypothesis: heteroskedasticity not present
Test statistic: LM = 58.4342
with p-value = P(Chi-square(5) > 58.4342) = 2.5591e-11

RESET test for specification -
Null hypothesis: specification is adequate
Test statistic: F(2, 748) = 5.65181
with p-value = P(F(2, 748) > 5.65181) = 0.00366278

Non-linearity test (squares) -
Null hypothesis: relationship is linear
Test statistic: LM = 7.85178
with p-value = P(Chi-square(2) > 7.85178) = 0.0197245

Test for normality of residual -
Null hypothesis: error is normally distributed
Test statistic: Chi-square(2) = 515.741
with p-value = 1.01897e-112

Chow test for structural difference with respect to OSh2 -
Null hypothesis: no structural difference
Test statistic: F(3, 747) = 0.818922
with p-value = P(F(3, 747) > 0.818922) = 0.483593

Chow test for structural difference with respect to OSh1 -
Null hypothesis: no structural difference
Test statistic: F(3, 747) = 0.836354
with p-value = P(F(3, 747) > 0.836354) = 0.474111

Chow test for structural difference with respect to BCE -
Null hypothesis: no structural difference
Test statistic: F(3, 747) = 0.135149
with p-value = P(F(3, 747) > 0.135149) = 0.939082

Chow test for structural difference with respect to DCB -
Null hypothesis: no structural difference
Test statistic: F(3, 747) = 0.501132
with p-value = P(F(3, 747) > 0.501132) = 0.681603

Chow test for structural difference with respect to FC -
Null hypothesis: no structural difference
Test statistic: F(3, 747) = 3.64469
with p-value = P(F(3, 747) > 3.64469) = 0.0124955

Chow test for structural difference with respect to Covid -
Null hypothesis: no structural difference
Test statistic: F(3, 747) = 1.22192
with p-value = P(F(3, 747) > 1.22192) = 0.30069

Chow test for structural difference with respect to RUKWar -
Null hypothesis: no structural difference
Test statistic: F(3, 747) = 6.06227
with p-value = P(F(3, 747) > 6.06227) = 0.000443851

Model 8: OLS regression of first difference wheat on its first lag and first differenced substitutes index in log-prices (structural breaks)

VAR system, lag order 2
OLS estimates, observations 1960:05-2022:11 (T = 751)
Log-likelihood = 1144.0356
Determinant of covariance matrix = 0.0027820243
AIC = -3.0360
BIC = -3.0114
HQC = -3.0266
Portmanteau test: LB(48) = 62.1603, df = 46 [0.0562]

Equation 1: d_l_Wheat

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.00125798	0.00193485	0.6502	0.5158	
d_l_Wheat_1	0.214462	0.0351211	6.106	<0.0001	***
d_l_Wheat_2	−0.0985516	0.0349098	−2.823	0.0049	***
d_l_Substitute_idx	0.420835	0.0461382	9.121	<0.0001	***

Mean dependent var	0.002578	S.D. dependent var	0.057588
Sum squared resid	2.089300	S.E. of regression	0.052886
R-squared	0.160017	Adjusted R-squared	0.156643
F(3, 747)	47.43447	P-value(F)	4.56e-28
rho	−0.002301	Durbin-Watson	2.002094

F-tests of zero restrictions:
All lags of d_l_Wheat F(2, 747) = 19.727 [0.0000]
All vars, lag 2 F(1, 747) = 7.9695 [0.0049]

For the system as a whole
Null hypothesis: the longest lag is 1
Alternative hypothesis: the longest lag is 2
Likelihood ratio test: Chi-square(1) = 7.96976 [0.0048]

Source: Author's work, Gretl

Model 9: VAR of first difference wheat on two FD lags and first differenced substitutes index in log-prices

VAR system, lag order 3
OLS estimates, observations 1960:05-2022:11 (T = 751)
Log-likelihood = 1145.4393
Determinant of covariance matrix = 0.0027716437
AIC = -3.0371
BIC = -3.0064
HQC = -3.0253
Portmanteau test: LB(48) = 58.2444, df = 45 [0.0889]

Equation 1: d_l_Wheat

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.00135380	0.00193338	0.7002	0.4840	
d_l_Wheat_1	0.209781	0.0351906	5.961	<0.0001	***
d_l_Wheat_2	−0.0838863	0.0359548	−2.333	0.0199	**
d_l_Wheat_3	−0.0585114	0.0350049	−1.672	0.0950	*
d_l_Substitute_idx	0.430508	0.0464449	9.269	<0.0001	***

Mean dependent var	0.002578	S.D. dependent var	0.057588
Sum squared resid	2.081504	S.E. of regression	0.052823
R-squared	0.163151	Adjusted R-squared	0.158664
F(4, 746)	36.35979	P-value(F)	8.68e-28
rho	0.001171	Durbin-Watson	1.995148

F-tests of zero restrictions:
All lags of d_l_Wheat F(3, 746) = 14.114 [0.0000]
All vars, lag 3 F(1, 746) = 2.794 [0.0950]

For the system as a whole
Null hypothesis: the longest lag is 2
Alternative hypothesis: the longest lag is 3
Likelihood ratio test: Chi-square(1) = 2.80745 [0.0938]

Test on the original VAR:

Null hypothesis: the regression parameters are zero for the variables
d_l_Fertilizer_idx, d_l_Oil
LR test: Chi-square(2) = 4.58801, with p-value = 0.100862

Source: Author's work, Gretl

Model 9B: VAR of first difference wheat on three lags and first differenced substitutes index in log-prices

Model 13: OLS, using observations 1960:05-2022:11 (T = 751)
Dependent variable: d_l_Wheat
HAC standard errors, bandwidth 6 (Bartlett kernel)

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.00125798	0.00182750	0.6884	0.4914	
d_l_Substitute_idx	0.420835	0.0720018	5.845	<0.0001	***
d_l_Wheat_1	0.214462	0.0393303	5.453	<0.0001	***
d_l_Wheat_2	−0.0985516	0.0394018	−2.501	0.0126	**

Mean dependent var	0.002578	S.D. dependent var	0.057588
Sum squared resid	2.089300	S.E. of regression	0.052886
R-squared	0.160017	Adjusted R-squared	0.156643
F(3, 747)	27.98089	P-value(F)	3.73e-17
Log-likelihood	1144.036	Akaike criterion	−2280.071
Schwarz criterion	−2261.586	Hannan-Quinn	−2272.949
rho	−0.002301	Durbin's h	NA

Test for omission of variables -
Null hypothesis: parameters are zero for the variables
d_l_Wheat_2
Test statistic: F(1, 747) = 6.25598
with p-value = P(F(1, 747) > 6.25598) = 0.0125905

Chow test for structural difference with respect to OSh1 -
Null hypothesis: no structural difference
Asymptotic test statistic: Chi-square(4) = 3977.55
with p-value = 0

Chow test for structural difference with respect to OSh2 -
Null hypothesis: no structural difference
Asymptotic test statistic: Chi-square(4) = 17.0132
with p-value = 0.00192159

Chow test for structural difference with respect to BCE -
Null hypothesis: no structural difference
Asymptotic test statistic: Chi-square(4) = 7.97754
with p-value = 0.0924044

Chow test for structural difference with respect to DCB -
Null hypothesis: no structural difference
Asymptotic test statistic: Chi-square(4) = 3.50592
with p-value = 0.476979

Chow test for structural difference with respect to FC -
Null hypothesis: no structural difference
Asymptotic test statistic: Chi-square(4) = 23.0072
with p-value = 0.000126209

Chow test for structural difference with respect to Covid -
Null hypothesis: no structural difference
Asymptotic test statistic: Chi-square(4) = 16.3589
with p-value = 0.00257348

Chow test for structural difference with respect to RUKWar -
Null hypothesis: no structural difference
Asymptotic test statistic: Chi-square(4) = 29.6209
with p-value = 5.84576e-06

Source: Author's work, Gretl

Model 10A: Structural Break tests for OLS of first difference wheat on two FD lags and first differenced substitutes index in log-prices (robust errors)

Model 16: OLS, using observations 1960:05-2022:11 (T = 751)
Dependent variable: d_l_Wheat

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.00125798	0.00193485	0.6502	0.5158	
d_l_Substitute_idx	0.420835	0.0461382	9.121	<0.0001	***
d_l_Wheat_1	0.214462	0.0351211	6.106	<0.0001	***
d_l_Wheat_2	−0.0985516	0.0349098	−2.823	0.0049	***

Mean dependent var	0.002578	S.D. dependent var	0.057588
Sum squared resid	2.089300	S.E. of regression	0.052886
R-squared	0.160017	Adjusted R-squared	0.156643
F(3, 747)	47.43447	P-value(F)	4.56e-28
Log-likelihood	1144.036	Akaike criterion	−2280.071
Schwarz criterion	−2261.586	Hannan-Quinn	−2272.949
rho	−0.002301	Durbin's h	−0.232388

Chow test for structural difference with respect to OSh1 -
Null hypothesis: no structural difference
Test statistic: F(4, 743) = 1.43065
with p-value = P(F(4, 743) > 1.43065) = 0.222015

Chow test for structural difference with respect to OSh2 -
Null hypothesis: no structural difference
Test statistic: F(4, 743) = 0.580633
with p-value = P(F(4, 743) > 0.580633) = 0.676762

Chow test for structural difference with respect to BCE -
Null hypothesis: no structural difference
Test statistic: F(4, 743) = 0.103803
with p-value = P(F(4, 743) > 0.103803) = 0.981175

Chow test for structural difference with respect to DCB -
Null hypothesis: no structural difference
Test statistic: F(4, 743) = 0.661862
with p-value = P(F(4, 743) > 0.661862) = 0.618644

Chow test for structural difference with respect to FC -
Null hypothesis: no structural difference
Test statistic: F(4, 743) = 2.68224
with p-value = P(F(4, 743) > 2.68224) = 0.0305786

Chow test for structural difference with respect to Covid -
Null hypothesis: no structural difference
Test statistic: F(4, 743) = 1.5224
with p-value = P(F(4, 743) > 1.5224) = 0.193768

Chow test for structural difference with respect to RUKWar -
Null hypothesis: no structural difference
Test statistic: F(4, 743) = 4.12649
with p-value = P(F(4, 743) > 4.12649) = 0.00258696

Source: Author's work, Gretl

Model 10B: Structural break tests for OLS of first difference wheat on two FD lags and first differenced substitutes index in log-prices (usual errors)

```
Augmented regression for Chow test
OLS, using observations 1960:04-2022:11 (T = 752)
Dependent variable: d_l_Wheat
HAC standard errors, bandwidth 6 (Bartlett kernel)

                      coefficient  std. error  t-ratio    p-value
  ---------------------------------------------------------------
  const                0.00115317  0.00185647    0.6212  0.5347
  d_l_Substitute_i~    0.437378    0.0776694     5.631   2.54e-08    ***
  d_l_Wheat_1          0.212435    0.0399881     5.312   1.43e-07    ***
  d_l_Wheat_2          0.109655    0.0413243    ,àí2.654   0.0081    ***
  OSh1                 0.100291    0.00185647   54.02    1.04e-259   ***
  OS_d_l_Substitut~    0.996047    0.0776694   ,àí12.82    3.71e-34  ***
  OS_d_l_Wheat_1       0.457165    0.0399881   ,àí11.43    5.35e-28  ***
  OS_d_l_Wheat_2       0.242357    0.0413243    ,àí5.865   6.76e-09  ***

Mean dependent var   0.002558   S.D. dependent var   0.057552
Sum squared resid    2.073394   S.E. of regression   0.052790
R-squared            0.166483   Adjusted R-squared   0.158641
Log-likelihood       1148.933   Akaike criterion     2281.866
Schwarz criterion    2244.884   Hannan-Quinn         2267.618
rho                  0.004936   Durbin-Watson        1.989899

Chow test for structural difference with respect to OSh1
  Chi-square(4) = 3984.87 with p-value 0.0000
  F-form: F(4, 744) = 996.219 with p-value 0.0000
```

Model 10C: Chow test augmented regression model using first Oil shock dummies

```
VAR system, lag order 2
OLS estimates, observations 1963-2020 (T = 58)
Log-likelihood = 132.64159
Determinant of covariance matrix = 1.2125668e-07
AIC = -3.3325
BIC = -2.0536
HQC = -2.8343
Portmanteau test: LB(14) = 194.224, df = 192 [0.4415]

Equation 1: d_l_Wheat
Heteroskedasticity-robust standard errors, variant HC1

coefficient   std. error   t-ratio   p-value
-----------------------------------------------------------------
const                  0.0166360     0.0250850      0.6632    0.5103
d_l_Wheat_1            0.361416      0.249998       1.446     0.1546
d_l_Wheat_2           ,àí0.522995     0.203659      2.568     0.0133  **
d_l_Fertilizer~_1      0.190640      0.229104       0.8321    0.4094
d_l_Fertilizer~_2     ,àí0.150130     0.130903      1.147     0.2570
d_l_Substitute~_1     ,àí0.154098     0.369004      0.4176    0.6781
d_l_Substitute~_2      0.163930      0.270430       0.6062    0.5472
d_l_Oil_1             ,àí0.0586093    0.120901      0.4848    0.6300
d_l_Oil_2              0.178583      0.100046       1.785     0.0804  *

Mean dependent var    0.022083   S.D. dependent var    0.183883
Sum squared resid     1.463823   S.E. of regression    0.172841
R-squared             0.240498   Adjusted R-squared    0.116498
F(8, 49)              2.142135   P-value(F)            0.049155
rho                   0.014139   Durbin-Watson         2.000201

F-tests of zero restrictions:

All lags of d_l_Wheat              F(2, 49) =   4.9379 [0.0111]
All lags of d_l_Fertilizer_i~      F(2, 49) =   1.3436 [0.2703]
All lags of d_l_Substitutes_~      F(2, 49) =  0.38211 [0.6844]
All lags of d_l_Oil                F(2, 49) =   1.6649 [0.1997]
All vars, lag 2                    F(4, 49) =   3.1471 [0.0222]
```

```
VAR system, lag order 2
OLS estimates, observations 1963-2022 (T = 60)
Log-likelihood = 21.602719
Determinant of covariance matrix = 0.02849668
AIC = -0.6201
BIC = -0.5154
HQC = -0.5791
Portmanteau test: LB(15) = 8.79005, df = 13 [0.7886]

Equation 1: d_l_Wheat
Heteroskedasticity-robust standard errors, variant HC1

                coefficient   std. error   t-ratio   p-value
  ----------------------------------------------------------
  const          0.0287822    0.0216968     1.327    0.1899
  d_l_Wheat_1    0.379112     0.105988      3.577    0.0007  ***
  d_l_Wheat_2   -0.335948     0.107844     -3.115    0.0029  ***

Mean dependent var    0.031661   S.D. dependent var    0.188076
Sum squared resid     1.709801   S.E. of regression    0.173195
R-squared             0.180729   Adjusted R-squared    0.151983
F(2, 57)             11.63737    P-value(F)            0.000058
rho                  -0.026476   Durbin-Watson         2.023534
```

Model 11: VAR of first difference wheat (Yearly data) on two FD lags and first differenced substitutes index in log-prices

```
VAR system, lag order 2
OLS estimates, observations 1963-2016 (T = 54)
Log-likelihood = 143.289
Determinant of covariance matrix = 5.8250778e-08
AIC = -3.2329
BIC = -1.1703
HQC = -2.4374
Portmanteau test: LB(13) = 208.373, df = 176 [0.0480]

Equation 1: d_l_Wheat
Heteroskedasticity-robust standard errors, variant HC1

                      coefficient   std. error   t-ratio   p-value
  ----------------------------------------------------------------
  const                 0.634027     0.863275      0.7344   0.4670
  d_l_Wheat_1           0.462888     0.289091      1.601    0.1172
  d_l_Wheat_2         ,àí0.485097     0.225808      2.148    0.0378  **
  d_l_Oil_1           ,àí0.167641     0.154874      1.082    0.2855
  d_l_Oil_2             0.148160     0.121596      1.218    0.2302
  d_l_Substitute~_1   ,àí0.413152     0.413191      0.9999   0.3234
  d_l_Substitute~_2     0.191137     0.270407      0.7068   0.4838
  d_l_Fertilizer~_1     0.308492     0.245269      1.258    0.2158
  d_l_Fertilizer~_2   ,àí0.143302     0.142587      1.005    0.3209
  GPA                   2.20442e-05  0.000753931   0.02924  0.9768
  GLTA                  0.0321658    0.0716617     0.4489   0.6560
  YAFFER              ,àí0.0175080    0.00970939   1.803    0.0789  *
  INFCPI                0.0286521    0.0132037     2.170    0.0360  **
  l_POPGR               0.156793     0.208338      0.7526   0.4561

Mean dependent var   0.017625   S.D. dependent var   0.188409
Sum squared resid    1.213393   S.E. of regression   0.174169
R-squared            0.355055   Adjusted R-squared   0.145448
F(13, 40)            2.365056   P-value(F)           0.018592
rho                  0.074878   Durbin-Watson        1.816421

F-tests of zero restrictions:

All lags of d_l_Wheat             F(2, 40) =   4.1088 [0.0238]
All lags of d_l_Oil               F(2, 40) =   1.4241 [0.2527]
All lags of d_l_Substitute_i~     F(2, 40) =  0.97909 [0.3845]
All lags of d_l_Fertilizer_i~     F(2, 40) =   1.7649 [0.1843]
All vars, lag 2                   F(4, 40) =   1.9984 [0.1133]
```

```
VAR system, lag order 2
OLS estimates, observations 1963-2016 (T = 54)
Log-likelihood = 22.100282
Determinant of covariance matrix = 0.025825161
AIC = -0.5222
BIC = -0.2276
HQC = -0.4086
Portmanteau test: LB(13) = 6.76638, df = 11 [0.8177]

Equation 1: d_l_Wheat

                coefficient    std. error    t-ratio    p-value
  ---------------------------------------------------------------
  const          0.255373      0.979019       0.2608    0.7954
  d_l_Wheat_1    0.257492      0.143370       1.796     0.0791   *
  d_l_Wheat_2   -0.339250      0.139327      -2.435     0.0188   **
  GPA           -0.000194873   0.000829688   -0.2349    0.8153
  GLTA           0.00743148    0.0748849      0.09924   0.9214
  YAFFER        -0.0148284     0.00959662    -1.545     0.1292
  INFCPI         0.0262072     0.0119820      2.187     0.0338   **
  l_POPGR        0.0625981     0.237016       0.2641    0.7929

Mean dependent var   0.017625   S.D. dependent var   0.188409
Sum squared resid    1.394559   S.E. of regression   0.174116
R-squared            0.258762   Adjusted R-squared   0.145964
F(7, 46)             2.294043   P-value(F)           0.043088
rho                 -0.005055   Durbin-Watson        1.987243
```

```
VAR system, lag order 2
OLS estimates, observations 1963-2021 (T = 59)
Log-likelihood = 22.961129
Determinant of covariance matrix = 0.026884096
AIC = -0.6428
BIC = -0.5019
HQC = -0.5878
Portmanteau test: LB(14) = 6.83009, df = 12 [0.8686]

Equation 1: d_l_Wheat

                coefficient   std. error   t-ratio   p-value
  ----------------------------------------------------------
  const         -0.0254937    0.0387565    -0.6578   0.5134
  d_l_Wheat_1    0.282078     0.132707      2.126    0.0380  **
  d_l_Wheat_2   -0.356947     0.126912     -2.813    0.0068  ***
  INFCPI         0.0139749    0.00858578    1.628    0.1093

Mean dependent var   0.026937   S.D. dependent var   0.186065
Sum squared resid    1.586162   S.E. of regression   0.169821
R-squared            0.210067   Adjusted R-squared   0.166980
F(3, 55)             4.875383   P-value(F)           0.004458
rho                  0.007138   Durbin-Watson        1.956892

VAR system, lag order 2
OLS estimates, observations 1963-2022 (T = 60)
Log-likelihood = 21.602719
Determinant of covariance matrix = 0.02849668
AIC = -0.6201
BIC = -0.5154
HQC = -0.5791
Portmanteau test: LB(15) = 8.79005, df = 13 [0.7886]

Equation 1: d_l_Wheat

                coefficient   std. error   t-ratio   p-value
  ----------------------------------------------------------
  const          0.0287822    0.0227134     1.267    0.2102
  d_l_Wheat_1    0.379112     0.126409      2.999    0.0040  ***
  d_l_Wheat_2   -0.335948     0.128993     -2.604    0.0117  **

Mean dependent var   0.031661   S.D. dependent var   0.188076
Sum squared resid    1.709801   S.E. of regression   0.173195
R-squared            0.180729   Adjusted R-squared   0.151983
F(2, 57)             6.287019   P-value(F)           0.003409
rho                 -0.026476   Durbin-Watson        2.023534
```

Source: Author's work, Gretl

Model 12: VAR of first difference wheat (yearly) instrumented with Precipitation, temperature, interest rates, inflation and population growth.

```
VAR system, lag order 1
OLS estimates, observations 1963-2022 (T = 60)
Log-likelihood = 21.602719
Determinant of covariance matrix = 0.02849668
AIC = -0.6201
BIC = -0.5154
HQC = -0.5791
Portmanteau test: LB(15) = 8.79005, df = 14 [0.8443]

Equation 1: d_l_Wheat
Heteroskedasticity-robust standard errors, variant HC1

                  coefficient   std. error   t-ratio   p-value
  ------------------------------------------------------------
  const            0.0287822    0.0216968     1.327     0.1899
  d_l_Wheat_1      0.0431642    0.153730      0.2808    0.7799
  d_d_l_Wheat_1    0.335948     0.107844      3.115     0.0029   ***

Mean dependent var   0.031661   S.D. dependent var   0.188076
Sum squared resid    1.709801   S.E. of regression   0.173195
R-squared            0.180729   Adjusted R-squared   0.151983
F(2, 57)             11.63737   P-value(F)           0.000058
rho                 -0.026476   Durbin-Watson        2.023534
```

Model 13: VAR of first difference wheat (yearly) on lagged second-difference wheat in log-prices.

Appendix C: Diagnostic Tests

Test for Normality

```
Test for normality of l_WHEAT_US_HRW:

 Doornik-Hansen test = 16.7667, with p-value 0.000228639

 Shapiro-Wilk W = 0.954578, with p-value 1.6485e-14

 Lilliefors test = 0.110744, with p-value ~= 0

 Jarque-Bera test = 11.984, with p-value 0.00249866
```

Test for Autocorrelation

```
Autocorrelation function for l_WHEAT_US_HRW
***, **, * indicate significance at the 1%, 5%, 10% levels
using standard error 1/T^0.5

  LAG      ACF          PACF         Q-stat. [p-value]

    1   0.9673  ***    0.9673 ***    237.6517  [0.000]
    2   0.9220  ***   -0.2115 ***    454.4353  [0.000]
    3   0.8808  ***    0.0806        653.0777  [0.000]
    4   0.8410  ***   -0.0329        834.9217  [0.000]
    5   0.8020  ***   -0.0054       1000.9480  [0.000]
    6   0.7665  ***    0.0320       1153.2198  [0.000]
    7   0.7327  ***   -0.0126       1292.9536  [0.000]
    8   0.7002  ***    0.0014       1421.0714  [0.000]
    9   0.6704  ***    0.0214       1539.0019  [0.000]
   10   0.6482  ***    0.0919       1649.6999  [0.000]
   11   0.6295  ***    0.0062       1754.5590  [0.000]
   12   0.6141  ***    0.0455       1854.7595  [0.000]
   13   0.5998  ***   -0.0050       1950.7382  [0.000]
   14   0.5839  ***   -0.0256       2042.1010  [0.000]
   15   0.5669  ***   -0.0071       2128.5622  [0.000]
   16   0.5542  ***    0.0700       2211.5527  [0.000]
   17   0.5423  ***   -0.0195       2291.3503  [0.000]
   18   0.5321  ***    0.0431       2368.5028  [0.000]
   19   0.5266  ***    0.0714       2444.4123  [0.000]
   20   0.5198  ***   -0.0454       2518.6842  [0.000]
   21   0.5065  ***   -0.0655       2589.5240  [0.000]
   22   0.4937  ***    0.0381       2657.1141  [0.000]
   23   0.4857  ***    0.0521       2722.8125  [0.000]
```

ADF Test for Stationarity (null of unit root)

```
 Augmented Dickey-Fuller test for l_WHEAT_US_HRW
testing down from 19 lags, criterion AIC
sample size 752
unit-root null hypothesis: a = 1

  test with constant
  including 2 lags of (1-L)l_WHEAT_US_HRW
  model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -0.00639371
  test statistic: tau_c(1) = -1.608
  asymptotic p-value 0.4785
  1st-order autocorrelation coeff. for e: -0.001
  lagged differences: F(2, 748) = 27.238 [0.0000]

  with constant and trend
  including one lag of (1-L)l_WHEAT_US_HRW
  model: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -0.0235776
  test statistic: tau_ct(1) = -3.40804
  asymptotic p-value 0.05024
  1st-order autocorrelation coeff. for e: 0.012
```

Johansen Test for Cointegration of Wheat with Oil Fertilizers and Substitutes

```
Johansen test:
Number of equations = 4
Lag order = 3
Estimation period: 1960:04 - 2022:11 (T = 752)

Coefficients, VAR in differences (9 x 4)

   0.0018381    0.0012413    0.0038534    0.0011149
     0.23731     0.091308     0.064322     0.048511
   -0.066534   -0.0074569     0.015636     0.062708
    0.053270      0.16485    -0.087080     0.056245
   -0.078090    -0.012692   -0.0097767     0.011751
    0.019872     0.052692      0.24354     0.012241
   0.0099388     0.066857    -0.048788    0.0044565
     0.13528      0.13775     -0.16042      0.32051
   -0.054159      0.20548      0.39296    -0.056580

Coefficients, eqns in lagged levels (9 x 4)

      4.9335       4.8775       2.7796       5.0187
     0.39104     -0.11071    -0.042089     -0.13191
     0.61843    -0.065729    -0.042833    -0.076366
     0.27984      0.74815      0.76913      0.11874
     0.51502       1.0771      0.79733      0.34556
    0.030920     -0.20613     0.075428     0.038415
   -0.011146     -0.21198      0.30121   -0.0016421
   0.0086925     -0.34741     -0.77971      0.19324
     0.18731     -0.15130     -0.90996      0.65909

Sample variance-covariance matrices for residuals

 VAR system in first differences (S00)

   0.0030325  5.0388e-05  0.00032931  0.00066193
  5.0388e-05   0.0027962   0.0014644  0.00025713
  0.00032931   0.0014644   0.0074805  8.9873e-05
  0.00066193  0.00025713  8.9873e-05   0.0015647

 System with levels as dependent variable (S11)

     0.25866     0.36726     0.65760     0.23071
     0.36726     0.57955     0.98744     0.34101
     0.65760     0.98744      1.9660     0.58821
     0.23071     0.34101     0.58821     0.22488

 Cross-products (S01)

  -0.0016297 -0.00052197 -0.00091145 -0.00080854
  0.00055684  -0.0012651  0.00045335 -0.00011701
  0.00014514  -0.0021333  -0.0054698 -5.5994e-05
 -0.00047689 -0.00080484  -0.0012905  -0.0013360

Case 3: Unrestricted constant

Log-likelihood = 6656.11 (including constant term: 4522.02)

Rank Eigenvalue Trace test  p-value  Lmax test  p-value
   0   0.067230      88.899 [0.0000]     52.337 [0.0000]
   1   0.033056      36.562 [0.0064]     25.278 [0.0103]
   2   0.013529      11.283 [0.1975]     10.244 [0.2001]
   3  0.0013818      1.0398 [0.3079]     1.0398 [0.3079]

Corrected for sample size (df = 739)
Rank Trace test p-value
   0     88.899 [0.0000]
   1     36.562 [0.0063]
   2     11.283 [0.1985]
   3     1.0398 [0.3085]

eigenvalue    0.067230     0.033056     0.013529    0.0013818

beta (cointegrating vectors)
l_Wheat                 8.2297     -0.63010      0.70566      0.83398
l_Fertilizer_idx       -1.7176      -3.6885       2.7941       1.0552
l_Oil                 -0.65167     0.064047      -1.9459   -0.0015849
l_Substitute_idx       -4.2881       6.7666      0.20166     -0.46776

alpha (adjustment vectors)
l_Wheat             -0.0084544   -0.0025772  -0.00099795   -0.0015303
l_Fertilizer_idx     0.0069618    0.0035527   -0.0040477  -0.00081652
l_Oil                0.0086631    0.0070480    0.0047741   -0.0020952
l_Substitute_idx     0.0040274   -0.0058534  -0.00034356  -0.00062003

renormalized beta
l_Wheat                 1.0000      0.17083     -0.36264      -1.7829
l_Fertilizer_idx      -0.20871       1.0000      -1.4359      -2.2559
l_Oil                -0.079185    -0.017364       1.0000    0.0033883
l_Substitute_idx      -0.52105      -1.8345     -0.10363       1.0000

renormalized alpha
l_Wheat              -0.069577    0.0095061    0.0019419   0.00071580
l_Fertilizer_idx      0.057294    -0.013104    0.0078763   0.00038193
l_Oil                 0.071295    -0.025996   -0.0092899   0.00098003
l_Substitute_idx      0.033144     0.021590   0.00066854   0.00029003

long-run matrix (alpha * beta')
                          l_Wheat l_Fertilizer_idx            l_Oil l_Substitute_idx
l_Wheat                 -0.069934         0.019624        0.0072887         0.019329
```

```
l_Fertilizer_idx         0.051518       -0.037233       0.0035684      -0.0062477
l_Oil                    0.068476       -0.029747       -0.014481       0.012485
l_Substitute_idx         0.036073        0.013059      -0.0023299      -0.056656
```

Johansen Test for Cointegration of Wheat with Oil Fertilizers and Substitutes

```
Johansen test:
Number of equations = 4
Lag order = 12
Estimation period: 1961:01 - 2022:11 (T = 743)
Case 3: Unrestricted constant

Log-likelihood = 6723.65 (including constant term: 4615.11)

Rank Eigenvalue Trace test  p-value  Lmax test  p-value
   0   0.049403     86.111 [0.0000]     37.644 [0.0010]
   1   0.033784     48.467 [0.0001]     25.536 [0.0093]
   2   0.027645     22.931 [0.0026]     20.829 [0.0031]
   3  0.0028248     2.1018 [0.1471]     2.1018 [0.1471]

Corrected for sample size (df = 694)
Rank Trace test p-value
   0     86.111 [0.0000]
   1     48.467 [0.0001]
   2     22.931 [0.0026]
   3     2.1018 [0.1475]

eigenvalue     0.049403     0.033784     0.027645     0.0028248

beta (cointegrating vectors)
l_Wheat                -10.464      0.12716       3.0476    -0.072645
l_Oil                  0.53707      -1.0490      -2.0052     -0.43122
l_Fertilizer_idx        2.7826      -1.8198       4.8423     -0.88937
l_Substitute_idx        5.0783       6.4320      -5.1789      0.61512

alpha (adjustment vectors)
l_Wheat              0.0087477   -0.0039201   -0.0032498    0.0011117
l_Oil              -0.00029800    0.0075372   0.00094268    0.0037484
l_Fertilizer_idx    -0.0064939   -0.0011031   -0.0053814    0.0011415
l_Substitute_idx   -0.00077774   -0.0055051    0.0023206   0.00096825

renormalized beta
l_Wheat                 1.0000     -0.12122      0.62938     -0.11810
l_Oil                -0.051325       1.0000     -0.41410     -0.70103
l_Fertilizer_idx      -0.26592       1.7348       1.0000      -1.4458
l_Substitute_idx      -0.48531      -6.1313      -1.0695       1.0000

renormalized alpha
l_Wheat              -0.091536    0.0041124    -0.015736   0.00068385
l_Oil                0.0031182   -0.0079068    0.0045647    0.0023057
l_Fertilizer_idx      0.067952    0.0011572    -0.026058   0.00070214
l_Substitute_idx     0.0081383    0.0057751     0.011237   0.00059560

long-run matrix (alpha * beta')
                         l_Wheat            l_Oil l_Fertilizer_idx l_Substitute_idx
l_Wheat                 -0.10202         0.014848         0.014750         0.036723
l_Oil                  0.0066773        -0.011574        -0.013315         0.044390
l_Fertilizer_idx        0.051328        0.0079683        -0.043136        -0.011501
l_Substitute_idx        0.014440       0.00028657         0.018230        -0.050781
```

{

Johansen Test for Cointegration of Wheat and Oil

```
Johansen test:
Number of equations = 2
Lag order = 12
Estimation period: 1961:01 - 2022:11 (T = 743)
Case 3: Unrestricted constant
Exogenous regressor(s): l_Fertilizer_idx l_Substitute_idx

Log-likelihood = 4023.16 (including constant term: 1914.62)

Cointegration tests, ignoring exogenous variables
Rank Eigenvalue Trace test  p-value  Lmax test  p-value
   0   0.059054     50.157 [0.0000]     45.226 [0.0000]
   1  0.0066142     4.9307 [0.0264]     4.9307 [0.0264]

Corrected for sample size (df = 716)
Rank Trace test p-value
```

```
   0     50.157 [0.0000]
   1     4.9307 [0.0265]

Note: in general, the test statistics above are valid only in the
absence of additional regressors.

eigenvalue    0.059054    0.0066142

beta (cointegrating vectors)
l_Wheat        -9.7964      0.10765
l_Oil           1.1769       1.8949

alpha (adjustment vectors)
l_Wheat       0.013271  -0.00033513
l_Oil      -0.00012602   -0.0068523

renormalized beta
l_Wheat         1.0000     0.056809
l_Oil         -0.12014       1.0000

renormalized alpha
l_Wheat       -0.13000  -0.00063501
l_Oil        0.0012345    -0.012984

long-run matrix (alpha * beta')
               l_Wheat        l_Oil
l_Wheat       -0.13004     0.014983
l_Oil       0.00049688    -0.013132
```

Augmented Regression for Chow test including oil shock dummies

```
Augmented regression for Chow test
OLS, using observations 1960:04-2022:11 (T = 752)
Dependent variable: d_l_Wheat
HAC standard errors, bandwidth 6 (Bartlett kernel)

                      coefficient  std. error  t-ratio    p-value
  ---------------------------------------------------------------
  const                0.00115317  0.00185647    0.6212  0.5347
  d_l_Substitute_i~    0.437378    0.0776694     5.631   2.54e-08    ***
  d_l_Wheat_1          0.212435    0.0399881     5.312   1.43e-07    ***
  d_l_Wheat_2          0.109655    0.0413243    ‚àí2.654   0.0081     ***
  OSh1                 0.100291    0.00185647   54.02    1.04e-259   ***
  OS_d_l_Substitut~    0.996047    0.0776694   ‚àí12.82    3.71e-34   ***
  OS_d_l_Wheat_1       0.457165    0.0399881   ‚àí11.43    5.35e-28   ***
  OS_d_l_Wheat_2       0.242357    0.0413243    ‚àí5.865   6.76e-09   ***

Mean dependent var   0.002558   S.D. dependent var   0.057552
Sum squared resid    2.073394   S.E. of regression   0.052790
R-squared            0.166483   Adjusted R-squared   0.158641
Log-likelihood       1148.933   Akaike criterion     2281.866
Schwarz criterion    2244.884   Hannan-Quinn         2267.618
rho                  0.004936   Durbin-Watson        1.989899

Chow test for structural difference with respect to OSh1
  Chi-square(4) = 3984.87 with p-value 0.0000
  F-form: F(4, 744) = 996.219 with p-value 0.0000
```

Additional tests:

Johansen Test for Cointegration of Wheat and Substitutes estimated separately

```
Johansen test:
Number of equations = 3
Lag order = 12
Estimation period: 1967:1 - 2022:3 (T = 223)
Case 3: Unrestricted constant
Exogenous regressor(s): l_CRUDE_PETRO l_DAP l_TSP l_UREA_EE_BULK

Log-likelihood = 1380.81 (including constant term: 747.965)

Cointegration tests, ignoring exogenous variables
Rank Eigenvalue Trace test  p-value  Lmax test  p-value
   0    0.16514     86.448 [0.0000]     40.249 [0.0000]
   1    0.14218     46.199 [0.0000]     34.198 [0.0000]
   2   0.052391     12.000 [0.0005]     12.000 [0.0005]

Corrected for sample size (df = 182)
Rank Trace test p-value
   0     86.448 [0.0000]
   1     46.199 [0.0000]
   2     12.000 [0.0006]

Note: in general, the test statistics above are valid only in the
absence of additional regressors.

eigenvalue     0.16514      0.14218     0.052391

beta (cointegrating vectors)
l_WHEAT_US_HRW        10.018       7.9313      -9.9116
l_MAIZE              -16.161     -0.85466       2.1157
l_RICE_05             5.0806       1.5558       7.0481

alpha (adjustment vectors)
l_WHEAT_US_HRW    -0.0016036    -0.031204    0.0089845
l_MAIZE             0.028759    -0.017079    0.0061022
l_RICE_05          0.0066180    -0.019769    -0.016792

renormalized beta
l_WHEAT_US_HRW        1.0000      -9.2801      -1.4063
l_MAIZE              -1.6132       1.0000      0.30018
l_RICE_05            0.50715      -1.8204       1.0000

renormalized alpha
l_WHEAT_US_HRW     -0.016065     0.026669     0.063323
l_MAIZE              0.28811     0.014597     0.043009
l_RICE_05           0.066300     0.016896     -0.11835

long-run matrix (alpha * beta')
               l_WHEAT_US_HRW        l_MAIZE      l_RICE_05
l_WHEAT_US_HRW       -0.35260       0.071593      0.0066284
l_MAIZE              0.092168       -0.43728        0.16255
l_RICE_05            0.075939       -0.12559       -0.11548
```

Johansen Test for Cointegration of Wheat and Fertilizers estimated separately

```
Johansen test:
Number of equations = 5
Lag order = 12
Estimation period: 1970:1 - 2022:3 (T = 211)
Case 3: Unrestricted constant
Exogenous regressor(s): l_CRUDE_PETRO l_RICE_05 l_MAIZE

Log-likelihood = 1843.12 (including constant term: 1244.33)

Cointegration tests, ignoring exogenous variables
Rank Eigenvalue Trace test  p-value  Lmax test  p-value
   0    0.34616     204.32 [0.0000]     89.652 [0.0000]
   1    0.25436     114.67 [0.0000]     61.931 [0.0000]
   2    0.14144     52.738 [0.0000]     32.178 [0.0005]
   3   0.055936     20.560 [0.0068]     12.145 [0.1052]
   4   0.039096     8.4149 [0.0037]     8.4149 [0.0037]

Corrected for sample size (df = 147)
Rank Trace test p-value
   0     204.32 [0.0000]
   1     114.67 [0.0000]
   2     52.738 [0.0000]
   3     20.560 [0.0073]
   4     8.4149 [0.0040]

Note: in general, the test statistics above are valid only in the
absence of additional regressors.
```

```
eigenvalue     0.34616      0.25436      0.14144     0.055936     0.039096

beta (cointegrating vectors)
l_WHEAT_US_HRW       -7.4457      -11.471      -1.7356      -5.0441      -6.1793
l_DAP                 2.6040      -21.420      -11.679       22.167       1.2468
l_TSP                -4.1144       16.675       19.061      -15.118     0.020027
l_UREA_EE_BULK       -2.9885       9.7987      -3.3473      0.48085       1.5373
l_POTASH              1.1489    -0.018934      -5.0395     -0.76584       4.1232

alpha (adjustment vectors)
l_WHEAT_US_HRW      0.040385     0.022430   -0.0056547   0.00011260    0.0032075
l_DAP               0.021047  -0.00056430   -0.0061210    -0.014134    -0.011339
l_TSP               0.030634   -0.0099380    -0.012503   -0.0039988    -0.011587
l_UREA_EE_BULK      0.060077    -0.039200     0.031130   -0.0086262   -0.0053866
l_POTASH           0.0016076     0.012141     0.012305    0.0058383    -0.011101

renormalized beta
l_WHEAT_US_HRW        1.0000      0.53554    -0.091055      -10.490      -1.4987
l_DAP               -0.34972       1.0000     -0.61272       46.099      0.30238
l_TSP                0.55258     -0.77850       1.0000      -31.440    0.0048572
l_UREA_EE_BULK       0.40137     -0.45747     -0.17561       1.0000      0.37284
l_POTASH            -0.15430   0.00088397     -0.26439      -1.5927       1.0000

renormalized alpha
l_WHEAT_US_HRW      -0.30070     -0.48043     -0.10778  5.4143e-05     0.013225
l_DAP               -0.15671     0.012087     -0.11667  -0.0067961    -0.046754
l_TSP               -0.22810      0.21287     -0.23831  -0.0019228    -0.047774
l_UREA_EE_BULK      -0.44732      0.83964      0.59336  -0.0041479    -0.022210
l_POTASH           -0.011970     -0.26005      0.23454   0.0028073    -0.045770

long-run matrix (alpha * beta')
               l_WHEAT_US_HRW        l_DAP        l_TSP l_UREA_EE_BULK       l_POTASH
l_WHEAT_US_HRW      -0.56856      -0.30274      0.098436        0.12300       0.087610
l_DAP              0.0017441      -0.18905    0.00076224      -0.072170       0.019109
l_TSP            -0.00062918       0.33557      -0.46985       -0.16682       0.053681
l_UREA_EE_BULK      0.025113       0.43458      -0.17717       -0.68028       -0.10272
l_POTASH            -0.13345      -0.28400       0.34189       0.058717       -0.11063
```

Appendix D: other datasets

{The following are present in the dataset but are to some extent less relevant to the study and thus discarder from it.

-The dataset contains coal prices in US Dollars per metric tonne from Australian and South Africa which are less relevant to the study.

-The dataset in the same fashion presents beverages, Cocoa, Coffee, Tea (Sri lanka,, Indian and Kenyan origin) in US Dollars per Kilograms. Then oils Coconut, Copra (Phippines/Indonesia),Groundnut (US), Palm, Palmkernel (Malaysia/Indonesia), Soybean and Sunflower oils (Dutch exchange). These will not be relevant to our study.

Other foods, like Bananas,Oranges, sugar along with Beef, Chicken, Lamb in Dollars per Kilogram, raw materials Logs, Plywood, Sawnwood, Woodpulp, Cotton, Rubber and Tobacco. Will not be included in the study.

Minerals and Metals, listed on the London Metals Exchange: Aluminium, Copper, Iron ore, Lead, Nickel, Tin and Zinc Dollards per metric ton and Steel from Japan in various forms. Precious metals are also included, as listed by the Lodon fix at 3PM of Gold, Platinium and Silver.}