Received: 16 June 2022 | Accepted: 21 December 2022

DOI: 10.1111/1467-8489.12504

The Australian Journal of Agricultural and Resource Economics

AARES AUSTRALASIAN AGRICULTURAL & RESOURCE ECONOMICS SOCIETY

ORIGINAL ARTICLE

Futures markets and price stabilisation: An analysis of soybeans markets in North America

Dragan Miljkovic[1] | **Cole Goetz**[2]

[1]Department of Agribusiness and Applied Economics, North Dakota State University, Fargo, North Dakota, USA

[2]CoBank, Denver, Colorado, USA

Correspondence
Dragan Miljkovic, Department of Agribusiness and Applied Economics, North Dakota State University, 500 Barry Hall, NDSU Dept. 7610, Fargo, ND 58108-6050, USA.
Email: dragan.miljkovic@ndsu.edu

Funding information
National Institute of Food and Agriculture, Grant/Award Number: Hatch grant #ND03300

Abstract

The objective of this paper was to determine whether the futures markets have a stabilising or destabilising impact on soybean's spot prices in North America. Directed acyclic graphs (DAGs) are used to test for causality between futures prices, spot prices and ending stocks, followed by time series econometric analysis. The DAGs point to the two-way causal link between futures and spot prices and a lack of a causal link between inventory/stocks and spot price volatility. Time series results, including cointegration, vector error correction, impulse response and variance decomposition analysis, indicate a large impact from futures markets on the level and volatility of soybean spot prices in both the short and long run. These results have potentially important implications, as the impact of commodity price volatility is typically asymmetric across different actors. Farmers, for example, unlike speculators, utilise price risk management (PRM) instruments such as futures markets to mitigate price risks and appear to suffer from intensified volatility precisely because of their use of these instruments. Therefore, additional policies to cope with commodity price volatility, such as direct price controls or mitigation of consequences, can have critical stabilising functions supporting farmers' welfare and regional (rural) development.

KEYWORDS

directed acyclic graphs, futures prices, price stabilisation, soybeans markets, spot prices

JEL CLASSIFICATION

C32, G13, Q11

1 | INTRODUCTION

The literature on the theory of price-stabilising versus destabilising impact futures trading has on the spot prices of storable commodities has been rather inconclusive (e.g. Deaton & Laroque, 1996; Kawai, 1983; Newbery, 1987; Tirole, 1985). Most recently, Goetz et al. (2021) refine Kawai's model and reaffirm his findings that the impact of futures markets on spot price volatility of storable commodities can be either stabilising or destabilising. That depends on whether the dominant/prevailing disturbance in the commodity market comes from consumption, production or inventory holding. We follow this theoretical result by Goetz et al. (2021) as the working hypothesis employed in our paper. In the empirical part of their paper, they determine destabilising impacts of futures markets on corn spot prices and stabilising impacts on oil spot prices in the United States. Hence, Goetz et al's (2021) empirical results are consistent with their model predictions. In the national oil markets, demand (consumption)-side disturbances were dominant during the period considered, hence a small and stabilising impact of futures markets on spot oil prices in the United States. This finding is reinforced in the companion paper (Miljkovic & Goetz, 2020a) that considers US regional oil markets.

While there is a lack of consensus in economic theory literature on the subject, the textbook agricultural economics literature on the interrelations between futures and spot markets in agricultural commodities is that futures markets allow for price discovery by market participants, the smoother allocation of commodities over time and the transfer of risk from hedgers to speculators (e.g. Ferris, 2005; Tomek & Kaiser, 2014). More specifically, futures markets are highly useful to all the segments of the economy. They are useful to the producers because they can get an indication of the price likely to prevail at a future point in time and therefore can decide between various competing commodities and choose the best that suits them. It enables the consumers to get an idea of the price at which the commodity would be available at a future point in time. Futures trading is also useful to exporters as it provides an advance indication of the price likely to prevail and thereby helps the exporter in quoting a realistic price and secure an export contract in a competitive market (Easwaran & Ramasundaram, 2008). Comprehensive literature reviews regarding empirical studies to date are provided by Irwin et al. (2009), or more recently by Dimpfl et al. (2017).

Empirical studies provide much less conclusive and decisive evidence in favour of this price-stabilising effect. These inconsistent findings could be Illustrated in detail by comparing Brorsen et al. (1989) and Weaver and Banerjee (1990): Brorsen et al. (1989) found that live cattle futures increased the volatility in the cash market, while Weaver and Banerjee (1990) found that live cattle futures did not affect cash market volatility, while considering the same time period. When considering storable commodities in global and developing economies, empirical studies also provide mixed results. For instance, Morgan et al. (1994) consider four commodities, cocoa, coffee, sugar and wheat, and analyse the efficiency of associated futures markets in terms of price discovery and risk reduction. In essence, all four markets exhibit efficiency and increase price stability, thereby providing, in theory, a viable policy alternative for developing economies. To the contrary, Easwaran and Ramasundaram (2008) analyse Indian agricultural commodity markets and find the impact of futures markets on spot/cash prices to be destabilising. Another interesting comparison and dichotomy of findings are the results by Goetz et al. (2021) that find oil futures markets to have a stabilising effect on spot prices, while corn futures markets have a destabilising effect on corn spot prices in the United States. These results at the national level are further affirmed at the US regional level for oil (Miljkovic & Goetz, 2020a) and US wheat markets (Miljkovic & Goetz, 2020b). In terms of sophistication, previous studies move from those using simple Granger causality (Irwin et al., 2009) to more complex information share methodology of Hasbrouck (1995) used by Dimpfl et al. (2017).

The objective of this paper was to determine whether futures markets have a stabilising or destabilising impact on spot soybean prices in North Dakota and to analyse dynamic interrelations between soybean futures and spot prices. Soybeans represent the single most traded agricultural commodity in the United States. North Dakota is chosen as a representative US soybeans market because soybeans represent the single most important crop in North Dakota with the value of its production exceeding 2 billion US dollars in 2020 (https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=north%20dakota). While North Dakota is not among the top soybeans producers in the United States, soybeans are the single most traded crop in that state and, as such, mimic its importance at the national level. Thus, North Dakota represents an appropriate representative region and market within the United States. Soybean production and significance in the state increased threefold over the last two decades, during the westward expansion of the Soybeans Belt (https://ndsoybean.org/soybean-stats/). Hence, understanding the (price stabilising) role of futures markets in the soybean price discovery process is critical for producers and supply chain participants. Equally important is to determine whether standard price stabilisation instruments, such as futures markets, are suitable when considering an emerging market, such as the soybean market in North Dakota, rather than a mature market.

Historically, Australia has imported soybean meal to support its large livestock sector. In recent years, Australia has imported between 800,000 and 1 million metric tonnes of soybean meal (indexmundi.com, 2022). Argentina and the United States have been the largest exporters of the meal to Australia historically. Hence, an improved understanding of the price discovery process in US soybean markets should be helpful to Australian soybean importers, in addition to the country's livestock and agricultural sector more broadly.

2 | DATA AND METHODS

2.1 | Data

The underlying economic model as presented by Goetz et al. (2021) implies that empirical analysis is conducted by studying the interrelations between soybean spot and futures prices, as well as storage (stocks) levels. For the sake of full comparability of empirical results, our constructed series for soybeans runs from January 2005 to December 2019. For spot price, we utilise the average monthly price received in dollars per bushel in North Dakota, obtained from the United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS). For futures prices, we use a series of monthly closing prices for the front contract of the Chicago Board of Trade (CBOT) soybean futures contract, which was obtained from Bloomberg. The construction of the futures price data is a time series of monthly closing prices for the contract nearest maturity, and upon maturity of that contract, the series continues on with the monthly closing prices of the next nearest dated contract. We were unable to obtain a previously compiled series of monthly stocks for soybeans. We construct the monthly ending stocks in much the same way that Goetz et al. (2021) did for corn using the same type of NASS and the Upper Great Plains Transportation Institute (UGPTI) data. The major differences are that there may be no significant users of soybeans (unlike the ethanol plants using corn) within the state that we are aware of (for the time period contemplated), so we have no inclusion of 'use' data in our monthly soybean's stocks. A full account of how the ending stocks data for soybeans were constructed is presented in Appendix A.

Most of the previous studies analysing the relationship between cash and futures prices use higher frequency data (e.g. daily and weekly), including the ones cited in this paper. The key reason for our use of monthly data is the information on ending stocks. Following our description of how the data on ending stocks were constructed, it should be obvious that there are no data on ending stocks that would be of higher frequency than monthly unless one concentrates on highly localised markets. Finally, to facilitate visual analysis, we include Figure 1

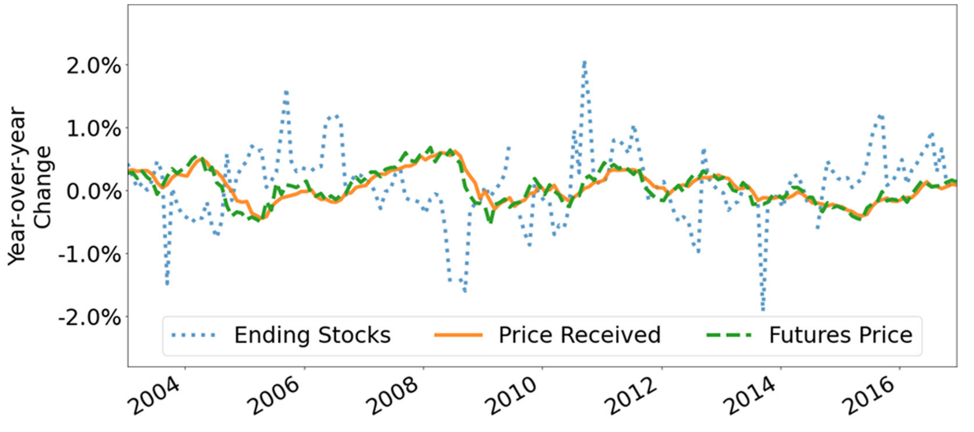

FIGURE 1 Year-to-year percentage change in futures and spot prices and ending stocks. [Colour figure can be viewed at wileyonlinelibrary.com]

with a plot of year-to-year percentage change as it makes it possible to observe all three variables simultaneously. A cursory glance at Figure 1 suggests that ending stock changes move in opposite directions from changes in either price, as one would expect from the theory. Cash and futures prices, meanwhile, exhibit changes that are aligned throughout the period, albeit volatilities may not have been obvious due to low frequency of the data.

2.2 | Causality analysis

A directed acyclic graph (DAG) is a method for determining contemporaneous causal relationships between (among) variables (e.g. Pearl, 1995, 2009). The DAGs are an alternative to Granger causality tests in that they look at nontime sequence asymmetry in causal interactions rather than the time sequence asymmetry used by Granger causality (Bessler & Yang, 2003). In causal structures, DAGs are used to represent researchers' *a priori* hypotheses about the relationships between and among variables. A DAG is a graphic illustration of a graph with directed edges (arrows), linking nodes (variables) and their paths. Computer algorithms make graphs that have nodes (variables) and edges (connections) between nodes to show these causal relationships (Bessler & Yang, 2003).

Let A, B and C represent nodes, which are variables. The edges can be directed or undirected, and they represent a causal relationship between nodes (indicated by the marks). A path is an unbroken sequence of distinct nodes connected by edges; a directed path, such as the path from A to C ($A \rightarrow B \rightarrow C$), follows the edges in the direction indicated by the arrows. An undirected path, such as the A to C path, does not follow the direction of the arrows. Kinship terms are usually employed in the representation of the relationship within a path. If a directed path exists from A to C, then A is C's ancestor and C is A's descendant. In the case of the directed path $A \rightarrow B \rightarrow C$, A is a direct cause or parent of B, and B is a child of A and parent of C, whereas A is an indirect cause or ancestor of C. As a node on the directed route, B is an intermediary or mediator variable. It is on the causal path between A and C.

Because no node may have an arrow pointing to itself and all edges must be directed (contain arrows), DAGs are acyclic (Greenland et al., 1999). In other words, there is no permissible directed path from any node to itself. The assumption that causes must come before effects is enforced by these rules. When assessing endogeneity from these graphs, variables with no causal input are exogenous, whereas variables with causal input are endogenous

(Spirtes et al., 2000). According to Miljkovic et al. (2016), a DAG is mathematically represented as the conditional independence by recursive product decomposition:

$$\Pr\left(v_1, v_2, \ldots, v_n\right) = \prod_{i=1}^{n} \Pr\left(v_i \mid p\pi_i\right) \tag{1}$$

where Pr is the probability of the variables $\left(v_1, v_2, \ldots, v_n\right)$, and $p\pi_i$ denotes the realisation of a subset of variables that produce v_i in the order (i = 1, 2, …, n). The product operator is denoted by Π. The work of Pearl (1995) on d-separation allows independencies and causes to be visually expressed. d-separation is a criterion for determining whether a set A of variables is independent of another set B, given a third set C, given a certain causal network. The concept is to identify 'dependency' with 'connectedness' (the presence of a connecting channel) and 'independence' with 'unconnected-ness' or 'separation'. Pearl (1995) suggests d-separation as a graphical representation of conditional independence. In other words, d-separation characterises the conditional independence relations defined by the equation. If we construct a DAG in which the variables corresponding to $p\pi_i$ are represented as the parents (direct causes) of v_i, we may read off the graph the independencies suggested by the equation using the concept of d-separation (Pearl, 1995).

Consider the three variable sets A, B and C while describing d-separation. We can say these variables are d-separated if the flow of information between these nodes is blocked. This is known as d-separation, and it can occur in two ways: first, if one variable, such as B in $A \leftarrow B \longrightarrow C$, is the cause of the other two variables, or if there is a passthrough variable, such as B in $A \rightarrow B \rightarrow C$; and second, when a variable is caused (influenced) by two variables, such as B in $A \rightarrow B \leftarrow C$. Spirtes et al. (2000) incorporated the concept of d-separation into the PC algorithm.

In comparison with the econometrics set-up in terms of employing instruments, DAG highlights the essential assumptions and structure of the relationship. The DAGs are clearer than the standard econometrics set-up, which presents the important assumptions in terms of the correlation between residuals and instruments. The DAGs can assist researchers define and share their opinions about the underlying data generation process, which may then assist in analysing the statistical relationships found in the data. Developing DAGs is not always simple, and it may need a heuristic approach in which assumptions are checked and amended based on observable statistical associations. A methodical approach to creating DAGs might be beneficial for presenting results and justifying covariate selection. The DAGs are also useful for causal modelling since they may infer identifiability from a complicated model.

The DAGs in this study were created using the PC and FGES algorithms in TETRAD software version 6.5.4. We first explore the PC approach (Spirtes et al., 2000) for learning DAG Markov equivalence classes. As a result of its use of conditional independence rules, the PC algorithm is called a constraint-based method. The PC algorithm begins with a fully connected network and determines whether an edge should be eliminated or preserved using conditional independence tests. However, there are two drawbacks to the PC algorithm, particularly when applied to large datasets: the runtime of the PC algorithm, which is exponential in terms of the number of nodes (variables) when applied to high-dimensional datasets (e.g. gene expression datasets), which was not a concern in our investigation. Second, the outcome of the PC method is variable-order dependent, that is, the result may change depending on the order of the variables in the input dataset. We then use the FGES algorithm (Colombo & Maathuis, 2014) to overcome this problem.

While DAGs explore causal relationships among variables, there are potential limitations of the DAG approach such as its nonparametric nature, lacking size of the associations, uncontrolled confounding biases or the inability to depict random errors. Potential implications of such limitations in the interpretation of results obtained by using DAGs could be significant. Thus, DAG analysis is often complemented with econometric analysis, to provide more informative outcomes (Imbens, 2020), in terms of both policy implications and business analytics.

2.3 | Time series analysis

The unit root test we employ is the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test (Kwiatkowski et al., 1992). The KPSS test has a null hypothesis that the given series is stationary, essentially testing for no unit root. As a robustness check, we employ the Dickey–Fuller generalised least squares (DF-GLS) test with the null hypothesis of a unit root. We test for cointegration using the Johansen test (Johansen, 1988; Johansen & Juselius, 1990).

One of the goals of this paper is to examine the dynamic relationships and interrelationships between spot and futures prices. Assuming that spot price and futures price are endogenous (the result ultimately determined by the DAGs), we can utilise a vector autoregression (VAR) or vector error correction model (VECM; Enders, 2010). An important clarifying note is in order here. Some of the previous literature suggests that DAGs should be applied to the shocks or innovations of these three variables after they are filtered in a three-variable VAR (e.g. Bessler & Yang, 2003; Bessler et al., 2003; Ji et al., 2018). We, however, utilise DAGs as an identification method only to specify the proper VAR, which is in line with Goetz et al. (2021) and Miljkovic et al. (2016). A VAR determined in such a way could consist of either three or two variables, or there could be an indication of complete absence of endogeneity among the variables, hence pointing to the inappropriateness of a VAR as an econometric modelling option. A VAR is a system of equations where all endogenous variables are a function of their own lagged values, lagged values of the other endogenous variables and any other exogenous explanatory variables that are deemed appropriate for the model. Note that the innovations are not correlated with their own lagged values and are uncorrelated with all explanatory variables but may be contemporaneously correlated (Wilson & Miljkovic, 2013). Since all right-hand-side explanatory variables are the same, ordinary least squares (OLS) yields efficient estimates (Enders, 2010).

When cointegration is present, the estimation approach must be changed. Since VAR models cannot deal with cointegration, we can restrict the VAR to achieve an error-correcting approach. A VECM allows us to examine the short-term adjustments of cointegrated variables to their long-run equilibrium. Similar to a VAR, in an ECM, the differenced endogenous variables are a function of lagged differenced values of itself, lagged differenced values of other endogenous variables, differenced exogenous variables and one or more cointegrating vectors which are the difference between the two cointegrated variables. Thus, a VECM can be represented mathematically as:

$$\Delta Y_t = B_0 + \pi z_{t-1} + B_1 \Delta Y_{t-1} + \ldots + B_p \Delta Y_{t-p} + A_0 \Delta X_t + \varepsilon_t \qquad (2)$$

where Δ is the difference operator, B_0 is an $n \times 1$ vector of intercept terms, $B_1, \ldots, B_p$ are $n \times n$ matrices of coefficients to be estimated for lagged endogenous variables, Y_t is an $n \times 1$ vector of endogenous variables, z_{t-1} is a $1 \times n$ vector containing the difference of cointegrating variables forming our cointegrating vector, π is an $n \times 1$ vector of adjustment coefficients related to our cointegrating vector, A_0 is $n \times n$ matrices related to coefficients of our exogenous variables (we may include lagged exogenous variables in a VECM as well), and ε_t is an $n \times 1$ vector of innovations. Note that if all elements of π are zero, we simply have a VAR in first differences (Enders, 2010).

Impulse response functions allow us to observe over time how an endogenous variable responds to an exogenous shock to itself and a shock to other endogenous variables. Based on the dynamic structure of a VAR or VECM, a shock to an endogenous variable will affect that variable but can also affect other endogenous variables. Thus, with impulse response functions, we can observe how a shock to one variable filters through the model to affect the other variables within the model (Pindyck & Rubinfeld, 1998). This technique ultimately allows us to examine what effect a shock in the futures market has on the time path of the spot market and vice versa.

Following Enders (2010), we can express a VAR as a vector moving average (VMA) in matrix form for two arbitrary variables y_t and z_t:

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \bar{z} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-i} \\ \varepsilon_{zt-i} \end{bmatrix} \tag{3}$$

Here, the coefficients $\phi_{11}(i), \ldots, \phi_{22}(i)$ are the impulse response functions. We can see that $\phi_{11}(0)$ is the instantaneous impact of a one unit change in ε_{yt} on y_t while $\phi_{11}(i)$ represents the ith period impact of a one unit change in ε_{yt-i} on y_t (Enders, 2010). We can plot the impulse response functions to see the time path of the responses to shocks.

Since an estimated VAR is underidentified, the impulse responses require additional restrictions to be identified (Enders, 2010). We utilise Cholesky decomposition to orthogonalise the innovations to obtain our impulse responses (Wilson & Miljkovic, 2013). Thus, the restriction alters the system so that y_t will not contemporaneously affect z_t. Again, following Enders (2010), we decompose the error terms in (3) such that:

$$e_{1t} = \varepsilon_{yt} - b_{12}\varepsilon_{zt} \tag{4}$$

$$e_{2t} = \varepsilon_{zt} \tag{5}$$

Thus, ε_{zt} has a contemporaneous direct effect on both z_t and y_t, while ε_{yt} has a direct effect on y_t and an indirect effect on z_t through lagged values of y_t. Hence, the impulse response functions allow us to observe how endogenous variables respond to shocks within the system.

The dynamic structure of the models can also be examined using variance decomposition, which breaks down the variance of the forecast errors for every endogenous variable into the percentage of the variance that can be credited to the other endogenous variables (Blanchard & Quah, 1989). This can be useful in identifying how large a role one variable has in affecting the variation of another variable. Blanchard and Quah (1989) variance decomposition can only be meaningfully applied to a two-variable VAR system; as our DAG results will indicate, it is appropriate in this case and hence the theoretical foundation of this procedure is presented below. Ultimately, more general impulse response analysis and resulting variance decomposition (Pesaran & Shin, 1998) is not necessary or superior in this case.

Following Enders (2010) and Blanchard and Quah (1989), we express the VMA in terms of its forecast errors where the n-period forecast error variance of y_t is $\sigma_y(n)^2$:

$$\sigma_y(n)^2 = \sigma_y^2\left[\phi_{11}(0)^2 + \phi_{11}(1)^2 + \ldots + \phi_{11}(n-1)^2\right] + \sigma_z^2\left[\phi_{12}(0)^2 + \phi_{12}(1)^2 + \ldots + \phi_{12}(n-1)^2\right] \tag{6}$$

Note that as the forecast horizon increases, so too will the forecast error variance due to the nonnegativity of all $\phi_{mn}(i)^2$ terms. The proportion of y_t's n-period forecast error variance due to shocks to ε_{yt} can be represented as follows:

$$\frac{\sigma_y^2\left[\phi_{11}(0)^2 + \phi_{11}(1)^2 + \ldots + \phi_{11}(n-1)^2\right]}{\sigma_y(n)^2} \tag{7}$$

while the proportion due to shocks to ε_{zt} can be represented similarly.

3 | RESULTS

The DAG results are presented in Figure 2. The presence of undirected edges in both PC and FGES algorithms for soybeans indicates that futures price and price received are endogenous, that is, there is bidirectional causality between these two prices. Ending stocks are not causally related to either price; hence, they are considered exogenous in subsequent econometric analysis. These results are even stronger than the results for corn in Goetz et al. (2021) as they report how, based on the PC algorithm but not the FGES algorithm, ending stocks and futures prices could be considered endogenous. They resort to Granger causality as the tie-breaking procedure to resolve this issue, while we are in this case confident that our results are robust.

To test for unit roots, we employ the KPSS test. The KPSS test has a null hypothesis that the given series is stationary, essentially testing for no unit root. KPSS test results suggest non-stationarity in the levels, while all three soybean variables are stationary in the first difference. To check for the robustness of this result, we use the DF-GLS test. The DF-GLS tests a null hypothesis that the given series has a unit root. Its results concur with the KPSS test findings. KPSS test results are presented in Table 1, and the results of the DF-GLS test are presented in Table 2.

The Johansen cointegration test reveals the presence of one cointegrating vector between the two prices, as ending stocks are considered exogenous. The resulting VEC model provides important results on short-term dynamics between the variables and the long-term adjustment as represented by the speed of adjustment coefficient. The number of lags is set at 3 based on the Akaike information criterion (AIC).

Next, Johansen test results are presented before the VEC results. Most importantly, we imposed a restriction on equality of speed of adjustment coefficients in two EC equations. This hypothesis was strongly rejected, and the result implies unequal long-run reaction or speed of adjustment in spot and futures prices to external shocks, pointing to some long-term inefficiencies in the price discovery process. Note that DAGs are the superior method in determining

FIGURE 2 Directed acyclic graph results. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Kwiatkowski-Phillips-Schmidt-Shin test – soybeans

Null hypothesis: The series is stationary				
Variables	Time period	Exogenous variables	LM-stat (level)	LM-stat (first diff.)
Soybeans futures price	2005–2019	Constant	1.055***	0.095
Soybeans price received	2005–2019	Constant	1.084***	0.153
Soybeans ending stocks	2005–2019	Constant	0.481**	0.188

Note: **5 % Significance; ***1% Significance.

TABLE 2 Elliott-Rothenberg-Stock DF-GLS test – soybeans

Null hypothesis: nonstationary series				
Variables	Time period	Exogenous variables	*t*-statistic (level)	*t*-statistic (first diff.)
Soybeans futures price	2005–2019	Constant	−1.014	−13.803***
Soybeans price received	2005–2019	Constant	−0.784	−9.553***
Soybeans ending stocks	2005–2019	Constant	0.157	−1.996**

Note: **5 % Significance; ***1% Significance.

endogeneity via causal linkages, as we referenced in the paper. Hence, we thought it more appropriate to restrict the speed of adjustment coefficients for equality and test for the efficiency of the price discovery process in spot markets via futures markets rather than to test for weak exogeneity of ending stocks, which would be a plausible alternative if not for the DAG results.

Short-term dynamics, as presented by the estimated coefficients in two EC equations, yield some interesting results as well. In explaining the first difference of FUTP, the third lag of FUTP is significant with a negative sign while the first two lags of SPOT are significant with positive signs. Thus, recent increases in spot prices lead to an increase in futures prices. In explaining the first difference of SPOT, the third lag of FUTP and the first difference of STOCK are significant at the 10% level. It is the speed of adjustment term that accounts for the bulk of change in spot price. However, it is important that even in the short-term dynamics we can see two-way impacts of SPOT on FUTP and vice versa, thus reinforcing the DAG bidirectional causality finding. VEC and cointegration results are presented in Table 3.

The soybean impulse responses are shown in Figure 3, while the variance decomposition is shown in Figure 4. A shock to SPOT prompts a permanent increase in FUTP by nearly $0.20 in the long run. Similarly, an innovation in FUTP elicits a permanently increased response in the standard deviation of SPOT by almost $0.80 in the long run. SPOT also comprises only a small portion of FUTP variance, accounting for only 4% in the 36th period. A large portion of SPOT variance is made up of FUTP, which accounts for 95% of the variance by the 36th month.

4 | CONCLUSIONS AND IMPLICATIONS

The theoretical model by Goetz et al. (2021) that we use as the reference point predicts that when production (supply side) is the dominant disturbance, the spot price is destabilised in the short run by futures markets but may or may not be stabilised in the long run. Agricultural commodity markets, including soybean markets, are subject to various production disturbances such as weather events (e.g. drought, flood and hail) or pest infestations (Tomek & Kaiser, 2014). Moreover, legislation on ethanol subsidies in recent decades further stimulated soybeans production within and outside the Soybean Belt, adding to the list of supply (and demand)-side disturbances on soybeans prices (e.g. McPhail & Babcock, 2012). In turn, soybeans represent a rotational crop with corn due to agronomic reasons (Bullock, 1992), hence amplifying the destabilising role futures markets may play in soybean spot markets. The empirical results indicate a large impact from futures markets on the levels and volatility of soybean spot prices in both the short and long run, which is consistent with the theoretical model. However, our empirical results point to a lack of a causal link between inventory/stocks and spot price volatility, another possible source of destabilisation in spot prices, based on the theoretical model.

The impact of commodity price volatility is typically asymmetric across different actors. Farmers, for example, unlike speculators, utilise price risk management (PRM) instruments such as futures markets to mitigate price risk and appear to suffer from intensified volatility

TABLE 3 Soybeans VECM

Cointegration restrictions		
$\pi_{11} = \pi_{21}$		
Convergence achieved after four iterations		
Not all cointegrating vectors are identified		
LR test for binding restrictions (rank = 1):		
Chi-square(1)		12.1863
Probability		0.0005***
Cointegrating Eq		**CointEq1**
FUTP(−1)		−2.4268
SPOT(−1)		2.7277
C		−0.6460
Error correction	**ΔFUTP**	**ΔSPOT**
CointEq1	−0.1414***	−0.1414***
	[−5.26514]	[−5.26514]
ΔFUTP(−1)	−0.0993	0.0972
	[−0.69228]	[1.55799]
ΔFUTP(−2)	−0.1580	0.0122
	[−1.20074]	[0.21322]
ΔFUTP(−3)	−0.3841***	−0.0829*
	[−3.52441]	[−1.74966]
ΔSPOT(−1)	0.5591***	0.0386
	[2.64322]	[0.42009]
ΔSPOT(−2)	0.6224***	0.0071
	[3.16849]	[0.08350]
ΔSPOT(−3)	−0.1346	−0.0596
	[−0.84841]	[−0.86412]
C	0.0199	0.0279
	[0.30284]	[0.97655]
ΔSTOCK	0.0000	0.0000*
	[0.22118]	[−1.79880]
R-squared	0.1519	0.4335

Note: *t*-statistics in []; *10% Significance; **5% Significance; ***1% Significance.

precisely because of their use of these instruments. Further, extreme price shocks can lead to irreversible negative welfare shocks when existing coping mechanisms are diminished or fail (Tröster, 2018). In combination, this can set in motion a downward spiral of rising vulnerability, affecting fragile systems and actors the most, for instance, farmers in food systems. Therefore, policies to cope with commodity price volatility, such as direct price controls or mitigation of consequences, can have critical stabilising functions supporting farmer welfare and regional (rural) development (Goetz et al., 2021).

The integration of US soybeans into the global economy, combined with competing uses for food, feed and fuel, has increased the number, frequency and impact of exogenous shocks to soybeans markets. Traditional theoretical models suggest price stabilisation attained through stockholding activities leads to a net welfare improvement to society, but there are gainers

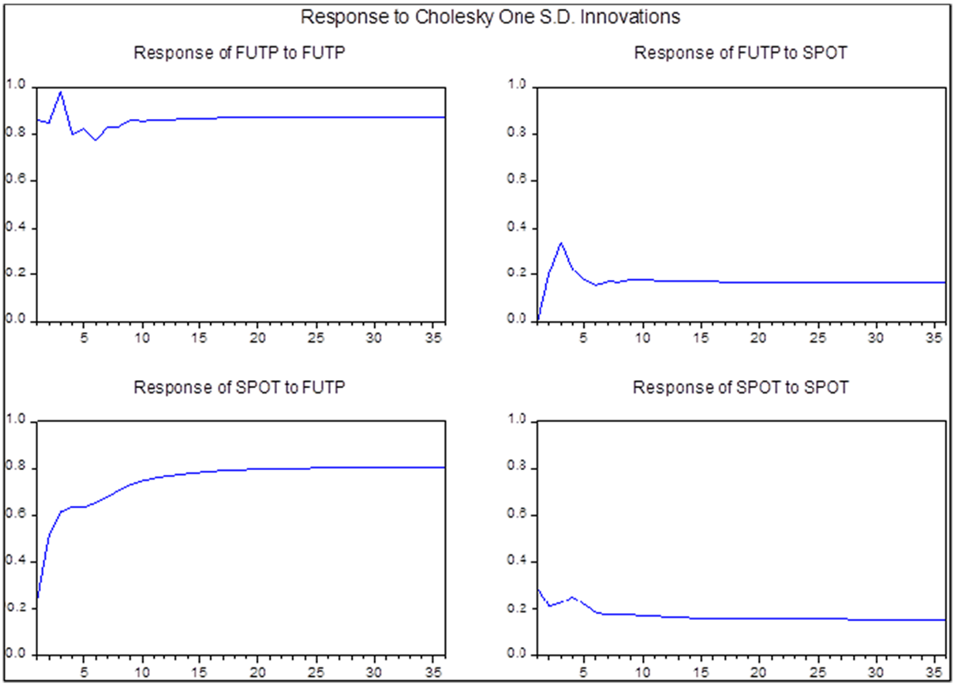

FIGURE 3 Soybeans impulse responses. [Colour figure can be viewed at wileyonlinelibrary.com]

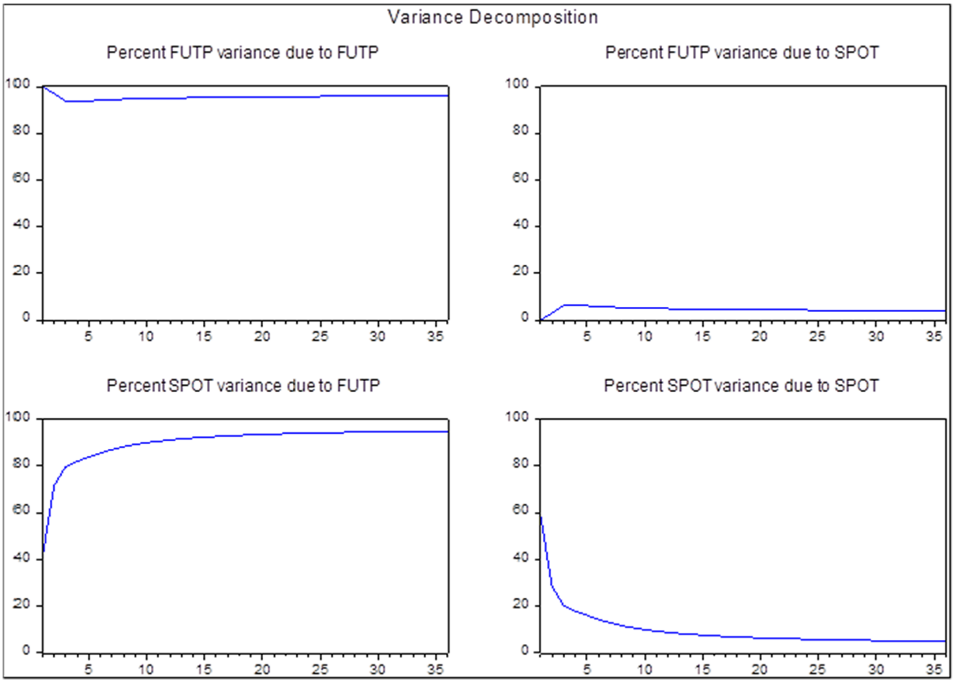

FIGURE 4 Soybeans variance decomposition. [Colour figure can be viewed at wileyonlinelibrary.com]

and losers from price stabilisation policies. Moreover, the effectiveness and cost of alternative price-stabilising systems, including operational futures markets for the commodities (soybeans in this case), in addition to more traditional domestic and international farm policies inventory management practices and supply chain coordination, have not been fully explored. More complex theoretical models reflecting new realities in commodity trading, including relevant technological developments (e.g. algorithmic trading or order filling algorithms), institutional factors (e.g. impacts of USDA or other scheduled reports) and changed philosophy of farm policy relevant to commodity markets, including soybeans, are needed to ensure a more credible and testable theory.

There are some potential caveats to our analysis, such as the sample period, frequency of the analysis and geographical area considered. Selecting major soybean production states, such as Illinois and Iowa, or using national-level data, one could consider superior to using North Dakota data. While there is merit to such a claim, North Dakota was selected for several reasons. Data were available for a relatively long period for all considered variables unlike for many other states. Note that data on ending stocks required somewhat of a creative approach as such data are not readily available, as described in the Appendix A. Many other states do not provide similar information at all. For those that provide aggregate information on ending stocks, it is unclear how this information was created as many factors are involved in producing ending stocks data time series. Also, if aggregate national data were to be used on ending stocks, serious aggregation problem would have been present. Finally, the reason for our use of monthly data rather than higher frequency data, which could be more adequate when analysing price volatilities, is the information on ending stocks. The data on ending stocks could be constructed as monthly data at the highest frequency, as the monthly data unless one is to look to highly localised markets.

Empirical analysis of the role futures markets play in cash market price stabilisation or destabilisation is also not definitive. Access to quality cash market datasets and accurate inventory levels is always a challenge, but private data sources for cash market price bids are becoming more accessible. A focus on not only the spot market but also forward pricing opportunities in both cash and futures markets may provide additional insights regarding the ability of cash market participants to manage price risk and adjust to changing conditions in an efficient way. Combining modern empirical analysis with event studies of extreme market shocks, for example, the 2008 financial crisis or the current impacts of the Russian invasion of Ukraine, may provide a more robust analysis of how market participants, both producers and consumers, adjust to price instability.

ACKNOWLEDGEMENT

The authors thank Nikita Barabanov, Lei Zhang and research seminar participants at North Dakota State University and Purdue University for their valuable feedback and comments. Remaining errors are solely ours. There are no competing interests to report. Data will be made available upon request from the contact author. Finally, we acknowledge financial support by USDA through Hatch grant #ND03300.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study. However, the data that support the findings of this study are available on request from the corresponding author.

ORCID

Dragan Miljkovic https://orcid.org/0000-0001-7511-7260

REFERENCES

Bessler, D.A. & Yang, J. (2003) The structure of interdependence of international stock markets. *Journal of International Money and Finance*, 22(2), 261–287.

Bessler, D.A., Yang, J. & Wongcharupan, M. (2003) Price dynamics in the international wheat market: modeling with error correction and directed acyclic graphs. *Journal of Regional Science*, 43(1), 1–33.

Blanchard, O.J. & Quah, D.D. (1989) The dynamic effects of aggregate demand and supply disturbances. *The American Economic Review*, 79(4), 655–673.

Brorsen, B.W., Oellermann, C.M. & Farris, P.L. (1989) The live cattle futures market and daily cash price movements. *Journal of Futures Markets*, 9(4), 273–282.

Bullock, D.G. (1992) Crop rotation. *Critical Reviews in Plant Sciences*, 11(4), 309–326.

Colombo, D. & Maathuis, M.H. (2014) Order-independent constraint-based causal structure learning. *Journal of Machine Learning Research*, 15(1), 3741–3782.

Deaton, A. & Laroque, G. (1996) Competitive storage and commodity price dynamics. *Journal of Political Economy*, 104(5), 896–923.

Dimpfl, T., Flad, M. & Jung, R.C. (2017) Price discovery in agricultural commodity markets in the presence of futures speculation. *Journal of Commodity Markets*, 5(1), 50–62.

Easwaran, R.S. & Ramasundaram, P. (2008) Whether commodity futures market in agriculture is efficient in price discovery?-an econometric analysis. *Agricultural Economics Research Review*, 21(347–2016-16671), 337–344.

Enders, W. (2010) *Applied econometric time series*, 3rd edition. Hoboken, NJ: John Wiley & Sons.

Ferris, J.N. (2005) *Agricultural prices and commodity market analysis*. East Lansing: Michigan State University Press.

Goetz, C., Miljkovic, D. & Barabanov, N. (2021) New empirical evidence in support of the theory of price volatility of storable commodities under rational expectations in spot and futures markets. *Energy Economics*, 100, 105375.

Greenland, S., Pearl, J. & Robins, J. (1999) Causal diagrams for epidemiologic research. *Epidemiology*, 1(10), 37–48.

Hasbrouck, J. (1995) One security, many markets: determining the contributions to price discovery. *The Journal of Finance*, 50(4), 1175–1199

Imbens, G.W. (2020) Potential outcome and directed acyclic graph approaches to causality: relevance for empirical practice in economics. *Journal of Economic Literature*, 58(4), 1129–1179.

Irwin, S., Sanders, D. & Merrin, R.P. (2009) Devil or angel? The role of speculation in the recent commodity price boom (and bust). *Journal of Agriculture and Applied Economics*, 41(2), 377–391.

Ji, Q., Zhang, H.-A. & Geng, J.-B. (2018) What drives natural gas prices in the United States? – a directed acyclic graph approach. *Energy Economics*, 69(1), 79–88.

Johansen, S. (1988) Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2–3), 231–254.

Johansen, S. & Juselius, K. (1990) Maximum likelihood estimation and inference on cointegration—with applications to the demand for money. *Oxford Bulletin of Economics and Statistics*, 52(2), 169–210.

Kawai, M. (1983) Price volatility of storable commodities under rational expectations in spot and futures markets. *International Economic Review*, 24(2), 435–459.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. & Shin, Y. (1992) Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1), 159–178.

McPhail, L.L. & Babcock, B.A. (2012) Impact of US biofuel policy on US soybeans and gasoline price variability. *Energy*, 37(1), 505–513.

Miljkovic, D., Dalbec, N. & Zhang, L. (2016) Estimating dynamics of US demand for major fossil fuels. *Energy Economics*, 55, 284–291.

Miljkovic, D. & Goetz, C. (2020a) The effects of futures markets on oil spot price volatility in regional US markets. *Applied Energy*, 273, 115288.

Miljkovic, D. & Goetz, C. (2020b) Destabilizing role of futures markets on north American hard red spring wheat spot prices. *Agricultural Economics*, 51(6), 887–897.

Morgan, C.W., Rayner, A.J. & Ennew, C.T. (1994) Price instability and commodity futures markets. *World Development*, 22(11), 1729–1736.

Newbery, D.M. (1987) When do futures destabilize spot prices? *International Economic Review*, 28(2), 291–297.

Pearl, J. (1995) Causal diagrams for empirical research. *Biometrika*, 82(4), 669–688.

Pearl, J. (2009) *Causality*, 2nd edition. Cambridge, MA: Cambridge University Press.

Pesaran, H.H. & Shin, Y. (1998) Generalized impulse response analysis in linear multivariate models. *Economics Letters*, 58(1), 17–29.

Pindyck, R.S. & Rubinfeld, D.L. (1998) *Econometric models and economic forecasts*, 4th edition. Boston, MA: McGraw-Hill.

Spirtes, P., Glymour, C. & Scheines, R. (2000) *Causation, prediction, and search*, 2nd edition. New York, NY: Springer-Verlag.

Tirole, J. (1985) Asset bubbles and overlapping generations. *Econometrica*, 53(6), 1499–1528.

Tomek, W.G. & Kaiser, H.M. (2014) *Agricultural product prices*. Ithaca: Cornell University Press.

Tröster, B. (2018) *Commodity price stabilization: the need for a policy mix that breaks the vicious cycle of commodity dependence and price volatility*. ÖFSE Policy Note, No. 20/2018. Vienna: Austrian Foundation for Development Research (ÖFSE).

Weaver, R.D. & Banerjee, A. (1990) Does futures trading destabilize cash prices? Evidence for U. S. Live beef cattle. *Journal of Futures Markets*, 10(1), 41–60.

Wilson, W.W. & Miljkovic, D. (2013) Dynamic interrelationships in hard wheat basis markets. *Canadian Journal of Agricultural Economics/Revue Canadienne d'Agroeconomie*, 61(3), 397–416.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Miljkovic, D. & Goetz, C. (2023) Futures markets and price stabilisation: An analysis of soybeans markets in North America. *Australian Journal of Agricultural and Resource Economics*, 67, 104–117. Available from: https://doi.org/10.1111/1467-8489.12504